

Національний університет
“Полтавська політехніка імені Юрія Кондратюка”

National University
“Yuri Kondratyuk Poltava Polytechnic”

СИСТЕМИ управління, навігації та зв'язку

Control, navigation and communication systems

Випуск 2 (72)

Issue 2 (72)

Щоквартальне видання

Засноване у 2007 році

У журналі відображені результати наукових досліджень з розробки та удосконалення систем управління, навігації та зв'язку у різних проблемних галузях.

Засновник і видавець:

Національний університет
“Полтавська політехніка імені Юрія Кондратюка”

Телефон:

+38 (050) 302-20-71

E-mail редколегії:

kuchuk_nina@ukr.net

Інформаційний сайт:

<http://journals.nupp.edu.ua/sunz>

Quarterly

Founded in 2007

Journal represent the research results on the development and improvement of control, navigation and communication systems in various areas

Founder and publisher:

National University
“Yuri Kondratyuk Poltava Polytechnic”

Phone:

+38 (050) 302-20-71

E-mail of the editorial board:

kuchuk_nina@ukr.net

Information site:

<http://journals.nupp.edu.ua/sunz>

За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор

*Журнал індексується міжнародними наукометричними базами: Index Copernicus (ICV = **82.05**),
General Impact Factor, Google Scholar, Academic Resource Index, Scientific Indexed Service*

*Затверджений до друку Вченою Радою Національного університету
“Полтавська політехніка імені Юрія Кондратюка” (протокол від 9 червня 2023 року № 6).*

Свідоцтво про державну реєстрацію КВ № 24464-14404 ПР від 27.03.2020 р.

Включений до “Переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії” до категорії Б – наказами МОН України від 17.03.2020 № 409 та від 09.02.2021 № 157

Полтава • 2023

Редакційна колегія

Головний редактор:

КОСЕНКО Віктор Васильович
(*д-р техн. наук, проф., Полтава, Україна*).

Заступники головного редактора:

НЕСТЕРЕНКО Катерина Сергіївна
(*д-р техн. наук, проф., Київ, Україна*);
ШЕФЕР Олександр Віталійович
(*д-р техн. наук, проф., Полтава, Україна*).

Члени редакційної колегії:

БЛАУНШТЕЙН Натан Олександрович
(*д-р техн. наук, проф., Ізраїль*);
БОГОМ'Я Володимир Іванович
(*д-р техн. наук, проф., Київ, Україна*);
ВАРБАНЕЦЬ Роман Анатолійович
(*д-р техн. наук, проф., Одеса, Україна*);
ВЕСОЛОВСЬКИЙ Кшиштоф
(*д-р техн. наук, проф., Польща*);
ГАВРИЛКО Євген Володимирович
(*д-р техн. наук, доц., Київ, Україна*);
ГАШИМОВ Ельшан Гіяс огли
(*д-р наук, проф., Баку, Азербайджан*);
ГЛИВА Валентин Анатолійович
(*д-р техн. наук, проф., Київ, Україна*);
ДАКІ Олена Анатоліївна
(*д-р техн. наук, доц., Ізмаїл, Україна*);
ЗАЩЕПКИНА Наталія Миколаївна
(*д-р техн. наук, проф., Київ, Україна*);
КАСАТКІНА Наталія Вікторівна
(*д-р техн. наук, доц., Київ, Україна*);
КОВАЛЕНКО Андрій Анатолійович
(*д-р техн. наук, проф., Харків, Україна*);
КОЛОМІЙЦЕВ Олексій Володимирович
(*д-р техн. наук, проф., Харків, Україна*);
КОРОБКО Богдан Олегович
(*д-р техн. наук, доц., Полтава, Україна*);
КРАСНОБАЄВ Віктор Анатолійович
(*д-р техн. наук, проф., Харків, Україна*);
КУЧУК Георгій Анатолійович
(*д-р техн. наук, проф., Харків, Україна*);
ЛЕВЧЕНКО Лариса Олексіївна
(*д-р техн. наук, доц., Київ, Україна*);
ЛУКОВА-ЧУЙКО Наталія Вікторівна
(*д-р техн. наук, доц., Київ, Україна*);
ЛУНТОВСЬКИЙ Андрій Олегович
(*д-р техн. наук, проф., Німеччина*);
ПИСАРЧУК Олексій Олександрович
(*д-р техн. наук, проф., Київ, Україна*);
ПОДКОПАЄВ Сергій Вікторович
(*д-р техн. наук, проф., Покровськ, Україна*);
СЕМЕНОВ Сергій Геннадійович
(*д-р техн. наук, проф., Харків, Україна*);
ТИМОЩУК Олена Миколаївна
(*д-р техн. наук, проф., Київ, Україна*);
ТРИСТАН Андрій Вікторович
(*д-р техн. наук, доц., Чернігів, Україна*);
ФРОЛОВ Євгеній Андрійович
(*д-р техн. наук, проф., Полтава, Україна*);
ЧОРНИЙ Олексій Петрович
(*д-р техн. наук, проф., Кременчук, Україна*);

Відповідальний секретар:

КУЧУК Ніна Георгіївна
(*д-р техн. наук, проф., Харків, Україна*).

Технічні секретарі:

ЗАХАРЧЕНКО Руслан Володимирович
(*канд. техн. наук, доц., Полтава, Україна*);
ПЕТРОВСЬКА Інна Юріївна
(*магістр комп. інж., Харків, Україна*).

Editorial board

Editor-in-Chief:

Viktor KOSENKO
(*Dr. Sc. (Tech.), Prof., Poltava, Ukraine*).

Associates editor:

Katerina NESTERENKO
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Oleksandr SHEFER
(*Dr. Sc. (Tech.), Prof., Poltava, Ukraine*).

Editorial board members:

Nathan BLAUNSTEIN
(*Dr. Sc. (Tech.), Prof., Israel*);
Volodymyr BOHOMYA
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Roman VARBANETS
(*Dr. Sc. (Tech.), Prof., Odesa, Ukraine*);
Krzysztof WESOŁOWSKI
(*Dr. Sc. (Tech.), Prof., Poland*);
Yevhen HAVRILKO
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Elshan Giyas oglu HASHIMOV
(*Dr. Sc., Prof., Baku, Azerbaijan*);
Valentyn GLYVA
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Olena DAKI
(*Dr. Sc. (Tech.), Ass. Prof., Izmail, Ukraine*);
Nataliia ZASHCHEPKINA
(*Dr. Sc. (Tech.), Ass. Prof., Kyiv, Ukraine*);
Natalia KASATKINA
(*Dr. Sc. (Tech.), Ass. Prof., Kyiv, Ukraine*);
Andriy KOVALENKO
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Oleksii KOLOMIITSEV
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Bohdan KOROBKO
(*Dr. Sc. (Tech.), Ass. Prof., Poltava, Ukraine*);
Viktor KRASNOBAYEV
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Heorhii KUCHUK
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Larysa LEVCHENKO
(*Dr. Sc. (Tech.), Ass. Prof., Kyiv, Ukraine*);
Natalia LUKOVA-CHUIKO
(*Dr. Sc. (Tech.), Ass. Prof., Kyiv, Ukraine*);
Andriy LUNTOVSKYY
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Oleksii PYSARCHUK
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Serhii PODKOPAIEV
(*Dr. Sc. (Tech.), Prof., Pokrovsk, Ukraine*);
Serhii SEMENOV
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Olena TYMOSHCHUK,
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Andrii TRYSTAN
(*Dr. Sc. (Tech.), Ass. Prof., Chernihiv, Ukraine*);
Yevhen FROLOV
(*Dr. Sc. (Tech.), Prof., Poltava, Ukraine*);
Oleksii CHORNYI
(*Dr. Sc. (Tech.), Prof., Kremenchuk, Ukraine*).

Responsible secretary:

Nina KUCHUK
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*).

Technical secretaries:

Ruslan ZAKHARCHENKO
(*PhD (Tech.), Ass. Prof., Poltava, Ukraine*);
Inna PETROVSKA
(*MSD of Comp. Eng., Kharkiv, Ukraine*).

З М І С Т

НАВІГАЦІЯ ТА ГЕОІНФОРМАЦІЙНІ СИСТЕМИ

<i>Пащенко Р. Е., Марюшко М. В.</i> Методика моніторингу сільськогосподарських земель і культур з використанням фрактального аналізу даних дистанційного зондування Землі	5
---	---

АВТОМОБІЛЬНИЙ, РІЧКОВИЙ, МОРСЬКИЙ ТА АВІАЦІЙНИЙ ТРАНСПОРТ

<i>Горбачов П. Ф., Любий Є. В.</i> Підхід до оцінювання затримок транспортних потоків у плані координат ..	15
<i>Куліш Р. В., Матюченко О. Г.</i> Модель маршрутизації об'їзду мобільних об'єктів безпілотним літальним апаратом	20
<i>Мураховська О. А.</i> Системно-спрямований підхід аналізу можливих методів дослідження аеродинамічної живучості літака	26
<i>Naghiyeva P. A.</i> Theoretical analysis of the level of communication security of UAVs operating as part of a group performing a combat mission	31
<i>Очеретенко С. В., Коровко В. А., Жатканбаєва Е. А.</i> Особливості формування замовлень автозапчастин при багатономенклатурному керуванні	35

УПРАВЛІННЯ В СКЛАДНИХ СИСТЕМАХ

<i>Baranenko R., Sokol H., Pustovoitov P., Zakharov R.</i> Information and analytical support of the education quality management system	39
<i>Dzheria T., Shevchuk V.</i> Forecasting electricity consumption using neural networks	42
<i>Зуєв А. О., Євсєєнко О. М.</i> Метод створення звукового оточення в імітаційно-тренажерних комплексах спеціальної техніки	45
<i>Ічанська Н. В., Шурник О. В.</i> Аналіз динаміки видобутку вугілля в Україні за допомогою методів математичного моделювання	49
<i>Караєва Н. В., Левченко Л. О.</i> Методологічні засади управління сталим розвитком енергетики в умовах глобальних ризиків і біфуркаційних зрушень	54
<i>Skakalina O., Kapiton A.</i> Implementation of the lighting function in the smart home concept	59
<i>Худов Г. В., Романенко І. О., Минко П. Є., Соломоненко Ю. С., Іванюк В. О.</i> Удосконалений метод прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму	63

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

<i>Бевзюк Р., Ляшенко О.</i> Система розпізнавання об'єктів для автоматизації складів з використанням згорткової нейронної мережі	68
<i>Vyshnivskiy D., Liashenko O., Yeromina N.</i> Human pose estimation system using deep learning algorithms	75
<i>Волк М. О., Гора М. В., Лабазов В. Г., Міщенко А. В., Барсуков А. І., Голець В. В.</i> Журналізація стану програм для самовідновлення паралельних програмних систем	80
<i>Гавриленко С. Ю., Зозуля В. Д., Омельченко В. В.</i> Дослідження методів підвищення якості класифікації на незбалансованих даних	87
<i>Ганпельманов А. В., Міхаль О. П., Щенка О. О.</i> Підвищення енергозбереження бездротових сенсорних мереж з використанням методів машинного навчання	92
<i>Golovko G., Ievliev D.</i> Enhanced authorization for secure management of sensitive data in hybrid applications ...	98
<i>Demydenko M., Kalnytskyi M.</i> Online platform prototype using microservice architecture and containerization for digitalization of the educational process	101
<i>Ляшшов О. А., Покора К. В., Дяченко В. О., Коваленко А. А.</i> Класифікація даних апаратними прискорювачами FPGA у центрах обробки даних та хмарах	106
<i>Каргін А. О., Жуков С. В., Сергєєв Д. А., Сілін Є. Л.</i> Модель базового рівня штучного оточення автономних інтелектуальних безлюдних систем на прикладі мобільного роботу що обслуговує	113
<i>Kuchuk N., Zdorenko Yu., Wang Shengquan.</i> A lightweight target detection algorithm based on Mobilenet Convolution	119
<i>Маловічко Є. В., Харченко В. С.</i> Інтерактивне мистецтво: мультисценарні AR-технології на підставі аналізу емоційного стану	125
<i>Матвєєв М. І., Кучук Г. А.</i> Оптимізація мобільної гри в середовищі Unity	131
<i>Рибальченко А. О.</i> Алгоритми рішення задачі оптимального розміщення даних в білінгових OLTP-системах на основі реалізації рангового підходу	135
<i>Росінський Д. М., Свірцевський К. О.</i> Оптимізація збору оперативних даних в піринговій мережі медичної установи	142
<i>Samadova S.</i> Algorithm for automatic recognition of cardiac arrhythmias	146
<i>Суліма С. В., Єрмолаєв О. Д.</i> Метод оптимізації SQL запитів системи управління базами даних	151
<i>Фомічов О. О., Бурцев В. С.</i> Дослідження імунних операторів в моделі штучної імунної мережі	158
<i>Чалий С. Ф., Лециньський В. О.</i> Оцінка чутливості пояснень в інтелектуальній інформаційній системі	165

ЦИВІЛЬНА БЕЗПЕКА

Вінтенко Б. Ю., Смірнов О. А., Коваленко О. В., Смірнов С. А., Коваленко А. С. Дослідження нормативних документів та галузевих стандартів розробки програмного забезпечення комп'ютерних систем управління АЕС, важливих для безпеки	170
Глива В. А., Тихенко О. М., Зозуля С. В., Козлітін О. О. Дослідження впливу електростатичних полів на концентрації аероіонів на комп'ютеризованих робочих місцях	179
Кочетов Г. М., Самченко Д. М., Ластівка О. В. Методологічні засади розроблення рідких сумішей для екранування електромагнітних полів	183
Ткаченко Т. М., Бурдейна Н. Б., Ченчева О. О. Екранування електромагнітних полів та шуму в будівлях і спорудах	186
Цопа В. А., Чеберячко С. І., Яворська О. О., Дерюгін О. В., Ченчева О. О., Сукач С. В., Резнік Д. В. Аналіз вимог рамкової стратегії ЄС до систем управління охороною здоров'я та безпекою праці	190

ЗВ'ЯЗОК, ТЕЛЕКОМУНІКАЦІЇ ТА РАДІОТЕХНІКА

Мордюк В. І., Синицький Ю. В. Методичні аспекти визначення основних параметрів сигнально-кодових конструкцій системи автоматизованого встановлення зв'язку у сеансах з кореспондентами	199
Peleshok Ye., Diedov M., Nikolaenko B. Incoherent demodulation of two synchronous mutually non-orthogonal digital signals with frequency shift keying	204
Сакович Л. М., Криховецький Г. Я., Курята Я. Е. Оцінка структурної надійності систем зв'язку	210

АЛФАВІТНИЙ ПОКАЖЧИК	216
---------------------------	-----

Організації авторів

Азербайджанський державний університет нафти та промисловості, Баку, Азербайджан
 Бакинський державний університет, Баку, Азербайджан
 Державний торговельно-економічний університет, Київ, Україна
 Державний НДІ випробувань і сертифікації озброєння та військової техніки, Черкаси, Україна
 Інститут проблем математичних машин і систем Національної академії наук України, Україна
 Інститут радіофізики та електроніки ім. О. Я. Усикова НАН України, Харків, Україна
 Казахський автомобільно-дорожній інститут імені Л. Б. Гончарова, Алмати, Казахстан
 Київський національний університет будівництва і архітектури, Київ, Україна
 Кременчуцький національний університет імені Михайла Остроградського, Кременчук, Україна
 Льотна академія Національного авіаційного університету, Кропивницький, Україна
 Науково-виробниче підприємство "Радій", Кропивницький, Україна
 Науково-дослідний інститут воєнної розвідки, Київ, Україна
 Національний авіаційний університет, Київ, Україна
 Національний технічний університет «Дніпровська політехніка», Дніпро, Україна
 Національний технічний університет «Харківський політехнічний інститут», Харків, Україна
 Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна
 Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна
 Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна
 Український державний університет залізничного транспорту, Харків, Україна
 Уманський національний університет садівництва, Умань, Україна
 Харківський національний університет радіоелектроніки, Харків, Україна
 Харківський національний автомобільно-дорожній університет, Харків, Україна
 Харківської національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна
 Центральноукраїнський національний технічний університет, Кропивницький, Україна

Authors affiliation

Azerbaijan State Oil and Industry University, Baku, Azerbaijan
 Baku State University, Baku, Azerbaijan
 State University of Trade and Economics, Kiev, Ukraine
 SR Institute of Testing and Certification of Weapons and Military Equipment, Cherkasy, Ukraine
 Institute of Mathematical Machine and Systems Problems of the NAS of Ukraine, Kiev, Ukraine
 Institute of Radiophysics and Electronics named after O. Ya. Usykov NAS of Ukraine, Kharkiv, Ukraine
 Kazakh Automobile and Road Institute named after L. B. Goncharov, Almaty, Kazakhstan
 Kyiv National University of Construction and Architecture, Kyiv, Ukraine
 Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk, Ukraine
 Flight Academy of the National Aviation University, Kropyvnytskyi, Ukraine
 «Radium» Scientific and Production Enterprise, Kropyvnytskyi, Ukraine
 Research Institute of Military Intelligence, Kyiv National Aviation University, Kyiv, Ukraine
 National Technical University "Dnipro Polytechnic", Dnipro, Ukraine
 National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine
 National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine
 National Aerospace University named after M.E. Zhukovsky "KHA1", Kharkiv, Ukraine
 National Technical University "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
 Ukrainian State University of Railway Transport, Kharkiv, Ukraine
 Uman National University of Horticulture, Ukraine
 Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
 Kharkiv National Automobile and Highway University, Kharkiv, Ukraine
 Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine
 Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

Р. Е. Пащенко¹, М. В. Марюшко²

¹ Інститут радіофізики та електроніки ім. О. Я. Усикова НАН України, Харків, Україна

² Національний аерокосмічний університет ім. М. С. Жуковського «ХАІ», Харків, Україна

МЕТОДИКА МОНІТОРИНГУ СІЛЬСЬКОГОСПОДАРСЬКИХ ЗЕМЕЛЬ І КУЛЬТУР З ВИКОРИСТАННЯМ ФРАКТАЛЬНОГО АНАЛІЗУ ДАНИХ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ ЗЕМЛІ

Анотація. Предметом дослідження є методика оцінки стану сільськогосподарських земель і культур з використанням фрактального аналізу космічних знімків супутника Sentinel-2, які є у вільному доступі у мережі Інтернет. **Об'єктом** дослідження є дані дистанційного зондування Землі (ДЗЗ) сільськогосподарських земель і культур для оцінки їх стану. **Метою** є розгляд методики моніторингу сільськогосподарських земель і культур з використанням онлайн ресурсів для отримання даних ДЗЗ, їх фрактального аналізу та представлення результатів аналізу за допомогою ГІС технологій. **Завдання:** аналіз різноманітних методів моніторингу стану сільськогосподарських земель і культур; визначення можливих шляхів отримання даних ДЗЗ, які є у вільному доступі у мережі Інтернет; розгляд можливості попередньої обробки космічних знімків за допомогою ГІС технологій; оцінка можливості використання фрактального аналізу для оцінки стану сільськогосподарських земель і культур; розробка методики моніторингу стану сільськогосподарських земель і культур з використанням фрактального аналізу космічних знімків супутника Sentinel-2. **Отримані наступні результати.** У результаті дослідження було розроблено методику моніторингу сільськогосподарських земель і культур, яка дозволяє отримувати дані ДЗЗ з онлайн ресурсів, здійснювати їх фрактальний аналіз та за результатами цього аналізу визначати їх стан. Наведено структурну схему методики, яка містить чотири етапи. Розглянуто зміст кожного з етапів та наведено основні операції, які виконуються на цих етапах. Показано, що застосування у методиці сучасних ГІС технологій дозволяє комплексно оцінювати стан земель і культур та прогнозувати рівень врожаю з використанням, як відомих вегетаційних індексів, так і нових методів фрактального аналізу космічних знімків супутників ДЗЗ. **Висновки.** Описана в даній роботі методика моніторингу дозволяє комплексно оцінювати стан сільськогосподарських земель і культур та прогнозувати рівень врожаю.

Ключові слова: дані дистанційного зондування Землі, моніторинг сільськогосподарських культур і земель, індекс NDVI, фрактальна розмірність, фрактальний аналіз.

Вступ

Як свідчить світовий досвід, дані дистанційного зондування Землі (ДЗЗ) все частіше використовують для рішення різноманітних завдань під час проведення сільськогосподарських робіт [1]. Зокрема основними з них є: картографування рослин; моніторинг сільськогосподарських культур і земель; прогнозування врожайності тощо [2]. Широке застосування даних ДЗЗ обумовлено перш за все оперативністю та наочністю під час обробки даних, що отримуються з великих територій, та в умовах швидкої зміни вегетаційного стану сільськогосподарських культур. Сучасні супутники ДЗЗ (Terra, Aqua, Landsat 8, 9 Sentinel-2) дозволяють отримувати космічні знімки територій, що досліджуються, з періодичністю від 1 до 8 діб та з просторовим розрізненням від 250 до 10 метрів [3]. Також сучасні методи обробки даних ДЗЗ дозволяють створювати наочні картографічні матеріали, що сприяє ефективному веденню агропромисловості [4]. Крім того, останнім часом спостерігається тенденція зростання доступності до даних дистанційного зондування.

Всі ці чинники є дуже важливими для організації моніторингу сільськогосподарських земель, але у більшій мірі достовірність одержуваних результатів залежить від способу обробки космічних знімків, а користувачів цікавить кінцевий результат та його представлення для прийняття управлінських рішень.

Основні методи моніторингу стану сільськогосподарських земель і культур розподіляються на [5]:

- статистичні методи на основі оцінок поточного стану, які використовують лінійну регресію для моделювання залежності врожайності від спектра поглинання рослинного покриву;

- динамічні методи на основі моделей росту, за допомогою яких досліджують зростання рослин під впливом метеорологічних параметрів, характеристик ґрунту і складу добрив, що вносяться;

- статистичні методи на основі оцінки фізичних факторів навколишнього середовища, які спрямовані на виявлення залежності між характеристиками довкілля і врожайністю із застосуванням методів лінійного і нелінійного регресійного аналізу та нейронних мереж.

Велика кількість методів моніторингу сільськогосподарських земель і культур також базується на використанні вегетаційних індексів (існує близько 160 варіантів таких індексів) [6], які визначаються емпірично і враховують фізичні властивості світла, відбиваючи здатність рослин та інші важливі характеристики. Серед всіх існуючих індексів найбільше розповсюдження отримав нормалізований різнице-вий вегетаційний індекс (індекс NDVI).

Але, під час визначення індексу NDVI необхідні дані двох спектральних каналів (ближнього інфрачервоного та червоного), що призводить до ускладнення обладнання для отримання вихідного зобра-

ження. А також, індекс NDVI у значній мірі залежить від кількості активної біомаси та стає неефективним в періоди, коли рослинність є послабленою або спостерігається низький рівень вегетації.

Перспективним напрямком підвищення інформативності космічних знімків, які отримуються тільки в одному діапазоні хвиль, є застосування методів фрактального аналізу зображень [7]. Різна складність поверхонь та рослинності приводить до різних форм цифрового зображення. Кількісно оцінити складність цих форм можна за допомогою величини фрактальної розмірності (ФР). У роботі [8] на основі даних дистанційного зондування та використання ФР показана можливість розподілу різних типів ґрунтів. У зв'язку з цим представляє теоретичний і практичний інтерес провести аналіз величин ФР космічних знімків під час зміни стану сільськогосподарських земель та застосування ФР під час прогнозування рівня врожаю. На відміну від використання вегетаційних індексів, коли необхідно застосовувати дані декількох спектральних каналів, під час фрактального аналізу достатньо одного каналу.

Метою статті є розгляд методики моніторингу сільськогосподарських земель і культур з використанням онлайн ресурсів для отримання даних ДЗЗ, їх фрактального аналізу та представлення результатів аналізу за допомогою ГІС технологій.

Методика моніторингу сільськогосподарських земель і культур з використанням фрактального аналізу

Для оцінки стану сільськогосподарських земель і культур розроблено методику з використанням фрактального аналізу космічних знімків супутника Sentinel-2, які є у вільному доступі у мережі Інтернет. Спрощену структурну схему методики наведено на рис. 1.

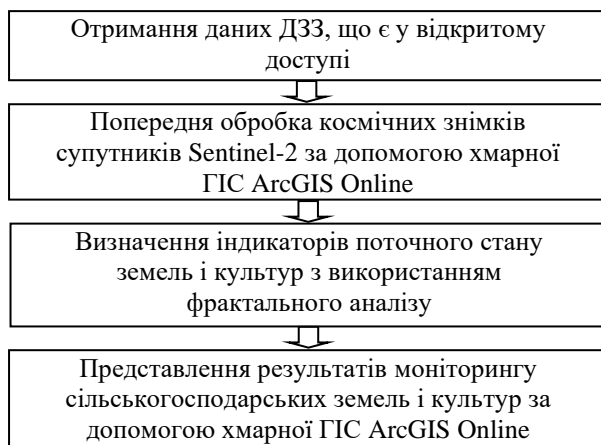


Рис. 1. Спрощена структурна схема методики моніторингу сільськогосподарських земель і культур з використанням фрактального аналізу

Як видно на рис. 1, методика містить декілька етапів. Розглянемо більш докладно зміст кожного з цих етапів. На першому етапі методики отримуються дані ДЗЗ, що є у відкритому доступі.

Спрощену структурну схему першого етапу наведено на рис. 2.

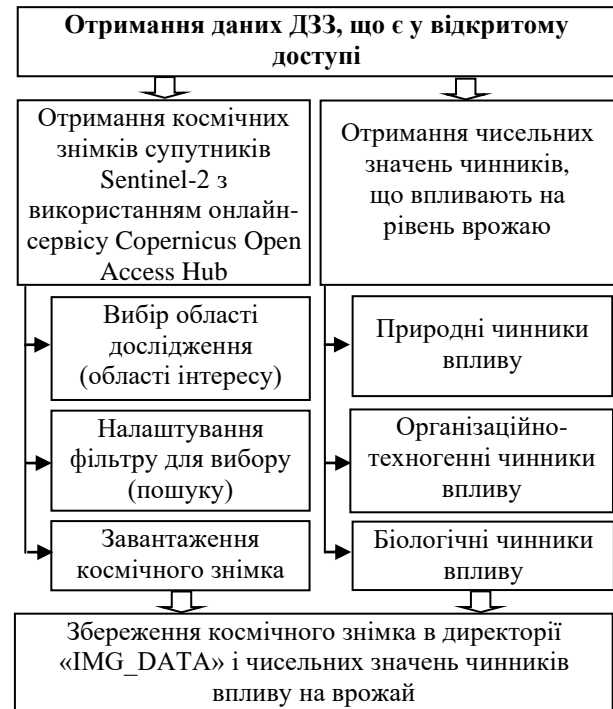


Рис. 2. Спрощена структурна схема першого етапу методики

Доступ до космічних знімків супутників Sentinel-2 здійснюється з використанням онлайн-сервісу Copernicus Open Access Hub, для чого переходять на сайт The European Space Agency “Copernicus Open Access Hub” [9] та виконують реєстрацію для нового користувача або авторизацію для вже зареєстрованого користувача. Інтерфейс сервісу Copernicus Open Access Hub наведено на рис. 3.

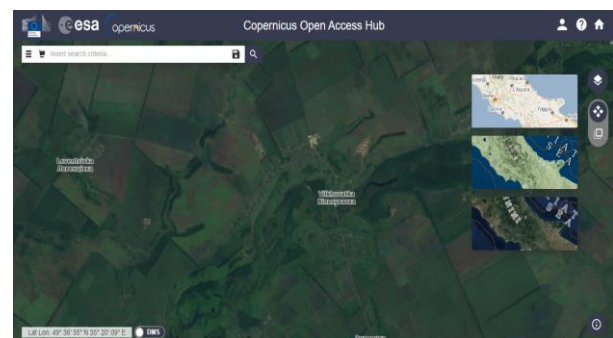


Рис. 3. Інтерфейс сервісу Copernicus Open Access Hub

Після успішної реєстрації/авторизації здійснюється вибір області дослідження (області інтересу). Для цього використовуються можливості вбудованого електронного атласу. Для прикладу, на рис. 4 показана карта на якій є область інтересу – землі Вільхуватської сільської ради Чутівського району Полтавської області.

На наступному кроці роботи користувач у ручному режимі приблизно вибирає (позначає прямокутником) область інтересу (дослідження), як це показано на рис. 5.

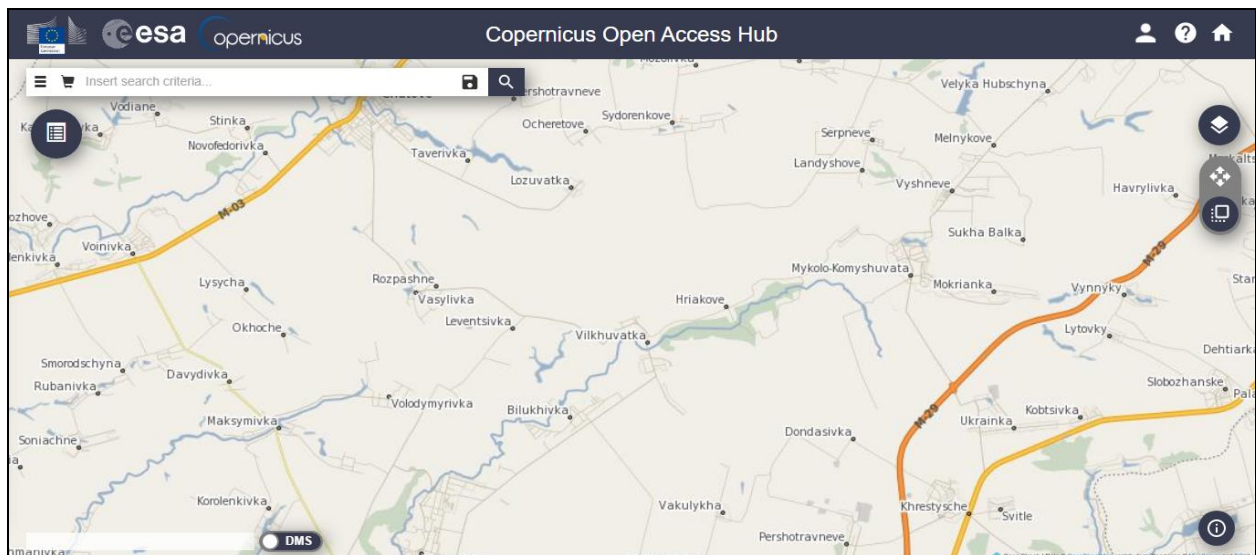


Рис. 4. Карта з областю інтересу (дослідження)

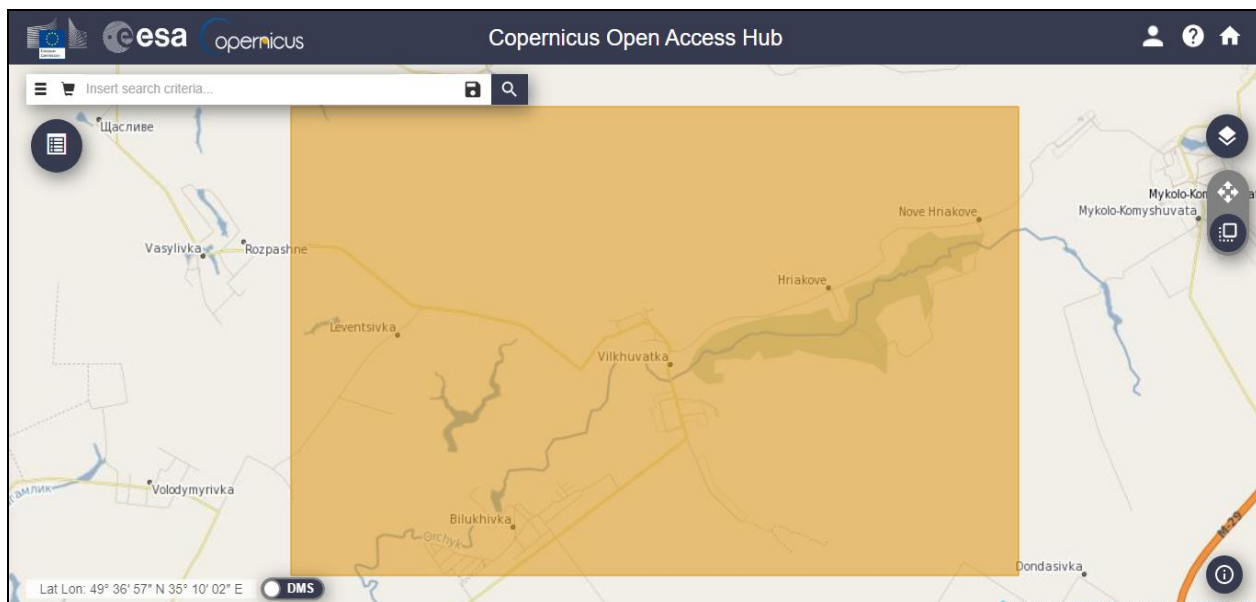


Рис. 5. Вибір області інтересу (дослідження)

Після вибору області інтересу користувач налаштовує фільтр для вибору (пошуку) необхідної колекції даних. Спочатку встановлюється тип сортування («Sort By: Sensing day»), за ним порядок сортування («Order By: Ascending»), а потім календарна дата початку періоду зондування («Sensing period: 2021/03/01»). Наступним кроком є вибір місії спостереження. Для цього встановлюється «галочка» напроти «Mission: Sentinel-2» та розпочинається пошук необхідної колекції даних. За результатами пошуку (обробки запиту) на сервері, користувачу надається поле з усіма можливими результатами пошуку відповідно до початкових налаштувань фільтру, як це показано на рис. 6.

Як видно на рис. 6, у вікні з результатами пошуку, можна отримати інформацію про кількість знайдених сцен та переглянути інформацію про сцену, яка містить дані про:

- тип апарату, з якого проводилося знімання – Mission Sentinel-2;

- апаратуру знімання – Instrument MSI;
- дату та час проведення знімання – Sensing Date 2021-03-10T08:48:01.024Z;
- розмір продукту – Size 762.19 MB;
- посилання для завантаження.

На наступному кроці роботи можна скористатися розширеною інформацією про обрану сцену – «View Product Details», інтерфейс вікна з цією інформацією наведено на рис. 7.

Розширена інформація про обрану сцену забезпечує отримання більш повного набору даних та попередню візуальну оцінку даних, що вибираються. Після ознайомлення зі всіма необхідними характеристиками сцени, що вибирається, її можна додати до кошику («Add Product to Cart»), в якій обрані дані будуть зберігатися для подальших маніпуляцій з ними. Проаналізувавши всі результати пошуку, користувач переходить до завантаження космічного знімка. Для цього відкривається корзина, вибирається необхідна сцена та натискається кнопка «Download product».



Рис. 6. Результати пошуку необхідної колекції даних

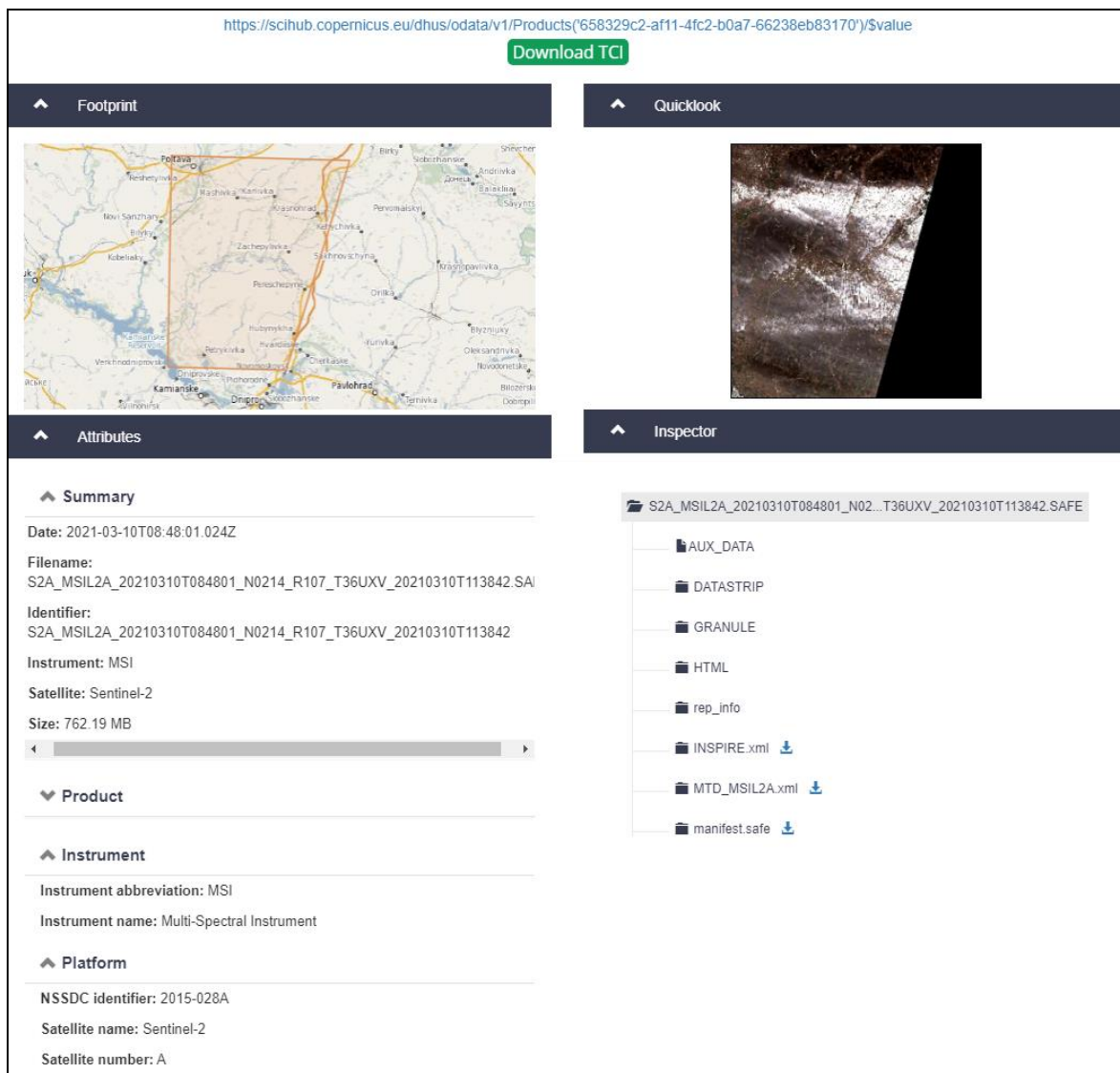


Рис. 7. Інтерфейс вікна «View Product Details»

Після цього вибирається необхідний каталог в якому буде зберігатися космічний знімок. За результатом роботи (завантаження) створюється архів з розширенням «*.zip». Отриманий архів містить різноманітні метадані про завантажену сцену та безпосередньо сам знімок. Знімок знаходиться в директорії «IMG_DATA». Директорія «IMG_DATA» містить 3 підкаталоги «R10m», «R20m» та «R60m» від-

повідно, що вказують на просторове розрізнення знімка.

В результаті проведених дій користувач отримує космічний знімок супутника Sentinel-2. На рис. 8 для прикладу наведено космічний знімок каналу b8 супутника Sentinel-2, на якому показані землі Вільхуватської сільської ради Чутівського району Полтавської області.



Рис. 8. Землі Вільхуватської сільської ради станом на 16.06.2018 в ближньому інфрачервоному діапазоні (канал b8)

Як видно на рис. 8, за допомогою космічного знімку можна виконати візуальний розподіл земель сільськогосподарського призначення, визначити їх межі, а також виділити елементи природного ландшафту (річки, яри та ін.) і штучні об'єкти, створені людьми (дороги, будинки та ін.). Але візуально неможливо визначити тип сільськогосподарських культур та стан їх вегетації. Для вирішення цих задач необхідна попередня (контактна) інформація про тип вирощуваних культур на конкретній ділянці та подальша обробка космічних знімків у різні часові інтервали. Для отримання чисельних значень чинників, що впливають на рівень врожаю, користувач використовує інформацію, яка є у мережі Інтернет, а також, яка є у сільгоспвиробників. Так інформацію про природно-кліматичні чинники (інтенсивність світла, температура навколишнього середовища, концентрація вуглекислого газу в атмосфері, вологість повітря, природна родючість ґрунту і т. д.) можна отримати у мережі Інтернет, а про біологічні (органічні добрива, сорт насіння, гібриди і т. д.) та організаційно-техногенні чинники (обробка ґрунту, терміни і способи посіву та посадки, мінеральні добрива, меліорація, засоби захисту рослин і т. д.) у сільгоспідприємстві.

На другому етапі методики проводиться попередня обробка космічних знімків супутників Sentinel-2 за допомогою хмарної геоінформаційної системи (ГІС) ArcGIS Online. Спрощену структурну схему другого етапу наведено на рис. 9.

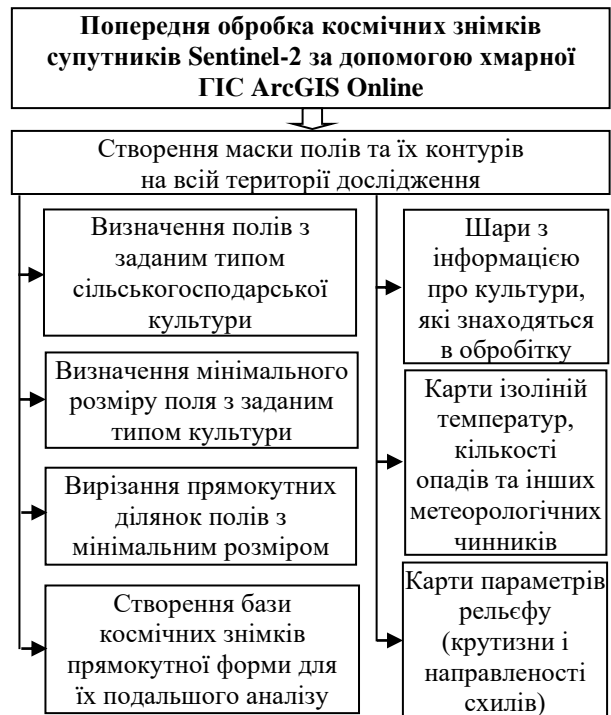


Рис. 9. Спрощена структурна схема другого етапу методики

Портал ArcGIS Online [10] представляє собою готову хмарну ГІС, в якій можна зберігати і публікувати свої просторові дані, карти, інструменти і сервіси, а також обмінюватися і керувати ними. Основні складові частини ArcGIS Online представлено на рис. 10.



Рис. 10. Хмарна інфраструктура ArcGIS Online

Для ефективного моніторингу стану сільськогосподарських земель і культур користувач створює низку картографічних продуктів: векторні шари з інформацією про культури які знаходяться в обробітку; індексні карти стану сільськогосподарських культур на протязі всього періоду вегетації; карти ізольованих температур, кількості опадів та інших метеорологічних чинників; карти параметрів рельєфу – крутизни і направленості схилів. Для цього користувач здійснює реєстрацію на сайті ArcGIS

Online та створює перераховані вище картографічні продукти, використовуючи існуючі дані ДЗЗ, статистичні дані з метеостанцій, картограми дослідження ґрунтів та ін. На рис. 11 наведено карту з маскою полів та їх контурів на всій території дослідження. Кінцевий вигляд карти залежить тільки від кількості доданих шарів, їх налаштування, порядку відображення, а також від вимог до інформації, що необхідна користувачу під час моніторингу стану сільськогосподарських земель і культур.

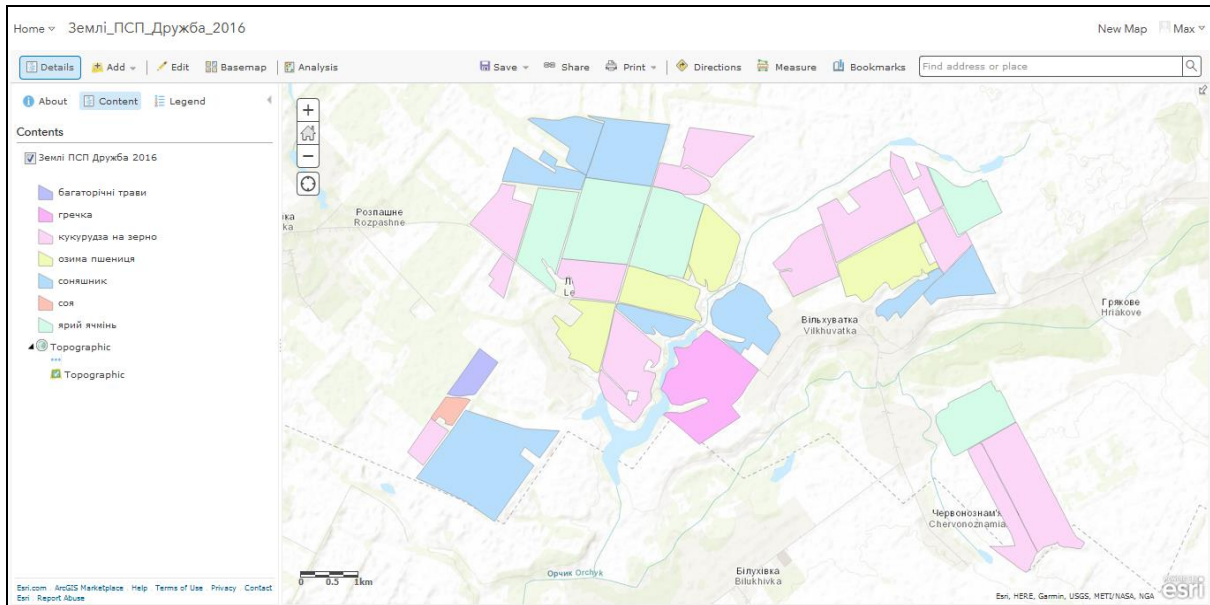


Рис. 11. Веб-карта території приватного сільськогосподарського підприємства (ПСП) «Дружба» Чутівського району Полтавської області

Аналізуючи створену карту, користувач визначає поля з заданим типом сільськогосподарської культури, наприклад, поля засіяні кукурудзою. У разі необхідності можна позначити всі типи культур на полях. Далі користувач визначає мінімальний розмір поля з заданим типом культури. Це необхідно, щоб отримати однакові за розміром «вирізки» прямокутних космічних знімків і забезпечити однакові умови їх аналізу. Прямокутні космічні знімки потрібні у зв'язку з тим, що вони у подальшому будуть оброблятися з використанням прямокутного «ковзного» або «стрибаючого» «вікна». Також формат знімків у разі потреби також може бути змінений, наприклад, з формату *.jp2 на формат *.bmp. Після вирізання прямокутних ділянок полів з мінімальним розміром, створюється база космічних знімків прямокутної форми з заданим типом культури на них, наприклад, база космічних знімків полів засіяних кукурудзою. Ці знімки можуть зберігатися в окремих папці з відповідним позначенням, наприклад, «Кукурудза», а назви файлів можуть містити інформацію про номер поля, спектральний канал знімку та його дату, наприклад, «r1_b8_2018.06.16». Такі бази космічних знімків можуть створюватися і періодично поповнюватися для всіх типів культур, що вирощуються у сільгоспдприємстві. Приклади прямокутних космічних знімків, поля, позначеного на рис. 8 колом, наведено на рис. 12.

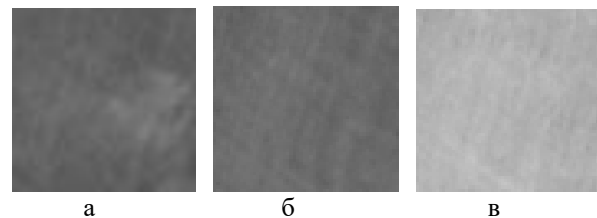


Рис. 12. Елементи космічних знімків (канал b8) з полем, засіяним кукурудзою, станом на: 16.06.2018 (а); 09.07.2018 (б); 29.07.2018 (в)

Як видно на рис. 12 невелика зміна градацій сірого кольору (від темно-сірого до світло-сірого) дозволяє зробити висновок, що за час спостереження пройшли якісь зміни досліджуваної культури.

Але здійснити кількісний аналіз і простежити зміни в структурі посівів на полях неможливо. Це обумовлено тим, що обробка знімку тільки каналу b8 не дозволяє розраховувати відомі числові показники аналізу підстильної поверхні та здійснювати за їх допомогою інтерпретацію даних.

На третьому рівні користувач розраховує індекси NDVI [11] та поля фрактальних розмірностей [12] космічних знімків супутників Sentinel-2, на яких зображені поля з заданим типом культури.

Спрощену структурну схему третього етапу наведено на рис. 13.

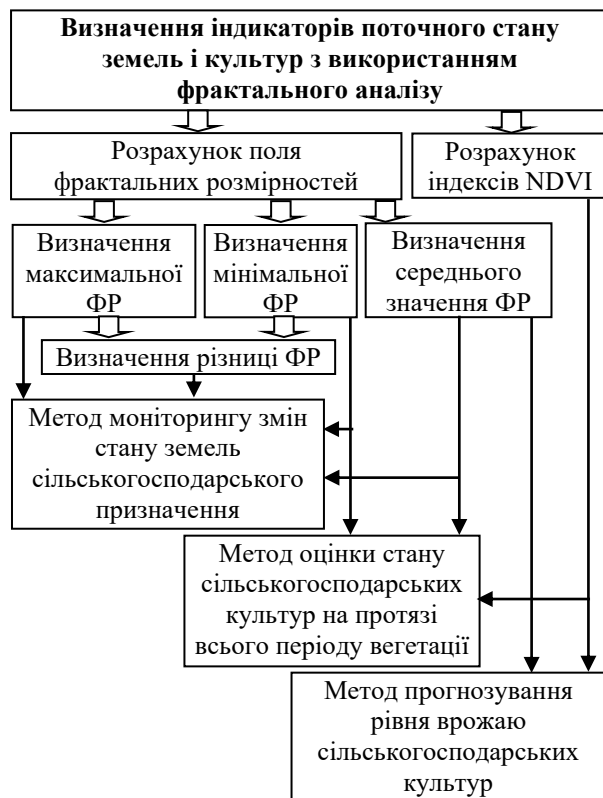


Рис. 13. Спрощена структурна схема третього етапу методики

Для визначення індексів NDVI може використовуватися хмарна ГІС ArcGIS Online, а для розрахунку поля фрактальних розмірностей (ПФР) космічний знімок обробляється з використанням «ковзального вікна» або «стрибаючого вікна» і для кожного положення «вікна» визначається чисельне значення фрактальної розмірності, що записується в окрему матрицю, яка називається ПФР.

За результатами розрахунку ПФР користувач також визначає та аналізує середні, максимальні і мінімальні фрактальні розмірності та їх різницю.

Індекси NDVI та середні фрактальні розмірності у подальшому використовуються як індикатори поточного стану посівів у методі прогнозування рівня врожаю сільськогосподарських культур. Суть цього методу докладно розглянуто у роботі [13]. Отримані середні, максимальні і мінімальні фрактальні розмірності та їх різниці використовуються у методі оцінки стану сільськогосподарських культур на протязі всього періоду вегетації та методі моніторингу змін стану земель сільськогосподарського призначення. Суть цих методів докладно розглянуто у роботах [12] і [14], відповідно.

На четвертому етапі методики оформлюються результати моніторингу сільськогосподарських земель і ці результати публікуються у хмарній ГІС ArcGIS Online. Спрощену структурну схему четвертого етапу наведено на рис. 14. Створюються карти стану земель сільськогосподарського призначення та вирощуваних культур на протязі всього періоду вегетації. При цьому використовується пошарова організація збереження даних у вигляді геореляційної моделі. Такій підхід забезпечує збереження однорідної просторової та ат-

рибутивної інформації про стан земель і культур в окремих таблицях реляційної бази даних.



Рис. 14. Спрощена структурна схема четвертого етапу методики

Далі проводиться візуалізація ПФР з подальшим його аналізом. Під час візуалізації ПФР вибирається кількість рівнів розбиття діапазону змін величин фрактальних розмірностей на ПФР. Кожен з рівнів позначається певним кольором або градацією сірого. Також може будуватися гістограма ПФР і з її використанням створюватися селективні зображення (зображення з позначеними діапазонами змін ФР).

Результати прогнозу рівня врожаю для кожного поля можуть також надаватися у вигляді таблиць реляційної бази даних. Карти ізольованих температур, кількості опадів та інших метеорологічних чинників і карти параметрів рельєфу можуть бути використані для прийняття інших управлінських рішень різними фахівцями сільськогосподарського підприємства. У разі необхідності візуалізації динаміки просторових процесів може створюватися набір карт (динамічний ряд карт) з заданою частотою дискретизації.

Завдяки інтеграції ArcGIS Online з настільними продуктами, готові картографічні продукти публікуються у вигляді сервісів даних за допомогою ArcGIS for Desktop. Першим кроком при публікації даних у вигляді сервісу здійснюється заповнення певних пунктів документу карти (Map Document Properties): назва (title); резюме (summary); докладний опис (description); автор (author); мітки (tags). На рис. 15 наведено приклад заповнення документу карти векторного шару для приватного сільськогосподарського підприємства (ПСП) «Дружба».

Необхідно зазначити, що перераховані вище пункти є обов'язковими для заповнення, тому що містять важливу описову інформацію для публікації даних. Після збереження опису карти проводиться аналіз даних на наявність помилок з використанням функції «Analyze Map». Якщо помилки існують необхідно здійснити їх виправлення на всіх рівнях, після чого здійснюється перехід до публікації даних.

Для публікації даних користувач виконує авторизацію, після чого отримує необхідний рівень доступу, який дозволяє виконувати публікацію даних і здійснити перехід до діалогового вікна «Publish a Service», яке наведено на рис. 16.

Як видно на рис. 16 для публікації даних встановлюється необхідне підключення і відповідно до правил надається назва сервісу. На завершальному

етапі, при публікації сервісу, здійснюється перехід до редактору параметрів сервісу даних – Service Editor, який наведено на рис. 17.

Рис. 15. Приклад заповнення документу карти векторного шару для ПСП «Дружба»

Рис. 16. Діалогове вікно Publish a Service

Рис. 17. Вікно редактору параметрів сервісу даних – Service Editor

За допомогою цього редактору проводяться завершальні налаштування даних, які будуть опубліковані. Також необхідно повторно виконати аналіз даних на наявність помилок, які будуть опубліковані. Завершається публікація даних діалоговим вікном, яке сповіщає про успішну публікацію сервісу на порталі. В разі невдачі користувачу буде сповіщено причини, які необхідно усунути і спробувати опублікувати дані знову.

Після публікації дані опиняються на порталі ArcGIS Online. Виконавши авторизацію, користувач у вкладці «Мої ресурси» може переглянути директорії з опублікованими даними. Інтерфейс вкладки «Мої ресурси» з опублікованими даними наведено на рис. 18.

Як видно на рис. 18, опубліковані дані представлені у вигляді двох файлів: індивідуального шару та опису сервісу. Файл опису індивідуального шару міс-

тять оглядову інформацію про веб-карт з інформацією про останні зміни, використані шари та ін. Також можливий перегляд регулярності використання даної карти в залежності від встановленого періоду. Налаштування дозволяють захистити карту від випадкового видалення, налаштувати екстент та ін. На рис. 19 представлено інформацію про опублікований індивідуальний векторний шар для ПСП «Дружба».

штування дозволяють захистити карту від випадкового видалення, налаштувати екстент та ін. На рис. 19 представлено інформацію про опублікований індивідуальний векторний шар для ПСП «Дружба».

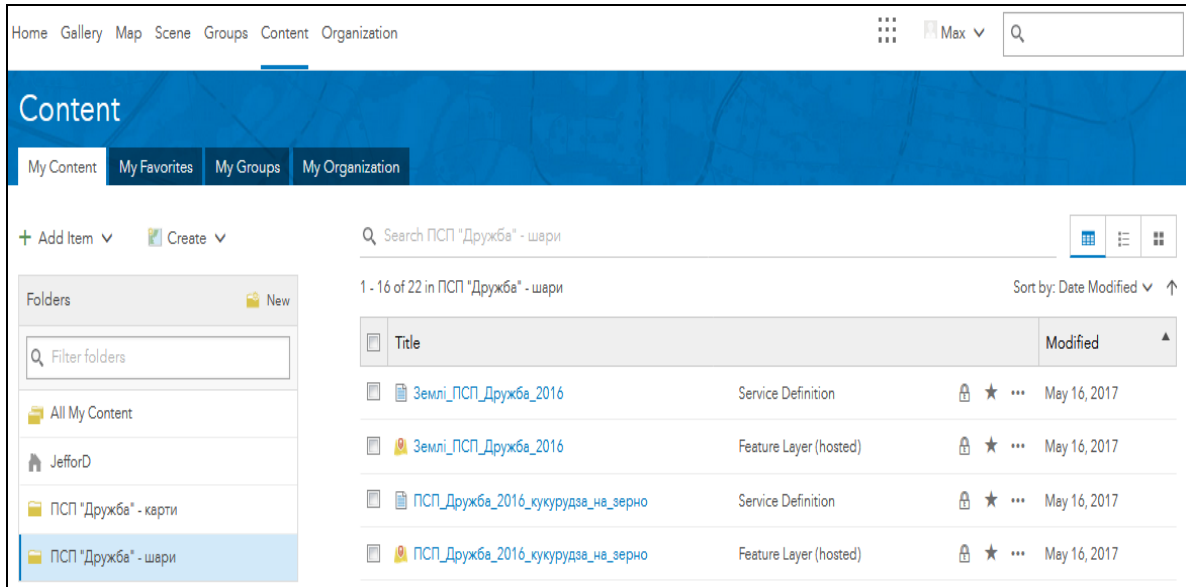


Рис. 18. Інтерфейс вкладки «Мої ресурси» з опублікованими даними

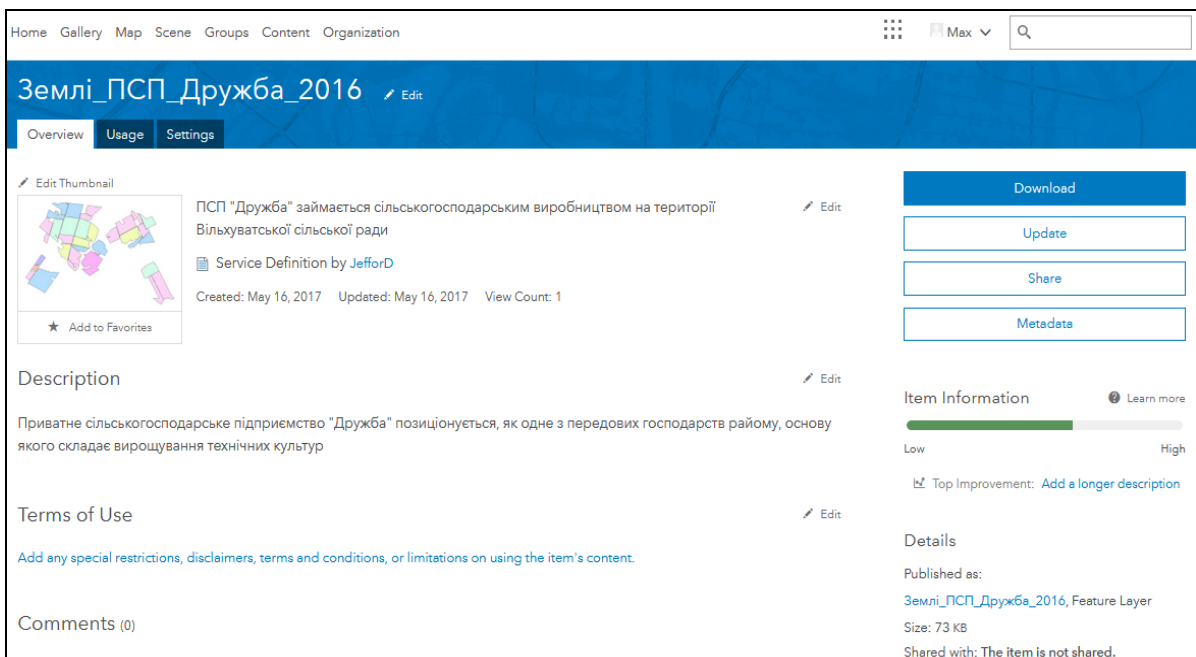


Рис. 19. Інформація про векторний шар ПСП «Дружба»

Кінцевий вигляд карти, яка публікується, залежить тільки від кількості доданих шарів, їх налаштування, порядку відображення, а також від вимог до інформації, що необхідна користувачу під час моніторингу стану сільськогосподарських земель і культур.

Висновки

Запропоновано методику моніторингу сільськогосподарських земель і культур, яка відрізняється від відомих використанням онлайн ресурсів для отримання даних ДЗЗ, їх фрактального аналізу та

представлення результатів аналізу за допомогою ГІС технологій.

Під час розробки методики була проведена оцінка характеристик супутників Sentinel-2 для моніторингу сільськогосподарських земель та проаналізовані основні методи оцінки стану сільськогосподарських земель з використання космічних знімків, а також визначені основні онлайн-ресурси, які представляють дані ДЗЗ в мережі Інтернет у вільному доступі, і показано можливість застосування фрактального аналізу для моніторингу сільськогосподарських земель і культур.

Застосування у методиці сучасних ГІС технологій дозволяє комплексно оцінювати стан земель та прогнозувати рівень врожаю з використанням, як відомих вегетаційних індексів, так і нових методів фрактально-го аналізу космічних знімків супутників ДЗЗ.

Під час подальших досліджень доцільно розглянути особливості застосування розробленої методики під час моніторингу конкретних сільськогосподарських земель і культур, а також використання космічних знімків різних супутників ДЗЗ.

СПИСОК ЛІТЕРАТУРИ

1. Remote Sensing and Cropping Practices: A Review [Text] / [Bégué A., Arvor D., Bellon B., Betbeder J. et al.] // Remote Sensing. – 2018. – Vol. 10, Iss. 1. – P. 1-32.
2. Sentinel-2 for agriculture national demonstration in Ukraine: Results and further steps [Text] / [Kussul N., Kolotii A., Shelestov A., Lavreniuk M. et al.] // IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23-28 July 2017. – IEEE, 2017. – P. 5842-5845.
3. Дрони і супутники: моніторинг стану посівів впродовж сезону [Електронний ресурс]. – Режим доступу: <https://www.smartfarming.ua/drony-i-suputnyku-monitorynh-stanu-posiviv-vprodovzh-sezonu/> – 19.12.2022 р.
4. Khaliq A. Land cover and crop classification using multitemporal Sentinel-2 images based on crops phenological cycle [Text] / Khaliq A., Peroni L., Chiaberge M. // IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS), Salerno, Italy, 21-22 June 2018. – IEEE, 2018. – P. 1-5.
5. Бальченко І.В. Сучасні методи управління станом аграрної екосистеми [Текст] / Бальченко І.В., Казимир В.В., Клименко В.П. // Вісник Чернігівського державного технологічного університету. Серія «Технічні науки» – 2012. – №3 (59). – С. 213-220.
6. Черепанов, А.С. Вегетационные индексы [Текст] / А.С. Черепанов // ГЕОМАТИКА. – 2011. – № 2 (11). – С. 98–102.
7. Фрактальный анализ процессов, структур и сигналов / коллективная монография [Текст] / [Доля Г.Н., Иванов В.К., Пашенко Р.Э., Стадник А.М. и др.]; под ред. Р.Э. Пашенко. – Харьков: ХООО НЭО «ЭкоПерспектива», 2006. – 348 с.
8. Распознавание типов почв с помощью фрактальных методов обработки космических изображений [Текст] / Пашенко Р.Э., Яцевич С.Е., Яцевич Е.И., Егорова Л.А. // Збірник наукових праць «Системи обробки інформації». – Х.: ХУПС. – 2014. – Вип. 6(122). – С. 74-76.
9. The European Space Agency [Електронний ресурс] / Copernicus Open Access Hub. – Режим доступу: <https://scihub.copernicus.eu/dhus/#/home> – 25.03.2021 р.
10. What is ArcGIS Online? [Електронний ресурс] / ArcGIS Online Help. – Режим доступу: <https://doc.arcgis.com/en/arcgis-online/reference/what-is-ago1.htm> – 25.03.2021 р.
11. Марюшко, М.В. Моніторинг сільськогосподарських культур із застосуванням космічних знімків SENTINEL-2 [Текст] / Марюшко М.В., Пашенко Р.Е., Коблюк Н.С. // Радіоелектронні і комп'ютерні системи. – 2019. – №1(89). – С. 99 – 108.
12. Марюшко, М.В. Фрактальний аналіз космічних знімків SENTINEL-2 для моніторингу сільськогосподарських культур [Текст] / Марюшко М.В., Пашенко Р.Е. // Радіоелектронні і комп'ютерні системи. – 2020. Вип. 4(96). С. 34-47.
13. Марюшко М.В. Прогнозування рівня врожаю сільськогосподарських культур із застосуванням нечіткого опису чинників впливу та індикаторів стану посівів [Текст] / Марюшко М.В., Пашенко Р.Е. // Системи управління, навігації та зв'язку. – К.: ЦНДІ НіУ. – 2022. – Вип. 3(69). – С. 78-
14. Марюшко М.В. Моніторинг змін стану сільськогосподарських земель за даними фрактального аналізу космічних знімків [Текст] / Марюшко М.В., Пашенко Р.Е. // Системи управління, навігації та зв'язку. – К.: ЦНДІ НіУ. – 2021. – Вип. 3(65). – С. 8-17.

Received (Надійшла) 21.02.2023

Accepted for publication (Прийнята до друку) 26.04.2023

Method monitoring of agricultural earths and cultures with the use fractal analysis of earth remote sensing data

R. Pashchenko, M. Mariushko

Abstract. The **subject** of the study in the article is the method estimation of agricultural earths and cultures state with the use fractal analysis of satellite Sentinel-2 spaces pictures, which are in free access in a network the Internet. The **object** of the study is the Earth remote sensing (ERS) date of agricultural earths and cultures for estimation of their state. The **goal** is consideration method monitoring of agricultural earths and cultures with the use on-line resources for the receipt of ERS data, their fractal analysis and presentation results of analysis by GIS technologies. The **tasks**: analysis various methods monitoring of agricultural earths and cultures state; determination possible ways receipt of ERS date, which are in free access in a network the Internet; consideration possibility rough-down of spaces pictures with the help GIS technologies; estimation possibility use fractal analysis for estimation of agricultural earths and cultures state; development method monitoring of agricultural earths and cultures state with use fractal analysis of satellite Sentinel-2 spaces pictures. **The following results were obtained.** As a result of research the method monitoring of agricultural earths and cultures, which allows to get ERS data with on-line resources, was developed, to carry out their fractal analysis and on by the results of this analysis to determine their state. Show the flow diagram of method, which contains four stages. Maintenance each stages is considered and basic operations, which are executed on these stages, are resulted. Show that application in the method modern GIS technologies allows complex to estimate earths and cultures state and to forecast the level of harvest with the use, both known vegetation indexes and new methods fractal analysis of satellites ERS spaces pictures. **Conclusions.** The method monitoring of prognostication described in this work allows complex to estimate of agricultural earths and cultures state and to forecast the level of harvest.

Keywords: Earth remote sensing data, monitoring of agricultural cultures and earths, index NDVI, fractal dimension, fractal analysis.

Автомобільний, річковий, морський та авіаційний транспорт

УДК 656.051

doi: 10.26906/SUNZ.2023.2.015

П. Ф. Горбачов, Є. В. Любий

Харківський національний автомобільно-дорожній університет, Харків, Україна

ПІДХІД ДО ОЦІНЮВАННЯ ЗАТРИМОК ТРАНСПОРТНИХ ПОТОКІВ У ПЛАНІ КООРДИНАЦІЇ

Анотація. У статті розроблені аналітичні моделі, що можна використовувати для оцінювання затримок транспортних засобів, які рухаються скоординованими ділянками міських магістралей в умовах низького трафіку та при їхньому груповому прибутті до чергового перехрестя в плані координації. Результати визначення затримок за представленими моделями доцільно використовувати при порівнянні альтернативних планів координації, формуванні методик їхнього створення, а також для пошуку дієвих і надійних інструментів для підвищення ефективності створюваних планів координації. В подальшому розроблені моделі потребують уточнення за рахунок розгляду інших можливих варіантів організації зсуву включення дозвольного сигналу для лівоповоротної фази в рамках плану координації.

Ключові слова: групове прибуття, зелена хвиля, пачка автомобілів, план координації, транспортні затримки.

Вступ

На стадії створення плану координації (ПК) світлофорного регулювання (СР) на міській магістралі (ММ) його ефективність визначається насамперед скороченням витрат часу учасників руху на очікування ними можливості проїзду чергового перехрестя зі СР. Оцінка таких витрат, які зазвичай просто називаються затримками, завжди базується на характеристиках вхідного потоку [1, 2]. Нині цей потік на підходах до ізольованих перехресть здебільшого приймають за Пуасонівський, або, у найпростішому варіанті, – за простий регулярний потік із рівними інтервалами прибуття між усіма учасниками руху. Другий варіант для ізольованих перехресть не заслуговує на пильну увагу через відсутність реальних транспортних об'єктів, які можуть йому відповідати, адже навіть для обмеженого в часі періодами високого навантаження колонного руху очевидним є той факт, що кожен водій самостійно обирає дистанцію безпеки, що порушує регулярність потоку автомобілів, які прибувають до перехрестя. Пуасонівський же, найпростіший потік добре підходить для опису випадкового прибуття транспортних засобів (ТЗ) до перехрестя, але і він у разі магістральної координації не відповідатиме реальній ситуації, оскільки вплив світлофорних циклів (СЦ) на характер прибуття автомобілів до перехрестя тут занадто великий.

Аналіз досліджень і публікацій. В існуючих методиках розроблення планів магістральної координації використовуються два основні підходи до опису групового прибуття автомобілів до скоординованих перехресть [3]. Найрозвиненішим можна вважати підхід циклічних профілів потоку Робертсона [4, 5], який в явному вигляді описує групи автомобілів, що прибувають до наступного перехрестя в ПК. Однак його реалізація в програмах TRANSYT і SCOOT не змогла забезпечити більш високих результатів координації порівняно з аналогічними ме-

тодами, в яких використовується простіший опис групового прибуття у вигляді стрічки руху автомобілів. Такий опис реалізовано в просторово-часовій діаграмі (ПЧД) [6] та американських програмах координації MAXBAND, MULTIBAND та ін. [7, 8]. Загальною ж причиною недостатньо високих результатів координації роботи світлофорів на ММ є використання узагальнених оптимізаційних підходів до розв'язання цієї задачі, без детального розгляду її окремих аспектів і, зокрема, у відсутності аналітичного опису втрат часу автомобілістами на очікування можливості проїзду чергового перехрестя в ПК.

Взагалі кажучи, у плані магістральної координації роботи світлофорів має враховуватися наявність не одного, а двох варіантів прибуття до регульованого перехрестя. Перший варіант стосується координованих автомобілів, які під'їжджають до чергового, але не першого перехрестя ділянки, що координується [9]. А другий – до всіх інших транспортних потоків (ТП) на перехрестях координованої ділянки ММ. До некоординованих належать потоки автомобілів на другорядних підходах до магістралі та на підході до першого перехрестя координованої ділянки магістралі, для яких параметри потоку, що входить на перехрестя, визначаються умовами руху на відповідній ділянці вулично-дорожньої мережі (ВДМ). При цьому відсутність координації з іншими світлофорними об'єктами зумовлює те, що найбільш підходящим тут видається припущення про доцільність розгляду прибуття ТЗ з цих напрямків до координованих перехресть як випадкового. Тоді для його опису добре підходить Пуасонівський найпростіший потік, а витрати часу учасників руху на подолання регульованого перехрестя можуть розраховуватися за відомими формулами затримки, заснованими на залежності Вебстера [10].

Постановка завдання. Для учасників руху, що пересуваються безпосередньо координованою магістраллю, час прибуття до чергового перехрестя багато

в чому визначається параметрами ПК, а випадковість цього процесу, яка зумовлена поведінкою водіїв, має враховуватися в налаштуваннях самого плану. За нормального проходження пачки через координовані світлофори на магістралі, час очікування дозвільного сигналу для пачки, коли ПК надає їй можливість проїзду без очікування початку дозвільного сигналу, розумно вважати рівним 0.

Своєчасне ввімкнення зеленого сигналу для координованої пачки автомобілів, власне, і є основною метою магістральної координації. Добротний ПК зобов'язаний створити таку можливість на всіх перехрестях в одному напрямку, обраному за прямий (основний), та на більшості перехресть, якщо не на всіх, у зворотному напрямку. Однак, жоден ПК не в змозі забезпечити його повне виконання всіма водіями в ТП, індивідуальна поведінка яких завжди призводить до випадкового характеру процесу руху автомобілів у потоці. При цьому затримки, що виникають через випадковий характер процесу руху автомобілів у пачці, передусім зумовлені саме поведінкою водіїв, унаслідок чого вони піддаються виключно емпіричній оцінці. Для неї потрібне спостереження за поведінкою багатьох пачок автомобілів в умовах функціонування конкретного реального плану магістральної координації.

Слід також відзначити, що роботи, які розвивають концепцію циклічних профілів потоку [11, 12] не можуть бути безпосередньо використані для емпіричного оцінювання закономірностей руху пачки, тому що, на жаль, сучасний рівень планування заходів з організації дорожнього руху (ОДР) ще не забезпечив у містах світу функціонування помітної кількості ефективних (добротних) планів магістральної координації. І це також ускладнює можливість для оцінювання випадкового елемента затримки автомобілів у складі координованої пачки на даному рівні досліджень. Така оцінка може бути елементом остаточного проекту із впровадження конкретного ПК. Але, оскільки отримати її аналітично не є можливим, то випадкові затримки можна врахувати на етапі імітаційного моделювання процесу функціонування нового ПК, проведення якого в сучасних умовах видається вельми бажаним.

Мета статті розробка аналітичних моделей для оцінки середніх затримок автомобілів, що рухаються в пачках скоординованими ділянками ММ за умов низького трафіку.

Виклад основного матеріалу

Слід розуміти, що аналітична оцінка затримок автомобілів у ПК є абсолютно необхідною для порівняння альтернативних планів, формування методик їхнього створення та пошуку дієвих інструментів для підвищення ефективності створюваних ПК, бо повністю уникнути затримок для автомобілів, які рухаються в координованій пачці, можна тільки у виняткових випадках – на фрагментах ВДМ з однаковим часом проїзду координованих ділянок. За стандартних підходів до планування, затримки здебільшого виникають у зворотному, по відношенню до основного напрямку руху ТП.

Вони викликані вимушеними відхиленнями розрахункового часу ввімкнення дозвільного сигналу з метою об'єднання дозвільних фаз зустрічних напрямків в одному СЦ на перехресті, а також під час повороту автомобілів, що координуються, ліворуч, коли виділена лівоповоротна фаза наявна у СУ ПК. Такі затримки виникають навіть за повного дотримання всіма учасниками руху умов, що надаються ПК, коли затримки для автомобілів у прямому напрямку руху пачки зеленої хвилі (ЗХ) можна вважати рівними 0.

Під час розгляду наслідків зміщення часу ввімкнення дозвільного сигналу відносно розрахункового часу прибуття пачки, її затримки та затримки лівоповоротних потоків, мають оцінюватися для умов низького трафіку, в яких уся пачка встигає проїхати чергове перехрестя за час дозвільного сигналу, виділеного для неї в циклі, у тому разі, якщо момент прибуття першого автомобіля в пачці на чергове перехрестя збігається з моментом ввімкнення дозвільного сигналу.

Навантаження на перехрестя не має враховуватися в затримці не тому, що час проїзду пачки підбирають під її розмір. Домогтися такої відповідності в жорстких ПК практично неможливо, оскільки фактичний розмір пачки є випадковим. Крім того, на магістралях завжди присутня випадкова кількість додаткових автомобілів з різних напрямків, які стають частиною пачки починаючи з того світлофора на магістралі, перед яким вони з'явилися. В умовах реалізації плану жорсткої координації зазвичай вважається, що поява додаткових автомобілів компенсується коливаннями швидкості пачки, за постійної середньої інтенсивності їхнього виникнення на шляху її руху (зараз було більше, на наступному перехресті або в наступному циклі стане менше). А самі витрати часу, викликані випадковим характером виникнення додаткових автомобілів, є предметом окремого вивчення, залежно від прийнятої стратегії виділення додаткового часу на їхній розгін до швидкості руху координованої пачки [9]. Ці витрати оцінюються окремо і під час оцінки витрат часу лівоповоротних потоків й пачки ЗХ на очікування дозвільного сигналу не враховуються.

Затримки пачки ЗХ та лівоповоротних потоків слід оцінювати в умовах низького трафіку внаслідок того, що вони безпосередньо і в явному вигляді мають відображати наслідки зміщення часу ввімкнення дозвільного сигналу відносно розрахункового часу прибуття пачки до перехрестя. Тому додаткові затримки, пов'язані з випадковістю процесу або високим навантаженням перехрестя, які безпосередньо не пов'язані з такими зсувами, тут враховувати не слід. Невід'ємна різниця між моментами ввімкнення дозвільного сигналу та розрахунковим часом прибуття пачки до перехрестя, власне, і є очевидною оцінкою часу очікування можливості проїзду пачки через регульоване перехрестя. Сама ж можливість проїзду пачки обмежена періодом роботи фази, що дозволяє рух пачки, і тому повністю визначається ПК.

Для таких умов достатньо припустити, що всі автомобілі в пачці розподілені по її довжині рівномірно, відомий час прибуття першого автомобіля в

пачці на перехрестя, яке розглядається, і довжина пачки в одиницях часу. Ці припущення багато в чому аналогічні представленню групового прибуття у вигляді стрічки, тому не потребують окремого обґрунтування, внаслідок широкого використання стрічки в практиці ОДР. Тоді, для розрахунку часу затримки, потрібно знати проміжок часу між моментом увімкнення дозвільного сигналу та розрахунковим моментом прибуття першого автомобіля в пачці, що, власне, й визначатиме час затримки на перехресті, яке розглядають, і якість ПК взагалі.

Також слід врахувати той факт, що оскільки різниця між цими двома моментами може набувати будь-яких значень у межах циклу, то в якихось випадках вона може ставати настільки великою, що частина дозвільної фази, яка залишилася для проїзду пачки, буде меншою за її довжину, а частина автомобілів буде змушена зупинитися й очікувати на ввімкнення дозвільної фази вже в наступному циклі. Для цих автомобілів час очікування збільшується на забороняючу рух частину циклу, яка разом із тривалістю дозвільного сигналу має бути доступною в будь-якому плані координаті, як обов'язкові для його формування елементи.

Першим етапом оцінювання має стати більш загальний, стосовно прямого руху пачки, випадок повороту автомобілів із пачки ліворуч, протягом виділеної для цього в циклі фази. Його спільність зумовлена вторинним характером фази лівого повороту щодо основного циклу координаті, що може породити найрізноманітніші поєднання між часом прибуття лівоповоротних автомобілів на чергове перехрестя в координаті та часом увімкнення дозвільного сигналу для них у межах СЦ. Природним (доречним) тут є припущення про рівномірний характер розподілу лівоповоротних автомобілів за довжиною пачки, що зумовлено цілковитою (повною) випадковістю процесу формування пачки координованих ТЗ з погляду напрямку подальшого руху кожного з них.

Звичайна для більшості наявних і потенційних ПК ситуація виникає, коли тривалість фази для пачки ЗХ перевищує тривалість фази для лівого повороту. Але як виняток, може бути й по-іншому, тому таку можливість необхідно враховувати під час формування відповідних аналітичних залежностей. Друга ситуація для лівоповоротних автомобілів, узагалі кажучи, сприятливіша, оскільки в ній обов'язково буде присутній період часу, коли автомобілі вже не прибувають, а виїжджати ще можуть, що також має бути враховано в аналітичному описі процесу.

Для виведення (розробки) відповідних залежностей, спочатку слід ввести поняття джерел виникнення часу затримки без навантаження, тобто для умов низького трафіку, в яких уся пачка встигає проїхати чергове перехрестя за час дозвільного сигналу, виділеного для неї в циклі, якщо момент прибуття першого автомобіля в пачці на чергове перехрестя збігається з моментом увімкнення дозвільного сигналу.

За цих умов затримки дорівнюють часу очікування ввімкнення дозвільного сигналу для лівоповоротних автомобілів, і воно обов'язково виникає для

тієї частини пачки, у складі якої перебувають і лівоповоротні автомобілі, і яка не має спільного періоду часу з дозвільним сигналом для лівого повороту.

Це завжди позитивні й не більші, ніж тривалість фази для ЗХ, одна або дві області з періоду прибуття пачки, положення і тривалість яких, у загальному випадку, визначають різницею часів початку й закінчення двох періодів часу: прибуття автомобілів (пачка хвили) та їхнього відправлення (фаза лівого повороту). Для розпізнання відповідних елементів часу вводяться такі позначення:

1. Область раннього прибуття ε , рис. 1

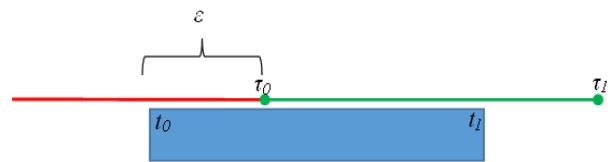


Рис. 1. Варіант раннього прибуття пачки, як джерела виникнення затримок

2. Область пізнього прибуття λ , рис. 2

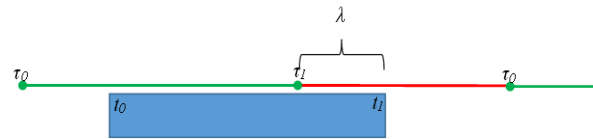


Рис. 2. Варіант пізнього прибуття пачки, як джерела виникнення затримок

Тут і далі використовуються наступні позначення: ε , λ – індекси, що позначають область раннього та пізнього прибуття пачки автомобілів; T_ε , T_λ – відповідно, середній час очікування, що виникає в області раннього і пізнього прибуття, с; t_0 , t_1 – відповідно, моменти початку і закінчення періоду планового прибуття пачки ЗХ до перехрестя, с; τ_0 , τ_1 – відповідно, моменти початку та закінчення годиння дозвільного сигналу для пачки автомобілів, що прибуває, с.

За скоординованої роботи світлофорів моменти планового прибуття пачки ЗХ до перехрестя t_0 і ввімкнення дозвільного сигналу для неї збігаються. Але зараз розглядається затримка лівоповоротних автомобілів, а внаслідок їхнього рівномірного розподілу в пачці ЗХ, вони також починають прибувати до перехрестя в момент часу t_0 . Але дозвільний сигнал для лівоповоротних автомобілів – τ_0 може вмикатися і не в момент їхнього прибуття до чергового перехрестя. Саме тому, що в ПК в загальному випадку моменти часу t_0 і τ_0 для лівоповоротних автомобілів не збігаються, вони є найзручнішим об'єктом для побудови шуканих залежностей часу очікування без навантаження під час групового прибуття – вони прибувають у складі координованої пачки, а зелене світло для них може вмикатися майже в будь-який момент у циклі. Усі інші варіанти групового прибуття є окремим випадком стосовно прибуття лівоповоротних автомобілів.

Щоб отримати загальний середній час очікування для всіх лівоповоротних автомобілів, що прямують у пачці ЗХ, середній час очікування для таких автомобілів, що прибувають на перехрестя впродовж періодів раннього і пізнього прибуття, потрібно зв'язати через відповідні їм частини пачки:

$$T = T_\varepsilon \cdot \Delta_\varepsilon + T_\lambda \cdot \Delta_\lambda, \quad (1)$$

де T – середній час очікування для всіх лівоповоротних автомобілів, що прямують у пачці ЗХ, с; Δ_ε , Δ_λ – частка в пачці автомобілів, що прибувають на перехрестя протягом періодів раннього і пізнього прибуття відповідно.

Значення часток розраховують з урахуванням того, що області раннього та пізнього прибуття не можуть перевищувати довжину всієї пачки:

$$\Delta_\varepsilon = \begin{cases} \frac{\tau_0 - t_0}{t_p}, & \text{якщо } \tau_0 - t_0 < t_p, \\ 1, & \text{якщо } \tau_0 - t_0 \geq t_p \end{cases}, \quad (2)$$

та

$$\Delta_\lambda = \begin{cases} \frac{t_1 - \tau_1}{t_p}, & \text{якщо } t_1 - \tau_1 < t_p, \\ 1, & \text{якщо } t_1 - \tau_1 \geq t_p \end{cases}, \quad (3)$$

де t_p – тривалість періоду планового прибуття пачки ЗХ з лівоповоротними автомобілями до перехрестя, $t_p = t_1 - t_0$, с.

Розрахунок середнього часу очікування для автомобілів, що прибувають за період раннього прибуття на перехрестя, як уже згадувалося вище, ґрунтується на припущенні про рівномірний розподіл за довжиною пачки автомобілів, що прибувають і повертають ліворуч.

Таке припущення зумовлює той факт, що середній час очікування дорівнює половині різниці між початком фази лівого повороту та початком прибуття пачки ЗХ в тому разі, якщо лівий поворот буде дозволено протягом дозвільного сигналу для ЗХ, як показано на рис. 1. Пояснюється це тим, що затримка першого автомобіля з області раннього прибуття пачки, який прибув у момент t_0 дорівнює $\tau_0 - t_0$, а останнього з тієї ж області, який прибув у момент τ_0 , дорівнює 0, відповідно середня затримка для них дорівнюватиме $\frac{\tau_0 - t_0}{2}$.

Сума затримок для всіх пар автомобілів, рівновіддалених від кінців проміжку часу між t_0 також дорівнюватиме $\tau_0 - t_0$, відповідно середня затримка для них дорівнюватиме $\frac{\tau_0 - t_0}{2}$. За рівномірного розподілу автомобілів у пачці, ймовірність виникнення для всіх таких пар дорівнює і становить $2/n$, де n – це кількість автомобілів у пачці. Тому й ймовірність у всіх середніх однакова і дорівнює $2/n$, що в підсумку призводить до загальної середньої затримки

автомобілів із розглянутої частини області раннього прибуття, що дорівнює $\frac{\tau_0 - t_0}{2}$.

Якщо ж початок лівого повороту відбувається поза фазою ЗХ, то частина області раннього прибуття пачки, що дорівнює $\tau_0 - t_1$, повністю, без поділу на 2, входить у час очікування дозвільного сигналу для лівоповоротних автомобілів, що прибувають до перехрестя у складі пачки ЗХ, рис. 3.

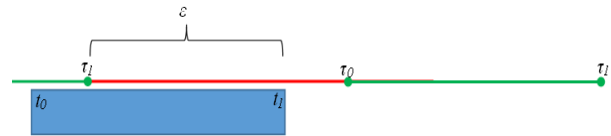


Рис. 3. Другий варіант раннього прибуття пачки, як джерела виникнення затримок

З урахуванням цих міркувань, вираз для середнього часу очікування автомобілів у період раннього прибуття матиме такий вигляд:

$$T_\varepsilon = \begin{cases} \frac{\tau_0 - t_0}{2}, & \text{якщо } \tau_0 - t_0 \leq t_p \\ \frac{t_1 - t_0}{2} + \tau_0 - t_1, & \text{якщо } \tau_0 - t_0 > t_p \end{cases}. \quad (4)$$

Наявність очевидних зв'язків між часовими характеристиками в формулі (4), зумовлених розташуванням обох фаз усередині СЦ, дає змогу подати цей вираз в іншому вигляді:

$$T_\varepsilon = \begin{cases} \frac{\tau_0 - t_0}{2}, & \text{якщо } \tau_0 - t_0 \leq t_p \\ \tau_0 - \frac{t_0 + t_1}{2}, & \text{якщо } \tau_0 - t_0 > t_p \end{cases}. \quad (5)$$

Область пізнього прибуття охоплює період прибуття пачки ЗХ, що триває після завершення дозвільної фази для лівого повороту, рис. 2. Автомобілі, що прибувають у цей період, змушені очікувати на початок фази лівого повороту вже в наступному циклі. У зв'язку з цим, а також з урахуванням властивостей середнього часу очікування дозвільного сигналу, зазначеними вище, вираз для часу очікування автомобілів з області пізнього прибуття матиме такий вигляд:

$$T_\lambda = \begin{cases} \frac{t_1 - \tau_1}{2} + T_y - \tau_l - (t_1 - \tau_1), & \text{якщо } 0 < t_1 - \tau_1 \leq t_p \\ 0, & \text{якщо } t_1 - \tau_1 \leq 0 \text{ або } t_1 - \tau_1 > t_p \end{cases}, \quad (6)$$

де τ_l – тривалість фази лівого повороту, $\tau_l = \tau_1 - \tau_0$, с.

Зв'язки між часовими характеристиками в формулі (6) дають змогу записати цей вираз у простішому вигляді:

$$T_\lambda = \begin{cases} T_y - \tau_l - \frac{t_1 - \tau_1}{2}, & \text{якщо } 0 < t_1 - \tau_1 \leq t_p \\ 0, & \text{якщо } t_1 - \tau_1 \leq 0 \text{ або } t_1 - \tau_1 > t_p \end{cases}. \quad (7)$$

Вирази, отримані для середнього часу очікування автомобілів з ділянок раннього і пізнього прибуття до перехрестя, є звичайними доданками

формули (1) не на всьому циклі світлофорного регулювання, оскільки можлива ситуація, коли вони існують одночасно, або одна з досліджуваних фаз – ЗХ чи лівоповоротної, розпочинається в одному циклі, а закінчується вже в наступному. Щоб визначити умови використання доданків формул (5) і (7) у (1), потрібно зафіксувати початок циклу, пов'язавши його з початком одного з основних тактів СЦ. Оскільки метою координації роботи світлофорів є безперешкодний пропуск пачки ЗХ, розумно вважати саме цю фазу першою в циклі. Тоді момент увімкнення дозвільного сигналу для фази ЗХ t_0 , він же початковий момент прибуття пачки автомобілів до перехрестя, стає ще й початком усього циклу, а старт дозвільного сигналу для лівого повороту τ_0 набуває додаткового сенсу – зсуву початку лівоповоротної фази відносно початку циклу.

Величина цього зсуву в реальному ПК може набувати будь-якого значення, яке не перевищує тривалості циклу, $0 \leq \tau_0 < T_{\text{ц}}$ і правила використання фор-

мули (1) мають бути визначені для всього цього діапазону.

Висновки

Одним із основних критеріїв, що характеризує ефективність впровадження ПК на ММ є затримки транспорту, які найчастіше характеризуються скороченням витрат часу на очікування учасниками руху можливості проїзду чергового перехрестя в ПК.

Розроблені моделі сформовані для умов низького трафіку і групового прибуття та можуть використовуватись при проведенні оцінювання витрат часу, що витрачається на очікування водіями, які рухаються скоординованими ділянками ВДМ, можливості проїзду регульованих перехресть, що включені до ПК.

Подальшим напрямком проведення досліджень є удосконалення отриманих аналітичних моделей за рахунок урахування в них інших можливих варіантів організації зсуву включення дозвільного сигналу для руху лівоповоротних потоків.

СПИСОК ЛІТЕРАТУРИ

1. Горбачов, П.Ф., Макарічев, О.В., Шевченко, В.В. (2019). Оцінка затримок руху на регульованих перехрестях міських вулиць із трифазним циклом регулювання. *Автомобільний транспорт*, 44, 30–39. <https://doi.org/10.30977/AT.2219-8342.2019.44.0.30>.
2. Горбачов, П.Ф., Макарічев, О.В., Шевченко, В.В. (2021). Імовірність виникнення неповної пачки автомобілів при координованому керуванні на міській магістралі. *Вісник ХНАДУ*, 92(1), 214–225. <https://doi.org/10.30977/BUL.2219-5548.2021.92.1.214>.
3. Любий, Є.В. (2022). Аналіз методів формування планів координації світлофорних об'єктів. *Сучасні технології в машинобудуванні та транспорті*, 2(19), 108–122. <https://doi.org/10.36910/automash.v2i19.909>.
4. Robertson, D.I. (1969). TRANSYT: A Traffic Network study Tool. *Gt. Brit. Road Research Laboratory, Rept. LR 253*, 22–41.
5. Zhao, L., Rilett, L. R., Tufuor, E. (2017). Calibrating the Robertson's Platoon Dispersion Model on a Coordinated Corridor with Advance Warning Flashers. *Transportation Research Record*, 2623(1), 10–18. <https://doi.org/10.3141/2623-02>.
6. National Academies of Sciences, Engineering, and Medicine. (2015). *Signal Timing Manual – Second Edition*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/22097>.
7. Little, J.D.C., Kelson, M.D., Gartner, N.H. (1981). MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks. *Transportation Research Record* 795, 40–46.
8. Gartner, N.H., Assmann, S.F., Lasaga, F., Hou, D.L. (1990). MULTIBAND – A Variable-Bandwidth Arterial Progression Scheme. *Transportation Research Record*, 1287, 212–222. URL: <https://onlinepubs.trb.org/Onlinepubs/trr/1990/1287/1287-022.pdf> (дата звернення 20.02.2023).
9. Горбачов, П.Ф., Любий, Є.В. (2022). Методика оцінки часу, необхідного для забезпечення рівномірного руху пачки автомобілів на координованій ділянці міської магістралі. *Автомобіль і електроніка. Сучасні технології*, 22, 81–91. <https://doi.org/10.30977/VEIT.2022.22.0.2>.
10. Webster, F.V. (1958). *Traffic Signal Settings*. Road Research Technical Paper No. 39. London: Department of Scientific and Industrial Research, 45 p.
11. Geroliminis, N., Skabardonis, A. (2005). Prediction of Arrival Profiles and Queue Lengths along Signalized Arterials by using a Markov Decision Process. *Transportation Research Record*, 1934(1), 116–124. <https://doi.org/10.1177/0361198105193400112>.
12. L., Shen, R., Liu, Z., Yao, W., Wu, H., Yang. (2019). Development of Dynamic Platoon Dispersion Models for Predictive Traffic Signal Control. In *IEEE Transactions on Intelligent Transportation Systems*, 20, 2, 431–440. doi: <https://doi.org/10.1109/TITS.2018.2815182>.

Received (Надійшла) 11.03.2023

Accepted for publication (Прийнята до друку) 24.05.2023

An approach to assessing transport delays in terms coordinated timing plan

Peter Horbachov, Yevhen Liubyi

Abstract. The article presents analytical models that can be used to estimate the delays of vehicles moving along coordinated sections of city highways in low traffic conditions and when they arrive in groups at the next intersection in terms of coordinated timing plan. The results of determining delays according to the presented models should be used for comparing alternative coordinated timing plan, developing methods for their creation, as well as for finding effective and reliable tools to improve the efficiency of the created coordinated timing plan. In the future, the developed models need to be improved by considering other possible options for organizing the offset in the inclusion of the permissive signal for the left-turn phase within the coordinated timing plan.

Keywords: group arrival, green wave, platoon, coordinated timing plan, transport delays.

Р. В. Куліш¹, О. Г. Матющенко²

¹ Льотна академія Національного авіаційного університету, Кропивницький, Україна

² Державний НДІ випробувань і сертифікації озброєння та військової техніки, Черкаси, Україна

МОДЕЛЬ МАРШРУТИЗАЦІЇ ОБ'ЄКТУ МОБІЛЬНИХ ОБ'ЄКТІВ БЕЗПІЛОТНИМ ЛІТАЛЬНИМ АПАРАТОМ

Анотація. Аналіз досвіду застосування безпілотних літальних апаратів доводить, що ефективність виконання завдань безпосередньо пов'язана із "якістю" планування маршруту. Більшість методів планування маршруту сфокусовані на пошуку та моніторингу стаціонарних об'єктів. Проте в даний час набагато більше значення має вирішення задачі спостереження за мобільними, рухомими об'єктами. В роботі вирішена задача планування маршруту об'єкту мобільних об'єктів, що зберігають напрямок свого руху протягом тривалого проміжку часу із використанням методу математичної оптимізації - динамічного програмування. При вирішенні завдання побудови оптимального маршруту враховувалась динаміка польоту безпілотного літального апарату, що дозволяє послідовно додавати об'єкти до маршруту.

Ключові слова: безпілотний літальний апарат, мобільні групи, маршрутизація, моніторинг, рухомий об'єкт динамічне програмування.

Вступ

Постановка проблеми. В сучасному світі важко уявити галузь людської діяльності яка хоча б дотично не пов'язана із застосуванням безпілотних систем. Завдяки відносно низькій вартості та доступності широкого розповсюдження набули безпілотні літальні апарати (БпЛА). Широке розповсюдження БпЛА викликало активний розвиток алгоритмів, методів управління БпЛА для вирішення широкого кола задач, однією з яких є планування маршруту.

Однак, незважаючи на різноманітність методів та алгоритмів щодо побудови маршрутів польоту БпЛА, метод об'єкту рухомих об'єктів потребує оптимізації та пошуку нових рішень враховуючи накладені обмеження.

Таким чином, основним шляхом у підвищенні ефективності застосування БпЛА є пошук методів планування маршрутів об'єкту рухомих об'єктів (груп) із врахуванням динаміки руху об'єкту спостереження та польоту самого БпЛА. Саме вирішенню цієї прикладної задачі присвячена дана робота.

Аналіз останніх досліджень і публікацій. У статті [1] проводиться порівняльний аналіз методів планування маршрутів БпЛА на графах та мультиагентні алгоритми, мурашиний алгоритм та алгоритм «запит-відповідь-угода». Розроблений метод, що базується на інтеграції пошукових походів на графах та мультиагентних, що працюють на паралельній основі. У статті [2] пропонується онлайн-алгоритм планування маршруту екстреної логістики на основі оптимізації подвійної колонії мурах. Результати експерименту свідчать, що запропонований алгоритм може підвищити ефективність розподілу виконавців між об'єктами. У статті [3] для генерації субоптимального маршруту БпЛА запроваджене використання нейронної мережі Хопфілда. Оптимальним в дослідженні вважається шлях найкоротшої довжини. Запропонований підхід спрямований, в першу чергу, на зниження навантаження оператора при виборі оптимального маршруту. У статті [4]

вирішена задача побудови польотного завдання - маршрутів руху для групи БпЛА. Розроблена мережа для груп БпЛА, яку можливо розгорнути у польових умовах. У статті [5] розроблено генетичний алгоритм для ефективного вирішення складної моделі оптимальної маршрутизації БпЛА. Для логістики яка є економічно вигідною та безпечною. У статті [6] описується децентралізоване управління групою БпЛА, об'єднаних у мережу, що ґрунтується на розв'язанні транспортно-комунікаційної задачі яка зводиться до класичної задачі маршрутизації транспортних засобів. Розроблено алгоритм розв'язку задачі маршрутизації транспортних засобів.

Метою статті є вирішення завдання маршрутизації об'єкту групи рухомих об'єктів, які мають сталу швидкість на інтервалі часу із врахуванням динаміки руху безпілотного літального апарату.

Виклад основного матеріалу

Постановка задачі. Розглянемо розв'язання задачі планування маршруту польоту за таких припущень:

1. Задано множину контрольованих точкових рухомих наземних об'єктів та їх координати x_{ij}, z_{ij}, z_j в даний час. Задані швидкості $V_{u,j}$ і напрямки $a_{u,j}$ руху цих об'єктів, що ускладнює або виключає попадання їх частини до плану. Це відрізняє цю постановку від класичного завдання комівояжера при якій кожний об'єкт додається до плану маршруту.

2. Розглядається рух БпЛА лише у горизонтальній площині, коли поточний стан польоту характеризується лінійними координатами колій x, z модулем швидкості V та кутом напрямку швидкості a . Швидкість V вважається заданою та незмінною, при цьому

$$K_j = V_{u,j}/V \ll 1. \quad (1)$$

3. Динаміка бічного руху БпЛА при перельоті до об'єкта j може бути описано диференціальними рівняннями. Якщо знехтувати різницею в довжині шляху по прямій і по кривій кола при перельоті в попереджену точку:

$$\begin{aligned} \dot{x} &= \frac{x_j - x}{\Delta t_j} \approx \frac{x_j - x}{r_j} V; \quad \dot{z} = \frac{z_j - z}{\Delta t_j} \approx \frac{z_j - z}{r_j} V; \\ \dot{\alpha} &= \frac{\frac{z_j - z}{r_j} - \alpha}{\Delta t_j} \approx \frac{\frac{z_j - z}{r_j} - \alpha}{r_j} V = \frac{\alpha_j - \alpha}{r_j} V, \end{aligned} \quad (2)$$

де j - номер пункту, що вибирається в поточний момент часу, r_j - відстань між БПЛА j -м об'єктом, x, z - координати БПЛА, α - шляховий кут БПЛА, V - швидкість БПЛА, x_j, z_j - координати j -го об'єкта, Δt_j - час перельоту в j -й пункт.

4. При бічному маневрі бічне перевантаження БПЛА обмежена, що визначає заданий мінімальний радіус R кола, за яким доступний розворот БПЛА.

5. Для досягнення чергового об'єкта, що не лежить на лінії прямолінійного польоту, здійснюється наведення шляхом пропорційної навігації. Процес перерахунку нового кута у роботі докладно не розглядається.

6. Як критерій оптимальності маршруту прийнято умову максимуму числа контрольованих об'єктів за заданий час T , чому відповідає умова мінімуму середнього часу перельоту від одного об'єкта до іншого, чому відповідає критерій

$$I = \sum_{j=1}^M \frac{r_j}{V} \approx \sum_{j=1}^M \frac{r_j}{V \Delta t} dt \approx \int_0^T \frac{r_j}{V \Delta t} dt \rightarrow \min. \quad (3)$$

Запропонований підінтегральний вираз $f_0 = \frac{r_j}{V \Delta t}$ функціоналу (1) свідчить, що у першому наближенні при постійній швидкості V без врахування динаміки БПЛА краще обрати найближчий черговий об'єкт на мінімальній відстані r_j .

При перерахованих припущеннях потрібно сформулювати такий алгоритм маршрутизації, який з урахуванням обмежень при бічному маневрі БПЛА забезпечив би найбільшу кількість контрольованих об'єктів, переважно розташованих по напрямку польоту, що змінюється.

Запропонований підхід до вирішення задачі за допомогою динамічного програмування. Оскільки маршрут, що обирається, є результатом багатокрокового вибору чергового об'єкта, а прийняття рішення на кожному кроці крім поточного штрафу у вигляді втраченого часу ще має наслідки для майбутніх кроків, найбільш підходящим методом оптимізації є динамічне програмування [5]. Тоді, враховуючи дискретний характер альтернативних рішень при виборі «куди летіти», рівняння Белмана в приєднаних похідних можна записати як

$$\begin{aligned} -\frac{\partial \mathcal{E}}{\partial t} &= \min_j \left\{ f_{0j} + \frac{\partial \mathcal{E}}{\partial z} \dot{z}_j + \frac{\partial \mathcal{E}}{\partial \alpha} \dot{\alpha}_j \right\} = \\ &= \min_j \left\{ f_{0j} + \frac{\partial \mathcal{E}}{\partial x} \frac{x_j - x}{r_j} V + \frac{\partial \mathcal{E}}{\partial z} \frac{z_j - z}{r_j} V + \frac{\partial \mathcal{E}}{\partial \alpha} \frac{\alpha_j - \alpha}{r_j} V \right\} = F_j(x, z, \alpha), \end{aligned} \quad (4)$$

де \mathcal{E} - функція Белмана, F_j - функція поточного ризику з урахуванням наслідків у майбутньому, j - номер, що обирається в поточний момент часу,

r_j - відстань між БПЛА і j -м об'єктом, V - швидкість БПЛА, x_j, z_j - координати j -го об'єкта, x, z - координати БПЛА, α - шляховий кут БПЛА. Наприклад, це може бути група рухомих об'єктів попереду у напрямку польоту ($j = 1.1, 1.2, 1.3, 1.4$), позаду ($j = 2.1, 2.2, 2.3, 2.4$), зліва ($j = 3.1, 3.2, 3.3, 3.4$) або праворуч ($j = 4.1, 4.2, 4.3, 4.4$), як показано на рис. 1.

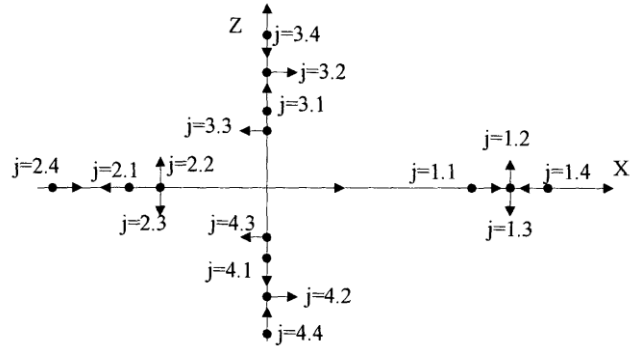


Рис. 1. Розташування 16 рухомих об'єктів у районі досяжності БПЛА, що мають приблизно однакову довжину шляху

Принциповою відмінністю розв'язуваної задачі від прийнятого в методі АКОР припущення про безперервність управління є те, що управління є альтернативним (зокрема, відповідно до рис. 1, $j = 1.1, 1.2, \dots, 4.4$). Тож наближеного синтезу закону управління скористаємося методом «робочої точки» [7], суть якого полягає у поверненні робочої точки в початковий стан після зниження впливу на окремий елемент.

Обчислення ординат мінімального ризику у різних польотних ситуаціях. Враховуючи особливості даної задачі подання функції Белмана є у квадратичній формі недостатньою, щоб отримати адекватний опис. Тому була використана наступна апроксимація:

$$\begin{aligned} \mathcal{E} &= \alpha + \beta_1 x + 0,5\gamma_1 x^2 + \beta_2 z + 0,5\gamma_2 z^2 + \beta_3 \alpha + \\ &+ 0,5\gamma_3 z^2 + \psi_{12} xz + \psi_{13} x\alpha + \psi_{23} z\alpha + \lambda xz\alpha. \end{aligned} \quad (5)$$

Розглянемо польотні ситуації у разі відхилення від робочої точки. При $x = x_0 + \Delta x_1 = x_0 + R$ мінімальний штраф буде при польоті БПЛА до об'єкта 1.1, а при $x = x_0 + \Delta x_2 = x_0 - 9R$ кращим варіантом є об'єкт 2.4, тому:

$$\begin{aligned} \frac{C_1^+}{V} &= \beta_1 + (x_0 + R)\gamma_1 + z_0\psi_{12} + \frac{7(1 + K)R}{V^2 \Delta t}; \\ \frac{C_1^-}{V} &= \beta_1 + (x_0 - 9R)\gamma_1 + z_0\psi_{12} + \\ &+ \frac{8(1 + 0,75K)R}{V^2 \Delta t}. \end{aligned} \quad (6)$$

Різні за модулем відхилення Δx_1 та Δx_2 були взяті для того, щоб напрямком польоту БПЛА був однаковим - вперед. Ілюстрація польотних ситуацій представлена на рис. 2.

Розглянемо відхилення Δz по другій координаті z . При $z = z_0 + \Delta z$ мінімальний шлях БПЛА буде при польоті до об'єкта 3.2, а при $z = z_0 - \Delta z$ - при польоті до об'єкта 4.2.

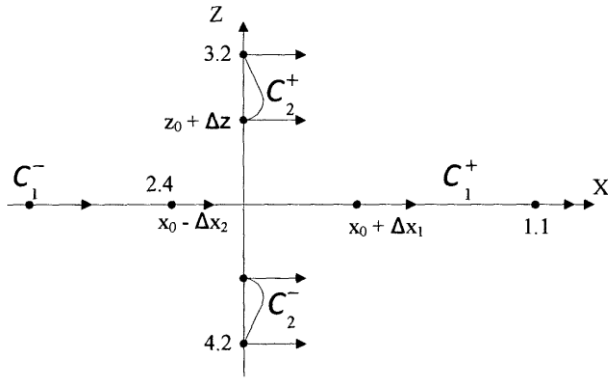


Рис. 2. Ілюстрація польотних ситуацій при почергових відхиленнях щодо та z від робочої точки

Тому при $\Delta z = R$ отримаємо:

$$\begin{aligned} \frac{C_2^+}{V} &= \frac{K}{1+3K} [\beta_1 + \gamma_1 x_0 + \psi_{12}(z_0 + R)] + \\ &+ \frac{2(1+5K)}{3(1+3K)} [\beta_2 + \gamma_2(z_0 + R) + \psi_{12}x_0] + \frac{\pi}{4R} \times \\ &\times [\beta_3 + \psi_{13}x_0 + \psi_{23}(z_0 + R) + x_0(z_0 + R)\lambda] + \\ &+ \frac{3(1+3K)R}{V^2\Delta t}; \\ \frac{C_2^-}{V} &= \frac{K}{1+3K} [\beta_1 + \gamma_1 x_0 + \psi_{12}(z_0 - R)] + \\ &+ \frac{2(1+5K)}{3(1+3K)} [\beta_2 + \gamma_2(z_0 - R) + \psi_{12}x_0] - \frac{\pi}{4R} \times \\ &\times [\beta_3 + \psi_{13}x_0 + \psi_{23}(z_0 - R) + x_0(z_0 - R)\lambda] + \\ &+ \frac{3(1+3K)R}{V^2\Delta t}. \end{aligned} \quad (7)$$

Ілюстрація для оцінок ординат C_2^+ і C_2^- також представлена на рис. 3.

Далі, розглянемо відхилення α , при $\Delta\alpha = \pm\pi/2$ отримаємо очевидний вибір альтернатив польоту, як показано на рис. 3, при досягненні об'єктів 3.4 та 4.4.

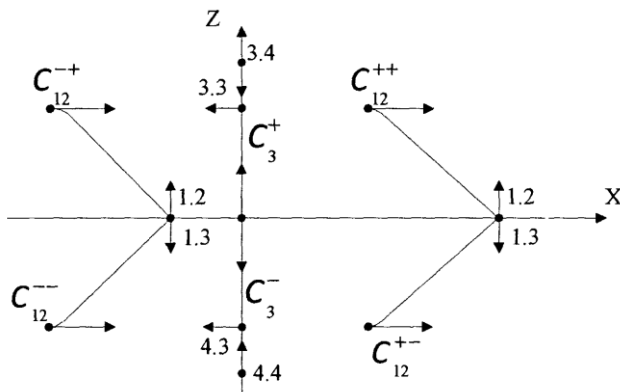


Рис. 3. Ілюстрація польотних ситуацій при відхиленні по α та попаданні до об'єктів 3.2 і 4.2 та попарних відхиленнях по x та z при попаданні до об'єктів 1.2, 1.3 та 2.2, 2.3

При $\Delta\alpha = \pi/2$ потрібно летіти до об'єкта 3.4, при $\Delta\alpha = -\pi/2$ - до об'єкта 4.4, і це забезпечить мінімальний час польоту.

Тому ординати C_3^+ і C_3^- дорівнюють:

$$\begin{aligned} \frac{C_3^+}{V} &= x_0\psi_{12} + \beta_2 + z_0\gamma_2 + \\ &+ \frac{\pi}{2}(\psi_{23} + x_0\lambda) + \frac{4R(1+K)}{V^2\Delta t}; \\ \frac{C_3^-}{V} &= [x_0\psi_{12} + \beta_2 + z_0\gamma_2 + \\ &+ (\pi/2) \cdot (\psi_{23} + x_0\lambda)] + \frac{4R(1+K)}{V^2\Delta t}. \end{aligned} \quad (8)$$

При відхиленнях по x і z виникають 4 ситуації, для яких очевидна перевага перельоту до об'єкта 1.2 при $x = x_0 + R$; $z = z_0 + 4R$, і тоді отримаємо:

$$\begin{aligned} C_{12}^{++} &= 0,97(1+K) \left[\beta_1 + (x_0 + R)\gamma_1 + \right. \\ &+ (z_0 + R)\psi_{12} \left. \right] + \frac{1-7K}{7,2} \times \\ &\times [\beta_2 + (z_0 + R)\gamma_2 + (x_0 + R)\psi_{12}] - 0,02 \times \\ &\times [\beta_3 + (x_0 + R)\psi_{23} + (z_0 + R)(z_0 + R)\lambda] + \\ &+ \frac{7,2(1+0,1K)R}{V^2\Delta t}. \end{aligned} \quad (9)$$

Аналогічним чином можна отримати оцінку ординат мінімального ризику C_{13} під час розгляду 4-х ситуацій, наведених на рис. 4.

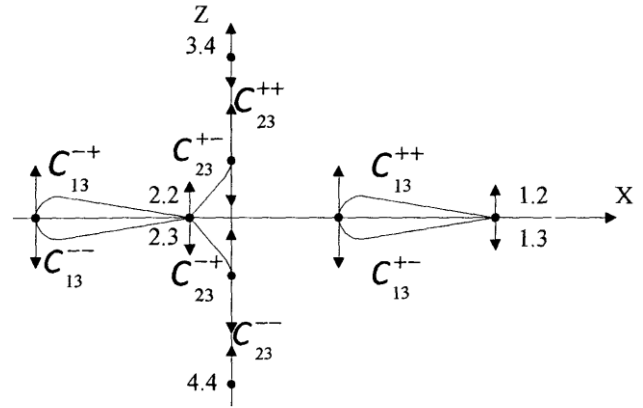


Рис. 4. Ілюстрація польотних ситуацій при попарних відхиленнях α і x і відповідно - α і z

У цих випадках краще летіти до об'єктів 2.2 та 2.3, якщо $x = x_0 - \Delta x_2$ і до об'єктів 1.2 та 1.3, якщо $x = x_0 + \Delta x_1$. Тому отримаємо, зокрема, при $x = x_0 + R$, $\Delta\alpha = -\pi/2$ оцінку C_{13}^{++} мінімального ризику у вигляді:

$$\begin{aligned} \frac{C_{13}^{++}}{V} &= 0,87 \left[\beta_1 + (x_0 + R)\gamma_1 + z_0\psi_{12} + \right. \\ &+ (\pi/2) \cdot (\psi_{13} + z_0\lambda) \left. \right] + \\ &+ K \left[\beta_2 + z_0\gamma_2 + (x_0 + R)\psi_{12} + \right. \\ &+ (\pi/2) \cdot (\psi_{23} + (x_0 + R)\lambda) \left. \right] + \\ &+ \frac{8R(1+K)}{V^2\Delta t} - \frac{0,225(1+1,7K)}{R} \times \\ &\times \left[\beta_3 + \gamma_2 \frac{\pi}{2} + (x_0 + R)\psi_{13} + z_0\psi_{23} + (x_0 + R)z_0\lambda \right]. \end{aligned}$$

При попарних відхиленнях α і z також виникають 4 ситуації, ілюстровані на рис 4, причому до об'єкта 2.2 краще летіти при $z = z_0 + R$, $\Delta\alpha = -\pi/2$, а при $z = z_0 - R$, $\Delta\alpha = \pi/2$ - до об'єкта 2.3, до об'єкта 3.4 - при $z = z_0 + R$, $\Delta\alpha = \pi/2$, до об'єкта 4.4 - $z = z_0 - R$, $\Delta\alpha = -\pi/2$.

Тому ординати ризику C_{23} рівні:

$$\begin{aligned} \frac{C_{23}^{++}}{V} &= \beta_2 + (R + z_0)\gamma_2 + x_0\psi_{12} + \frac{\pi}{2}\psi_{23} + \frac{x_0\pi}{2}\lambda + \frac{3R(1+K)}{V^2\Delta t}; \\ \frac{C_{23}^{--}}{V} &= -\beta_2 + (R - z_0)\gamma_2 - x_0\psi_{12} + \frac{\pi}{2}\psi_{23} + \frac{x_0\pi}{2}\lambda + \frac{3R(1+K)}{V^2\Delta t}; \\ \frac{C_{23}^{\pm}}{V} &= -\frac{\pi}{2}(1 + 0,6K) \begin{bmatrix} \beta_1 + x_0\gamma_1 + (z_0 + R)\psi_{12} - \\ -\frac{\pi}{2}\psi_{13} - \frac{\pi}{2}(z_0 + R) \end{bmatrix} - \frac{2}{\pi}(1 - K) \begin{bmatrix} \beta_2 + (z_0 + R)\gamma_2 + x_0\psi_{12} \\ -\frac{\pi}{2}\psi_{23} - \frac{\pi}{2}\lambda x_0 \end{bmatrix} - \\ &\quad \frac{1-K}{R} \left[\beta_3 - \frac{\pi}{2}\gamma_3 + x_0\psi_{13} + (z_0 + R)\psi_{23} - (z_0 + R)x_0\lambda \right] + \frac{\pi R(1+K)}{V^2\Delta t}; \\ \frac{C_{23}^{\mp}}{V} &= -\frac{\pi}{2}(1 + 0,6K) \left[\beta_1 + x_0\gamma_1 + (z_0 - R)\psi_{12} + \frac{\pi}{2}\psi_{13} + \frac{\pi}{2}(z_0 + R) \right] - \\ &\quad - \frac{2}{\pi}(1 - K) \left[\beta_2 + (z_0 - R)\gamma_2 + x_0\psi_{12} + \frac{\pi}{2}\psi_{23} + \frac{\pi}{2}\lambda x_0 \right] - \\ &\quad - \frac{1-K}{R} \left[\beta_3 + \frac{\pi}{2}\gamma_3 + x_0\psi_{13} + (z_0 - R)\psi_{23} - (z_0 - R)x_0\lambda \right] + \frac{\pi R(1+K)}{V^2\Delta t}. \end{aligned} \tag{11}$$

Залишилося розглянути останній випадок одночасно трьох можливих відхилень $\Delta x, \Delta z, \Delta \alpha$, яким відповідає 8 можливих комбінацій, показаних на рис. 5.

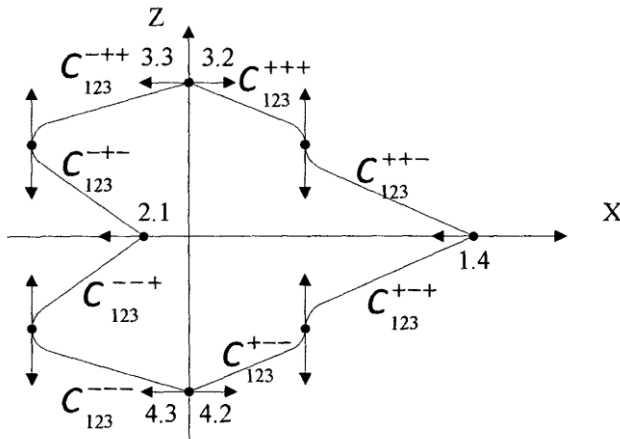


Рис. 5. Ілюстрація 8 польотних ситуацій при одночасному відхиленні по x, z та α від робочої точки

Зокрема, за $x = x_0 + R, z = z_0 + R, \Delta \alpha = \pi/2$, потрібно летіти до об'єкта 3.2, і відповідно ордината ризику C_{123}^{+++} дорівнює:

$$\begin{aligned} C_{123}^{+++} &= \frac{1 - 8K}{3,2} \left[\beta_1 + (x_0 - R)\gamma_1 + (z_0 - R)\psi_{12} + \right. \\ &\quad \left. + \frac{\pi}{2}\psi_{13} + \frac{\pi}{2}\lambda(z_0 + R) \right] + \\ &\quad + 0,9 \left[\beta_2 + (z_0 + R)\gamma_2 + (x_0 + R)\psi_{12} + \right. \\ &\quad \left. + \frac{\pi}{2}\psi_{23} + \frac{\pi}{2}\lambda(z_0 + R) \right] + \\ &\quad + \frac{0,1 - K}{R} \left[\beta_3 + \frac{\pi}{2}\gamma_3 + (x_0 + R)\psi_{13} + \right. \\ &\quad \left. + (z_0 - R)\psi_{23} + \lambda(x_0 + R)(z_0 - R) \right] + \\ &\quad + \frac{3,2R(1 + 5K)}{V^2\Delta t}. \end{aligned} \tag{12}$$

Таким чином, якщо проаналізувати всі розглянуті вище польотні ситуації, то можна побачити, що загалом 27 різних ординат, а прирівнювання їх один одному для оцінки 10 шуканих коефіцієнтів $\beta_i, \gamma_i, \psi_{ik}, \lambda$ функції Белмана представляє важко осяжне число систем рівнянь, що дорівнює кількості

поєднань C_{27}^{10} . Тому був використаний спрощений підхід, для якого додатково будуть потрібні так звані «середні» ординати ризику C_{13} та C_{23} , обчислені за формулами:

$$\begin{aligned} C_{13} &\approx 0,25(C_{13}^{++} + C_{13}^{\pm} + C_{13}^{\mp} + C_{13}^{--}); \\ C_{23} &\approx 0,25(C_{23}^{++} + C_{23}^{+-} + C_{23}^{-+} + C_{23}^{--}). \end{aligned}$$

Наближений аналітичний синтез функції Белмана в задачі маршрутизації польоту. Прирівнюючи один одному ординати C_{13} та C_{23} згідно з викладеною в [8] методикою, можна відразу обчислити один з коефіцієнтів функції Белмана:

$$\gamma_3 = \frac{(2,7 - 4K)R^2}{V^2\Delta t}. \tag{13}$$

Оскільки при прирівнюванні один одному ординат ризику існує надмірна кількість варіантів складених рівнянь, використовуємо спочатку лише ординати ризику при одиночних відхиленнях по x, z і α випадках, які пов'язані з подальшими спрощеннями.

Це дозволяє відразу обчислити коефіцієнти γ_1 та γ_2 і отримати співвідношення, що полегшують подальші розрахунки.

Прирівнюємо ординати $C_1^+ = C_1^-$. Тоді за допомогою формул (12) та (13) отримуємо:

$$\gamma_1 = \frac{0,1(1 - K)}{V^2\Delta t}. \tag{14}$$

Тоді за умови $C_1^* = C_0$ отримуємо рівність:

$$\beta_1 + \gamma_1 x_0 + \psi_{12} z_0 = \frac{-1,65R(1 + K)}{V^2\Delta t}. \tag{15}$$

Прирівнюючи ординати $C_2^+ = C_2^-$ за допомогою формул (15) отримуємо:

$$\begin{aligned} &0,8(x_0\psi_{12} + \beta_2 + z_0\gamma_2) + \\ &\frac{1}{R} [\beta_3 + x_0\psi_{13} + z_0(\psi_{23} + x_0\lambda)] = -2KR^2\psi_{12}. \end{aligned} \tag{16}$$

За умови рівності $C_3^+ = C_3^-$, за використовуючи формулу (16) отримуємо:

$$x_0\psi_{12} + \beta_2 + z_0\gamma_2 = 0. \tag{17}$$

Тоді з рівностей (16) та (17) отримуємо:

$$\beta_3 + x_0\psi_{13} + z_0(\psi_{23} + x_0\lambda) = -2KR^2\psi_{12}. \quad (18)$$

Прирівнюючи ординати $C_1^* = C_3^+$; $C_1^* = C_2^+$ отримаємо рівність:

$$\psi_{23} + \lambda x_0 = \frac{0,9R(1+K)}{V^2\Delta t}. \quad (19)$$

Звідси можна визначити коефіцієнт γ_2 функції Белмана:

$$\gamma_2 = \frac{1,4 - 4,6K}{V^2\Delta t}. \quad (20)$$

Рівності (15 – 19) будуть потрібні для обчислення решти 7 коефіцієнтів функції Белмана $-\beta_1, \beta_2, \beta_3, \psi_{12}, \psi_{23}, \psi_{13}, \lambda$ знайдених таким чином.

Прирівнюючи ординати

$$C_{13}^{++} = C_1^+; C_{123}^{++} = C_1^+; C_{12}^{++} = C_1^+,$$

отримаємо систему трьох лінійних рівнянь, щодо невідомих коефіцієнтів $\psi_{12}, \psi_{13}, \lambda$.

Рішення цієї системи дає такі результати:

$$\psi_{13} = \frac{z_0(1,3 - 5,6K) - R(0,34 + 4K)}{V^2\Delta t};$$

$$\psi_{12} \cong 0; \lambda = \frac{-1,3 + 6,4}{V^2\Delta t}. \quad (21)$$

Обчислення коефіцієнтів не становить труднощі. Знаючи значення ψ_{12} і γ_1 з рівності (15) знаходимо коефіцієнт β_1 , з рівності (16) при відомих значеннях ψ_{12} і γ_2 знаходимо коефіцієнт β_2 . Знаючи значення ψ_{13} та умова (19), знаходимо коефіцієнт β_3 за допомогою рівності (18), а коефіцієнт ψ_{23} при відомому значенні λ визначаємо з рівності (19).

Тоді отримаємо:

$$\beta_1 = -\frac{1,65R(1+K) + 0,1(1-K)x_0}{V^2\Delta t};$$

$$\beta_1 = -\frac{4,65K - 1,4}{V^2\Delta t};$$

$$\beta_3 = -\frac{(x_0R(0,35 + 4K) - 0,9z_0R \times (1+K) - x_0z_0(1,3 - 5,6K))}{V^2\Delta t};$$

$$\psi_{23} = \frac{0,9R(1+K) + (1,3 - 5,6K)x_0}{V^2\Delta t}. \quad (22)$$

Цим завершується аналітичний розрахунок всіх 10 коефіцієнтів функції Белмана, які обчислюються за формулами (13-14), (20-22). Потрібно підкреслити, що всі ці формули входять як параметри V і R , що характеризують властивості БПЛА, і характеризують зовнішнє середовище параметри: x_0 – середнє значення координат x_j ($j = 1 \dots M$) рухомих об'єктів, z_0 – середнє значення координат z_j ($j = 1 \dots M$) цих об'єктів, Δt – середній час перельоту з одного об'єкта в інший як фактор насиченості контрольованого регіону, $K_j = \frac{V_{uj}}{V}$ параметри відносної швидкості руху об'єктів.

Нарешті, якщо підставити знайдені коефіцієнти до виразу (4) функції ризику F_j прирівняти цю функцію деякій константі, то можна отримати межу області тяжіння, де значення ризику однакове. Отримана геометрична інтерпретація, подана на рис. 6, вказує на те, що знайдена фігура дуже схожа на

«скручений» еліптичний циліндр, переріз якого є еліпс, повернутий на кут, що дорівнює куту $\alpha + \Delta\alpha_j K$, де характеризує політ БПЛА, $\Delta\alpha_j$ – рух об'єкта.

Тому при виборі чергового об'єкта очевидна перевага надається тим об'єктам, які знаходяться та рухаються у напрямку польоту БПЛА, і в другу чергу – об'єктам, рух яких відповідає зближенню з БПЛА.

Об'єкти ліворуч і праворуч менш переважні, а останньою за пріоритетом є частина регіону в задній півсфері, що повністю відповідає фізичному змісту завдання, що розв'язується.

Тому у випадку області тяжіння – як «скрученого» еліптичного циліндру, показаний на рис. 6.

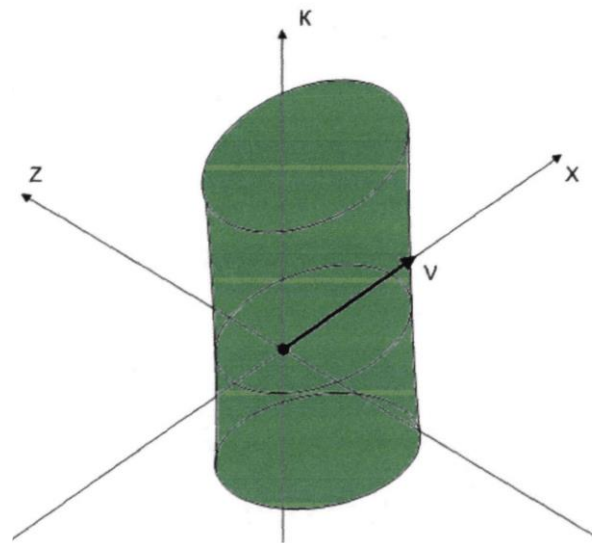


Рис. 6. Геометрична інтерпретація області тяжіння у вигляді «скрученого» еліптичного циліндра

Апроксимація функції Белмана та приклад роботи отриманого алгоритму маршрутизації. Отриманий аналітичний результат передбачається використовувати на кожному кроці вибору чергового наземного об'єкта, додаючи до маршруту той, що має мінімальне значення F_j . Приклад роботи отриманого алгоритму наведено на рис. 7.

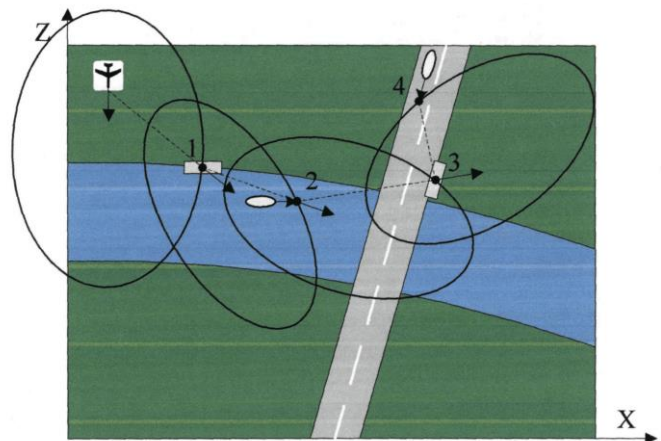


Рис. 7. Приклад послідовного вибору об'єктів, що спостерігаються

Отже, обчислення функцій ризику відбувається багаторазово. Тому цю процедуру доцільно спростити. Скористаємося тригонометричною формулою для опису межі однакового значення ризику у вигляді еліпса:

$$F_j^* = \sqrt{[(\Delta x_j - a) \cos \varphi + (\Delta z_j - b) \sin \varphi]^2 + M[(\Delta z_j - b) \cos \varphi + (\Delta x_j - a) \sin \varphi]^2 - 1}, \quad (23)$$

де a і b – координати центру, M – коефіцієнт «стиснення» його головних осей, φ – кут збільшення, Δx_j , Δz_j – збільшення координат розташування БПЛА при перельоті до об'єкту.

Якщо функцію F_j^* спростити, то при її розкладанні до ряду Тейлора можна отримати таку апроксимацію:

$$F_j^* = r_j + (\Delta x_j / r_j) \cdot [-\alpha + (M - 1)b\varphi] + \frac{\Delta z_j}{r_j} [-Mb + (M - 1)a\varphi] - \frac{ab\varphi(M - 1)}{r_j}. \quad (24)$$

Формула (24) зручна для зіставлення з виразом (4), у результаті параметри «повернутого» еліпса виявляються рівними

$$\begin{aligned} a &= 1,5R(1 + K) + 0,1x_0(1 - K); \\ b &= z_0; \varphi = \alpha + K\Delta\alpha_j; \\ M &\approx 1,4 - 4,6K. \end{aligned} \quad (25)$$

Залежність параметрів еліпса a, b, φ, M від x_0, z_0 також відповідає характеру переваги у прий-

нятті рішень. У формулах (10) кут повороту еліпса скоригований на значення $K\Delta\alpha_j$, що направляє до випередженої точки зустрічі з рухомих об'єктом.

Висновки

В статті вирішено завдання маршрутизації об'єкту групи рухомих об'єктів БПЛА. При вирішенні завдання враховувались: стала швидкість об'єкту моніторингу на інтервалі часу та динаміка руху БПЛА.

Критерієм оптимальності маршруту прийнято умову максимуму числа контрольованих об'єктів за заданий час.

Оскільки прийняття рішення для вибору точки польоту здійснюється на кожному кроці крім поточного, використовується система штрафів у вигляді втраченого часу, що впливає на вибір наступної точки-об'єкту. З огляду на зазначені обмеження застосовано метод оптимізації - динамічне програмування.

При прирівнюванні один одному ординат ризику виникає надмірна кількість варіантів складених рівнянь.

В роботі застосовано спрощений підхід, для якого додатково визначаються «середні» значення ординат ризику. При побудові маршруту на кожному кроці вибору чергового наземного об'єкта додаються до маршруту ті, що мають мінімальне значення штрафу.

В роботі отримано аналітичний результат послідовного вибору об'єктів, що спостерігаються.

СПИСОК ЛІТЕРАТУРИ

1. Интеграционный подход к задаче выбора маршрута группы БПЛА / А.Н. Козуб, Д.П. Кучеров // Искусственный интеллект. — 2013. — № 4. — С. 333–343
2. Research on efficient online planning of emergency logistics path based on doublelayer ant colony optimization algorithm / D.Liu, S.Li // International Journal of Computers and Applications. — 2018. — № 33. — P. 1–7
3. Генерація субоптимальних маршрутів безпілотного літального апарата з використанням нейронної мережі Хопфілда / Журавська І.М. // Проблеми інформаційних технологій. — 2018. — № 1. — С. 181–185
4. Розробка алгоритмів розв'язання задачі маршрутизації транспортних засобів з часовими вікнами / М.І. Огурцов, О.М. Ходзінський // Комп'ютерна математика. — 2016. — №1. — С. 134–142
5. Realizing the UAV potential through vehicle routing optimization, Applied Energy. — 2019. — № 242. — P. 1164–1175
6. Транспортно-комунікаційна задача для груп безпілотних апаратів / В.Ю. Корольов, М.І. Огурцов // Математичні машини і системи. — 2017. — № 1. — С. 82–89
7. Hank Zumbahlen. Linear Circuit Design Handbook Analog Devices Inc. — 2008. — 954 p.
8. Теорія оптимальних розгалужених траєкторій / О. І. Лисенко, О. М. Тачиніна, С. О. Пономаренко, О. Г. Гуйда. — К: КПІ ім. Ігоря Сікорського, 7БЦ. — 2023. — 260 с.

Received (Надійшла) 06.02.2023

Accepted for publication (Прийнята до друку) 26.04.2023

Routing model of flight of mobile objects by unmanned aircraft

R. Kulish, O. Matiushchenko

Abstract. An analysis of the experience of using unmanned aerial vehicles proves that the efficiency of performing tasks is directly related to the quality of route planning. Most route planning methods are focused on finding and monitoring stationary objects. However, at present, the solution of the problem of monitoring mobile, moving objects is of much greater importance. The paper solves the problem of planning a route for flying around mobile objects that maintain their direction of motion for a long period of time using the method of mathematical optimization - dynamic programming. When solving the problem of constructing an optimal route, the dynamics of the flight of an unmanned aerial vehicle was taken into account, which makes it possible to sequentially add objects to the route. When the ordinates of risk are equated to each other, there is an excessive number of variants of compound equations. The work uses a simplified approach, for which the "average" values of the risk ordinate are additionally determined. When building a route, at each step of selecting the next ground object, those with the minimum penalty value are added to the route.

Keywords: unmanned aerial vehicle, mobile groups, routing, monitoring, moving object, dynamic programming.

О. А. Мураховська

Національний аерокосмічний університет ім. М. Є. Жуковського «ХАІ», Харків, Україна

СИСТЕМНО-СПРЯМОВАНИЙ ПІДХІД АНАЛІЗУ МОЖЛИВИХ МЕТОДІВ ДОСЛІДЖЕННЯ АЕРОДИНАМІЧНОЇ ЖИВУЧОСТІ ЛІТАКА

Анотація. При створенні нового покоління високо маневрених літаків однією з основних вимоги є висока надійність і живучість, зокрема, він повинен зберігати працездатність при втраті ділянок несучих і керуючих поверхонь. На основі системно-спрямованого підходу здійснено аналіз існуючих досліджень аеродинамічних характеристик і динаміки польоту літального апарату з урахуванням специфічних вимог, пов'язаних з дослідженням особливих польотних ситуацій при втраті ділянок несучих і керуючих поверхонь. Розглянуто можливий варіант побудови узагальненого ранжування альтернатив та визначення вагових коефіцієнтів експертів із застосуванням методів статистичної обробки експертних оцінок при груповому узгодженні для удосконалення аналізу ефективності заходів щодо підвищення цього виду відмовостійкості.

Ключові слова: літальний апарат, аеродинамічна живучість, методи дослідження особливих польотних ситуацій, експертні оцінки, ранжування, вагові коефіцієнти.

Вступ

Постановка проблеми. Розрахунково-теоретичні методи дослідження аеродинамічної живучості літака характеризуються оперативністю, економічною ефективністю і низькою трудомісткістю, можливістю проведення досліджень при неповному комплекті вихідних даних і цілим рядом інших переваг. У той же час, забезпечення високої достовірності результатів розрахунків можливо лише за умови попередньої апробації застосовуваних математичних моделей по дослідним даним і подальшої експериментальної перевірки найбільш характерних результатів (аналіз особливостей різних методів розрахунку особливих польотних ситуацій виходить за рамки даної роботи). Таким чином, розрахунково-теоретичні та експериментальні методи дослідження особливих польотних ситуацій не виключають, а взаємодоповнюють один одного [1]. Кожен з цих методів має свої переваги і недоліки, які визначають раціональну область його застосування, однак можливість їх застосування для специфічних досліджень, що аналізуються, як правило, не розглядається.

Основним завданням (метою) даної статті є аналіз з єдиних позицій існуючих експериментальних методів дослідження аеродинамічної живучості літака з урахуванням специфічних вимог, пов'язаних з дослідженням даного виду особливих польотних ситуацій. Побудова узагальненого ранжування альтернатив, що забезпечує вибір найбільш раціонального методу дослідження або їх сукупності, а також можливість оперативного аналізу великої кількості варіантів.

Основна частина

Продувки в трубах вважаються в даний час основним методом визначення аеродинамічних характеристик ЛА [2]. Зокрема, цей метод може бути використаний при дослідженні аеродинамічної живучості та вразливості. Його перевагами є:

- наочність, візуальна спостережуваність;
- можливість для випробувача втручатися у виконання програми експерименту, управляти ходом процесу;

- відносно невисока вартість і можливість довести широкі параметричні дослідження;
- можливість точного виходу на заданий режим (швидкість, кут атаки, підйомну силу) і, як наслідок, можливість строго проаналізувати повторюваність одержуваних результатів;
- можливість ізольованого аналізу впливу окремих факторів і параметрів;
- детальна розробка методу, багаторічний досвід проведення таких досліджень.

Недоліками застосування цього методу для вирішення завдань дослідження аеродинамічної живучості є:

- наявність підвіски (закріплення) моделі, що стискає її переміщення і обмежує можливості моделювання маневрів;
- обмежені розміри потоку і, як наслідок, вузькі межі переміщення моделі при модельованих маневрах; інтерференція з межами потоку і стінками труби;
- труднощі моделювання навколосвукових швидкостей польоту;
- неможливість проведення досліджень зі змінною густини повітря (моделюється висоти польоту) в ході експерименту;
- труднощі відділення в ході експерименту частин моделі.

Експерименти на "літаючих" в аеродинамічній трубці моделях [3] і на моделях в вертикальній штопорній трубці не знаходять в даний час широкого поширення. Їх основні недоліки:

- малі розміри моделі, що не дозволяють провести експеримент в області автомоделності за числами Рейнольдса;
- неможливість розмістити на моделі необхідний комплект вимірювальних приладів і апаратури системи керування;
- труднощі моделювання процесу звалювання (переходу від вихідного горизонтального польоту або маневру до режиму штопора).

Методи продувки в аеродинамічній трубці викривлених моделей і застосування ротативних машин не дозволяють змоделювати процес маневру літака. В ході одного експерименту адекватно відображається

єдина точка траєкторії. Моделювання наступної точки з іншої кривизною траєкторії зажадало б нової викривленої моделі або перенастроювання ротативних машини.

Методи дослідження коефіцієнтів аеродинамічних похідних несталої руху ЛА найчастіше засновані на вимушених коливаннях жорсткої моделі в потоці (кінематичні і динамометричні методи). У деяких випадках відхилення оточуючого потоку забезпечується поворотом розташованої перед моделлю решітки профілів. У цих методах моделювання маневру літака (обтікання моделі викривленим потоком) досягається лише для коротких проміжків часу. Повністю відсутній взаємозв'язок режиму польоту (кривизни траєкторії), що моделюється, і діючих на модель аеродинамічних сил. Крім того, у всіх варіантах цього методу рух моделі обмежено підвіскою тієї чи іншої конструкції.

Аналогічний висновок можна зробити про метод дослідження аеродинамічних характеристик на моделях, закріплених на літаку-носії або на ракетній візку, що рухається по рейковому шляху (випробування на ракетному треку). Перевагами в обох випадках є:

- практично необмежений обсяг і маса використовуваної вимірювальної і реєструючої апаратури;
- збереження об'єкта після випробувань, можливість його подальшого аналізу.

Основний недолік, який не дозволяє досліджувати цими методами динаміку польоту літака в особливих польотних ситуаціях — жорстке кріплення моделі до візка або літаку-носія, що виключає або обмежує свободу її переміщення; відсутність взаємозв'язку діючих на модель сил і траєкторії її руху.

Метод буксирних моделей забезпечує трохи більшу свободу переміщень моделі, однак моделювати процес звалювання і штопора літака він також не дозволяє.

В принципі, найбільш достовірні дані про характер руху ЛА в особливих польотних ситуаціях можуть бути отримані в льотних експериментах на натурних літаках, зокрема — при стрільбі реальними засобами ураження по натурних літаках в польоті, керованим мішеням і переобладнаним в мішені літакам [1]. Незважаючи на значну вартість експериментальної частини програм, інформативність проведених випробувань виявилася порівняно невисокою. Дійсно, неможливо проводити систематичне вивчення такої проблеми, маючи такий складний комплекс параметрів. На процес ураження аеродинамічної цілі істотно впливають особливості аеродинамічної схеми і компонування літака, закони керування, що реалізуються бортовою системою керування, тип і калібр використовуваних боєприпасів, ймовірний промах, швидкість і розмір осколків і, нарешті, який з агрегатів або систем літака уражено. З огляду на високу вартість кожного окремого експерименту, при такому широкому комплексі впливають параметрів практично неможливо набрати обсяг даних, необхідний для статистичної обробки і аналізу.

Таким чином, до недоліків цього методу дослідження аеродинамічної живучості слід віднести:

- високу вартість;
- широкий комплекс одночасно впливаючих і трудноураховуваних чинників;
- неможливість контролю повторюваності через складності отримання заданого, чітко визначеного пошкодження;
- неможливість точного визначення по уламках отриманого в польоті пошкодження — величини втраченої частини несучого агрегату, супутніх ушкоджень системи керування і т. п.;
- неможливість виконання випереджальних випробувань для проєктованих вітчизняних літаків і обмежені можливості проведення експериментів на зразках зарубіжної авіаційної техніки.

Джерелом інформації про аеродинамічну живучість літаків є також статистика бойових дій в локальних конфліктах і дані про льотних події [4-6]. Однак одержувані цими способами відомості характеризуються дуже низькою інформативністю. Окремі розрізнені випадки, по яких є дані, трудно зіставляти при аналізі. Недостатнім є для дослідження динаміки польоту літака і обсяг інформації, що реєструється бортовими накопичувальними пристроями. Практично дані про режими польоту літака зазвичай обмежені записом швидкості і нормального перевантаження. Відомості про кутові швидкості, кути відхилення рульових поверхонь, діях екіпажу, кутах крену, курсу і тангажу відсутні. Крім того, одержувані таким способом дані, як правило, відносяться до типів ушкоджень, що дозволяє літаку повернутися на свій аеродром, а найбільший інтерес при дослідженні бойової живучості літаків і ефективності систем ППО представляють близькі до критичних типи пошкоджень.

Вільнолітаючі динамічно подібні моделі (ВДПМ) набули широкого поширення в якості інструменту для випереджальних досліджень характеристик проєктованого (неушкодженого) літака і апробації технічних рішень, що характеризуються високим ступенем технічного ризику. ВДПМ широко використовують в практиці зарубіжного авіабудування [7], але особливості їх конструкції і методики випробувань зі зрозумілих причин не публікуються. Методичні проблеми проведення таких випробувань, але без тематики аеродинамічної живучості, розглянуті в книзі [8]. Застосування методів системного аналізу, здійснене в роботі [9], дозволяє істотно скоротити кількість потрібних випробувальних пусків при складанні програми льотних досліджень критичних величин ушкоджень несучих і керуючих агрегатів літака. Переваги цього методу при дослідженні проблем підвищення аеродинамічної живучості ЛА:

- необмежені переміщення моделі і, як наслідок, достовірність моделювання процесів звалювання і штопора;
- відсутність ризику загибелі екіпажу і втрати дорогих дослідних екземплярів майбутнього літака;
- відносно короткий термін проєктування та підготовки до випробувань ВДПМ;
- простота відстеження удосконалень в конструкції і роботі систем основного літака;
- відносно невисока вартість;

- багаторазовість застосування моделі та можливість проаналізувати повторюваність результатів;
- відпрацьованість методики підготовки і проведення таких випробувань на вільнолітаючих моделях.

Як показано в роботах [9-11], методи, що застосовуються при дослідженні аеродинамічної живучості ЛА повинні забезпечувати можливість обліку складного комплексу факторів, що включають тривимірний просторовий рух, зривне обтікання, явища аеропружності та автоаеропружності, відділення частин ЛА і аварійний скидання підвісних вантажів, роботу екіпажу та його помилки пілотування. Крім того, проведений аналіз показав надзвичайно високий рівень багатоваріантності досліджень. Тому вибір найбільш раціонального методу дослідження або їх сукупності – відповідальний етап, на якому необхідно врахувати мінімізацію витрат часу та коштів на аналіз одного варіанту, а також можливість оперативного аналізу великої кількості варіантів. Ступінь відповідності кожного методу сукупності критеріїв оцінюється ваговими коефіцієнтами, що призначаються групою експертів, та включає фахівців різного профілю. Для обробки та узгодження результатів їх роботи ефективно використовувати, запропоновану у роботі [12] методику побудови узагальненого ранжування альтернатив та визначення вагових коефіцієнтів експертів із застосуванням методів статистичної обробки експертних оцінок при груповому узгодженні. Блок-схему алгоритму приведено на рис. 1.

Метод складається з реалізації чотирьох етапів. На першому етапі за допомогою рекурентної процедури (див. рисунок 1) здійснюється розрахунок вагових коефіцієнтів експертів та групової оцінки

об'єктів. Нехай m експертів оцінили n альтернатив за l показниками. В результаті отримаємо величини x_{ij}^h та k_j^h ($i = \overline{1, n}$; $j = \overline{1, m}$; $h = \overline{1, l}$).

На другому етапі знаходиться кількісна оцінка ступеня узгодженості думок експертів. Для нестрогого ранжирування при наявності зв'язаних рангів застосовується дисперсійний коефіцієнт конкордації

W (рис. 1), де $S = \sum_{i=1}^n \left(\sum_{j=1}^m r_{ij} - \bar{r} \right)^2$, $\|r_{ij}\|$ – матриця

ранжування альтернатив; p – кількість експертів, у ранжируваннях яких є рівні ранги; T_j – показник зв'язаних рангів у j -му ранжуванні,

$$T_j = \sum_{k=1}^{H_j} (h_k^3 - h_k); \quad h_k - \text{число рівних рангів у } k\text{-й}$$

групи зв'язаних рангів при ранжируванні j -м експертом. Якщо всі експерти дають однакові оцінки, $W = 1$, за умов повної відсутності узгодженості оцінок – $W = 0$.

На третьому етапі оцінюється рівень узгодженості думок усіх експертів щодо рангу однієї альтернативи. Для слабо прогнозованих ситуацій необхідно враховувати інформацію про відстань між рангами, тому було обрано коефіцієнт варіації Беккера V_i^B (рис. 1), де C_m^2 – число сполучень з n по 2. Збільшення значення коефіцієнта означає розкид оцінок експерта щодо відповідної альтернативи.

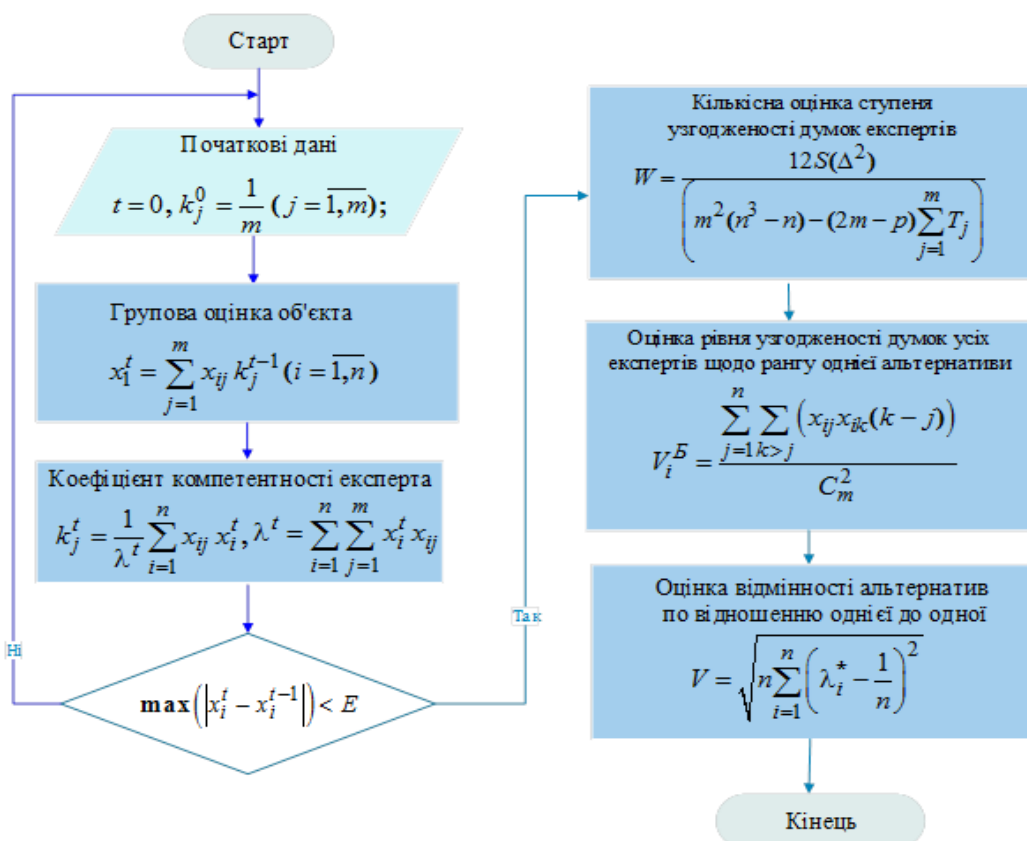


Рис. 1. Блок-схема алгоритму побудови узагальненого ранжування

На четвертому етапі здійснюються оцінка відмінності альтернатив по відношенню однієї до одної.

Таку оцінку дає коефіцієнт варіації V (рис. 1), розрахований за елементами узагальненого вектор значущості, де λ_i^* – узагальнена значущість i -го об'єкта

Чим ближче значення коефіцієнта V до одиниці, тим більше відчутна різниця між значущістю провідних альтернатив та значущістю альтернатив-аутсайдерів.

Для апробації модельної задачі використано фрагмент критеріальної бази, яку розроблено в [10]. В табл. 1 представлено вхідні дані та результати розрахунків за алгоритмом, приведеним на рис. 1. Результати розрахунків свідчать про досить високу узгодженість оцінок експертів (коефіцієнт конкордації $W = 0,81$). Результати групової оцінки факторів та коефіцієнти компетентності експертів досить добре узгоджуються з результатами отриманими в роботі [10] і дозволяють удосконалити результати отримані в попередніх дослідженнях.

Таблиця 1 – Вхідні дані та результати розрахунків за алгоритмом побудови узагальненого ранжування

Можливість урахування факторів	Експерт					Групова оцінка факторів		
	1	2	3	4	5	Номер ітерації		
						1	2	3
Початковий режим польоту: швидкість,	0,9	0,8	1	0,8	1	0,9	0,907	0,907
висота польоту,	0,9	0,8	1	0,8	1	0,9	0,907	0,907
кут тангажу	1	0,9	1	0,9	1	0,96	0,963	0,963
кут крену	0,9	0,8	1	0,7	1	0,88	0,889	0,889
перевантаження у момент ушкодження	0,8	0,8	0,9	0,7	0,9	0,82	0,826	0,826
Ступінь (величина) пошкодження агрегату чи системи	0,8	0,8	0,8	0,7	0,9	0,8	0,804	0,804
Дії екіпажу	1	0,9	0,9	0,8	1	0,92	0,932	0,924
Помилки екіпажу	0,9	0,7	0,7	0,8	0,8	0,78	0,779	0,779
Робота САУ штатна	0,7	0,6	0,6	0,7	0,8	0,68	0,682	0,682
Ушкодження САУ	0,5	0,55	0,6	0,5	0,8	0,59	0,597	0,597
Робота САУ з урахуванням роботи бортової системи діагностики та коригування законів управління	0,8	0,8	0,9	0,7	0,8	0,8	0,804	0,804
Жорсткість несучих агрегатів	0,7	0,75	0,8	0,7	0,8	0,75	0,753	0,753
Жорсткість корпусних агрегатів	0,8	0,7	0,9	0,7	0,8	0,78	0,785	0,785
Жорсткість проводки управління	0,7	0,5	0,9	0,7	0,8	0,72	0,727	0,727
Частотні характеристики несучих агрегатів	0,3	0,2	0,5	0,3	0,5	0,36	0,368	0,368
Частотні характеристики корпусних агрегатів	0,5	0,7	0,7	0,5	0,7	0,62	0,625	0,625
Частотні характеристики проводки управління	0,3	0,8	0,7	0,3	0,7	0,56	0,569	0,568
Номер ітерації	Коефіцієнт компетентності експерта					W = 0,81		
	1	2	3	4	5			
1	0,198	0,189	0,216	0,177	0,220			
2	0,198	0,189	0,215	0,179	0,219			
3	0,198	0,189	0,216	0,177	0,220			

Висновки

Виконаний аналіз застосовності різних методів для дослідження динаміки польоту в особливих польотних ситуаціях показав, що для отримання експериментальних даних високого ступеня достовірності ефективним є випробування великомасштабних вільнолітаючих динамічно подібних моделей літаків. Раціональним є спільне застосування ВДПМ (для отримання інформації про найбільш складні та характерні ситуації) і апробованих за експериментальними даними математичних моделей (для проведення широких параметричних досліджень). Однак при застосуванні літаків можуть виникати ушкодження, що викликають не втрату агрегату, а

значне зниження його жорсткості, яке може привести до розвитку небезпечних явищ аеропружності. Такі ситуації також підлягають теоретичному і експериментальному дослідженню.

Апробована на модельній задачі методика побудови узагальненого ранжування альтернатив та визначення вагових коефіцієнтів експертів із застосуванням методів статистичної обробки експертних оцінок при груповому узгодженні показала досить добрі результати, які узгоджуються з результатами попередніх досліджень. Її можна використовувати при виборі найбільш раціонального методу дослідження або їх сукупності, а також можливість оперативного аналізу великої кількості варіантів.

СПИСОК ЛІТЕРАТУРИ

1. Энциклопедия безопасности авиации / Н.С. Кулик, В.П. Харченко, М.Г. Луцкий и др.; Под. ред. Н.С. Кулика. – К.: Техніка, 2008. – 1000 с.
2. Летные исследования и испытания. Фрагменты истории и современное состояние: Научно-техн. сборник. – М.: Машиностроение, 1993. – 496 с.
3. Методы исследований на летающих моделях [Текст] / Миронов А. Д., Владычин Г. П., Кондратов А. А. и др. – М.: Машиностроение, 1988. – 144 с.
4. Новожилов Г. В., Неймарк М. С., Цесарский Л. Г. Безопасность полета самолета. Концепция и технология. – М.: Машиностроение, 2003. – 144 с.
5. Рыженко А. И. Живучесть авиационных силовых установок : учебник для студентов высших учебных заведений (направление "Авиация и космонавтика") / А. И. Рыженко, В. С. Кривцов. — Х.: Нац. аэрокосм. ун-т "ХАИ", 2004. — 659 с.
6. Рыженко А. И., Рябков В. И. Причины авиационных происшествий с гражданскими и военными самолетами / А.И. Рыженко, В.И. Рябков — Учеб. пособие по курсу "Надежность и живучесть самолетов". — Харьков: Харьк. авиац. ин-т, 1997. — 208 с.
7. Концепции развития современной авиационной техники основных назначений [Текст] : инновац. учеб. для неавиационных специальностей аэрокосм. ун-та / В. А. Богуслаев, А. И. Рыженко, Е. А. Мураховская, Р. Ю. Цуканов. — Запорожье : Просвіта, 2020. – 707 с.
8. Chambers, J. R. Modeling flight: The Role of Dynamically Scaled Free-Flight Models in Support of NASA's Aerospace Programs [Text] / J. R. Chambers // NASA SP 2009-575. – USA, Washington, 2010. – 202 p.
9. Murahovska, O. A. Designing the Set of Flight Experiments for Exploration of Critical Flight Conditions in Case Failure and Damage of Lifting and Control Aircraft Components with a Point of View of Systems Approach (Системный подход к разработке программ экспериментов по исследованию критических режимов полета при отказах и повреждениях несущих и управляющих агрегатов) / О. А. Murahovska, О. І. Ryzhenko // Открытые информационные и компьютерные технологии. – Харьков: НАУ «ХАИ», 2018. – Vol. 79. Pp. 140–147.
10. Вознюк С. Н. Фактор многовариантности при выборе метода исследования проблем выживаемости в условиях неопределенностей проектирования самолетов военного назначения / С. Н. Вознюк, Л. И. Курпа, Е. А. Мураховская, А. И. Рыженко // Вопросы проектирования и производства конструкций летательных аппаратов: сб. науч. тр. Нац. аэрокосм. ун-та им. Н.Е. Жуковского «ХАИ». – Вып. 4(72). – Х., 2012. – С.21-30
11. Макаричев В. А. Критерияльная база ранжирования альтернативных методов исследования проблем безопасности полетов в условиях неопределенности проектирования самолетов гражданского назначения / В. А. Макаричев, Е. А. Мураховская, А. И. Рыженко, Ю. А. Щербакова // Открытые информационные и компьютерные технологии. – 2012. – Вып. 57. – С. 197–206.
12. Мураховська О. А. Аналіз застосування методів статистичної обробки експертних оцінок на етапі попереднього проектування складних технічних пристроїв / О. А. Мураховська, Н. А. Українець // Системи управління, навігації та зв'язку. – 2022. – № 3 (69). – С. 39-44. doi: 10.26906/SUNZ.2022.3.039

Received (Надійшла) 21.02.2023

Accepted for publication (Прийнята до друку) 03.05.2023

A system-oriented approach to the analysis of possible methods of studying the aerodynamic survivability of an aircraft

Olena Murahovska

Abstract. When creating a new generation of highly maneuverable aircraft, one of the main requirements is high reliability and survivability, in particular, it must maintain operational efficiency in the event of the loss of parts of the bearing and control surfaces. On the basis of a system-oriented approach, an analysis of existing studies of aerodynamic characteristics and flight dynamics of the aircraft was carried out, taking into account the specific requirements associated with the study of special flight situations when areas of the bearing and control surfaces are lost. Considered a possible option of building a generalized ranking of alternatives and determining weighting coefficients of experts using methods of statistical processing of expert evaluations with group agreement to improve the analysis of the effectiveness of measures to increase this type of fault tolerance.

Keywords: aircraft, aerodynamic survivability, research methods of special flight situations, expert evaluations, ranking, weighting factors.

Naghiyeva Parvin Abdulla

Baku State University, Baku, Azerbaijan

THEORETICAL ANALYSIS OF THE LEVEL OF COMMUNICATION SECURITY OF UAVS OPERATING AS PART OF A GROUP PERFORMING A COMBAT MISSION

Abstract. The research relevance lies in the issue of information impact on the AI of the drone and its vulnerability in order to perform sabotage actions by the enemy in local military conflicts. The article provides a theoretical analysis of the security of the information transmitted to a group of drones in order to establish the susceptibility of drones to false data and further organizational actions that can prevent the sabotage actions committed by the enemy and his allies, as well as analyzes situations when an aircraft is misinformed, which in turn entails the failure of the task by the entire group of drones. The subject and the purpose of the research is to establish the existence of the fact of misinformation by the enemy of the UAV, controlled remotely and performing a combat mission as part of a group without the direct participation of the operator. In the process of modeling the situation, theoretical analyzes of the behavior of the UAV were carried out, since the artificial intelligence in one aircraft is only able to influence its behavior and is not able, under false information impact, to establish the fact of false interference by the enemy in another drone in-group in the course of combat operations and warn other elements of the system in its group. Research methods are theoretical analyses of the communicative interaction of a swarm of drones based on the study of the research of algorithms of operator control. Obtained results and conclusion: encrypted false information transmitted with the aim of disorienting the UAV cannot be detected in advance by the classical approach to preventing cyber-attacks. In this regard, the problem of information security of drone AI remains relevant and requires further comprehensive theoretical and practical research with the involvement of IT specialists, linguists, engineers, and military experts.

Keywords: unmanned aerial vehicle, group of unmanned aerial vehicles, artificial intelligence, information security, data transmission.

Introduction

Currently, armed UAVs play an essential role in military operations during any local military conflicts. The situation in the Second Karabakh war in 2020 clearly showed and gave the idea to AI developers, engineers, and UAV designers that today remotely controlled aircraft can perform the combat missions which a few decades ago required numerous ground combat equipment, a professional army, and people's lives [1-6].

Unmanned aviation of states with a small professional army can not only participate in local confrontations with minimal damage to military personnel but also act as allies of other states in the international arena [7].

Ground operations are gradually moving to combat missions conducted in the air, namely, drones can solve the issue of the speed of the military conflict, reduce losses on both sides, and provide comprehensive surveillance and reconnaissance of military areas.

For the purposes of this study, the author sets the task to determine whether the enemy has a remotely controlled UAV performing combat tasks in a group without direct participants, the fact of sabotage against a drone that can provoke disruption in the work of the entire drone group, and, consequently, losses and failure in completion of the task.

The subject of the study is false information transmitted to a drone operating in a group that performs combat missions with no human factor.

The relevance of the study lies in the issue of information influence on the AI of drones and their vulnerability. The research method consists in a theoretical analysis of the communicative relationship of a drone group based on operator control algorithms. To

determine if there is a vulnerability in the system of a remotely controlled drone that performs combat missions in a group the analysis of the transmitted data has been carried out.

The results of the research [8] show that the issue of informational support of the drone performing the task in the group, as well as the entire group of drones, is the most important factor influencing the performance of the task, and requires developers to strengthen security systems against various enemy programs. Encrypted false data transmitted for the purpose of disinformation of UAVs cannot reveal the classic approach to preventing cyber-attacks in advance. The problem of drone information security currently remains relevant and will require further comprehensive theoretical and practical research with the participation of IT experts, engineers, linguists, and military specialists.

Practical value. The theoretical assessment of drone behavior in the situation, when it receives false information and its further interaction with other drones in the group, can be used for deeper practical research and the development of new methods for the security of transmitted information in order to eliminate vulnerabilities in the AI control system in the future local wars.

The second Karabakh war in 2020 clearly showed the civilized world that the speed of high-tempo combat engagements increased due to AI-enabled ISR and rapid command and control. Like any robotic equipment, UAVs have a number of vulnerabilities, and the efficiency of the equipment is to obtain reliable data and true information in order to complete the tasks assigned to the AI. At each stage of software development, there are a number of information security issues, so the main aim of this article is to analyze simulated situations when a drone receives false information and starts to perform a task based on that false information.

To reduce the risk of obtaining false data channels are analyzed for the presence of the threat of sabotage in the papers [9]. The results of the research confirm that in order to reduce the risk of obtaining false data, it is necessary to use a cryptographic encryption method.

The study [10] analyzed standard data management between drones, as well as between a drone and a human operator. According to the results of the research, AI is able to control remotely in real-time, but the standard way of transmitting information without any proper data security cannot be applied to military purposes, that is for reconnaissance, etc. due to the rather vulnerable transmission system.

Research [11] considers the issue of a group of drones, as well as an unmanned surveillance and reconnaissance aerial vehicle controlled from a ground control station (GCS). The analysis showed the vulnerabilities of communication channels, as well as the swarm of drones, that indicate cyberattacks of opponents aimed at diversification. The information received is intended for centralized control of the swarm, so it is not relevant for decentralized or single UAV control.

The author also notes dissertation research [12], which analyzes the drone network on its information vulnerability and its susceptibility to various cyberattacks. Namely, the paper discusses the methods of organizing unmanned networks in civil and military aviation, but there is no analysis of the decentralized control of UAVs in combat situations.

Analyzing the above researches on ensuring information security, as well as on the level of security, the author assumes that identifying a vulnerability in the field of information security of combat UAVs is a key task at the development stage, as well as an integrated approach to the development of AI systems used for military purposes.

Therefore, the following tasks are formulated:

1. Analysis of existing approaches and research studies in the field of developing drone information security within the group;

2. Develop an algorithm for modeling the interaction of UAV groups, taking into account the data and conclusions of existing experiments;

To accomplish the stated tasks, it is necessary to apply new approaches to data transmission, and a comprehensive analysis of the existing interaction system, since the proposed and existing concept is not able to protect drones from cyberattacks, and the main task of the drone is to perform the task transmitted via the communication channel.

Centralized control of a drone swarm. The main difference between centralized and decentralized control is the presence of a system that comprehensively controls the performance of drones while conducting combat missions. This element controls all the elements in the swarm system. The main advantage of this information and communication interaction of the drone is the synchronization of tasks using AI algorithms, but this control method has a number of vulnerabilities.

Let us consider the vulnerabilities:

1. If a swarm consists of a significant number of UAVs, the swarm central control system needs sufficient

communication channels to synchronize the data of each drone in the swarm, and a powerful system for encrypting and decrypting transmitted data to avoid interception and false information impacts.

2. With the information destruction of the control center, all drone systems disrupt the synchronization of the assigned tasks, since a swarm of drones controlled by one drone system performs only the tasks assigned from the main system.

3. In case of a cyber-attack or sabotage, a swarm controlled centrally and remotely from one place disrupts communication with the center and other drones, since the AI of a single drone in a swarm can only influence its own behavior, and does not have the function of transmitting information to other drones in the group.

Particular attention here can be paid specifically to armed drones since they perform combat missions. The advantage of centralized collective drone control can be called the speed of completing tasks and ensuring the effectiveness of a sudden strike on the enemy, and this in turn is an example of military tactics during combat operations.

A drone system built on the basis of the swarm algorithm due to the lack of communication channels has the highest viability, and, as a result, a high percentage of task completion. A minimum load of operators regardless of the number of drones can be considered another advantage of a drone swarm controlled by means of centralized control.

The disadvantage of this UAV control method is the complexity of the algorithms for controlling the control element and transmitting data, which can lead to a disruption in the service of the control system during sabotage intervention or the transmission of false information.

Moreover, the information and communication interaction of a drone in a group has a high level of vulnerability. In order to consider the disadvantages and the means, of using them to decipher the swarm, it is necessary to consider the general model for transmitting information to devices that are controlled via the centralized control system.

To achieve the goal of this study, the author sets the task of establishing the presence of disinformation by the enemy of the UAV, which is controlled remotely, and performs a combat mission in a group without its direct participation, since the fact of sabotage against one drone can provoke a disruption in the operation of the entire drone group, and, consequently, can cause losses and failure to achieve the goal. Therefore, the author assumes that information and communication interaction in a group of drones, without the direct participation of operators, is vulnerable when the drone perceives false information. To analyze in more detail the information and communication interaction of the UAV, the external and internal factors that affect the efficiency of the control system are considered.

Internal information and communication interactions of UAVs include the following elements:

- Sensors that collect information about changes in the area, weather conditions, and barriers;

- The processor, which takes part in the decoding of the received commands from the CGC;
- Engine and other technical elements that ensure the work of UAVs.

In order to fulfill tasks during the operation of a drone in a swarm, information is transmitted through internal information and communication channels. For example, information about coordinates, targets, the roll and pitch angles, area, the technical condition of the drone, a command to adjust the position in space, the location of the drone in relation to other drones in the group, the state of the drone and the command for combat missions.

The external information and communication interaction of UAVs consists in the exchange of

information about their groups of drones and other groups of drones, about the location, changes in combat missions, losses, technical conditions and other data that ensure efficient distribution of tasks in real time.

Let's simulate the operation of a group of drones of the AnyLogic simulator (Fig. 1).

The swarm is represented in a 10 x 10 square cells, consisting of 5 UAVs. Each UAV has its own cell, and identification number from 1 to 5.

The goal of the group is to complete 10 tasks in the form of moving the "flags" to the zone, schematically indicated in gray in Fig. 1a. In order to fulfill the assigned tasks, the control center conducts centralized swarm management, distributing common tasks for each drone.

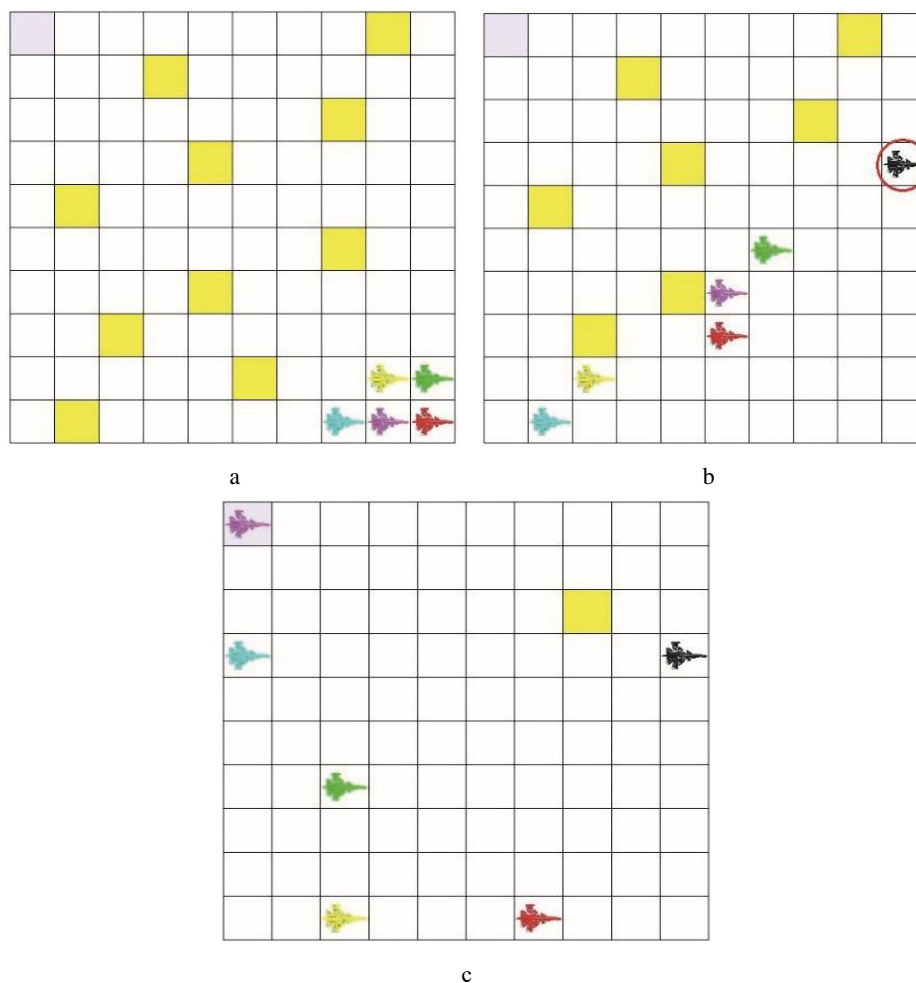


Fig. 1. Disruption in the operation of a group of unmanned aerial vehicles: group at the start of functioning (a); introduction of a false agent into a functioning group (b); disruption in the operation of the group (c)

A drone receives a task to perform if it is free from other tasks and is located closer than others drones to the system control center.

Tasks have a serial number and may have the status "failed" when data is accepted and not executed. Also, the task is considered failed if one of the UAVs does not accept it. Only the command executed by all drones gets the status of the task in progress. The status of the «completed» task is received only by those commands that are completed by all drones in the swarm.

To experimentally prove the vulnerability of the

information and communication security of a group of drones, a drone, which will create conflict situations in the swarm by not accomplishing the task is launched. Since the vulnerability of swarm drones lies in their inability to influence other drones, and they are only capable of performing central control tasks, the swarm becomes disoriented and does not perform the task supplied by the central control system. The launched drone, which had false information about not completing the assigned tasks, provoked the entire swarm to fail to complete the assigned tasks.

Conclusion

The theoretical research is aimed at analyzing the security of data transmitted to UAVs in order to misinform and disrupt the combat mission performed remotely, without the direct involvement of operators. A theoretical analysis of scientific research on the chosen topic was carried out, namely, an analysis of the security of data transmitted to UAVs, for misinformation and disruption of the combat mission conducted in a group without the direct participation of an operator. The disadvantages of controlling swarms of drones by a single control center are analyzed. It has been established that external and internal

data exchange can expose all elements of the system and information to false information since a drone controlled by one center cannot detect violations of the behavior of other drones in its group. An experiment in which the built-in drone accepted the task but did not complete it, and therefore the task was not completed by the entire swarm was conducted.

The conclusion of the research paper is that cryptographic methods cannot protect the drone from hidden hacker influences, sabotage, and cyber-attacks on data, so it is necessary to develop new methods to counter external influences with the involvement of IT specialists, linguists, and experts in the field of military security.

REFERENCES

1. Kirichenko V.V. Information security of communication channels with UAV electronics and control systems. 2015. H.3. P.23-27. Doi: 10.18372, 1990, 5548.45.9892.
2. Hooper M, Tian Y, Zhou R, et al. Securing commercial WiFi-UAVs from common secure attacks Proc. IEEE International Communications Conference. 2016. P.1213-1218, id: 10. 1109 MILCOM. 2 Kirichenko V.V. Information security of communication channels with UAV electronics and control systems. 2015. H.3. P.23-27. doi: 10.18372, 1990, 5548.45.9892.
3. Bayramov, A.A., Hashimov, E.G. Application SMART for small Unmanned Aircraft System of Systems. Chapter // Handbook of Research on Artificial Intelligence Applications in the Aviation and Aerospace Industries. IGI Global, PA, USA, 2019. Eds. T. Shmelova, Y.Sikirda, A.Sterenharz, 390 p. Chapter 8. Application SMART for Small Unmanned Aircraft System of Systems. pp.193-213.
4. Rivera E., Baikov R., Gu G. Analysis of randomness and cybersecurity. Texas, United States, 2014.
5. Chung T. H., Jones K. D., M. A. Day, M. M. Day, M. M. Day 50 in 2015: Swarm vs. Team UAV Live-Fly Contest at the postgraduate school on the Volga. AUWSI, 2013, p. 1792-1811.
6. Hashimov, E.G., Khudeynatov E.K. Evaluation of the effectiveness of the application of UAV systems in modern wars // Baku: Military knowledge, - 2022. No. 1 (January-March), p. 11-17.
7. Javaid A.Y. Cyber security analytics and testing of an attack on an unauthorized aerial machine. PhD dissertation. University of Togliatti, 2015.016. 7895496.
8. Zikratov I.A., Zikratova T.V., Lebedev I.S., Gurtov A.V. Model of trust and reputation for objects of multi-agent robotic systems with decentralized control. Scientific and Technical Journal of Information Technology, Mechanics and Optics, 2014, no. 3. pp. 30-38.
9. Yakimenko O. A., Chung T. H. Using autonomous capabilities for systems with CORRUSER. Prospectus of the 28th Meeting of the International Union of Aeronautical Physics, ICAS 2012, 2012, p. 47-49.
10. Vatkin L., Li Kirichenko V.V. Information security of communication channels with UAV electronics and control systems. 2015. H.3. P.23-27. doi: 10.18372, 1990, 5548.45.9892.
11. Yang J. H., Kapolka M., Chung T. H. Autonomy in combat with the MUT command during a joint assault. Scientific and technical technologies and applications for robots, 2012, p. 561-5569. doi: 10" 1007 978-3642-3737-4-9 54".
12. Sidorov V., N. V. K., Lam K. I., Salle M. F. B. M. Cyberthreat analysis of UAV transport management for the city airport of the International Conference on Air Transportation. 2017.

Received (Надійшла) 13.02.2023

Accepted for publication (Прийнята до друку) 26.04.2023

Теоретичний аналіз рівня безпеки зв'язку БПЛА у складі групи виконання бойового завдання

П. А. Нагієва

Анотація. Актуальність дослідження полягає в питанні інформаційного впливу на ШІ безпілотної та його уразливості для здійснення противником диверсійних дій у локальних військових конфліктах. У статті проведено теоретичний аналіз захисту інформації, що передається групі БПЛА, з метою встановлення вразливості БПЛА до неправдивих даних та подальших організаційних дій, які можуть запобігти диверсійним діям противника та його союзників, а також аналізує ситуації, коли літальний апарат дезінформований, що в свою чергу тягне за собою невиконання завдання всією групою дронів. **Предметом та метою дослідження** є встановлення факту дезінформації противником БПЛА, що керується дистанційно та виконує бойове завдання у складі групи без безпосередньої участі оператора. У процесі моделювання ситуації були проведені теоретичні аналізи поведінки БПЛА, оскільки штучний інтелект в одному літальному апараті здатний лише впливати на його поведінку і не в змозі при впливі неправдивої інформації встановити факт неправдивої дії. втручання противника в іншу групу безпілотної у ході бойових дій та попередження інших елементів системи у своїй групі. **Метами дослідження** є теоретичний аналіз комунікативної взаємодії рою дронів на основі вивчення дослідження алгоритмів операторського керування. **Отримані результати та висновок:** зашифрована неправдива інформація, що передається з метою дезорієнтації БПЛА, не може бути заздалегідь виявлена класичним підходом до запобігання кібератакам. У зв'язку з цим проблема інформаційної безпеки ШІ безпілотної залишається актуальною та потребує подальших комплексних теоретичних і практичних досліджень із залученням IT-спеціалістів, лінгвістів, інженерів, військових фахівців.

Ключові слова: безпілотний літальний апарат, група безпілотної літальних апаратів, штучний інтелект, інформаційна безпека, передача даних.

С. В. Очеретенко¹, В. А. Коровко¹, Е. А. Жатканбаєва²

¹ Харківський національний автомобільно-дорожній університет, Харків, Україна

² Казахський автомобільно-дорожній інститут ім. Л. Б. Гончарова, Алмати, Казахстан

ОСОБЛИВОСТІ ФОРМУВАННЯ ЗАМОВЛЕНЬ АВТОЗАПЧАСТИН ПРИ БАГАТОНОМЕНКЛАТУРНОМУ КЕРУВАННІ

Анотація. Розглядаються заходи, які спрямовані на підвищення ефективності функціонування системи замовлень при багатономенклатурному управлінні. Розглянути параметри замовлень у багатономенклатурних партіях. Встановлено параметри замовлень у системі кратних періодів. Встановлені моделі, що дозволяють визначити параметри замовлень за системою кратних періодів. Встановлені моделі при використанні яких витрати на транспортування та зберігання запасів будуть найменшими. Критерієм вибору моделі визначення та інтервалу між замовленнями та коефіцієнтами кратності є найменші втрати на виконання замовлень.

Ключові слова: прибуток, замовлення, витрати, запаси, попит, автозапчастини, кратний період.

Вступ

Сучасне підприємство – це складна виробнича система, що включає такі елементи, як основні фонди, сировину та матеріали, трудові та фінансові ресурси. Найважливішим завданням управління на підприємстві є використання перелічених елементів таким чином, щоб забезпечити ефективне функціонування всієї виробничої системи, виживання та розвиток в умовах жорсткої конкурентної боротьби. В умовах ринкових відносин та конкурентного середовища управління товарними запасами розглядається, як здатність підприємств завжди бути в змозі задовольнити вимоги клієнтів, покупців товарів, що зберігаються.

В даний час у зв'язку з необхідністю скорочення витрат, пов'язаних з процесом руху товарів є управління запасами, як різницею швидкостей потоків надходження і вибуття товарів, є одним з найбільш важливих завдань, що стоять перед комерційними службами підприємств.

Крім того, стохастичний характер ринкових процесів призводить до того, що в ринкових умовах будь-якому підприємству (фірмі) доводиться функціонувати в умовах яскраво вираженої невизначеності довкілля, і вони змушені швидко реагувати на ці зміни. І тому використовується ціла система методів управління запасами. У логістичній системі сучасного підприємства товарні запаси грають роль вирішального чинника, який впливає на кінцеві результати діяльності будь-якого підприємства. Товарний запас є товари які знаходяться у каналах сфери звернення, які призначені для продажу з метою задоволення попиту споживачів. Товарні запаси на підприємствах можуть бути у вигляді готової продукції у постачальників, а також в оптовій та роздрібній торгівлі. Товарні запаси є важливим елементом ресурсного потенціалу підприємства, становлячи значну частину його активів. Товарні запаси забезпечують безперервність процесу обігу товарів, при цьому необхідно, щоб частина товарних ресурсів постійно перебувала у стані товарного запасу [1-4].

Таким чином, товарні запаси здійснюють, насамперед, функції безперервного процесу товарного обігу:

– забезпечують безперервність розширеного виробництва та обігу, у процесі яких відбувається їх систематичне утворення та витрачання;

– задовольняють платоспроможний попит населення, оскільки є формою товарної пропозиції;

– характеризують співвідношення між обсягом попиту та товарної пропозиції.

Ефективне управління запасами є запорукою фінансової стійкості підприємства. Індикатором фінансового «здоров'я» будь-якого комерційного підприємства чи фірми є структура та рівень товарних запасів. Це один із показників ступеня ефективності використання матеріальних ресурсів та виробленого продукту, досконалості економічних взаємин у цілому.

Товарні запаси є базою для забезпечення надійності процесу відтворення і можуть приносити фірмі прибуток на основі прискорення їх оборотності. В основі створення запасів на торгових підприємствах лежать фактори економічного, організаційно-технологічного та соціального характеру [5-8].

На сучасних підприємствах у структурі їх управління головною проблемою є неефективна система управлінського обліку, яка, за спотвореною інформацією, може легко підірвати зусилля компаній із чудовими розробками, виробництвом та маркетингом.

Наслідком цієї проблеми є невикористання сучасних методів управління виробничими запасами. Для забезпечення ритмічної роботи підприємства необхідно створити систему сучасного управлінського урахування, яка може підготувати своєчасну інформаційну базу для основних елементів управління витратами та собівартістю продукції підприємства, якими є:

- прогнозування та планування,
- нормування витрат,
- організація обліку витратта калькулювання собівартості продукції,
- аналіз, контроль та регулювання діяльності у процесі її здійснення.

Аналіз публікацій. Управління запасами – традиційна сфера практичної роботи, яка стала розвиватися як самостійний напрямок на початку 20 років минулого століття (Ф. Харріс, Е. Тафт).

У 1917 році Ф. Раймонд випускає першу монографію, присвячену управлінню запасами. У 30-40 рр. минулого століття активно розробляються алгоритми вирішення завдань лінійного програмування для керування запасами (Л.В. Канторович, Дж. Данциг, М. Вуд).

Класичний апарат оптимізації рівня запасів був розроблений у 50-60-ті роки ХХ століття в рамках розділу математичної економіки, який отримав назву «Теорія управління запасами» та був орієнтований на оптимізацію рівня запасів в організації.

Розробляються основні класифікації показників величини запасів та описуються моделі їх визначення та оцінки (С. Хольт, П. Вінтерс, Р. Феттер, Е. Смайкей, Д. Бауерсокс, Ф. Моссман).

У подальших вітчизняних та зарубіжних дослідженнях акцент робився виключно на розрахункову складову запасу [1, 10]. У зв'язку з процесами перебудови та розвитку ринкової економіки в Україні проблеми безпосереднього управління запасами, здавалося, пішли на другий план, почали розглядатися лише на рівні виконавців та низової ланки управління.

Вивченню теоретичних аспектів та проблем управління запасами присвячено багато робіт зарубіжних та вітчизняних авторів.

Серед них слід відзначити фундаментальні дослідження проблем управління товарним запасом підприємства (Д.С. Джонсон, І.А. Кеншин, М.Р. Ліндерс), аналіз складових запасу (В. Дж. Стівенсон, Р. Чейз).

Проблема оптимізації матеріального потоку розглянута у Д. Бауерсокса та Д. Клосса, М. Ліндерса та С. Харольда, В. Стівенсона.

Розглянуто модифікації задач оптимізації (М. Крістофер, Ю.М., С.І. Bodenstab).

Тим часом, зарубіжна наука та практика управління (насамперед завдання управління логістичного, пов'язаного з управлінням матеріальними потоками) за останні 20 років зробила значний крок від удосконалення методів і моделей розрахунку рівнів до вдосконалення процесу управління запасами, заснованого на аналізі результатів розрахунку рівнів запасів (Е. Ф. Харольд, П. Ф. Друкер, J. L. Gattona).

Проте проблема управління динамікою зміни показників запасу підприємства з багатономенклатурним асортиментом в умовах нестаціонарного попиту на продукцію у вітчизняній та зарубіжній літературі досліджено недостатньо, що й визначило дослідження.

Дослідження виконано на статистичних даних окремих позицій номенклатури ТОВ «Омега».

Суть пропозиції

Проведений аналіз літературних джерел дозволив виявити декілька методик при управлінні багатономенклатурними запасами.

Подальшим удосконаленням методики управління багатономенклатурними запасами є управління з системою кратних періодів.

Поставка по системі «кратних періодів» передбачає, що хоча б одна номенклатурна позиція замовляється в кожному базисному періоді (T), а інші товарні позиції поставляються через інтервали, кратні базисному періоду (kT , де $k = 1, 2, 3, \dots, n$).

При поставках по системі «кратних періодів» необхідно таким чином вибрати базисний період T і так закріпити номенклатурні позиції за групами різної кратності, тобто для кожної позиції визначити свій коефіцієнт кратності k_i , щоб сумарні витрати на управління запасами були мінімальними [10]

$$C_{\Sigma} = \frac{D}{T} \left(C_0 + \sum_{i=1}^n \frac{C_i}{k_i} \right) + \frac{T}{2D} \sum_{i=1}^n S_i C_{збі} k_i, \quad (1)$$

де C_{Σ} – сумарні витрати на управління запасами, грн.;

D – аналізований період, дні;

T – базисний період, дні;

C_0 – постійні витрати на виконання багатономенклатурної поставки, грн.;

C_i – змінні витрати на виконання багатономенклатурної поставки для i -го виду продукції, грн.;

k_i – кратність;

S_i – обсяг споживання i -го виду продукції в періоді розрахункового періоду, од.;

$C_{збі}$ – витрати на зберігання, грн.

При аналізі літератури так само встановлено, що існує декілька доходів до визначення базисного періоду. Який із існуючих підходів дозволяє скоротити витрати на зберігання і транспортування невідомо.

У зв'язку з цим проведемо аналіз і встановимо який із існуючих підходів дозволяє забезпечити найбільший прибуток для підприємства.

В першому підході за основу початкового значення періоду групування приймається період між незалежними поставками першої номенклатурної позиції, якої відповідає мінімальна величина відносини

$$\frac{D^2 C_i}{S_i C_{збі}}.$$

Тоді початкове наближення для базисного періоду розраховується за формулою

$$T_0 = D \sqrt{\frac{2(C_0 + C_1)}{S_1 C_{збі}}}. \quad (2)$$

Базисному періоду відповідає коефіцієнт кратності $k_i = 1$ Для інших позицій необхідно послідовно перебирати значення k_i починаючи з 1 і продовжуючи до тих пір, поки виконується умова

$$S_i C_{збі} \leq \frac{2C_i D^2}{k_i (k_i + 1) T_0^2}. \quad (3)$$

Тоді найменше k_i при якому починає виконуватися умова (2) і є початковим коефіцієнтом кратності для i -ї позиції номенклатури.

Другий підхід описаний в роботі М. Харіга.

Вибір початкового значення базисного періоду T_0 можна виконати за багатьма номенклатурами, які

вони замовляють в кожному періоді, при цьому перші j -й номенклатурні позиції поставляються одночасно.

У цьому випадку формула для розрахунку T_0 матиме вигляд

$$T_0 = D \sqrt{\frac{2 \left(C_0 + \sum_{i=1}^j \frac{C_1}{k_i} \right)}{\sum_{i=1}^j S_i C_{збi}}}. \quad (4)$$

Приєднувати до зазначеної безлічі j спільних товарів, що замовляються, наступні позиції виявиться недоцільно в той момент, коли буде виконуватися наступна нерівність

$$C_{j+1} D^2 > S_{j+1} C_{зб,j+1} \cdot T^2. \quad (5)$$

При підстановці (4) в формулу (5) отримаємо умову припинення накопичення даної безлічі, що має вигляд нерівності

$$\frac{C_{j+1}}{S_{j+1} C_{зб,j+1}} > \frac{2 \left(C_0 + \sum_{i=1}^j \frac{C_1}{k_i} \right)}{\sum_{i=1}^j S_i C_{збi}}. \quad (6)$$

Перевірка рівнянь (5) і (6) починається з другої позиції товарної номенклатури, при цьому в правій частині нерівностей підставляються значення $C_0 + C_1$ і $S_1 C_{зб1}$.

При виконанні умови (6) для j -ї товарної позиції для всіх наступних, для яких $i > j$, обчислюється оптимальний період між поставками T_i при незалежних поставках і початкових значеннях коефіцієнтів кратності k_i за формулою

$$k_i = \frac{T_i}{T_0}. \quad (7)$$

Третій підхід являє собою синтез першого і другого підходів. Початкове значення базисного періоду приймається рівним періодичності при незалежних поставках першої позиції номенклатури в ранжированному ряду на етапі 1 і розраховується за формулою (2).

Відповідно до базисного періоду набір коефіцієнтів, за допомогою яких проводиться формування базового варіанту груп різної кратності, визначається за формулою (7).

У відповідності з представленими методиками проведе розрахунок базисних періодів і проведено розрахунок коефіцієнтів кратності. У табл. 1 представлена частина отриманих даних.

Таблиця 1 – Результати розрахунку початкового значення базисного періоду T_0 і початкових коефіцієнтів кратності

Коефіцієнт	Інтервал, T_0 , дні	Спосіб розрахунку		
		1	2	3
		95,2	84,2	95,2
k_1		1	1	1
k_2		1	1	1
k_3		1	1	1
...	
k_{37}		2	4	2
k_{38}		2	4	2

Аналіз табл. 1 показав, що значення базових періодів T_0 для 1 та 3 підходу значно відрізняється від другого.

Для того щоб визначити підхід, який дозволяє отримати більший прибуток, необхідно провести подальший розрахунок допоміжних величин.

Для стратегії поставок по системі «кратних періодів» оптимальний базисний період може бути розрахований за формулою

$$T^* = D \sqrt{\frac{\beta_0}{\beta_1}}, \quad (8)$$

де β_0, β_1 – параметри які визначаються за формулами

$$\beta_0 = 2 \left(C_0 + \sum_{i=1}^n \frac{C_i}{k_i} \right), \quad (9)$$

$$\beta_1 = \sum_{i=1}^n S_i C_{збi} k_i. \quad (10)$$

Даному періоду відповідають мінімальні сумарні витрати на управління запасами

$$C_{\Sigma}^* = \sqrt{\beta_0 \beta_1}. \quad (11)$$

Дали проведемо розрахунки. Результати обчислень допоміжних параметрів, а так же оптимального періоду групування та сумарних витрат на управління запасами для базового варіанту груп різної кратності наведемо в табл. 2. Аналіз таблиці 2 дозволив встановити, що найменші витрати на управління запасами отримані за допомогою третього методу визначення базового варіанту груп різної кратності, то саме цим методом доцільно виконувати подальші розрахунки та визначати оптимальні параметри багатноменклатурної поставки.

Таблиця 2 – Результати розрахунку допоміжних параметрів, оптимального базисного періоду і оптимальних витрат

Спосіб розрахунку	Допоміжний параметр, β_0 , грн.	Допоміжний параметр, β_1 , грн.	Інтервал, T^* , дні	Витрати, C_{Σ}^* грн.
1	416000	4466038	111,4	1363037
2	232000	8033397	62,03	1365191
3	332000	5440111	90,17	1343918

Висновки

Аналіз отриманих результатів дозволив зробити висновок, що використання запропонованої методики по визначенні тривалості базисного періоду та значення коефіцієнтів кратності забезпечує найменші витрати на зберігання запасів на складі та їхнє

транспортування на підприємство. Таким чином, використання даної методики є доцільним.

Отримані знання дозволяють визначити який економічний ефект буде мати підприємство, якщо використовувати запропоновану методику, в порівнянні з класичним методом визначення тривалості базисного періоду та коефіцієнтів кратності.

СПИСОК ЛІТЕРАТУРИ

1. Тюріна Н.М., Гой І. В., Бабій І. В. Лоїстика : навч. посіб. К.: Центр навчальної літератури, 2021. 392 с.
2. Coyle J., VAardi E., Langlely C. Zarzadzanie logistyczne Warszawa: PWE, 2002. 734 с.
3. Очеретенко С.В. До питання про управління запасами автомобільних запчастин на торгових підприємствах // Комунальне господарство міст, 2018. Вип. 142. С. 114-117.
4. Крикавський Є.В., Чернописька Н.В. Логістичні системи : навч. посіб. Львів: Львівська політехніка, 2019. 288 м.
5. Gwynne R. Warehouse management: a complete guide to improving efficiency and minimizing costs in the modern warehouse. London : Kogan Page Limited, 2014. 449 p.
6. Mahroof K. A human-centric perspective exploring the readiness towards smart warehousing: The case of a large retail distribution warehouse. Int. J. Inf. Manag. 2019. Vol. 45. P. 176–190.
7. Bowersox D., Closs D. Logistical Management : The Integrated Supply Chain Process. McGraw Hill Education, 2017. 752 p.
8. Bowersox D., Closs D., Cooper M. B. Supply Chain Logistics Management : 5th edition. McGraw-Hill Education, 2019. 480 p.
9. Чухрай Н.І., Патора Р. Інновації та логістика товарів : навчальний посібник. Львів: «Львівська політехніка», 2001. 262 с.
10. Окландер М.А. Логістика : навчальний посібник. К.: Центр навчальної літератури, 2018. 346 с.

Received (Надійшла) 15.03.2023

Accepted for publication (Прийнята до друку) 16.05.2023

Features of forming auto parts orders at Multiple nomenclature management

S. Ocheretenko, V. Korovko, E. Zhatkanbaeva

Анотація. The subject of the study is the change in costs at the enterprise when ordering a batch of auto parts with multi-item management according to the system of multiple periods in modern conditions. The goal is to develop is the development of a method for determining the optimal size of a multi-item order according to the system of multiple periods, taking into account real parameters reflecting a large number of factors, which will improve the efficiency of the enterprise by reducing general logistics costs. **Research objectives:** analysis of existing methods aimed at improving multi-item orders, determination of order parameters according to the system of multiple periods; definition of mathematical models allowing to determine the optimal parameters of a multi-item order: the interval between orders, the multiplicity factor. As an optimality criterion, the minimum costs for transportation and storage are taken. **The following results are received.** A mathematical model is proposed that allows determining the interval between multi-item deliveries and the multiplicity factor between deliveries for each item. A mathematical model has been established that takes into account the costs of storage and transportation, taking into account the order of a different number of items. A decision-making scheme is proposed for determining the system of multi-item orders depending on the storage parameters and the cost of delivery. **Conclusions.** The issue of minimizing the costs caused by the process of purchasing, transporting and storing car spare parts directly affects the efficiency of an enterprise engaged in car repair. Therefore, it is necessary to look for various options aimed at reducing costs. Therefore, when calculating batches of an order, it is necessary to consider several options for organizing deliveries. With the help of the proposed model, it is possible to determine such batch parameters at which the costs of storage and transportation of a multi-item delivery will be minimal.

Keywords: profit, order, expenses, stocks, optimal order size, demand, auto parts, multiple period.

R. Baranenko¹, H. Sokol², P. Pustovoitov², R. Zakharov

¹ Uman National University of Horticulture, Uman, Ukraine

² National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

³ State University of Trade and Economics, Kiev, Ukraine

INFORMATION AND ANALYTICAL SUPPORT OF THE EDUCATION QUALITY MANAGEMENT SYSTEM

Abstract. The article analyzes the problem of informational and analytical provision of the education quality management system. The purpose of the article is to determine the role of modern information technologies in the process of improvement and modernization of the educational system. The components of the information system of quality management, which is defined as a set of interconnected or interacting elements of the organization for the development of policies and goals, as well as processes for achieving goals in the field of education quality, have been studied. It is proven that the formalized information system of education quality management creates a basis for planning, implementation, monitoring and improvement of education quality management activities.

Keywords: information system, information technologies, quality of education, quality management.

Introduction

The process of informatization is inextricably linked with the status of today's society, namely with the status of the information society, which is dominated by information, its quality, freedom, publicity and accessibility. Informatization is a large-scale process that affects all spheres of social life, aimed at satisfying people's information needs, as well as building a powerful telecommunications infrastructure.

The process of informatization of education includes a system of measures: equipping educational institutions and educational management bodies with hardware and software tools of information technologies; connection via high-speed channels to regional, national and international computer educational networks, to the global Internet; creation and placement of educational information resources, integration of various databases at the regional and state level. formation of information culture among all participants of the educational process: employees, teachers, future specialists, their parents, creation of a system of continuous training of teachers in information technologies (courses, express courses, mini-seminars, permanent seminars, conferences) [1, 2].

Modern society is characterized by constantly developing information technology (IT) tools. Objectively, the process of informatization of society significantly affects the goals and content of education, presents new requirements for the professional training of specialists in the field of using IT tools.

Analysis of recent research and publications.

The analysis of the works of leading experts in the field of research into the issues of information and analytical support of the education quality management system allows us to establish the relevance of the identified issues. The problem of determining the role of modern information technologies in the process of improvement and modernization of the education system is being

investigated by a number of leading foreign and Ukrainian scientists. The researches of such experts as M. Bublyk, A. Karpyak, O. Rybyska are dedicated to the study of the components of the modern informational educational space, in particular, the analysis of the structure of the standard and the study of its content [3, p.116]. A. Kapiton, O. Skakalina, R. Baranenko, T. Franchuk consider the relevance of using relevant search in the process of developing an information system [1, p.64]. N. Chukhrai and T. Shcherbata indicate the need for cooperation between higher education institutions and enterprises working in the field of information technologies [4, p.161]. N. Popadynets, E. Korcelli-Oleynichak, M. Melnyk, and N. Chorna analyze the leading methods of designing an information system for managing the quality of education [5, p. 19]. The analysis of the main components of the quality management information system is devoted to the work of S. Alyoshin, O. Borodina, A. Hafyak, O. Nosach [2, p.364].

Main part

It is necessary to understand and evaluate the possibilities of information technologies for the fuller development of the personality of students, to see how it is possible to most organically integrate learning information technologies in the educational process. The question of the role of modern information technologies in improving and modernizing the existing educational system has remained relevant for several decades. Successful implementation of the education modernization program will require not only modern technical equipment of universities, but also appropriate training of teachers and organizers of the education system. Thus, for every teacher, the main goal is to ensure the quality of education, which can be facilitated to a greater extent by the use of information technologies. IT in the educational society can be considered as a means of self-realization and self-

affirmation of teachers, which significantly increases the level of their professional culture, which expands the possibilities of sharing accumulated experience, own views, and also helps to move from the role of a teacher-translator of knowledge in class to the position of a teacher-tutor who organizes and directs the process of independent cognitive activity of students. Information technologies play an important role in the development of teaching methods, because the work of each teacher is important for methodical and pedagogical science as a whole.

A project is an event that includes a concise description of a defined need for public space and requires its implementation. The objects of quality management of the developed project are products and processes. The evolution of the development of the approach from product quality management to the general management of the quality of the organization's activities involves the creation and ensuring the effective functioning of the quality system, which includes the necessary organizational structure, procedures, processes and resources [1, 2, 3]. A quality management information system (QMIS) can be defined as a set of interrelated or interacting elements of an organization for the development of policies and goals, as well as processes for achieving goals in the field of quality. A formalized QMIS creates a basis for planning, implementing, monitoring and improving quality management activities.

The standards contain requirements and explanations for their application to the following QMIS processes:

- 1) product life cycle processes (group of main processes);
- 2) resource management (a group of resource provisioning processes);
- 3) management responsibility (group of management processes);
- 4) measurement, analysis and improvement (a group of processes of information and analytical support for management and development).

The operational structure of the "Quality Management at the Enterprise" process includes the following operations:

1. Creation of quality requirements (determination of indicators critical to quality; determination of preventive measures for quality assurance; verification of compliance with the quality plan; verification of possibilities for ensuring compliance with requirements; completion of the quality plan);
2. Assessment of compliance with requirements (checking compliance with the quality plan; assessment of test results);
3. Management of non-conformities (assessment of potential impact; determination of immediate measures; identification of causes; elimination or taking of preventive measures; closure of non-conformity);
4. Implementation and support of the quality management system at the enterprise (definition of the management strategy; planning and deployment of specialized software for the quality management system (scope, goals and objectives); definition of the key

processes of the quality management system, management bodies and indicators; development and documentation of regulations, procedures, standards and indicators of quality management, evaluation of the performance of the quality management system, creation of the environment and competencies for the improvement of the quality management system.

In essence, management of the quality of higher education refers to situational management, which is based on: formation of options - alternatives of management decisions; solving problems of current and prospective preparation of plans. The growing importance of the use of modern information technologies in the education process is currently caused by many factors, and first of all: the complication of the pedagogical process in the educational institution in the conditions of the integration of special disciplines, as well as the integration of the educational institution with advanced companies and research organizations; the expansion of the subject world of the student, which leads to an increase in the volume of educational material and the need for its generalization; expansion of spheres of activity, which leads to the need to solve various professional tasks: project, research, technological; inclusion in the learning process of promising technologies, including the base of modern telecommunications and computing tools.

Thus, information technologies in education are a kind of response to changes in the system of higher professional education related to the optimization of management of students' cognitive activities. Fulfillment of the listed conditions contributes to the achievement of the main goal of education modernization - improving the quality of education, increasing the availability of education, ensuring the needs of the harmonious development of an individual and the information society as a whole [4, 5].

The application of the developed project begins with "identifying the need" by studying official documents, surveys, focus groups, etc. At the first stage, the results of project implementation are determined. On the second stage, the consequences of specific transformations are specified, which is a step for monitoring and evaluation even during the planning of the project itself.

The third stage of direct project implementation includes: project improvement planning; implementation of certain processes; carrying out certain operations: promotions, trainings, consultations, publications, etc. The fourth stage is the monitoring of the project implementation process, as well as the evaluation of short-term results.

The fifth stage involves the assessment of long-term results or "consequences".

The results of this final assessment help to determine a "new need" and to start a new project.

Constant quality control of the project is necessary at all stages of its planning, preparation, direct implementation and until the stage of achieving the final goal of the project were taken into account, looks as shown in Fig. 1.

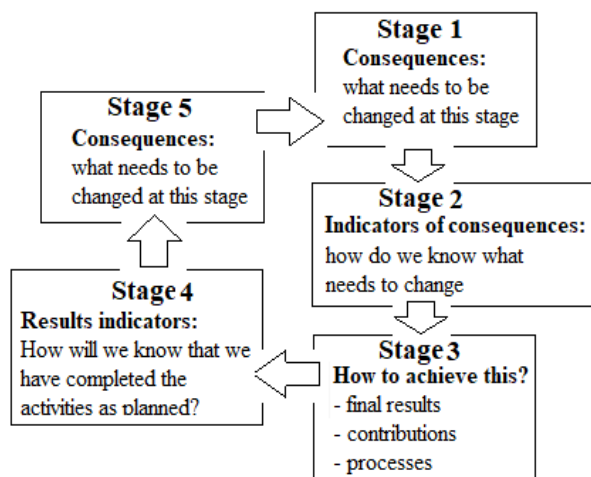


Fig. 1. Monitoring in the project cycle

The information system has the following recommended hardware requirements:

Server: 1 GHz processor, 512 Mb RAM, 500 Mb RAM;

Client: 800MHz processor, 256MB RAM;

Software requirements: OS Windows XP, Vista, 7; Linux; iOS;

Browser: Opera, Google Chrome, IE 8, FireFox, Flash Player 11.

The advantages include: the possibility of implementing almost any site (from a business card site to a portal and an online store); a huge number of ready-made design solutions; relative simplicity of content management; ease of setup. Administration of the designed system is carried out using CMS. To access the administrator panel, you must log in by entering the system administrator's personal data, namely: enter the administrator's name; enter the password specified

during registration; choose the language of the administrative section management interface from the list of installed ones (English, Russian or Ukrainian); human, material and organizational resources can be allocated on the basis of data on the progress of work.

Conclusions

The main feature of the result-oriented management model is the need to make assumptions, identify risks and overcome these risks. As already mentioned, the process of achieving results is based on a cause-and-effect relationship. If causality holds, then the assumptions made during design will also hold. That is why, in order to achieve success, it is necessary to have a plan that recognizes the risks and identifies strategies to reduce the risks to the minimum possible. The input data for the planning and development of the project of educational activity are : educational and professional program; curriculum by specialty for the appropriate level of training (bachelor, specialist, master); working curriculum for the academic year. Input data in the planning and development of a project of scientific activity are: customer requirements; requirements of the current legislation of Ukraine and regulatory and technical documentation.

In the presence of complete information about the course of project implementation, it is possible to know: directions in which the achieved results can be strengthened; difficult directions can be strengthened, different approaches can be tried or measures can be added; measures and system results, which are recognized as unnecessary, can be removed; information about the results of the work progress will help to understand what is working and what is not; it is possible to increase assistance to target groups or add new such groups.

REFERENCES

1. Kapiton A. M., Skakalina O.V., Baranenko R.V., Franchuk T.M. Development of information system using relevant search. *Electronic modeling*. №1. 2023. 63-79.
2. Hafiiak A.M., Borodina O.O., Alyoshin S.P., Nosach O.B. The Information Society and Informatization Development Problems of Economy. *International Journal of Engineering & Technology*. 7. (4.8). 2018. 364 – 369.
3. Bublyk M.I., Karpiak A.O., Rybytska O.M. The perspectives of IT industry development in Ukraine on the basis of data analysis of the world economic forum. *Innovative management: theoretical, methodical, anapplied grounds/ Pražský Institut zvyšování kvalifikace Prague institute for Qualification Enhancement*. 2018. 115 -128.
4. Chukhrai, N. I., & Shcherbata, T. S. Cooperation between IT development enterprises and universities. *Marketing And Management Of Innovations*. (3). 2016. 161-169.
5. Melnyk, M., Korcelli-Olejniczak, E., Chorna, N., & Popadynets, N. Development of Regional IT clusters in Ukraine: institutional and investment dimensions. *Economic Annals-XXI*, (173). 2018. 19-25.

Received (Надійшла) 14.03.2023

Accepted for publication (Прийнята до друку) 24.05.2023

Інформаційно-аналітичне забезпечення системи управління якістю освіти

Р. Бараненко, Г. Сокол, П. Пустовойтов, Р. Захаров

Анотація. У статті аналізується проблема інформаційно-аналітичного забезпечення системи управління якістю освіти. Метою статті є визначення ролі сучасних інформаційних технологій у процесі вдосконалення та модернізації системи освіти. Досліджено компоненти інформаційної системи управління якістю, яка визначається як сукупність взаємопов'язаних або взаємодіючих елементів організації для розробки політики та цілей, а також процесів досягнення цілей у сфері якості освіти. Доведено, що формалізована інформаційна система управління якістю освіти створює основу для планування, реалізації, моніторингу та вдосконалення діяльності з управління якістю освіти.

Ключові слова: інформаційна система, інформаційні технології, якість освіти, управління якістю.

T. Dzheria, V. Shevchuk

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" Kyiv, Ukraine

FORECASTING ELECTRICITY CONSUMPTION USING NEURAL NETWORKS

Abstract. The problem of early and accurate forecasting of electricity consumption is acute for the unified energy system of Ukraine. With successful forecasting of consumption, which is based on many aspects, it is possible to buy electricity/losses in different market segments much more profitably, saving large amounts of money, which can then be directed to the development and modernization of electricity networks. This has always been an urgent issue, but today, when a large part of Ukraine's energy equipment has been destroyed by Russian missiles, it has become even more painful. The use of the method of artificial neural networks (ANN) for short-term forecasting of electricity consumption is considered. It was established that ANN can be used to make a forecast of electricity consumption a day ahead with an error of 4.86% compared to the actual amount of electricity consumption. Performing a comparison of forecast values with actual values allows us to talk about the adequacy of the selected forecasting model and its application in practice for the successful operation of energy supply companies in the electricity market.

Keywords: forecasting, artificial neural network, neural network architecture, Neural Network Toolbox module, hourly consumption database, Levenberg-Marquardt gradient algorithm, inverse error propagation algorithm.

Introduction

Forecasting is a very popular topic, given the large impact on business profitability, by minimizing costs. For the energy sector, consumption forecasts are the basic source information for decision-making in the process of planning optimal modes of operation and development of the power system. As the main tasks solved on the basis of the received forecasts of consumption of the electric power, it is possible to allocate the following:

- 1) planning of development of generating capacities and electric networks of electric power system;
- 2) tariff planning;
- 3) planning the loading of power plants for the next day, energy and power generation, fuel needs;
- 4) planning of repair works of the main equipment of power plants and networks.

Traditional statistical models (regression and time series models) and models based on expert systems and neural networks can be used to solve the problem of forecasting electricity consumption [1]. The most commonly used models are neural networks. This is due to the fact that it is not necessary to build a model of the object, does not lose performance with incomplete input information. Neural networks are resistant to interference, have high speed. Predicting electricity consumption using artificial neural networks is one of the most actively developed in the energy sector. This is due to the fact that this structure is a universal approximator and is able to build complex nonlinear dependencies, which allows you to successfully predict. The difficulty of creating a neural network is the unavailability of data for its training [2, 3].

In [4–7] the issue of forecasting the electrical load using neural networks was considered. Namely, neural network learning algorithms, neural network architecture and forecasting for different terms for different conditions. For our purposes on the basis of works the training of a neural network by a method of back propagation was chosen. The advantages of this method are the simplicity and speed of its application. The solution of [7] is suitable, as the paper performs short-term forecasting of electricity consumption, which

depends on changes in air temperature. Based on this work, a neural network was created and training was conducted on our own data sample.

The main part of the article

1. Analysis of the time series of electricity consumption and the impact of weather factors on it.

Knowledge of the characteristics of daily charts $P(t)$ system helps to design adequate forecasting models that work effectively in different situations. Various factors that affect the change in the load of the power system can be divided into the following main categories: meteorological, seasonal (cyclical), economic, random disturbances. Ambient temperature is the factor that has the greatest impact on electricity consumption. This is due to the fact that maximum human productivity is achieved only when comfortable living conditions are created. The most comfortable for humans is the temperature range from 18 to 25 °C.

2. Neural network model. The multilayer perceptron of an artificial neural network (ANN) is today one of the most common neural network models due to its ability to display complex nonlinear relationships between input and output parameters. The network consists of several layers of neurons and weights that reflect the relationships between them. The transfer of information is based on the method of direct dissemination. Presented in Fig. 1 model consists of one input layer, one hidden and one output layer as the ANN of this type allows to solve practically any problems with nonlinear relations [8, 9].

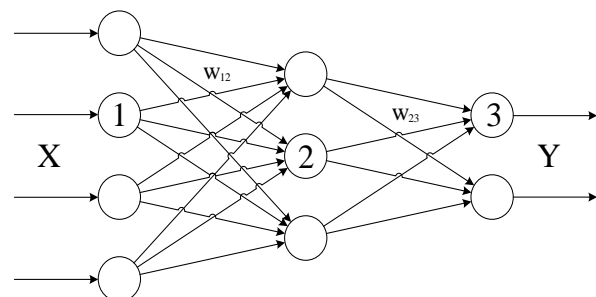


Fig. 1. Neural network of a multilayer perceptron

In Fig. 1: "X" - input vector; "W12" - vector of scales between the input and hidden layers; "W23" - vector of scales between the hidden and output layers; "Y" is the output vector (end result).

ANN training is performed to approximate the nonlinear function to the predicted) and input variables.

The MATLAB software product was selected to create a forecasting model based on ANN. This choice is due to the experience of MATLAB, as well as the module Neural Network Toolbox for the creation and training of ANN. When creating the ANN model, a database of hourly electricity consumption from 2016 to 2018 inclusive was available. This amount of data is sufficient for training and network operation on test data samples.

A network of direct data dissemination and reverse error propagation was used to forecast electricity consumption. Fig. 2 shows a diagram of this network. The algorithm of the Levenberg-Markwardt gradient method was chosen for the study of ANN.

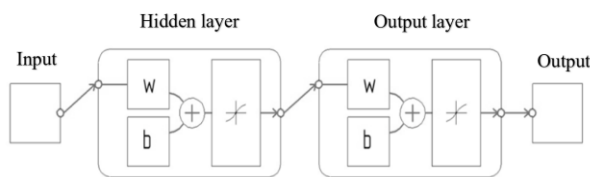


Fig. 2. ANN diagram of direct data dissemination and reverse error propagation

With this network architecture for learning, you need to set time series targets, which are the benchmark when creating a forecast, and the input data to be processed.

3. Generalization of the network training algorithm. In the first step, the scales are initiated by small random variables. After passing the signal from input to output, the value d , is entered and calculated, which is equal to the difference between the required or target output T_i and the actual output Y :

$$d = (T - Y). \tag{1}$$

In the case when $d = 0$ in the network nothing changes, it means that the weight coefficients are chosen correctly. At $d \neq 0$ the correction of the value of the weight of each connection between the layers of the neural network in the direction from output to input is calculated.

$$D_i = hdx_i, \tag{2}$$

$$w_i(n + 1) = w_i(n) + D_i, \tag{3}$$

where D_i – is the correction associated with the i -th input; $w_i(n+1)$ – weight value i after correction; $w_i(n)$ – weight value i before correction; x_i – the magnitude of the i -th output; h – coefficient of learning speed, which allows you to control the average amount of weight change. The process is repeated until the stop condition is fulfilled.

Weight coefficients are selected during training according to the training sample, that is in different places - in electric networks different volume of consumption at the same air temperature, accordingly selected weight coefficients for one network, will not be actual for any other.

- The ANN model has the following parameters:
- number of layers - 2 (hidden, output);
 - the number of neurons in the hidden layer – 24;
 - the number of neurons in the output layer – 24;
 - the function of activation of the hidden layer is sigmoid;
 - learning algorithm - reverse propagation of error;
 - speed of learning – 0,3;
 - impulse coefficient – 0,3;
 - number of learning epochs – 1000.

The forecast of electricity consumption in the network is considered in Fig. 3. It should be noted that ANN correctly recognized the regularity of the graph and the result was the construction of a similar graph with an accuracy of 4.86% deviations modulo the total volume per day, which in quantitative terms amounted to 1,081,367 kWh.

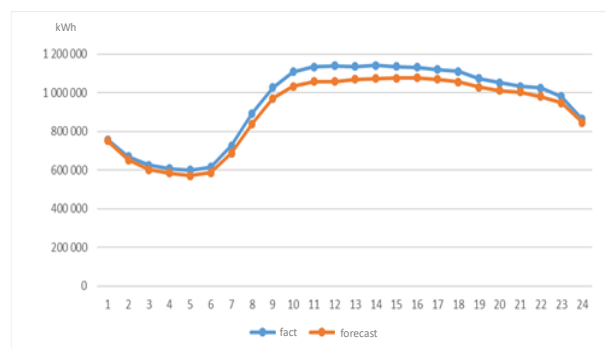


Fig. 3. The result of forecasting electricity consumption on July 17, 2019 (Wednesday) in Kyiv Electric Networks

A forecast for the holiday was also made to check the operation of the network during accidents and disturbances. The results of forecasting on the holiday are shown in Fig. 4.

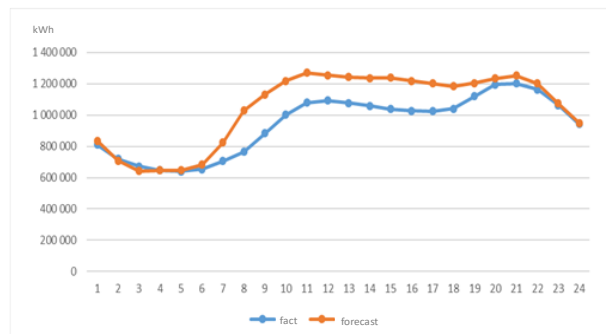


Fig. 4. Result of forecasting electricity consumption on October 14, 2019 (Monday, public holiday) in Kyiv Electric Networks

The graph of imbalance between the actual and projected volume of electricity consumption is given in Fig. 5 and 6 in absolute units and percentages, respectively, to compare the results.

As the experiment showed, the network is not adapted to forecasting on holidays. To do this, you need to enter another factor that will affect the forecast - it's a holiday/weekend or weekday.

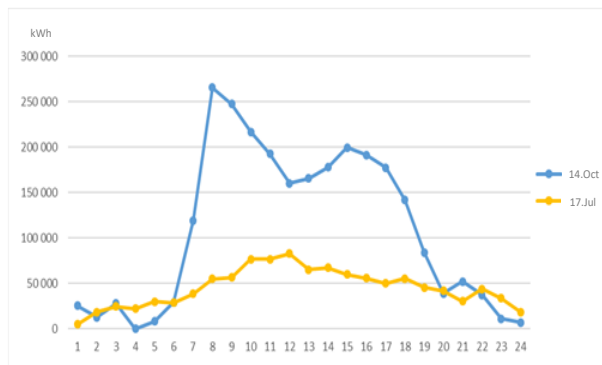


Fig. 5. Graphs of imbalances between actual and projected volume in absolute units

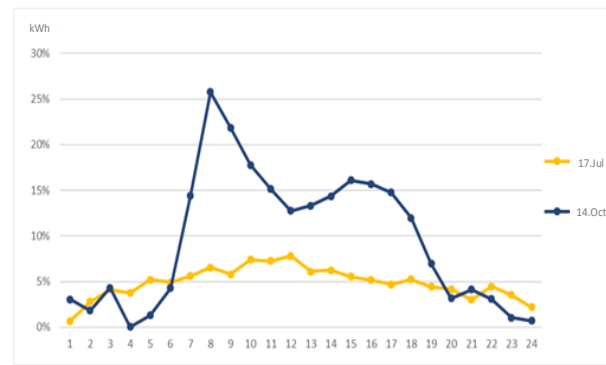


Fig. 6. Graphs of imbalances between actual and projected volume in percent

Conclusions

The influence of air temperature on electricity consumption is analyzed and the features that need to be taken into account when building a forecast consumption model using ANN are identified.

On the basis of the built model of ANN the forecast of consumption of the electric power for 2 days (weekday

and holiday) of the Kiev electric networks is created. A significant increase in the error of forecasting electricity consumption on the holiday was revealed. Therefore, you need to introduce an additional factor for forecasting - a holiday/weekend or weekday.

Therefore, it can be concluded that the introduction of another such factor can significantly improve the quality of the ANN forecast.

REFERENCES

1. Zhezhenenko I.V. Indicators of electricity quality and their control at industrial enterprises. M. : Higher school, 1986. 168 p.
2. Karpova T. Databases: models, development, implementation. SPb. : Peter, 2001. P. 286-289.
3. Haikin S. Neural networks: a full course. 2nd ed. M. : Williams, 2006. P. 89-102.
4. Filipe Rodrigues, Carlos Cardeira, J.M.F. Calado: The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal // Energy Procedia 62 (2014) 220 – 229 URL: <http://www.dem.ist.utl.pt/~cardeira/papers/1-s2.0-S1876610214034146-main.pdf>.
5. Ekonomou, L. (2010): Greek long-term energy consumption prediction using artificial neural networks. Energy, 35(2), pp. 512-517. doi: 10.1016/j.energy.2009.10.018//URL:<https://openaccess.city.ac.uk/id/eprint/13250/3/>.
6. Derya Aydın, Hüseyin Toros: Prediction of Short-Term Electricity Consumption by Artificial Neural Networks Using Temperature Variables // European Journal of Science and Technology No. 14, p. 393-398, December 2018, URL: <https://dergipark.org.tr/en/download/article-file/616792>.
7. Prasanna Kumar, Dr. Mervin Herbert, Dr. Srikanth Rao: Demand forecasting Using Artificial Neural Network Based on Different Learning Methods: Comparative Analysis // International journal for research in applied science and engineering technology, URL: <https://www.ijraset.com/files/serve.php?FID=381>.
8. Galushka V.V., Fathi V.A. Formation of a training sample when using artificial neural networks in database error retrieval problems // Engineering Bulletin of the Don, 2013, №2 URL: ivdon.ru/ru/magazine/archive/n2y2013/1597/.
9. Puchkov E.V. Comparative analysis of training algorithms for artificial neural network // Engineering Bulletin of the Don, 2013, №4 URL: ivdon.ru/ru/magazine/archive/n4y2013/2135/.

Received (Надійшла) 13.03.2023

Accepted for publication (Прийнята до друку) 26.04.2023

Прогнозування споживання електричної енергії за допомогою нейронних мереж

Т. Е. Джеря, В. В. Шевчук

Анотація. Проблема завчасного та точного прогнозування споживання електричної енергії гостро стоїть для об'єднаної енергетичної системи України. При вдалому прогнозуванні споживання, яке базується на багатьох аспектах, можна значно вигідніше купувати електроенергію/втрати на різних сегментах ринку, заощаджуючи великі кошти, які потім можна направити на розвиток та модернізацію електричних мереж. Це завжди було актуальним питанням, але сьогодні, коли значна частина енергетичного обладнання України знищена російськими ракетами, воно стало ще більш болючим. Розглянуто використання методу штучних нейронних мереж (ШНМ) для короткострокового прогнозування електроспоживання. Встановлено, що ШНМ може бути використана для складання прогнозу споживання електричної енергії на добу наперед із похибкою 4,86% порівняно з фактичним обсягом споживання електроенергії. Виконання порівняння прогнозних значень із фактичними дозволяє говорити про адекватність обраної моделі прогнозування та її застосування на практиці для успішної роботи енергопостачальних компаній на ринку електроенергії.

Ключові слова: прогнозування, штучна нейронна мережа, архітектура нейронної мережі, модуль Neural Network Toolbox, база даних погодинного споживання, алгоритм градієнтного методу Левенберга-Марквардта, алгоритм зворотного поширення помилки.

А. О. Зуєв, О. М. Євсеєнко

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

МЕТОД СТВОРЕННЯ ЗВУКОВОГО ОТОЧЕННЯ В ІМІТАЦІЙНО-ТРЕНАЖЕРНИХ КОМПЛЕКСАХ СПЕЦІАЛЬНОЇ ТЕХНІКИ

Анотація. У статті розглянуто метод створення звукового оточення в імітаційно-тренажерних комплексах з урахуванням дистанції до спостерігача та внеску окремого звукового фрагменту. **Метою** статті є аналіз та оптимізація процесу вибору звукових фрагментів, що програються у процесі моделювання звукового оточення. Проведено аналіз існуючих програмних рішень та математичних моделей, що застосовуються у тренажерних комплексах та системах віртуальної реальності. Наведено алгоритм з використанням схеми резервування та функціональну залежність для оцінювання необхідності програвання фрагменту в залежності від його відстані до спостерігача. Розглянуто питання практичної реалізації методу. За **результатами** дослідження визначено, що запропонована функція оцінювання, забезпечує зменшення кількості фрагментів, що одночасно програються при одночасному збільшенні сумарного внеску чутних спостерігачем фрагментів.

Ключові слова: симуляція, імітаційно-тренажерний комплекс, звукове оточення, обробка звуку.

Вступ

Розвиток науки і техніки вплинув на появу нових видів транспортних засобів, зокрема й техніки спеціального призначення. Одночасно значно зросла їх складність та збільшилися вимоги до якості керування ними, що в свою чергу потребує якісного підвищення рівня підготовки операторів цієї техніки. Навчання персоналу із застосуванням класичних методів навчання (різноманітні інструкції, наочні посібники, плакати, відеофільми, макети, що демонструють один або декілька аспектів експлуатації чи використання техніки) є неефективним, а безпосередньо на техніці є витратним і не завжди можливе у достатньому обсязі [1-2]. Одночасно, робота недостатньо підготовлених операторів може призвести до виходу з ладу різних вузлів та агрегатів і, відповідно, коштовного ремонту, і навіть до трагічних наслідків, що загрожують життю людей.

Деякі види транспортних засобів спеціального призначення використовуються в екстремальних та позаштатних ситуаціях, тому повноцінна підготовка операторів до роботи у разі безпосереднього використання реальної техніки або неможлива, або вкрай ускладнена. Ці два аспекти – вартість та якість, на яку впливає "реалістичність" тренування, а також урахування умов безпеки курсантів, визначають потреби в якісному та одночасно дешевому навчанні навичкам управління складною технікою і як наслідок необхідність дослідження альтернативних варіантів навчання, без безпосереднього використання техніки.

Таким засобом є імітаційно-тренажерні комплекси (ІТК), що використовують комп'ютерний синтез навколишнього середовища у процесі навчання [3]. Вони позбавлені недоліків традиційних засобів навчання, мають невисоку вартість як самого комплексу, так і його експлуатації, безпечні для людини й теоретично дозволяють моделювати будь-які ситуації, що виникають при експлуатації техніки [4]. Найважливішою складовою будь-якого ІТК, є аудіо-візуальна складова процесу навчання. Саме вона дає змогу донести до майбутнього оператора принципи

керування транспортним засобом, дозволяє оперативно реагувати на різні ситуації, що виникають у процесі експлуатації техніки, відображає результати його дій, що робить можливим проведення контролю процесу навчання.

1. Огляд існуючих технічних рішень

Роботи в галузі створення та вдосконалення комп'ютерних систем та методів ІТК мають велике значення: створення нових, досконаліших ІТК дозволяє підвищити якість навчання операторів одночасно знизивши його вартість. Істотним обмеженням комп'ютерного синтезу позакабінної обстановки ІТК досі є недостатня реалістичність як візуальної, так і звукової частини моделювання. Методи та алгоритми що відносяться до візуальної частини систем моделювання ІТК досліджуються науковцями [5-7] вже протягом багатьох років. Цьому також сприяє значний розвиток індустрії розваг, у тому числі і комп'ютерних ігор. Алгоритми та методи візуалізації, які застосовуються в них, практично повністю ідентичні таким, що використовуються в системах візуалізації ІТК [8-9]. Але звукова частина все ще опрацьована недостатньо.

У роботах [10, 11] наведено методи синтезу звуків двигуна автомобіля з використанням суміщення та змішування заздалегідь заготовлених звукових фрагментів з урахуванням швидкості обертання двигуна. Такий підхід до моделювання може бути застосований і в ІТК, але фактично тільки для моделювання внутрішньокабінної обстановки або як локальна частина системи звукового моделювання, тому що вимагає значних обчислювальних ресурсів для синтезу звукового середовища. У статті [12] також описано алгоритм синтезу звуку двигуна автомобіля з урахуванням його частоти обертання, але наведений метод можна використовувати у системах, що працюють у реальному масштабі часу. Однак, як і в попередніх роботах, не розглянуто питання оптимізації та роботи звукової підсистеми в цілому. У роботі [13] застосовано підхід до синтезу звуків, пов'язаний з використанням таблиць звукових фрагментів, з урахуванням швидкості обертання різних

частин автівки. Проблеми пов'язані з оптимізацією комплексного відтворення звукового оточення в процесі моделювання в цій роботі не розглянуті, фрагменти відтворюються за допомогою звукових API та комерційних бібліотек для рендерінгу звуків (наприклад, FMOD). У роботах [14, 15] описано досить цікаві методи синтезу звуків навколишнього середовища (вогонь, вітер, дощ) з використанням системи частинок, а також наведено процедурні моделі для синтезу звуків, за рахунок використання яких досягається додаткова різноманітність та варіативність відчуття звукового оточення.

У роботах [16, 17] розглянуто методи визначення та трасування звукових шляхів за допомогою перевірки взаємної видимості поверхонь об'єктів з використанням різних спеціалізованих структур даних, які можна застосувати для побудови аудіосистеми, що працює в реальному масштабі часу. У статтях [18, 19] наведена система рендерінгу звуку в реальному масштабі часу, яка поєднує повну симуляцію невеликого приміщення та багатоканальний бінауральний синтез акустичного "зображення". Система може бути використана для моделювання приміщення будь-якої форми.

У проаналізованих роботах розглядаються питання відтворення та синтезу локалізованих у невеличкому просторі звуків, або спеціальні дослідження пов'язані з перевіркою шляхів поширення звуків для синтезу ефектів реверберації. Проблема вибору звукових фрагментів, що відтворюються, і мінімізація їх кількості, в розглянутих роботах не розглянута.

Метою статті є оптимізація процесу відтворення звукових фрагментів у системі моделювання акустичного оточення імітаційно-тренажерних комплексів спеціальної техніки. Розглянуто вирішення цього завдання з урахуванням особливостей таких систем моделювання – наявність великої кількості короточасних звуків, що одночасно відтворюються, і можливості вільного переміщення спостерігача в просторі моделювання.

2. Визначення внеску звукового фрагмента щодо спостерігача

Всі звукові фрагменти в системі моделювання ІТК можна розділити на позакабінні та внутрішньо-кабінні звуки. Кількість внутрішньокабінних звуків з одного боку обмежена, так як спостерігач може чути звуки лише з кабіни одного єдиного транспортного засобу. З іншого боку, такі звуки найважливіші для процесу навчання і добре локалізовані у просторі, тому немає сенсу їх якимось чином обмежувати. Зовсім інакша ситуація зі звуками позакабінної обстановки, для яких ключовими відмінностями є:

1) кількість звуків що потрібно відтворювати одночасно – відносно велика, та може сягати 100 і навіть 1000 фрагментів;

2) різні види звуків (двигуни, природні звуки, мовлення, спеціальні ефекти) потребують різних умов виведення і мають різну важливість;

3) необхідність брати до уваги швидкість поширення звукової хвилі, та можливість переміщення спостерігача у просторі.

З урахуванням програмно-апаратних обмежень на кількість звуків, що можуть бути відтворені одночасно, особливу проблему становлять звуки спеціальних ефектів, які мають наступні особливості:

1) велика гучність і, як наслідок, дистанція чутності, тому фактично можна вважати що такі звуки не мають локалізації у просторі;

2) непередбачувано велика кількість звуків, що виникають одночасно або з невеликою перервою, яка за тривалістю не дає звільнитися черзі звуків що відтворюються;

3) необхідність просторової обробки, врахування відстані та напряму до спостерігача, що веде до значного споживання обчислювальних ресурсів.

Звуки спецефектів недоцільно перервати, оскільки вони мають відносно коротку тривалість (1-3 секунди). Також, через невелику тривалість для них не має сенсу робити плавне згасання гучності. Ці звуки мають велику гучність, що в більшості випадків експоненційно зменшується, і тому потрібен ще один звуковий канал для плавного згасання, інакше будуть чутні клацання та тріск, при миттєвому перериванні звукового фрагменту.

Звуки цієї групи неможливо відсортувати та вибрати найбільш пріоритетні, оскільки звукове оточення динамічно змінюється і немає накопичення звукових фрагментів за якийсь відчутний період (1 секунда та більше), як для звуків з інших груп (фонів, мовлення, двигунів, та ін.). Затримувати звуки цієї групи для накопичення недоцільно, тому що, зазвичай, є наочний, помітний візуальний ефект до якого вони відносяться, і будь-яка затримка викликає тиме дисонанс і розсинхронізацію звукової та візуальної складової у системі моделювання ІТК.

3. Метод вибору звукових фрагментів, що програються, з урахуванням резервування

Джерела спецефектів і пов'язані з ними звуки, мають довільний порядок виникнення, тому більш віддалені звуки, що з'явилися в момент t_0 , можуть вичерпати вільне місце у черзі звукових фрагментів – для ближчих звуків, що з'являються в момент $t_1 = t_0 + dt$, не буде вільного місця, і, як наслідок, вони не відтворюються. Для того, щоб у черзі залишалося вільне місце для більш відчутних звуків, потрібно додати резервування за наступним алгоритмом, який включає такі кроки:

1) Розділити всі звуки на групи за їх характером та параметрами відтворення, для кожної групи виділити незалежну кількість фрагментів P_{max} , що відтворюються одночасно та залежно від пріоритетності групи.

2) Для кожної групи:

- відсікати новий звук, якщо у безпосередній близькості (менше за 1-2 м) вже програється звук із цієї ж групи;

- відсікати звук, якщо відстань до спостерігача $d > d_{max}$, перевищує максимальну відстань відтворення встановлену для групи;

- розрахувати коефіцієнт завантаження групи a , :

$$a = P/P_{max}, \quad (1)$$

де P – кількість звуків, що відтворюються зараз, P_{max} – кількість фрагментів, що можуть незалежно відтворюватися в межах групи.

3) Поки $a < a_0$, менше деякої межі $0 < a_0 \leq 1$, для відтворення звуку обирається вільний фрагмент і він починає програватися.

4) Інакше визначається необхідність відтворення звуку:

- якщо $a = 1$, то всі фрагменти у черзі зайняті, немає можливості програти звук;

- якщо $a < 1$, та звук розташований на відстані ближче ніж d_0 до спостерігача, тобто $d < d_0$, де d відстань до спостерігача, в цьому випадку, звук завжди відтворюється;

- інакше, для того щоб визначити, чи потрібно відтворити звук, потрібно перевірити додаткові умови, з урахуванням резервування фрагментів у черзі, за якими звук відсікається.

4. Визначення необхідності відтворення звукового фрагмента

Визначимо коефіцієнт відтворення:

$$a_k = \frac{a - a_0}{1 - a_0}. \quad (2)$$

Умови, за якими звук відсікається, визначимо як $d_k < a_k$, де d_k буде пропорційним відносній відстані від спостерігача до позиції звука[^]

$$d_k = (1 - d_1)^2, \quad (3)$$

де $d_1 = \frac{d - d_0}{d_{max} - d_0}$ – відносна відстань, $d_0 \ll d_{max}$.

Таким чином, наведений алгоритм дозволяє зарезервувати місце в черзі звукових фрагментів, для важливих (найбільш чутних) звуків, якщо черга заповнена більш ніж на $100a_0$ відсотків.

5. Експериментальне дослідження розробленого методу

Проведено порівняльне тестування для запропонованого методу з резервуванням і методу з вибором звуку для програвання без резервування. Оцінку проведено за допомогою функції чутності F_h , з урахуванням відстані від i -го джерела звуку до спостерігача d_i :

$$F_h = \frac{1}{N} \sum_{i=0}^N f(d_i), \quad (4)$$

де $d_i = e^{-\frac{d_i}{10}}$.

Підраховується сума гучностей всіх N звуків, які програватимуться (загальний внесок) за період (приблизно 400 секунд) з інтервалом в 1 секунду. Оцінка проводиться виходячи з припущення, що найбільш ефективним є той метод, для якого значення оцінної функції F_h буде більшим – спостерігач чути переважно звуки з великим внеском по гучності. При цьому, очікувано, що для більш ефективного методу кількість звуків, що програватимуться, за період часу буде меншою, за рахунок відкидання звуків, які мають незначний внесок.

Випробування було проведено при симуляції з участю понад 100 активних об'єктів, що безперервно взаємодіють. Спостерігач переміщався у просторі

моделювання таким чином, щоб досягалося близьке до максимального заповнення черги звукових фрагментів, що програватимуться, таким чином імітувалася ситуація, коли за технічними обмеженнями неможливо програти всі звуки які має чути спостерігач. На рис. 1 показана залежність внеску всіх чутних для спостерігача звуків, з урахуванням відстані для методу без резервування та запропонованого методу з резервуванням.

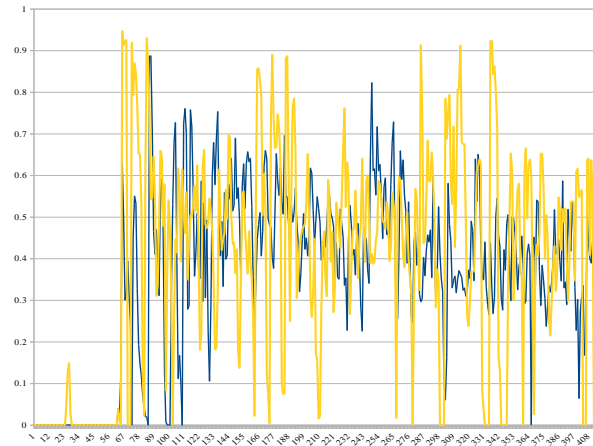


Рис. 1. Залежність внеску всіх чутних спостерігачем звуків з урахуванням відстані: синій – для методу без резервування, померанчевий – для запропонованого методу з резервуванням; по осі абсцис – час з початку моделювання в секундах

На рис. 2 наведені залежності внеску всіх чутних звуків, нормовані щодо кількості звуків, що одночасно програватимуться.

За розглянутий період часу, метод з резервуванням відтворює в середньому в 2 рази менше звуків (для наведеного прикладу це 902, проти 1884 для методу без резервування). При цьому сумарний внесок по гучності для запропонованого методу з резервуванням на 9.5% більше, ніж для методу без резервування (162 проти 148).

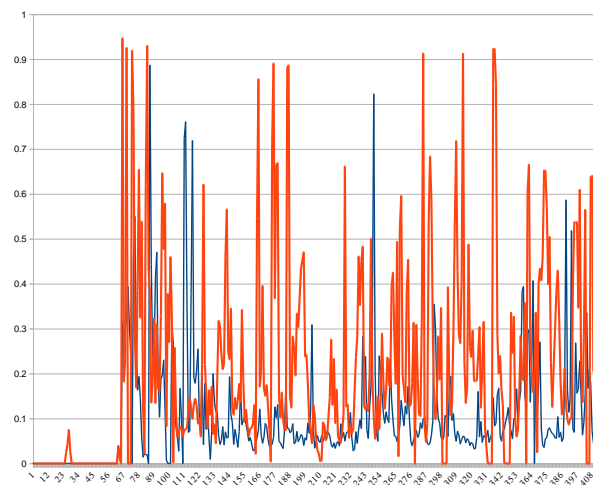


Рис. 2. Залежність внеску для всіх чутних спостерігачем звуків, нормована за кількістю звуків що програватимуться одночасно: синій – для методу без резервування, червоний – для наведеного методу з резервуванням; по осі абсцис – час з початку моделювання в секундах

Висновки

За результатами дослідження запропонованого методу, можна зробити такі висновки:

- 1) Запропонований метод вимагає в середньому у 2 рази менше ресурсів для обробки звукових фрагментів.
- 2) Сумарний внесок чутних спостерігачем звуків стає на 10 % більше.
- 3) Програмна реалізація методу не вимагає значних змін у структурі підсистеми обробки звуків та

її додавання до ІТК не веде до збільшення обчислювальних ресурсів необхідних для обробки звукового потоку.

Запропонований метод дозволяє значно знизити навантаження з розрахунку та обробки звукових фрагментів на обчислювальну систему, а також знизити різні шуми та артефакти, що виникають при змішуванні великої кількості звукових фрагментів, що програватимуться одночасно, за рахунок виключення тих звуків, які вносять найменший внесок у звукову картину.

СПИСОК ЛІТЕРАТУРИ

1. Українські тренажери: теорія і практика // Defence Express. 2010. №1/2. С. 32–52.
2. K.Yao and S.Huang. Simulation Technology and Analysis of Military Simulation Training// Kai Yao and Shaoluo Huang, J. Phys.: Conf. Ser. 2021.**doi: 10.1088/1742-6596/1746/1/012020**
3. Page, E. H. and Smith R. (1998). "Introduction to military training simulation: a guide for Discrete Event Simulationists", Winter Simulation Conference (WSC'98), USA.
4. Бусяк Ю.М., Васильченко О.Г. Побудова структур даних обміну інформацією між підсистемами тренажерів транспортних засобів // Вісн. Нац.техн. ун-ту «ХПІ». Харків, 2002. № 9, т.7: Автоматика та приладобудування. С. 27–30.
5. Larsen M., Gruendell F. A Visual Systems Display for Full-Mission Flight Simulator Training // Presented at the IMAGE VII Conference. Tucson, Arisona, 1994.
6. Pharr M., Green S. Ambient occlusion // GPU Gems: Programming Techniques, Tips and Tricks for Real-Time Graphics. Addison-Wesley Professional, 2004. P. 279–292.
7. Mittring M. Finding Next Gen: Cryengine 2 // SIGGRAPH '07: ACM SIGGRAPH 2007 courses: Special Interest Group on Computer Graphics and Interactive Techniques Conference, San Diego, August 5–9, 2007: Proc.. New York, 2007. P. 97–121.
8. Sennnersten C. and Lindley C. (2009). "An Investigation of Visual Attention in FPS Computer Gameplay", VS-GAMES'09: Games and Virtual Worlds for Serious Applications, March 23–24, Coventry, UK.
9. Spruill M. (2010). Technical Evaluation Report, NATO Modelling & Simulation Group MSG-078 Workshop on Exploiting Commercial Games and Technology for Military Use 8th Workshop, Sept. 2009, NATO Report RTO-MP-MSG-078.
10. D. Miljkovic, "Sample based synthesis of car engine noise," in 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), pp. 1012–1017, Sep. 2020. **doi: 10.23919/mipro48935.2020.9245323**
11. F.Chen, X. Zhang. Synthesising the sound of a car engine based on envelope decomposition and overlap smoothing. Journal of Vibroengineering. August 2021, Vol. 23, Issue 5. pp.1254-1266. **doi:10.21595/jve.2021.21920**
12. J.Jagla, J.Maillard, N.Martin. Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-And-Add method// November 2012. The Journal of the Acoustical Society of America 132(5):3098-108. **doi:10.1121/1.4754663**
13. David A. Heitbrink, Steve Cable. Design of a Driving Simulation Sound Engine// DSC 2007. September 2007.
14. C.Verron, G.Drettakis. Procedural audio modeling for particle-based environmental effects. 133rd AES Convention, Oct 2012, San Francisco, United States. 2012.
15. Y. Dobashi, T. Yamamoto, and T. Nishita, "Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics," ACM Transactions on Graphics (Proc. SIGGRAPH 2003), vol. 22, no. 3, pp. 732–740, 2003. **doi:10.1145/882262.882339**
16. F. Antonacci, M. Foco, A. Sarti, and S. Tubaro. Real time modeling of acoustic propagation in complex environments. Proceedings of 7th International Conference on Digital Audio Effects, pages 274–279, 2004.
17. A.Chandak, C.Lauterbach, M.Taylor, Z.Ren, D.Manocha. Ad-frustum: Adaptive frustum tracing for interactive sound propagation. IEEE Transactions on Visualization and Computer Graphics, 14(6):1707–1722, 2008. **doi: 10.1109/TVCG.2008.111**
18. N.Raghuvanshi, J.Snyder, R.Mehra, M.C. Lin, N.K. Govindaraju. Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. ACM Transactions on Graphics (proceedings of SIGGRAPH 2010), 29(3), July 2010. **doi: 10.1145/1833349.1778805**
19. Tobias Lentz, Dirk Schroder, Michael Vorlander, and Ingo Assenmacher. Virtual reality system with integrated sound field simulation and reproduction. EURASIP J. Appl. Signal Process., 2007(1):187, 2007. **doi: 10.1155/2007/70540**

Received (Надійшла) 14.02.2023

Accepted for publication (Прийнята до друку) 19.04.2023

Method of creating a sound environment in simulation-training complexes of special vehicles

A. O. Zuev, O. M. Yevseienko

Abstract. A method of creating a sound environment in simulation-training complexes, taking into account the distance to the observer and the contribution of a separate sound fragment is considered. The **purpose** of the article is the analysis and optimization the process of selecting sound fragments that are played in the process of modeling the sound environment. An analysis of existing software solutions and mathematical models used in training complexes and virtual reality systems was carried out. An algorithm using a reservation scheme and a functional dependency for evaluating the decision to play a fragment depending on its distance to the observer is given. The issue of practical implementation of the method is considered. According to the **results** of the study, was determined that the proposed evaluation function provides a reduction in the number of the simultaneous played fragments while at one time increasing the total contribution of the fragments heard by the observer.

Keywords: simulation, simulation-training complex, sound environment, sound processing.

Н. В. Ічанська, О. В. Шурпик

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава. Україна

АНАЛІЗ ДИНАМІКИ ВИДОБУТКУ ВУГІЛЛЯ В УКРАЇНІ ЗА ДОПОМОГОЮ МЕТОДІВ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ

Анотація. Сучасний статус промислового розвитку характеризується вивченням статистики, аналізом негативних явищ у розвитку та розміщенні промислового виробництва, оцінкою технологічного рівня та економічної ефективності. У роботі розглядається математична модель, що базується на застосуванні методів апроксимації до заданих даних шляхом побудови кривих регресій. Побудова математичних моделей, що описують тенденції та напрями розвитку різних галузей народного господарства є актуальною проблемою потенціалу продуктивних систем енергоменеджменту та енергосервісу забезпечить реалізацію промислового потенціалу енергозбереження та сприятиме підвищенню енергоефективності. Гірничодобувна промисловість є основою обробної промисловості і вагомою складовою економіки країни. Її сучасний стан та розвиток тісно пов'язані з загальним станом національної економіки, що підкреслює актуальність розглянутої авторами тематики. У даній статті проаналізовано динаміку видобутку вугілля за 2003-2020 рр. та проведено обробку статистичних даних за допомогою сучасних математичних методів. Авторами побудовано математичну модель, дослідження якої може стати основою для пошуку ефективних рішень важливих проблем у видобувній галузі. Аналізуючи побудовані графіки, можна зробити висновки, що найбільш точно апроксимує квадратична залежність і тому повне вивчення даної моделі може стати фундаментом для пошуку ефективних шляхів вирішення важливих проблем добувної галузі промисловості. Проведене в роботі дослідження є ефективним для побудови алгоритмів розв'язання задач видобувної галузі та підвищення ефективності сталого розвитку вугільного сектору промисловості. Основним фактором прогресивного розвитку видобувної галузі є забезпечення зростання обсягів видобутку вугілля. Вивчення шляхів вирішення, проаналізованих у роботі проблем, є запорукою ефективного розв'язку проблем розвитку промисловості, первинного сектору економіки та господарства України в цілому.

Ключові слова: математична модель, математичні методи, добувна промисловість, вугілля, вугільна промисловість, математичне моделювання.

Вступ

Промисловість - це один з ключових секторів економіки країни, її прогрес є вагомим показником визначення рівня економічного розвитку, конкурентоспроможності та рівноправної участі у світовому ринку. Видобувна галузь є вагомою ланкою промисловості України, тому її стабільність та прибутковість є першочерговим завданням розвитку національної економіки, враховуючи значний потенціал по запасам природних ресурсів. Україна може забезпечити обсяги споживання та експорту і має утримувати лідерство за багатьма позиціями на міжнародному ринку.

Паливно-енергетичні ресурси є особливо важливими для нинішньої національної економіки. Умови економічного розвитку держави залежать від масштабів і ефективності використання енергетичної мінеральної сировини та продуктів її переробки, що забезпечує, з одного боку, роботу провідних промислових підприємств та агропромислових комплексів, а з іншого – енергетичну безпеку країни.

За динамікою видобутку твердих корисних копалин в Україні переважає кам'яне вугілля. Проте слід зазначити, що саме видобуток вугілля належить до найбільш проблемних галузей у промисловості України. Важливою обставиною є те, що вугільна індустрія держави дуже строката як за природними умовами, так і за виробничим та технічним рівнем підприємств. Сталий економічний розвиток країни вимагає забезпечення ефективного енергетичного розвитку, вагомою складовою якого

є видобувна галузь. Впровадження ефективних систем енергоменеджменту та енергосервісу забезпечить реалізацію промислового потенціалу енергозбереження та сприятиме підвищенню енергоефективності країни [1]. Сучасні реалії підкреслили вагомість відбудови вугільних шахт, які було пошкоджено з 2014 року і на сьогодні не функціонують. Ремонт і подальша експлуатація цих шахт може забезпечити вихід України на новий економічний рівень, стати стартом стабільного розвитку вугільної промисловості.

Енергетична стратегія України передбачає, що після 2020 року вугільна галузь не буде підпорядковуватись державі, а повністю функціонуватиме в рамках ринкових відносин. Зміни у регулюванні галузі стануть впевненим підґрунтям для залучення приватних інвестицій, що сприятиме її розвитку [2].

Аналіз останніх досліджень і публікацій. Шляхи розвитку видобувної галузі України завжди викликали жвавий інтерес науковців через актуальність проблеми. Питання особливості функціонування видобувної галузі України досліджено у роботах Т.М. Артюха, В.В. Федія, І.В. Григоренка, М.Д. Красножона, А.А. Макурина, З.Я. Шацька, Т.Д. Ганзюка, К.В. Слюсаренка, О.В. Бондар-Підгурської та інші [3, 4]. Питання енергозабезпечення та енергетичної безпеки у своїх працях вивчено А.К. Шидловським, М.П. Ковалком, М.М. Куликом, Б.С. Стогнієм, В.А. Жовтянським, О.М. Суходольським [5]. Зацікавленість цією проблемою пояснюється тим, що розвиток промислового сектору України є вагомим фактором високотехно-

логічного розвитку рівня добувної галузі та сталого розвитку економіки країни.

Математичне моделювання — метод дослідження процесів або явищ шляхом створення та опису відповідних математичних моделей. Застосування методів математичного моделювання до розв'язання прикладних задач присвячено багато робіт (наприклад, [6–8] та інші).

У цій роботі проведено аналіз обсягів видобутку вугілля та побудовано математичну модель, що якнайближче описує вивчене дослідження. Авторами за допомогою методів прогнозування побудовано та описано моделі лінійної, параболічної та гіперболічної регресії, що апроксимують дані обсяги видобутку вугілля в Україні у 2003–2020 роках.

Метою статті є дослідження обсягів й аналіз динаміки виробництва видобутку вугілля, побудова та опис математичних моделей апроксимації, вивчення яких дає можливість знаходження ефективних шляхів розв'язку проблем розвитку промислового комплексу та первинного сектору економіки України, базовою складовою якої є видобувна галузь.

Виклад основного матеріалу

Економіка України складна та багатогранна комплексна структура, основним завдання якої є підвищення рівня життя громадян та сталий розвиток конкурентоспроможного бізнесу. Для виконання поставлених цілей потрібно підтримувати баланс між секторами економіки, які залежать від рівня економічного розвитку.

Основним сектором економіки від якого залежить подальший розвиток інших двох секторів є первинний. Ця секція відповідає за видобуток сировини, які використовуються у інших галузях. Однією із найважливіших корисних копалин, яка активно видобувається в нашій країні залишається вугілля.

Вугільна промисловість є базовою галуззю економіки України та основним енергоносієм. Проте видобуток даної корисної копалини характеризується надзвичайно складними умовами розробки родовищ. Коли йде мова про великі запаси вугілля в надрах країни, слід зауважити, що більше 80% цих запасів зосереджено в пластах потужніс-

тю 1,2 м, які є дуже газоносними, часто небезпечними за раптовими викидами вугілля і газу, вибуховістю пилу, схильністю до самозагорання, які переважно залягають на великих глибинах [1].

У теперішній час видобуток вугілля відбувається на старих шахтах та у складних гірничо-геологічних умовах. Приблизно 96% шахт більше 20 років працюють без модернізованого обладнання, понад 50% машин і спорядження повністю зношені [9–11]. Складним залишається питання із східною частиною вугільної промисловості України. Ситуація, яка складається на окупованих територіях, які в свою чергу є базовими регіонами вугільної промисловості, руйнує систему функціонування видобутку та реалізації горючої корисної копалини. Руйнування інфраструктури та зупинка вугільного бізнесу в окремих східних районах України викликає ризики втрати ролі вугільної промисловості у забезпеченні енергетичної безпеки держави.

Динаміка відновлення розвитку вугільної галузі буде залежати від ефективного відновлення функціонування вугільних підприємств та реформування взаємовідносин на ринку вугільної діяльності.

У цій роботі вивчено й проаналізовано динаміку обсягів видобутку горючої корисної копалини, а саме вугілля, в Україні. Зауважимо, що горючі корисні копалини (г.к.к) – це природні сполуки, які мають здатність горіти і використовуються в народному господарстві як джерело теплової енергії.

Вугілля є тверда осадова порода, яка утворилася під час вуглефікації рослинних залишків під дією тиску та високих температур без доступу кисню. Воно слугує цінним паливом для металургійної та хімічної промисловості [4].

Проведене в цій роботі дослідження базується на даних обсягів видобутку вугілля, які взято сайту державної служби статистики України, з розділу економічна статистика та економічна діяльність, у підрозділі промисловість та пункті виробництво окремих видів промислової продукції [12].

Наведемо вибірку даних державної служби статистики України щодо видобутого вугілля за 2003–2020 роки [12].

Таблиця 1 – Обсяг виробленої продукції добувної промисловості (вугілля, млн т)

Рік	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Вугілля (млн. т)	59,8	59,4	60,4	61,7	58,9	59,5	55	55	62,7	65,7	64,4	45,9	30,2	31,6	24,2	26,3	25,5	24,2

Виклики сучасного життя, суспільно-політичні проблеми суттєво вплинули на обсяги видобутку вугілля в тих підприємствах галузі, які розташовані на сході України, особливо з 2014 року. Наслідком цих змін стало стрімке зменшення загальних обсягів видобутку горючих корисних копалин, зокрема й вугілля. Тому авторами, з метою підвищення точності апроксимації, у дослідженні

використовуються дані табл. 2, які отримано з відповідних показників табл. 1 шляхом усереднення значень з 2014 року та застосування формули (1). Зауважимо, що для усереднення за основу було взято дані за останні чотири роки.

$$\bar{x} = x_0 + \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}, \quad (1)$$

Таблиця 2 – Обсяг виробленої продукції добувної промисловості (вугілля, млн т з перерахунком)

Рік	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Вугілля (млн. т)	59,8	59,4	60,4	61,7	58,9	59,5	55	55	62,7	65,7	64,4	61,95	46,25	47,65	40,45	42,35	41,55	40,25

Зобразимо дані табл. 2 у вигляді графічної залежності (рис. 2). Зауважимо, що значення 1 по вісі абсцис відповідає обсягу виробленої продукції за

2003 р. і всі інші значення вісі абсцис (OX) – це обсяги за відповідний рік. По вісі ординат (OY) подано значення видобутого вугілля (млн. т) по рокам.

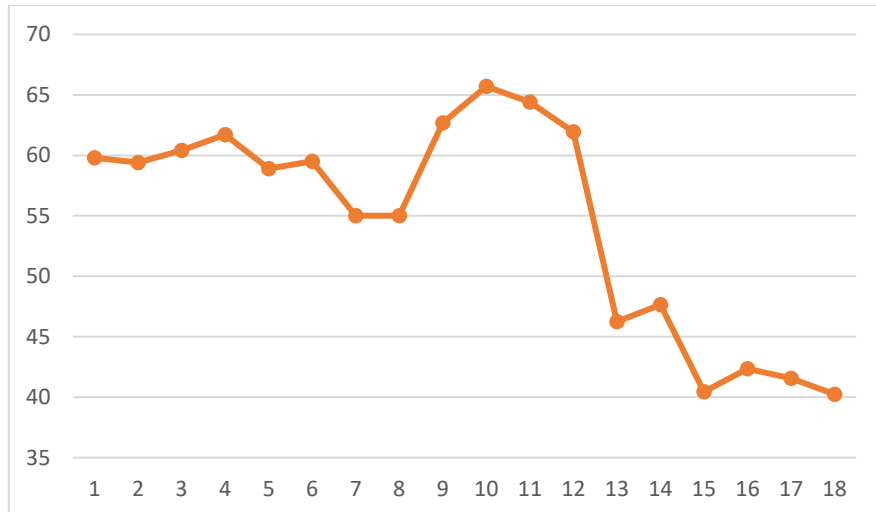


Рис. 1. Обсяг виробленої продукції добувної промисловості (вугілля, млн т)

На графіку суттєво виокремлюється значення обсягів в околі 2010 року, що можна пояснити стрімким ростом обсягів видобутку вугілля згідно до програми розвитку Енергетичної стратегії України [9].

Для аналізу динаміки видобутку вугілля побудуємо математичну модель за допомогою лінійної, квадратичної та гіперболічної залежностей. Щоб побудувати дані залежності використовуємо метод найменших квадратів.

Метод найменших квадратів (МНК) має широкий спектр застосувань, особливе місце займає в методах математичної статистики. Завданням МНК є оцінка спостережуваної закономірності на фоні стохастичних коливань, використовуючи її для подальших розрахунків, зокрема для прогнозування. МНК вплинув на методи опису, відновлення і прогнозування даних самої різної природи. Сутність даного методу полягає в визначенні параметрів моделі тренда, яка найкраще буде описувати тенденцію розвитку якого-небудь явища, в нашому випадку видобутку нафти. Завданням методу найменших квадратів (МНК) – знаходження оптимальної моделі. Ця модель буде оптимальною, якщо сума квадратичних відхилень між спостережуваними фактичними величинами й відповідними їм розрахунковими величинами тренда буде мінімальною (найменшою): $\sum_{i=1}^n (y_i - \bar{y}_x)^2 \rightarrow \min$, де $(y_i - \bar{y}_x)^2$ – квадратичне відхилення між спостережуваною фактичною величиною та відповідною їй розрахунковою величиною тренда; y_i – фактичне (спостережуване) значення явища, що вивчається; \bar{y}_x – розрахункове

значення моделі тренда; n – число спостережень за явищем, що вивчається.

МНК використовується як необхідний технічний прийом при кореляційних дослідженнях, інформаційною основою якого може бути лише достовірний статистичний ряд, причому мінімальне число спостережень 4.

Розв'язання задачі апроксимації та побудови математичної моделі завжди цікавило науковців, а особливо актуально є зараз, коли сучасне суспільство переживає глобальні виклики. У цій роботі авторами знайдено розв'язок цієї задачі методом побудови моделей, які використовують дані обсягів видобутку вугілля за 2003-2020 роки і базується на застужанні МНК.

Першим етапом МНК ми приймаємо, що x – номер дослідження, у даному випадку 2003 рік – 1 спостереження, 2020 рік – 18 спостереження; y – кількість видобутої вугілля. Перевірку гіпотези про існування залежності між « x » та « y » авторами виконано за допомогою комп'ютерних програм, що дозволяє працювати з широкою базою даних та автоматизує проведення обчислень.

Другим етапом методу є визначення траєкторії, яка краще всього може описати тенденцію зміни видобутку обсягів вугілля.

Третій етап – це розрахунок параметрів регресійного рівняння, яке характеризує дану лінію. Знаходження значень параметрів рівняння регресії, у нашому випадку параметрів a і b , є головною частиною МНК. Цей процес зводиться до розв'язання відповідних систем нормальних рівнянь:

- лінійна залежність;

$$\begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i; \end{cases} \quad (2)$$

- параболічна залежність;

$$\begin{cases} na + b \sum_{i=1}^n x_i + c \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i; \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i^3 = \sum_{i=1}^n y_i x_i; \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^4 = \sum_{i=1}^n y_i x_i^2; \end{cases} \quad (3)$$

- гіперболічна залежність.

$$\begin{cases} na + b \sum_{i=1}^n \frac{1}{x_i} = \sum_{i=1}^n y_i; \\ a \sum_{i=1}^n \frac{1}{x_i} + b \sum_{i=1}^n \frac{1}{x_i^2} = \sum_{i=1}^n \frac{y_i}{x_i} \end{cases} \quad (3)$$

Результати розрахунків з параметрами регресійного рівняння, в залежності від вигляду траєкторії опису подаємо у вигляді наступних трьох графіків (рис. 2 – 4). Усі ці траєкторії описують тенденцію зміни видобутку обсягів вугілля. Аналізуючи побудовані графіки, можна зробити висновки, що найбільш точно апроксимує квадратична залежність і тому повне вивчення даної моделі може стати фундаментом для пошуку ефективних шляхів вирішення важливих проблем добувної галузі промисловості.

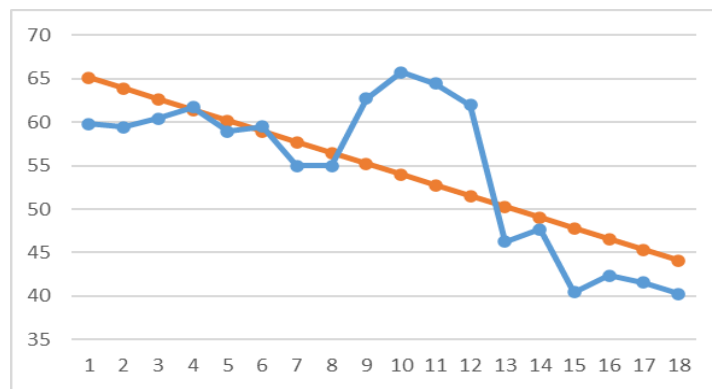


Рис. 2. Лінійна залежність

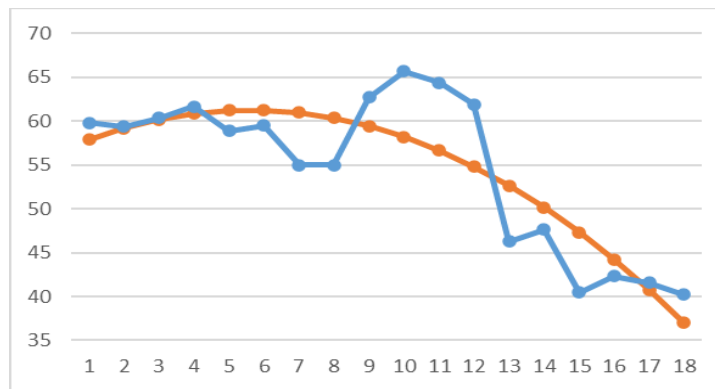


Рис. 3. Параболічна залежність

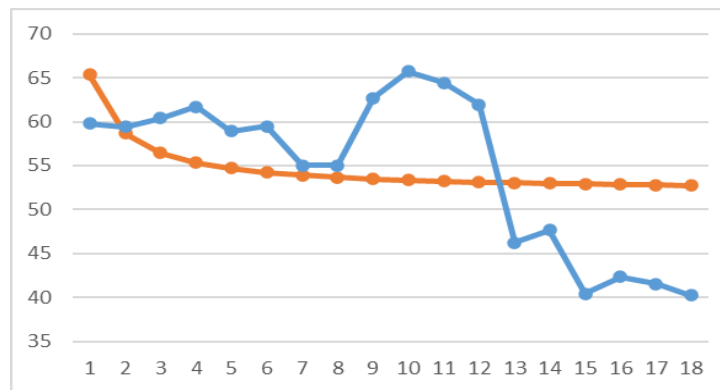


Рис. 4. Гіперболічна залежність

Зауважимо, що проведені в роботі дослідження базуються на даних до 2020 року. На жаль авторам не вдалося знайти статистику по теперішній час, та наведені в [12] дані за 2021 рік про видобуток вугілля ілюструють зменшення обсягу видобутку вугілля на 0,44 млн т, що підтверджує правильність зроблених авторами висновків щодо апроксимації обсягів видобутку вугілля.

Висновки

Важливою складовою загального комплексу економіки України є добувна промисловість, стабільність і прибутковість якої суттєво впливає на функ-

ціонування всіх секторів національної економіки. Основним фактором прогресивного розвитку видобувної галузі є забезпечення зростання обсягів видобутку вугілля. Знаходження ефективних шляхів розв'язання проблем видобувної галузі дасть змогу Україні повернути належний статус країни-імпортера на світовому ринку та забезпечити працевлаштування значної частини населення, успішно розвивати національну економіку. У роботі авторами проведено дослідження обсягів й аналіз динаміки виробництва, побудовано описи траєкторії тенденції зміни видобутку обсягів вугілля у вигляді лінійної, параболічної та гіперболічної залежності.

СПИСОК ЛІТЕРАТУРИ

1. Красножон М.Д. Мінерально-сировинна база України. Стаття 4. Паливно-енергетичні ресурси й перспективи їх нарощування / М.Д. Красножон // Мінеральні ресурси України. – 2015. – № 4. – С. 3–6 [Електронний ресурс]. – Режим доступу : http://nbuv.gov.ua/UJRN/Mru_2015_4_2
2. OECD (2020), Ukraine Реформа підприємств державної форми власності у секторі електроенергетики України, <http://www.oecd.org/corporate/soe-reform-electricity-sector-ukraine.htm>
3. Сьомак О. М. Аналіз сучасного стану видобувної галузі. ISSN 1994-1749. 2020. Вип. 1 (45) Проблеми теорії та методології бухгалтерського обліку, контролю і аналізу. DOI: [http://dx.doi.org/10.26642/pbo-2020-1\(45\)-90-95](http://dx.doi.org/10.26642/pbo-2020-1(45)-90-95)
4. Бондар-Підгурська О.В. Науково-методичні аспекти інноваційного розвитку гірничодобувної промисловості в контексті сталого зростання. Науковий вісник НГУ. 2014. № 1. С. 143-152
5. Кулибаба С. В. Сучасний стан та тенденції енергозбереження в Україні та світі / С. В. Кулибаба, В. Г. Дем'янишин // «Економічні науки». – Серія «Облік і фінанси». Збірник наукових праць. – Луцьк: ЛНТУ, 2010. Вип. 7. – Ч. 4. – С. 163-171. «Економічні науки». – Серія «Облік і фінанси». – Випуск 7 (25). – Ч. 4. – 2010
6. Онищенко В., Ічанська Н., Скриль В., Фурманчук О. (2022). Економіко-математичне моделювання інноваційного розвитку підприємств будівельної галузі В: Онищенко В., Мамедова Г., Сівицька С., Гасимов А. (ред.) Матеріали 3-ї Міжнародної конференції з будівельних інновацій. ICBI 2020. Конспект лекцій з цивільного будівництва, тому 181. Springer, Cham. https://doi.org/10.1007/978-3-030-85043-2_65
7. Горік А.В., Піскунов В.Г., Серов Н.І. Аналітичне рішення задачі про вигин неоднорідного некруглого циліндричного тіла. Міжнародна прикладна механіка 38, 1261-1271 (2002). <https://doi.org/10.1023/A:1022218731489>
8. Горік А.В., Піскунов В.Г., Серов М.І. та співавт. Аналітичне рішення задачі про вигин складеної балки на основі вдосконаленої моделі деформування. Сила Матер 31, 85-98 (1999). <https://doi.org/10.1007/BF02509745>
9. Енергетична стратегія України на період до 2030 року. Схвалено розпорядженням Кабінету Міністрів України від 15 березня 2006 р. № 145. Режим доступу: <http://search.ligazakon.ua>
10. Про схвалення Енергетичної стратегії України на період до 2035 року «Безпека, енергоефективність, конкурентоспроможність»: Розпорядження Кабінету Міністрів України від 18.08.2017 р. № 605-р. Дата оновлення: 18.08.2017. <https://zakon.rada.gov.ua/laws/show/605-2017-%D1%80/sp:dark?dark=0#Text>
11. Міністерство енергетики та вугільної промисловості України. <https://mev.gov.ua/>
12. Офіційний сайт державної служби статистики [Електронний ресурс]. — Режим доступу: <https://www.ukrstat.gov.ua/>

Received (Надійшла) 16.03.2023

Accepted for publication (Прийнята до друку) 10.05.2023

Analysis of the dynamics of coal mining in Ukraine using the methods of mathematical modeling

N. Ichanskaya, O. Shurpik

Abstract. The current status of industrial development is characterized by a study of statistics, an analysis of negative phenomena in the development and location of industrial production, an assessment of the technological level and economic efficiency. The paper considers a mathematical model based on the application of approximation methods to given data by constructing regression curves. The construction of mathematical models that describe the trends and directions of development of various sectors of the national economy is an urgent problem today. The introduction of productive energy management and energy service systems will ensure the realization of industrial energy saving potential and will contribute to energy efficiency. The mining industry is the basis of the manufacturing industry and a significant component of the country's economy. Its current state and development are closely related to the general state of the national economy, which emphasizes the relevance of the topic considered by the authors. This article analyzes the dynamics of coal production for 2003-2020. and processing of statistical data using modern mathematical methods. The authors have built a mathematical model, the study of which can become the basis for finding effective solutions to important problems in the extractive industry. Analyzing the constructed graphs, we can conclude that the quadratic dependence is most accurately approximated, and therefore a complete study of this model can become the foundation for finding effective ways to solve important problems in the extractive industry. The study carried out in the work is effective for constructing algorithms for solving the problems of the extractive industry and increasing the efficiency of the sustainable development of the coal sector of the industry. The main factor in the progressive development of the mining industry is to ensure the growth of coal production. The study of the ways of solving the problems analyzed in the work is the key to an effective solution to the problems of the development of industry, the primary sector of the economy and the economy of Ukraine as a whole.

Keywords: mathematical model, mathematical methods, extractive industry, coal, coal industry, mathematical modeling.

Н. В. Караєва, Л. О. Левченко

Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна

МЕТОДОЛОГІЧНІ ЗАСАДИ УПРАВЛІННЯ СТАЛИМ РОЗВИТКОМ ЕНЕРГЕТИКИ В УМОВАХ ГЛОБАЛЬНИХ РИЗИКІВ І БІФУРКАЦІЙНИХ ЗРУШЕНЬ

Анотація. Проаналізована проблематика забезпечення сталого розвитку енергетики в умовах глобальних ризиків, зростання глобальної невизначеності, непередбачуваності та біфуркаційних зрушень, пов'язаних із Четвертою промисловою революцією та зеленим «енергетичним» переходом. Доведена необхідність розгляду сучасної енергетики як складної емерджентної системи в умовах багатоваріантності та нелінійності векторів розвитку глобальної економіки, енергетики. На основі аналізу методологічних положень теорій лінійної та нелінійної (синергетичної) парадигм обґрунтовано доцільність використання синергетично-кібернетичного підходу до управління сталим розвитком енергетики, враховуючи, що траєкторії функціонування і розвитку енергетики притаманні синтез еволюційного (лінійного, впорядкованого) та хаотичного (нелінійного) процесів. Визначено, що основним напрямом кібернетичного управління енергетичною системою є пошук та реалізація таких керуючих впливів, які за наявності зовнішніх і внутрішніх флуктуацій забезпечать стійкий, безпечний, гомеостатичний режим функціонування системи, у якому вона досягне максимуму своєї ефективності. Зазначено, що проблема гомеостазу (стійкості) в кібернетичі розглядається з чисто функціональної точки зору, і тому в ній (на відміну від синергетики) не аналізуються конкретні механізми самоорганізації. Синергетика досліджує нелінійні процеси самоускладнення систем, виникнення «порядку із хаосу» (і навпаки) та вивчає процеси, коли можуть виникати біфуркаційні зрушення, декілька станів рівноваги (атрактор) і коли керуючі впливи не регулюють безпосередньо поведінку системи. Доведено, що синергетично-кібернетичний підхід до управління сталим розвитком енергетики необхідно розглядати як міждисциплінарний, що ґрунтується на широкому застосуванні апарату синергетичного моделювання (нечітка логіка, нейронні мережі, еволюційні алгоритми, фрактали, елементи теорії хаосу тощо), економіко-математичних, балансових, соціо-психологічних та й інших моделей та методів.

Ключові слова: сталий розвиток енергетики, глобальні ризики, лінійна парадигма, нелінійна парадигма, біфуркація, синергетично-кібернетичний підхід.

Вступ

Ще у 2012 році резолюцією № 67/215 Генеральної Асамблеї ООН «Сприяння розширенню використання нових та відновлюваних джерел енергії» (Promotion of new and renewable sources of energy), 2014–2024 роки оголошено «Десятиліттям сталої енергетики для всіх» [1]. В Енергетичній стратегії України на період до 2035 року зазначено, що сталий розвиток енергетики має стати першим кроком для оздоровлення та зростання економіки країни в цілому [2]. Також сталий розвиток енергетики є важливою передумовою досягнення Цілей сталого розвитку (ЦСР), визначених ООН у документі «Порядок денний в галузі сталого розвитку на період до 2030 року» [3] та виконання Паризької угоди про зміну клімату [4]. Підвищення енергоефективності, збільшення частки відновлювальних джерел енергії (ВДЕ) та перехід на більш екологічно чисті енерготехнології мають важливе значення для сталого розвитку, в тому числі для боротьби зі зміною клімату.

Сучасний світ входить у нову еру глобальних ризиків, зростання глобальної невизначеності, непередбачуваності та біфуркаційних зрушень (зокрема перехід на шостий технологічний уклад у рамках Четвертої промислової революції та перехід до зеленої економіки й енергетики). На думку експертів Стокгольмського міжнародного інституту досліджень проблем миру (Stockholm International Peace Research Institute, SIPRI) глобальні кризи у сферах безпеки людства (геополітичні, воєнні, енергетичні, соціальні тощо) та екології посилюють одна одну та генерують конфліктність, нестабільність, створюють взаємо-доповнювані нові ризики і

загрози, яким притаманна багатоаспектність прояву та масштабність наслідків [5]. Ключовими глобальними ризиками і загрозами у сфері безпеки людства є війни та інші збройні конфлікти на міжетнічному та релігійному ґрунті; нарощування ядерних озброєнь; міжнародний тероризм; суспільно-політичні заворушення насильницького характеру; геноцид населення тощо. Наприклад, війна Російської Федерації (РФ) проти України створила комплекс загроз воєнного, соціально-економічного, техногенно-екологічного характеру, небезпеку ядерної війни, ризики світовій енергетичній і продовольчій безпеці. Згідно дослідження компанії Economist Intelligence Unit (EIU), саме агресія РФ стала вагомим ризик-чинником для 81 % країн, породжуючи найбільші глобальні ризики саме у сфері енергетичної безпеки для країн Європейського Союзу та інших держав [6]. Сьогодні локальні війни ведуться, головним чином, за володіння енергоресурсами, за контроль над чужими запасами енергоресурсів та магістральними трубопроводами транспортування нафти і газу.

Різноманітність множини факторів глобальних, національних ризиків, їх проявів ускладнюють процес прийняття стратегічних рішень в енергетиці, враховуючи, що за змістовною сутністю ризик є слабо структурованим, оскільки виникає в результаті невизначеності подій та ситуацій, зовнішнього середовища, неповноти і неточності формалізації задачі, нечіткості як самих вхідних даних, процесів, так і моделей, що їх описують.

Постановка проблеми та аналіз останніх досліджень і публікацій. Дослідження з розробки обґрунтованих прогнозів розвитку енергетики та формування на їх основі стратегії державної політики

ґрунтується, переважно, на лінійній парадигми класичної і неокласичної економічної науки, яка є найбільш традиційною для індустріального ХХ століття та слугує базисом кібернетичного підходу до управління енергетикою, під терміном якої, як правило, розглядається "енергетична галузь" або "паливно-енергетичний комплекс", тобто сукупність складних, ієрархічних, керованих техніко-економічних систем (галузей, підприємств), що забезпечують країну паливом, електричною і тепловою енергією, функціонування яких відбувається за фізичними, детермінованими законами. Основними групами методів дослідження у межах кібернетичного підходу є: балансові, економетричні та дослідження операцій (методи оптимізації та математичного моделювання).

В Україні енергетичний баланс є необхідним інформаційним джерелом для оперативного управління та стратегічного планування розвитку енергетики і використовуються, зокрема, для аналізу і зміни прогнозування потреби в енергоресурсах [7, 8]. Але в умовах глобальних ризиків і біфуркаційних зрушень використання балансових моделей значно ускладнює в результаті спотворення надійної інформаційної бази щодо руху енергетичних потоків, зокрема, достовірності розрахунку натуральних обсягів споживання, показників імпорту, експорту та вартісних параметрів, значення яких залежать від стану економіки інших країн та міжнародних фінансових і енергетичних ринків.

Економетричні методи, що базуються на побудові економіко-математичних функцій (моделей) і засновані на регресійному аналізі та виявленні тенденцій розвитку (екстраполяції тренду), доцільно використовувати лише для керованих екзогенних параметрів (зокрема, облікова ставка, тарифи на енергетичні ресурси та послуги, енергоефективність нового обладнання або норми викидів забруднюючих речовин), які є "нормативно" встановлені і для короткострокових прогнозів.

Швидкий розвиток комп'ютерних технологій дозволило зосередити економіко-математичні дослідження вже на розробленні оптимізаційних моделей. Методи лінійного (цільова функція і область пошуку визначаються лінійними залежностями), нелінійного (цільова функція задана як лінійними так і нелінійними залежностями при допустимій області пошуку), стохастичного (цільова функція задається імовірнісною функцією) програмування в енергетиці стають найбільш поширеними. Наприклад, при формуванні сценаріїв для плану дій реалізації Енергетичної стратегії України на період до 2035 року, використовується комплексна енергетична модель програмування «Times-Україна» [9], яка за методологічним підходом є лінійною оптимізаційною квазідинамічною економіко-математичною моделлю виробничого типу часткової рівноваги з фіксованим попитом, у якій окремі модельні розрахунки проводяться із підключенням спеціальних модулів, реалізованих з елементами цілочисельного, нелінійного та стохастичного програмування [10]. Але функціонування і розвиток енергетики здебільшого визначається станом зовнішнього середовища (економічної, соціальної, політичної, екологічної), тому сьогодні стохастичні (випадкові події) фактори розвитку енергетичної системи, у значній

мірі, спотворюють результати, отримані за допомогою детермінованих економіко-енергетичних моделей.

Особливо після світової фінансової кризи 2008-09 рр. у зарубіжному і вітчизняному науковому колі категоричність суджень щодо застарілості методології лінійної парадигми та кібернетичного підходу управління системами посилюється, що є, на наш погляд, достатньо дискусійним питанням враховуючи, що процес управління сталим розвитком енергетики, з однієї сторони має бути спрямовано на забезпечення сталого (стійкого, стабільного) стану енергетики а, з іншої – на розвиток, тобто перехід її в якісно новий стан.

Метою статті є розгляд методологічних засад управління сталим розвитком енергетики в умовах глобальних ризиків і біфуркаційних зворушень.

Виклад основного матеріалу

Загалом проблематика управління сталим розвитком енергетики, на наш погляд, лежить у площині дуалістичності категорій сталість (стійкість) і розвиток в контексті одного часового простору. Дефініція поняття "сталість", змістовна сутність якої більш конкретизується терміном "гомеостаз" (від грецьких *homios* – однаковий і *statis* – стан), розглядається як здатність системи зберігати динамічну рівновагу і розширювати свої структурно-функціональні властивості при впливі на неї збурених дій навколишнього середовища. Механізм гомеостазу направлений на досягнення оптимальної структури елементів системи і визначає умови безпечного її існування. У вищезазначеному контексті категорія «сталій (стійкій)» та «безпечний» нами розглядається як синоніми. Безпека підвищує стійкість системи – здатність системи майже не відхилитися від впорядкованого (обраного) руху під впливом випадкових зовнішніх та внутрішніх подій (флуктуацій, небезпек). Саме на забезпечення енергетичної безпеки, що є важливою умовою досягнення ЦСР, у першу чергу, спрямована енергетична політика і стратегії України на період до 2035 року. Найвищий рівень безпеки досягається за умови, що весь комплекс показників стану системи перебуває в певних допустимих, порогових межах і має свої граничні (нормативно визначені державними органами) значення. Наприклад, у моделі "TIMES-Україна" сценарні значення таких ендегенних і керованих показників як облікова ставка, тарифи на енергетичні ресурси та послуги, енергоефективність нового обладнання або норми викидів, що визначаються відповідними державними органами, можуть бути чітко обґрунтовані й прогнозовані лише за умов стійкого, безпечного стану енергетики.

Дослідження системи у стійкому стані ґрунтуються, переважно, на підходах лінійної парадигми, концептуальними положеннями та принципами якої є: детерміністське уявлення про стійкість; раціональна поведінка "людини економічної" (*homo economicus*); ідея ринкової рівноваги; постулат про декілька станів рівноваги в системі; присутність у системі локальних оптимумів; лінійність взаємозв'язків між параметрами; прогнозованість процесів розвитку системи; пошук найбільш значущих, сильних чинників, тоді як слабкі чинники переважно виключаються з аналізу чи розглядаються як випадкові; керованість системою. Фізичний

зміст принципу лінійності зводиться до твердження, згідно з яким відгуки системи на відносно малі впливи на неї лінійно (пропорційно) залежать від їх сили.

Критерієм раціональної поведінки при виборі оптимальної альтернативи є максимізація корисності в умовах обмеженості ресурсів. Домінуючою теорією дослідження раціональної економічної поведінки досі залишається неокласична, яка доповнена аналізом прийняття рішень за умов ризику (теорія очікуваної корисності Неймана-Моргенштерна) та невизначеності (теорія суб'єктивної очікуваної корисності Л. Севіджа). Принцип максимізації очікуваної корисності сприймається як основний критерій раціональності рішення та вибору.

В енергетичній сфері вихідні положення лінійної парадигми (лінійної динаміки) полягають у наступному: суб'єкти енергетичних ринків (енергокомпанії (виробники), споживачі, інвестори, державні і регіональні органи влади) діють раціонально та намагаються досягти власних цілей, зокрема мінімізація витрат (збитків), максимізація прибутковості; існує прозорість конкурентних ринків факторів виробництва; ринкова рівновага (баланс попиту та пропозиції) досягається в результаті балансу цілей суб'єктів ринку тощо. З математичної точки зору цілі формуються у вигляді цільової функції та системи різноманітних обмежень (екологічних, економічних, політичних тощо), а в якості методичного інструментарію використовуються моделі міжгалузевого балансу, різноманітні виробничі функції (наприклад, виробничі функції Кобба–Дугласа чи Р. Солоу та Е. Денісону, економічні побудови Кейнса), лінійні диференціальні рівняння, у яких невідомі величини входять до ступеня не вище одиниці (наприклад, рівняння Максвелла, рівняння Гамільтона та ін).

У межах кібернетичного підходу до управління енергетичною системою головна увага приділяється вирішенню основного завдання – знаходження оптимального стаціонарного (рівноважного) стану системи, а категоріями прийняття оптимальних рішень є максимізація споживання окремого виду енергоресурсу, максимізація прибутку, мінімізація загальних витрат, мінімізація або досягнення визначеного рівня викидів парникових газів та ін. Але у багатьох випадках оптимальне рішення неможливо вибрати однозначно не лише для ситуацій з ризиками, невизначеністю, але й для детермінованих, враховуючи, що у процес прийняття оптимальних рішень залучаються багато суб'єктів енергетичних ринків із різними інтересами. В умовах конфлікту інтересів різних груп суб'єктів глобального і національного енергетичних ринків (а це поширене явище), ефективність рішень повинні оцінюватися за багатьма, як правило, суперечливими критеріями, зокрема: надійність, безпечність функціонування енергосистем – екологічні вимоги – економічна ефективність – соціальна стабільність. Ігрові підходи управління системою в умовах конфлікту є базовими серед класів задач теорії прийняття рішень. Однак, при використанні стандартних математичних алгоритмів рішення багатокритеріальних задач (методи головної компоненти, Парето-оптимальність, лінійне звертання вагових коефіцієнтів та ін) розв'язок детермінованих моделей не завжди може бути якісно та

кількісно економічно інтерпретовано, тому при моделюванні енергетичних систем намагаються уникати багатокритеріальної оптимізації.

Загалом кібернетика та різноманітні напрямки загальної теорії систем і системного аналізу, на основі функціонального підходу «сигнал-відгук», формують гіпотези про внутрішній склад і будову систем, намагаючись описувати нелінійні процеси еволюції систем за допомогою лінійних моделей (коли це можливо). При цьому акцент робиться на статистичні системи, на морфологічному та функціональному опису. Але інструментарій моделювання, прогнозування лінійної динаміки не спроможна адекватно описати поведінку енергетичної системи чи її складників за нестационарного (нерівноважного) стану, враховуючи, що можливість застосування регресійного аналізу обмежується вимогами до стаціонарності прогнозованого процесу, тобто стійкості у часі його ймовірнісних характеристик.

Основним напрямом кібернетичного управління енергетичною системою є пошук та реалізація таких керуючих впливів, які за наявності зовнішніх і внутрішніх збурень (флуктуацій) забезпечать стійкій, безпечний, гомеостатичний режим функціонування системи, у якому вона досягне максимуму своєї ефективності, найбільш продуктивного режиму функціонування. Іншими словами, кібернетичний підхід, у значній мірі, спрямований на вивчення гомеостатичних процесів, спираючись в основному на негативні зворотні зв'язки і на формалізованому абстрактноматематичному забезпеченні. Але проблема гомеостазу (стійкості) в кібернетичі розглядається з чисто функціональної точки зору, і тому в ній не аналізуються конкретні механізми самоорганізації (самодії) та виникнення якісно нових систем в результаті хаотичних (біфуркаційних) процесів.

Змістовна сутність поняття "розвиток" не може бути пояснена, виходячи лише з логіки лінійної парадигми (неокласичної економічної теорії). Стратегічне управління сталим розвитком енергетикою в умовах глобальних ризиків, невизначеності, біфуркації вимагає використання підходів нелінійної (як синонім синергетичної) парадигми та коригування сутності дефініції поняття "енергетика". З погляду синергетики, як методології постнеокласичної науки, у роботі під поняттям «енергетика» розуміється енергетична галузь, яка розглядається як складна емерджентна система, що складається з взаємопов'язаних підсистем (технологічної, соціально-економічної, екологічної) і охоплює весь комплекс економіко-господарських відносин, пов'язаних з оборотом енергоносіїв (видобуток, переробку, транспортування енергоресурсів та постачання енергії у зручних для споживання формах). Розгляд сучасної енергетики саме як емерджентної системи необхідне в умовах багатоваріантності та нелінійності векторів розвитку глобальної економіки, енергетики, що визначаються як фактором впливу зовнішнього середовища (глобальні кризи у сферах екології та безпеки людства, зростання різного роду конфліктів), так і зміною структурно-функціональних властивостей самої енергетичної системи в результаті прояву наслідків збурюючих подій зовнішнього середовища.

У контексті синергетичної парадигми функціонування і розвиток енергетики слід розглядати як послідовність періодів, що відповідають стійким (гомеостатичним) станам системи, які перериваються періодами біфуркації (хаосу), після чого відбувається перехід до якісно нового стійкого, рівноважного стану (атрактору) або руйнування. Вибір атрактору в точці біфуркації визначається випадковими флуктуаціями, тобто збуджуючими імпульсами (події типу "Чорного Лебедя" Н. Таллеби), які викликають ланцюгову реакцію, що здатна охопити багато елементів системи та сприяють виведенні системи за поріг стійкості (гомеостазу) і початку фази руйнівного хаосу. Пандемія COVID-19, військова агресія РФ стали спусковим гачком сучасної хаотизації світового порядку, глобальної економіки й енергетики. Лише у життєздатних системах із достатнім рівнем потенціалу розвитку хаос виконує конструктивну функцію, породжуючи якісно нові форми порядку. Так, глобальний хаос в енергетичній галузі, спричинений військовою агресією РФ і наслідками цієї агресії (руйнуванням газопроводів і української енергетичної інфраструктури, введенням санкцій), сприяє пришвидшенню виходу енергетики багатьох країн із минулої траєкторії використання викопного палива (вугілля, нафта, природний газ) і ядерного палива (урану) та забезпечує можливість швидкого зеленого «енергетичного переходу» (тобто переходу до якісно нового стійкого, рівноважного стану енергетики). Взаємопов'язані переходи системи від порядку до хаосу і від хаосу до якісно нового порядку здійснюються лише у стані нестійкості (кризи, катастрофи). Слід зазначити, що якісно новий не обов'язково значить кращий.

Саме синергетика (на відміну від кібернетики), досліджує нелінійні процеси самоускладнення систем, виникнення "порядку із хаосу" (і навпаки) та вивчає процеси, коли можуть виникати біфуркаційні зрушення, декілька станів рівноваги (атрактору) і коли керуючі впливи не регулюють безпосередньо поведінку системи, а через використання інструментів "м'якої сили" (soft power), здатні "підштовхнути систему" до формування внутрішніх механізм її самоорганізації.

Синергетична самоорганізація забезпечується позитивним зворотним зв'язком, який відповідає за вибір атрактору в точці біфуркації та переведення системи на якісно новий тренд розвитку. В результаті розвитку змінюється не тільки структура системи, але і її поведінка, функціонування елементів та взаємозв'язки між ними (цим обмежується використання методичного інструментарію лінійної парадигми, зокрема методів теорії дослідження операцій, моделей міжгалузевого балансу тощо). Іншими словами, випадкові флуктуації є необхідною умовою соціально-економічних, політичних, інституціональних якісних змін. У відсутності флуктуацій, за умов чистої рівноваги чи квазі-рівноваги економічна, енергетична системи, як на макро-, так і мікрорівні поступово деградуують.

Вищезазначені положення свідчать, що аналіз і управління енергетичною системою в умовах глобальних ризиків, невизначеності і біфуркаційних зрушень додатково потребує використання методологічного апарату теорій нелінійної парадигми, що враховують відкриті постнеокласичною наукою синергетичні ефекти. Мова йде про теорії синергетики

(Г. Хакена), дисипативних структур, самоорганізації (І. Пригожина), динамічного (детермінованого) хаосу (М. Фейгенбаума), самоорганізуючої критичності або самоорганізації на "краю хаосу" (П. Бака) та дослідження фракталів (Б. Мандельброта). У постнеокласичному (синергетичному) науковому середовищі вище наведені теорії об'єднанні міждисциплінарною "Complexity Science" ("Наука про складне").

У синергетичній парадигмі теорія обмеженої раціональності Г. Саймона стає базисом прийняття управлінських рішень. Ідеї теорії Г. Саймона спонукали до стрімкого розвитку дослідження на стику економіки, психології, біології, соціології та формуванні нових галузей (теорій) постнеокласичної економіки: поведінкова (біхевіористська) економіка і нейроекономіка. Основним синергетичним інструментарієм моделювання і прогнозування енергетики в умовах відсутності стаціонарності аналізованих показників і незмінності умов та мультиплікативного впливу глобальних ризик-чинників невизначеності, конфліктності є нечітка логіка, нейронні мережі, еволюційні алгоритми, фрактали, елементи теорії хаосу тощо.

Але слід не забувати, що моделювати енергетичну систему у термінах синергетики потрібно з обережністю: будь-яка система має "пам'ять", тобто не забуває початкових (граничних) умов. Крім того, складна структура енергетичної системи має не тільки пам'ять, але й обмежені можливості передбачення майбутнього стану: в одних периферійних фрагментах цієї структури (наприклад, нафтовій, вугільній галузі) процеси протікають ще так, як у всій структурі (наприклад, паливно-енергетичному комплексі) у минулому, в інших – як будуть в майбутньому. Тобто, якщо говорити мовою синергетики, стійкий стан окремих підсистем виявляється підірваним, але хаотизація самої системи ще не настала. Це пояснюється тим, що аттрактори мають інваріантні рішення, в яких простір і час залежать і тісно пов'язані один з одним.

Вибір нового атрактору сталого розвитку енергетики можуть визначитися початковими умовами (попереднім станом системи), пам'яттю суб'єктів економіко-господарських відносин (енергетичного ринку), зворотними зв'язками, накопиченим соціально-економічним, екологічним, інституціонально-управлінським потенціалом розвитку, а також флуктуаціями типу "Чорного Лебедя".

Висновки

Процес управління сталим розвитком енергетики має ґрунтуватися на синергетично-кібернетичному підході, враховуючи, що траєкторії функціонування і розвитку енергетики притаманні синтез еволюційного (лінійного, впорядкованого) та хаотичного (нелінійного) процесів.

Використання методологічного інструментарію лінійної парадигми класичної і неокласичної економічних теорій, які слугують базисом кібернетичного підходу до управління енергетикою є достатньо логічним в задачах розробки державної політики забезпечення стійкого, безпечного функціонування енергетики.

Основним напрямом кібернетичного управління енергетичною системою є пошук та реалізація таких керуючих впливів, які за наявності зовнішніх і внутрішніх збурень (флуктуацій) забезпечать стійкий, безпечний, гомеостатичний режим функціонування системи, у якому вона досягне максимуму своєї ефективності.

Основними групами методів дослідження у межах кібернетичного підходу є: балансові, економетричні та дослідження операцій (методи оптимізації та математичного моделювання). Але проблема гомеостазу (стійкості) в кібернетичі розглядається з чисто функціональної точки зору, і тому в ній (на відміну від синергетики) не аналізуються конкретні механізми самоорганізації. Саме синергетика досліджує нелінійні процеси самоускладнення систем, виник-

нення «порядку із хаосу» (і навпаки) та вивчає процеси, коли можуть виникати біфуркаційні зрушення, декілька станів рівноваги (атрактору) і коли керуючі впливи не регулюють безпосередньо поведінку системи, а через використання інструментів "м'якої сили", здатні "підштовхнути систему" до формування внутрішніх механізм її самоорганізації.

Синергетично-кібернетичний підхід до управління сталим розвитком енергетики необхідно розглядати як міждисциплінарний, що ґрунтується на широкому застосуванні апарату синергетичного моделювання (нечітка логіка, нейронні мережі, еволюційні алгоритми, фрактали, елементи теорії хаосу тощо), економіко-математичних, балансових, енергетичних, соціо-психологічних та й інших моделей та методів.

Список літератури

1. UNGA resolution 67/215. 'Promotion of new and renewable sources of energy' UN doc. A/RES/67/215, of 21 December 2012. URL: https://archive.unescwa.org/sites/www.unescwa.org/files/un_resolutions/a_res_67_215_e.pdf (assessed 10 March 2023).
2. Енергетична стратегія України на період до 2035 року «Безпека, енергоефективність, конкурентоспроможність». URL: www.zakon.rada.gov.ua/signal/kr06145a.doc (дата звернення 10.03. 2023).
3. Transforming our world: The 2030 Agenda for Sustainable Development. Resolution adopted by the UN General Assembly (2015). URL: <https://sdgs.un.org/2030agenda> (assessed 10 March 2023).
4. Paris Agreement on Climate Change. – UN (2015). URL: <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement> (assessed 10 March 2023).
5. World stumbling into a new era of risk, concludes SIPRI report / Stockholm International Peace Research Institute (SIPRI). 2022. 23 May. URL: <https://www.sipri.org/media/press-release/2022/world-stumbling-new-era-...> (assessed 10 March 2023).
6. Global operational risk review. How war is fuelling geopolitical uncertainty / Economist Intelligence Unit (EIU), 2022. URL: <https://www.eiu.com/n/campaigns/operational-risk-review-2022> (assessed 10 March 2023).
7. Розпорядження Кабінету Міністрів України «Про утворення Міжвідомчої координаційної ради з питань розроблення енергетичного балансу держави» від 1 березня 2006 року № 121-р. URL: <http://kmu.gov.ua/control/uk/cardnpd?docid=30531237> (дата звернення 10.03. 2023).
8. Рубан-Максимець О. О. Оптимізаційні моделі прогнозування потреби в енергоресурсах на основі синтезу методів формування перспективного міжгалузевого та паливно-енергетичного балансів з урахуванням екологічних обмежень. *Проблеми загальної енергетики*, 2010. № 2 (22). С. 12–17.
9. Довгострокове енергетичне моделювання та прогнозування в Україні: сценарії для плану дій реалізації Енергетичної стратегії України на період до 2035 року. Київ-Копенгаген 2019. URL: https://ens.dk/sites/ens.dk/files/Globalcooperation/long-term_energy_modelling_and_forecasting_in_ukraine_ukrainian.pdf.
10. Подолець Р. З., Дячук О. А. Стратегічне планування у паливно-енергетичному комплексі на базі моделі «TIMES-Україна»: наук. доп. / НАН України; ІЕП. Київ, 2011. 150 с.

Received (Надійшла) 24.02.2023

Accepted for publication (Прийнята до друку) 17.05.2023

Methodological principles of management of sustainable energy development in the conditions of global risks and bifurcation shifts

Nataliia Karaieva, Larysa Levchenko

Abstract. The problem of ensuring the sustainable energy development in the conditions of global risks, the growth of global uncertainty, unpredictability and bifurcation shifts associating with the Fourth Industrial Revolution and the green "energy" transition is analyzed. The necessity of consider modern energy as complex emergent system in conditions of multivariate and non-linear development vectors of the global economy, energy is proven. Based on the analysis of the methodological provisions of the theories of linear and non-linear (synergistic) paradigms, the expediency of synergistic-cybernetic approach using for sustainable energy development managing is substantiated, taking into account that the energy trajectories of functioning and development is inherent in the synthesis of evolutionary (linear, orderly) and chaotic (non-linear) processes. It is determined that the main direction of cybernetic management of the energy system is the search and implementation of such controlling influences that, in the presence of external and internal fluctuations, will ensure a stable, safe, homeostatic mode of functioning of the system, in which it reaches its maximum efficiency. It is noted that the problem of homeostasis (stability) in cybernetics is considered from a purely functional point of view, and therefore (unlike synergetic) it does not analyze specific mechanisms of self-organization. Synergetic investigates non-linear processes of self-complexity of systems, the emergence of "order from chaos" (and vice versa) and studies processes when bifurcation shifts, several states of equilibrium (attractor) can occur, and when controlling influences does not directly regulate the system behavior. It is proved that the synergistic-cybernetic approach for managing the sustainable energy development must be considered as interdisciplinary, based on the wide application of the apparatus of synergistic modeling (fuzzy logic, neural networks, evolutionary algorithms, fractals, elements of chaos theory, etc.), economic-mathematical, balance, socio-psychological and other models and methods.

Keywords: sustainable energy development, global risks, linear paradigm, non-linear paradigm, bifurcation, synergistic-cybernetic approach.

O. Skakalina, A. Kapiton

National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine

IMPLEMENTATION OF THE LIGHTING FUNCTION IN THE SMART HOME CONCEPT

Abstract. The principle of "System of intelligent management of the building" provides a completely new approach in the organization of life support of the building, in which the efficiency of operation and reliability of management of all systems and executive devices of the building increases significantly due to the combination of hardware and software. A smart building should be understood as a system that should be able to recognize specific situations occurring in the building and react accordingly: one of the systems can control the behavior of others according to a previously developed algorithm. The main feature of such a building is the unification of separate subsystems into a single controlled complex. An important feature and property of a smart house is the way of organizing the living space - this is the most progressive concept of human interaction with the living space, when a person sets the desired environment with one command, and the automation, in accordance with external and internal conditions, sets and monitors the modes of operation of all engineering systems and electrical devices. In this case, there is no need to use several remotes when watching TV, many switches when controlling lighting, separate blocks when controlling ventilation and heating systems, video surveillance and alarm systems.

Keywords: smart home, smart lighting, remote control, software, ESP8266, microcontroller, WI-FI, HTTP.

Introduction

The development of the "smart home" concept was largely due to the rapid development of innovative technologies and the improvement of sensors. The concept of "smart light" is one of the components of the global IoT system - the "Internet of things" - the global infrastructure of the information society. The IoT system provides real-time services by connecting physical and virtual things based on existing and evolving interoperable information and communication technologies. The interaction between physical things is implemented through built-in elements with the appropriate software, which is responsible for collecting and analyzing data, device interaction, communication with the cloud. The method of sending and receiving information, as well as communication with external devices, is determined by the choice of communication solution. The system recognizes what is happening around and informs the user about it through IoT - a platform on which the received data is analyzed, then they are converted into a form convenient for the end user.

"Smart light" takes into account the presence of people in the location, the time of day, the presence or absence of natural light. These parameters can be controlled remotely - turn on or change the brightness, simulate the presence of people in the house. Types of lighting devices can be very different - incandescent, gas discharge, diode. The control elements responsible for adjusting the light are microcontrollers, sensors, motion sensors, dimmers. The controller for a smart home with gsm is essentially an intelligent system that coordinates the work of connected devices, makes logical decisions, the algorithms of which are embedded in the corresponding program.

Analysis of recent research and publications.

People tend to follow certain patterns in their daily lifestyle. In the context of a smart home, the daily actions of the user generate patterns that play an important role in predicting future events in the smart home. The purpose of the smart home environment is to

help the user in daily life; thus, a smart home must find repetitive patterns in user activity and predict user behavior in order to receive assistance [1].

User activity monitoring is used to observe and record human actions in order to achieve the "comfort and efficiency goal" that a smart home can offer. User behavior refers to the range of actions, actions and responses made by the user. Therefore, a smart home must be able to learn and apply the knowledge gained in order to adapt the home to the user's behavior. Since the user generates a pattern, abnormal user behavior can be detected by building a normal user behavior pattern. Typically, smart home sensors and cameras are used to track or identify user activity and perform analysis of human behavior. User behavior can be used to predict and determine future user actions. Thus, the activity recognition method implemented by a smart home should be as accurate as possible to control the system [2].

Context awareness is an important step in the concept of user behavior and behavior. The context-awareness system in this concept is one that "adapts depending on the place of use, the accumulation of nearby people and objects, and changes in these objects over time." Day A.K. [3] defined context as "any information that can be used to characterize a subject's position", and a context-aware system is one that "uses context to provide 8 relevant information and/or services to the user, where appropriateness depends on the user's task".

Smart home devices can be controlled both through voice assistants built into smart speakers, and through websites and applications connected to the smart home control center. Unfortunately, in the "many home" market there is no single standard for pairing devices, and therefore, large technology manufacturers are forming their own ecosystem with a list of devices that can be connected. To interface many devices without being tied to the manufacturer of the smart home control center, there are open-source DIY solutions based on the Raspberry single-board computer, discussed in the following sections. However, it should be understood

that this type of solution requires technical knowledge from the user and time costs.

Statement of the research problem

The wireless method uses a local network and a data transfer protocol such as HTTP, WebSocket or its own. In this way, the smart lighting connects to the home local network and tells the user its address. In turn, the user needs to take any device with software for interacting with smart lighting and actually control it. Typically, these control devices are smartphones, computers, and laptops. To implement such interaction, you need to create a set of commands, implement software that uses the specified protocol and processes user requests to send these commands to the microcontroller. Commands are sent in most cases using the HTTP protocol.

HTTP is a protocol that allows you to receive various resources, such as HTML documents. The HTTP protocol is the basis of data exchange on the Internet, and is also a protocol of client-server interaction, which means that requests to the server are initiated by the recipient, usually a web browser [11]. The result obtained has its own body. The request or response body has a specific content type, such as text, HTML, JSON. Obviously, a set of commands must contain a certain set of data. In order for the data to be transferred and analyzed conveniently, you need to use one of the known standards, as this eliminates the need to spend time on the implementation of your own type of content. One of the most common options is JSON.

JSON (JavaScript Object Notation) is a text format for exchanging data between computers. JSON is text-based, human-readable. A format allows you to describe objects and other data structures. This format is mainly used to transmit structured information over a network (thanks to a process called serialization).

JSON is built on two structures:

1. A set of name/value pairs. In various languages, this is implemented as an object, record, structure, dictionary, hash table, keyed list, or associative array;
2. Ordered list of values. In many languages, this is implemented as an array, vector, list, or sequence.

These are universal data structures. In theory, all modern programming languages support them in one form or another. Since JSON is used to exchange data between different programming languages, it makes sense to use this format to implement the communication mechanism between the control device and the smart lighting itself [4,5,6].

Android is the most popular OS in the world for smartphones since 2011 and for tablets since 2013. As of May 2021, it has more than two billion monthly active users, the largest installed base of any operating system, and as of January 2022, Google's Play Store features over 2.9 million apps.

The Android source code is released by Google under an open source license, and its open nature encourages a large community of developers and enthusiasts to use the open source code as the basis for community-driven projects that provide updates for older devices, add new features for users, or port Android on devices originally shipped with other

operating systems. These releases often provide new features and device updates faster than through official manufacturer channels, with a comparable level of quality [7, 8].

So, the Android OS is a platform with an open architecture, it has convenient tools for creating software. The extremely large number of users of this OS means that almost every smartphone user uses Android. Openness, popularity, support for HTTP, JSON, HTML, JavaScript standards are sufficient arguments for recognizing this platform as appropriate for the development of software for controlling smart lighting devices.

A microcontroller for "smart lighting" must have certain functionality - work with WI-FI, interact with HTTP protocols and be powerful. Therefore, among a large number of microcontrollers, those created on the basis of ESP were chosen. Today, there are two options that satisfy the conditions - ESP8266 and ESP32.

ESP8266 is a popular platform for "Internet of Things" that need to transmit or receive data to the Internet using WI-FI technologies. In addition to WI-FI, the microcontroller differs from others in the absence of flash memory in a single-chip system, and user programs are executed in flash memory via a serial peripheral interface. In turn, the microcontroller attracts with a pleasant price and low power consumption. One of the popular boards based on ESP8266 is NodeMCU V3.

The ESP32 is an inexpensive microcontroller that replaced the ESP8266. Like its predecessor ESP8266, it has low power consumption. It represents a system on a crystal with integrated WI-FI and Bluetooth controllers and antennas. It exists in two types, single-core with a clock frequency of 160 MHz and dual-core, respectively, with a frequency of 240 MHz. Based on ESP32, the DOIT ESP32 DEVKIT V1 board is most often used.

The NodeMCU V3 board was chosen as the board for installing the microcontroller.

The next step is to choose the programming language and development environment - Arduino IDE. The structure of the software (software) of the microcontroller was divided into separate modules:

- ✓ Interaction with a WI-FI access point;
- ✓ Server part;
- ✓ Initialization of controller devices;
- ✓ Description of effects;
- ✓ Effects manager;
- ✓ Utilities.

First of all, the controller must be connected to the WI-FI network. There are a number of means for this. The Arduino IDE suggests using the ESP8266WiFi library. It provides an easy way to interact with access points. Accordingly, we get the ability to connect to the network, obtain an IP address from the code, and also have the ability to monitor the connection status. Thus, when power is supplied to the controller, it starts up, which in turn initiates the connection to the WI-FI access point. Next, it is necessary to find out the status of the connection and, if successful, to display the IP address on the matrix so that the client part can find out which address to send requests to.

The next module is the server part. This refers to the code that implements API access via a defined protocol with a defined content type and content. An available tool for implementing these principles is the ESP8266Web Server library. One of the main possibilities is to create a REST API with request handlers and sending responses using the HTTP protocol. In this way, the developer gets rid of the need to implement his own protocols. So, to ensure client-server interaction, you need to specify which types and how the microcontroller will respond. Regarding the type and format of the content, it should be noted that the body in requests and (or) responses is necessary only in the case of execution of commands for obtaining the current state of the configuration. When using the API to change the state of the controller, you only need to specify the command parameters.

The microcontroller initialization module is a block of code in which variables are described, objects are created, and methods are executed to ensure the start of work. In this case, you need to describe the constants that characterize the numbers of matrix effects, the size of the matrix itself, the number of LEDs, as well as create objects for interaction with WI-FI, HTTP, JSON, FastLED and perform methods for starting the corresponding processes.

The Description of Light Effects module will be organized as a tab in the Arduino IDE. Structurally, it will be a sequence of descriptions of each effect together with its auxiliary methods.

The effects manager in this case is a module with the task of checking the current effect and calling the appropriate method for its execution. The manager will be used when starting the controller, as well as when processing state change commands received from the client part of the smart lighting device.

The utility module is a set of auxiliary classes and methods to simplify the interaction of the use of libraries, as well as to optimize the code as a whole. In this way, large logical chains of code will be placed in separate methods, so they will be available for reuse from anywhere in the controller software, and as a result, the total amount of code will be significantly reduced.

Basic material and results

At the first stage of software implementation, it is necessary to create a reference architecture to ensure a scaling strategy. For this, the most adequate step is to use a modular structure [9]. This architecture provides opportunities for the independence of individual parts of the software in the event of the need for modifications or changes. For Android OS, the most common architectural design pattern is MVVM. According to the conditions of its implementation, the following directories must be created: model, view, viewmodel, repository, utils. Each of them contains a set of classes to perform the corresponding role:

- Model – entities, classes that represent data sets and an interface for interacting with them according to the principles of encapsulation;
- Repository – classes for interaction with the server, database or other sources of information. These classes contain the code for interaction methods with the

back-end part and the organization of their work in a multi-threaded environment;

- View – graphic shell screen handlers;
- ViewModel – classes describing the logic of interaction with data for view components;
- Utils – utility classes to facilitate work and also to avoid redundant code duplication.

Strict rules are imposed on the communication between these classes. First of all, to avoid conflicts and memory leaks, you should avoid references to visual elements inside any module other than view. This need exists as each visual element, as well as the viewModel, has its own life cycles, the work of which must be managed by the system without interference.

For the functionality of the command system in the lighting device control software, all operations with the web server must be performed in a separate thread. For this purpose, we will create the corresponding classes and interfaces. Since a separate stream does not have the ability to directly transfer data to the main one, it is necessary to create a "Callback" interface with a callback method of the same name. It is impossible to create an interface object due to the limitation of the JAVA programming language standard. Creating a large number of classes that implement it is not a good solution, because such an approach introduces the permanence of the method's behavior. Therefore, anonymous classes are used for flexibility in the use of the callback mechanism. Thus, when the use of this interface occurs, the developer must create an anonymous class with the implementation of the callback method. This implementation can then be passed to a separate thread. Accordingly, at a certain point in time, this thread will execute a call to this method.

The next stage of development of the mechanism of interaction with the web server is the development of the repository class with the appropriate methods. To do this, we will create the necessary types of requests and response handlers. Based on the microcontroller software, we can perform the following requests:

- Downloading data of the current state of the controller;
- Setting up new settings.

First of all, we will form the general structure of the methods of executing requests. We will distinguish three main components, such as operations on input data, request formation, sending for execution in a separate stream. The first component consists in performing simple type conversion manipulations and (or) extracting values from entity objects. The second component is the creation of a request by creating the "Request" object of the OkHTTP library. Then the processed data is placed in the headers and (or) body of the request, the content type, HTTP method and other settings are specified. The third component is the execution of a request to the web server. To achieve this goal, the implementation of multithreading processors is required. For this, it is worth creating a class that implements the "Runnable" interface, to the input of which Request and Callback objects will be submitted. The inner class will create a new thread, implementing the request execution and callback, after

receiving the response from the server. Next, we implement certain types of operations.

The request to receive the state of the controller as input data involves only the index of the mode whose data needs to be received. According to the web server specification, the request does not need additional headers and bodies. It is enough to form the correct URL. Processing the response consists of converting the JSON into a "Setting" entity object and then saving it to the appropriate storage.

A configuration change request requires 4 fields as input, namely mode index, brightness, speed, and special property value. All interaction, except for processing the response, is similar to the algorithm described above. Since there is nothing in the response except the status code, its processing is reduced to the output of a corresponding message in the debugging log.

Conclusions

An analysis of the technical characteristics of existing analogs presented on the "smart lighting" market was performed. A number of shortcomings were identified - low connection stability, weak interaction speed and high price for not always justified quality and functionality. As a result of the review of the subject area, the main electronic components necessary for the

implementation of a smart lighting device were found - a microcontroller, a matrix, a power supply unit.

Considered ways for software implementation, which indicated the availability of appropriate libraries for connecting to a WI-FI network and deploying a web server within the microcontroller software using the ESP8266WebServer library. The development of the logic of the microcontroller, namely the modes of operation, sequence of actions, modularity, was carried out. As a result, the overall structure of the software was divided into three separate modules: initial initialization, interaction with the network, display of effects on the LED matrix.

The "smart lighting" software is implemented in the form of a mobile application. The mobile application includes an architectural design template and basic visual components. To communicate with the web server, mechanisms for building requests and processing responses from the OkHTTP library are used. The implemented structure of the "smart lighting" device has the possibility of further scaling without additional costs, which in turn increases its competitiveness on the commercial market, can be used as a prototype of a hardware and software complex for the implementation of a budget version of "smart lighting".

REFERENCES

1. Adim A.O. Big Sensed Data Meets Deep Learning for Smarter Health Care in Smart Cities / A. O. Adim, B. Kantarci // Journal of Sensor and Actuator Networks. – 2017. – Volume 6. – Issue 4. - PP. 1-22. – URL: <https://doaj.org/article/1067f3eaf1d94db38796b306f62c692f>.
2. Fowler R.J. Electricity; Principles and Applications / R.J. Fowler - New-York: Delmar Cengage Learning. 2017. 247 p.
3. Haritaoglu I., Ghost: A human body part labeling system using silhouettes / Haritaoglu I., Harwood D., Davis L. S. // Proc. 14th International Conference on Pattern Recognition, Australia. 2015. pp. 16-20.
4. Format JSON [Electronic resource].URL: <https://timeweb.com/ru/community/articles/format-json> // (date of access: 15.02.2023).
5. JavaScript Object Notation.URL: <https://www.php.net/manual/ru/json.installation.php> // (date of access: 15.02.2023).
6. Autonomous JSON Database [Electronic resource].URL: <https://www.oracle.com/autonomous-database/autonomous-json-database/get-started> // (date of access: 15.02.2023.)
7. Android 13 Highlights.URL: <https://www.android.com/android-13/#a13-beyond-the-phone> // (date of access: 15.02.2023.)
8. Economic Benefits of the Global Positioning System (GPS) [Electronic resource].URL: https://economic-definition.com/Technology/Android_Android_eto.html // (date of access: 15.02.2023.)
9. Froiz-Míguez, I.; Fernández-Caramés, T.M.; Fraga-Lamas, P.; Castedo, L. Design, Implementation and Practical Evaluation of an IoT Home Automation System for Fog Computing Applications Based on MQTT and ZigBee-WiFi Sensor Nodes. Sensors 2018. 18, 2660. doi.org/10.3390/s18082660

Received (Надійшла) 30.03.2023

Accepted for publication (Прийнята до друку) 10.05.2023

Реалізація функції освітлення у концепції «розумного будинку»

О. В. Скакаліна, А. М. Капитон

Анотація. Принцип «Системи інтелектуального управління будинком» передбачає абсолютно новий підхід до організації життєзабезпечення будинку, в якому за рахунок сукупності програмно-апаратних засобів значно зростає ефективність функціонування і надійність управління всіх систем і виконавчих пристроїв будівлі. Під розумним будинком слід розуміти систему, яка повинна вміти розпізнавати конкретні ситуації, що відбуваються в будівлі, і відповідним чином на них реагувати: одна з систем може управляти поведінкою інших за задалегідь виробленим алгоритмом. Основною особливістю такої будівлі є об'єднання окремих підсистем в єдиний керований комплекс. Важливою особливістю і властивістю розумного будинку є спосіб організації життєвого простору - це найбільш прогресивна концепція взаємодії людини з житловим простором, коли людина однією командою задає бажану обстановку, а вже автоматика відповідно до зовнішніх і внутрішніх умов задає і відстежує режими роботи всіх інженерних систем і електроприладів. У цьому випадку виключається необхідність користуватися кількома пультами при перегляді ТВ, багатьма вимикачами при управлінні освітленням, окремими блоками при управлінні вентиляційними і опалювальними системами, системами відеоспостереження та сигналізації.

Ключові слова: розумний дім, розумне освітлення, віддалене керування, програмне забезпечення, ESP8266, мікроконтролер, WI-FI, HTTP.

Г. В. Худов¹, І. О. Романенко², П. Є. Минко³, Ю. С. Соломоненко¹, В. О. Іванюк¹

¹ Харківський національний університет Повітряних Сил імені І. Кожедуба, Харків, Україна

² Інститут проблем математичних машин і систем Національної академії наук України, Україна

³ Харківський національний університет радіоелектроніки, Харків, Україна

УДОСКОНАЛЕНИЙ МЕТОД ПРОГНОЗУВАННЯ ТРАЄКТОРІЇ ПОЛЬОТУ КРИЛАТИХ РАКЕТ НА ОСНОВІ МУРАШИНОГО АЛГОРИТМУ

Анотація. Предметом вивчення в статті є метод прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму. **Метою** є отримання адекватних результатів при прогнозуванні траєкторії руху крилатих ракет при здійсненні планування застосування сил та засобів протиповітряної оборони та вибору їх раціональної побудови. **Завдання:** аналіз відомих методів оцінки траєкторії польоту крилатої ракети; моделювання траєкторії польоту крилатих ракет як пошуку послідовності дій крилатої ракети, яка може переміщати з одного місця в інше, уникаючи будь-яких перешкод на траєкторії польоту; удосконалення методу прогнозування траєкторії польоту крилатих ракет. Використовуваними **методами** є: методи системного аналізу, теорії ймовірності, методи ройового інтелекту, математичної статистики. Отримані такі **результати**. Удосконалено метод прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму, в якому, на відміну від відомих, розрахунок траєкторії польоту крилатої ракети передбачає розрахунок цільової функції, визначення сукупності ділянок руху агентів, концентрації феромону на маршрутах руху агентів, переміщення агентів по визначених ділянках руху. Проведена перевірка працездатності удосконаленого методу прогнозування дій повітряного противника на основі мурашиного алгоритму. **Висновки.** Встановлено, що в усіх випадках метод прогнозування дій повітряного противника на основі мурашиного алгоритму забезпечує найменшу довжину маршруту від вихідної точки маршруту до об'єктів ураження. Напрямою подальших досліджень є побудова системи радіолокаційної розвідки за результатами моделювання варіантів дій повітряного противника.

Ключові слова: крилата ракета, повітряний противник, траєкторія польоту, об'єкт ураження, мурашиний алгоритм, радіолокаційна розвідка.

Вступ

Постановка проблеми у загальному вигляді. Відомо [1], що застосування ракетного озброєння (крилатих, балістичних, аеробалістичних ракет) несе небезпеку для цивільного населення. Сучасні ракети постійно маневрують та змінюють напрямок руху [2]. Це, безумовно, ускладнює завдання виявлення, визначення координат крилатої ракети та її знищення. Прогнозування замислу дій повітряного противника є важливим завданням органів військового управління, що вирішується в ході оцінки обстановки та визначення вихідних даних для прийняття рішення та планування протиповітряної оборони (ППО). Правильно відповісти на питання, якими маршрутами полетять крилаті ракети, у складі яких ешелонів та груп, де буде створюватися смуга прориву протиповітряної оборони – значить правильно оцінити обстановку та вірно обрати вихідні дані для прийняття рішення та вибору раціональної структури системи ППО та її підсистем.

Відомі методи прогнозування траєкторії польоту крилатих ракет пов'язані з проведенням ручних розрахунків, мають певну долю суб'єктивізму тощо. Відомі методи прогнозування траєкторії польоту крилатих ракет на основі автоматизованих розрахунків не задовольняють сучасним вимогам та є застарілими. Отже, виникає невідповідність між можливостями існуючих методів прогнозування траєкторії польоту крилатих ракет та вимогами до прогнозування такої траєкторії. Для усунення вказаної невідповідності у роботі запропоновано метод прогнозування траєкторії руху крилатої ракети на основі мурашиного алгоритму.

Мета статті – розробка методу прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму.

Аналіз останніх досліджень і публікацій. Методики оцінки дій повітряного противника орієнтовані, в основному, на проведення ручних розрахунків. В загальному випадку крилаті ракети до досягнення об'єкту удару можуть використовувати певну множини можливих маршрутів і профілів польоту [3]. Можливо стверджувати, що з цієї множини противник буде використовувати такі маршрути, які забезпечать високу ефективність повітряних ударів [4].

При прогнозуванні траєкторії польоту крилатої ракети в найпростішому випадку у якості найкращого маршруту обирається найкоротший [4]. Для цього визначаються точки пуску крилатих ракет, елементи системи об'єктів прикриття ППО з вказанням їх зон бойових дій і бойового застосування [5].

Далі визначається перелік об'єктів, що досягаються, складається список об'єктів, які ймовірно можуть бути уражені та оцінюється можливість ураження даних об'єктів [6].

Наступним кроком відомих методів оцінки траєкторії польоту крилатої ракети є визначення глибини та ширини удару. В результаті обирається найкоротший маршрут з точки пуску крилатої ракети до об'єкту удару [7-9].

Відомі методики забезпечують оцінку параметрів удару на початок операції (бойових дій), не є адаптивними та не можуть бути використані для оцінки параметрів наступного удару крилатих ракет. Основними недоліками відомих методик є [10-13]:

– постульована в апіорі та незмінна побудова ракетного удару;

– заданий задалегідь та незмінний в часі процентний розподіл крилатих ракет в ударі.

Зазначені недоліки унеможливають отримання адекватних результатів при прогнозуванні траєк-

торії руху крилатих ракет при здійсненні планування застосування сил та засобів ППО та вибору їх раціональної побудови.

Отже, існуючі методи прогнозування траєкторії польоту крилатих ракет потребують удосконалення.

Основна частина

Планування траєкторії польоту крилатих ракет полягає в пошуку послідовності дій крилатої ракети, яка може переміщати з одного місця в інше, уникаючи будь-яких перешкод на траєкторії польоту. Це може бути інтерпретовано як завдання пошуку оптимальної траєкторії крилатої ракети в статичному середовищі. Це, в свою чергу, передбачає досягнення мети руху при оптимізації визначених витрат.

Точки траєкторії польоту крилатої ракети позначимо як $W_i(\mathbf{X}) \in \mathbf{R}^n$, для $i = 1, 2, \dots, N$, де N – загальна кількість точок маршруту; \mathbf{X} – n -мірний вектор координат руху; \mathbf{R}^n – простір руху розміру n . Точки маршруту крилатої ракети поділяються на:

- вихідну точку маршруту (ВТМ);
- поворотні точки маршруту (ПТМ);
- кінцеву точку маршруту (КТМ).

ПТМ будемо позначати як $C_i(\mathbf{X}) \in \mathbf{R}^n$, для $i = 1, 2, \dots, M$, де M – загальна кількість ПТМ.

Будемо враховувати мінімальну безпечну відстань між крилатими ракетами та розміри ракет. Мінімальну безпечну відстань позначимо як $R_{safe}(\mathbf{X}) \in \mathbf{R}^n$. Також будемо враховувати перешкоди, що визначають зміну траєкторію польоту крилатої ракети.

Перешкоди будемо позначати $Obst_i(\mathbf{X}) \in \mathbf{R}^n$, для $i = 1, 2, \dots, Obs$, де Obs – загальна кількість перешкод. Будемо враховувати обмеження та допущення, що при плануванні траєкторії польоту крилатої ракети кількість ПТМ дорівнює кількості перешкод ($M = Obs$).

При визначенні траєкторії польоту крилатої ракети будемо враховувати конфігурацію простору руху. Під конфігурацією простору розуміється місцевість, погодні умови та перешкоди. Фізичні особливості місцевості будемо визначати за допомогою цифрової карти місцевості з урахуванням цифрової моделі висот (матриці точок висот). Будемо враховувати обмеження щодо можливого напрямку руху крилатої ракети. Будемо вважати, що крилата ракета може рухатися тільки вперед і не можуть робити різкі повороти або підйоми (рис. 1).

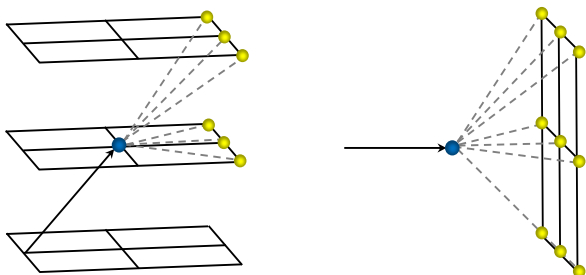


Рис. 1. Можливі напрямки руху крилатої ракети

Залежно від призначення ракет можна розглядати різні види цільової функції, відносно якої потрібно вирішити оптимізаційне завдання. Крилаті ракети, як правило, мають обмежену дальність та обмежений запас пального.

Цільова функція повинна бути обрана такою, що мінімізує відстань між точками траєкторії і, відповідно, витрати пального. Отже, цільова функція при забезпеченні мінімальної відстані між точками W_i траєкторії і мінімальних витрат пального може бути записана виразом (1):

$$\varphi(\mathbf{X}) = \sum_{i=1}^{N-1} \|W_{i+1}(\mathbf{X}) - W_i(\mathbf{X})\| \rightarrow \min. \quad (1)$$

Завдання планування траєкторії польоту крилатої ракети може бути зведено до пошуку траєкторії крилатої ракети в дискретному середовищі. Таке завдання визначається наступними елементами:

- простір станів \mathbf{X} , який є скінченною або зліченою кінцевою множиною станів;
- набір доступних дій, $u \in U(x)$ для кожного стану $x \in \mathbf{X}$;
- функція переходу стану f , яка створює стан $f(x, u) \in \mathbf{X}$ для кожного $x \in \mathbf{X}$ і $u \in U(x)$;
- початковий стан $x_1 \in \mathbf{X}$;
- стан цілі $x_G \in \mathbf{X}$.

Загальний алгоритм пошуку оптимальної траєкторії працює шляхом систематичного пошуку в графі, застосовуючи функцію переходу та вибираючи стани, які мінімізують цільову функцію, одночасно відстежуючи відвідувані вузли, щоб не було зайвих розрахунків (дій).

З урахуванням вищезазначеного оптимізаційне завдання визначення траєкторії польоту (визначення вектору координат \mathbf{X}) крилатої ракети в загальному вигляді може бути сформульовано у вигляді:

$$\begin{aligned} \varphi(\mathbf{X}) &= \sum_{i=1}^{N-1} \|W_{i+1}(\mathbf{X}) - W_i(\mathbf{X})\| \rightarrow \min; \\ W_i(\mathbf{X}) &\in \mathbf{R}^n, i = 1, 2, \dots, N; \\ R_{safe}(\mathbf{X}) &\in \mathbf{R}^n; \\ C_j(\mathbf{X}) &\in \mathbf{R}^n, j = 1, 2, \dots, M; \\ Obst_k(\mathbf{X}) &\in \mathbf{R}^n, k = 1, 2, \dots, Obs; \\ \forall \mathbf{X} &\wedge (\mathbf{X} \notin Obst_i(\mathbf{X})); \\ \forall \mathbf{X} &\wedge (\mathbf{X} > R_{safe}(\mathbf{X})); \\ M &= Obs. \end{aligned} \quad (2)$$

Для вирішення завдання (2) доцільно використання мурашиного алгоритму, перевагами якого є спроможність вирішення оптимізаційних завдань без знання простору пошуку [14-17]. Мурашиний алгоритм реалізує прямий випадковий пошук можливих рішень оптимізаційного завдання, що є оптимальними або близькими до оптимальних, доки не буде виконана деяка умова або досягнута задана кількість ітерацій.

Узагальнений опис мурашиного алгоритму можна представити виразом:

$$AS = \{S, M, A, P, I, O\}, \quad (3)$$

де S – множина агентів; M – об'єкт для обміну досвідом між агентами (мурахами); A – правила роботи мурашиного алгоритму (створення, поведінка, модифікація агентів); P – параметри (евристичні коефіцієнти), що використовуються в правилах A ; $I = \{I_1 \cdot I_{33}\}$, I_1 – вхід системи, на який подається цільова функція, обмеження, I_{33} – вхід для зворотного зв'язку; $O = \{O_1 \cdot O_{33}\}$, O_1 – вихід (найкраще знайдене рішення задачі), O_{33} – вихід для зворотного зв'язку.

Розглянемо особливості застосування мурашиного алгоритму для прогнозування маршруту польоту крилатої ракети.

В найпростішому випадку маршрут польоту крилатої ракети можна представити як сукупність наступних ділянок: ділянки зльоту з ВТМ та набору висоти, горизонтальних ділянок, на яких відбувається крейсерський політ, подолання зони ППО, вихід на об'єкт удару, ділянка зниження до рубежу виконання завдання або КТМ. Горизонтальні ділянки проходять через ПТМ, в яких відбувається зміна курсу, а в загальному випадку, і висоти польоту. У подальшому вважаємо, що положення ВТМ, КТМ та ПТМ повністю визначає маршрут польоту крилатої ракети.

Політ по кожній з ділянок маршруту, а також здійснення маневру в вибраних ПТМ, має певні небезпеки для крилатої ракети та вимагає певних витрат ресурсів, що призводить до наявності переваги одного маршруту польоту перед іншим. Оскільки варіантів просторового розташування ПТМ може бути дуже багато, кількість можливих маршрутів польоту ЗПН буде надзвичайно великою, що ускладнює вибір маршруту ЗПН найпростішим методом – перебору.

Послідовність етапів методу прогнозування дій повітряного противника на основі мурашиного алгоритму може бути представлена наступним чином.

1. Ініціалізація початкових положень агентів (крилаті ракети) на першій ітерації ($j=1$). $X_{i1}(x_{i1}, y_{i1})$ – вектор положень агентів на першій ітерації, $i=1, 2, \dots, S$; S – загальна кількість агентів. Загальна кількість агентів S дорівнює кількості крилатих ракет.

2. Розрахунок цільової функції $\phi_j(X)$ на j -й ітерації.

У якості цільової функції на j -й ітерації будемо визначати наступну функцію:

$$\phi_j(X) = \sum_{k=1}^S \sum_{i=1}^{N_k} (P_i^k(j) \cdot D_i^k(j)), \quad (4)$$

де k – поточний номер агента; N_k – загальна кількість точок маршруту для k -го агента; $P_i^k(j)$ – ймовірність переходу k -го агента в i -ту ПТМ на j -й ітерації:

$$P_i^k(j) = (F_i^k(j))^\alpha (L_i^k(j))^\beta / \sum_{m=1}^M (F_m^\alpha(j) \cdot L_m^\beta(j)), \quad (5)$$

де $\alpha \geq 0$ і $\beta \geq 0$ – параметри, що задають вагу феромона і “жадібність” методу відповідно; M – кількість можливих ПТМ; $L_i^k(j)$ – привабливість ділянки маршруту для k -го агента в i -й точці маршруту на j -й ітерації; $F_i^k(j)$ – концентрація феромону k -го агента в i -й точці маршруту на j -й ітерації; функція $D_i^k(j)$ визначає довжину ділянки маршруту для k -го агента в i -й точці на j -й ітерації та для тривимірного простору визначається за виразом:

$$D_i^k(j) = |\Delta x_i^k(j)| + |\Delta y_i^k(j)| + |\Delta z_i^k(j)|, \quad (6)$$

де $|\Delta x_i^k(j)|$, $|\Delta y_i^k(j)|$, $|\Delta z_i^k(j)|$ – елементарні переміщення k -го агента в i -й точці на j -й ітерації по осям x , y та z відповідно.

3. Переміщення агентів. В простішому мурашиному алгоритмі в кожній ітерації ітераційного процесу k агентами здійснюється пошук рішення та оновлення феромонів на знайденому маршруті. Кожний k -й агент починає шлях з ВТМ, послідовно проходить вибрані методом ПТМ і завершує шлях в одній з КТМ. Переміщення агентів проводиться по критерію мінімуму цільової функції (4), яка, з урахуванням дев'ятизв'язності переміщення агентів (рис. 1):

$$|\Delta x_i^k(j)| + |\Delta y_i^k(j)| + |\Delta z_i^k(j)| = 1, \quad (7)$$

приймає вигляд:

$$\phi_j(X) = \sum_{k=1}^S \sum_{i=1}^{N_k} (P_i^k(j)) = \sum_{k=1}^S \sum_{i=1}^{N_k} \left((F_i^k(j))^\alpha (L_i^k(j))^\beta / \sum_{m=1}^M (F_m^\alpha(j) \cdot L_m^\beta(j)) \right), \quad (8)$$

а задача (2) може бути сформульовано у вигляді:

$$\phi_j(X) = \sum_{k=1}^S \sum_{i=1}^{N_k} \left(\frac{(F_i^k(j))^\alpha (L_i^k(j))^\beta}{\sum_{m=1}^M (F_m^\alpha(j) \cdot L_m^\beta(j))} \right) \rightarrow \min;$$

$$W_i(X) \in \mathfrak{R}^n, i = 1, 2, \dots, N;$$

$$R_{safe}(X) \in \mathfrak{R}^n; \quad (9)$$

$$C_j(X) \in \mathfrak{R}^n, j = 1, 2, \dots, M;$$

$$Obst_k(X) \in \mathfrak{R}^n, k = 1, 2, \dots, Obs;$$

$$\forall X \wedge (X \notin Obst_i(X));$$

$$\forall X \wedge (X > R_{safe}(X));$$

$$M = Obs.$$

Привабливість ділянки маршруту L_i в (9) обернено пропорційна затратам на подолання ділянки C_i :

$$L_i = 1/C_i, \quad (10)$$

а затрати на подолання ділянки в мурашиному алгоритмі залежать тільки від його довжини, тобто:

$$C_i = D_i, \quad (11)$$

де D_i – довжина i -ї ділянки маршруту.

В більш складних випадках значення L_i може бути розраховане в залежності від наявності зон небезпеси (вираз (12)):

$$L_i = \begin{cases} 1 & \text{якщо } i\text{-й маршрут пролягає} \\ D_i & \text{за межами "зон небезпеси";} \\ 0, & \text{в інших випадках.} \end{cases} \quad (12)$$

На початку ітераційного процесу кількість феромону на ділянках маршруту приймається однаковою і рівною деякому невеликому числу F_0 . Після кожної ітерації концентрація феромонів на вибраних агентами ділянках оновлюється за правилом (13):

$$F_i(t+1) = (1-\rho)F_i(t) + \sum_{k=1}^m \Delta F_i^k, \quad (13)$$

де $\rho \in [0,1]$ – швидкість випаровування феромону; ΔF_i^k – концентрація феромону на i -й ділянці маршруту, що створюється проходженням k -го агента.

У тому випадку, якщо у поточній ітерації по ділянці не пройшов жодний з агентів, правило (13) перетворюється на (вираз (14)):

$$F_i(t+1) = (1-\rho)F_i(t), \quad (14)$$

тобто оновлення феромону полягає в його випаровуванні із швидкістю ρ .

4. Перевірка виконання умови зупинки. Якщо умова виконана, то отримується траєкторія руху крилатої ракети (X). В іншому випадку – здійснюється перехід до другого пункту. Параметри P (5) методу визначаються як $P(\alpha, \beta, \rho, F_0)$.

Таким чином, удосконалено метод прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму, в якому, на відміну від відомих, розрахунок траєкторії польоту крилатої ракети передбачає розрахунок цільової функції, визначення сукупності ділянок руху агентів, концентрації феромону на маршрутах руху агентів, переміщення агентів по визначених ділянках руху.

Проведемо перевірку працездатності методу прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму на контрольному прикладі. Для наочності представлення результатів застосована прямокутна система координат, у якій ВТМ, КТМ і ПТМ знаходяться в одній горизонтальній площині.

Вихідні дані:

- кількість ВТМ $N_{ВТМ}=1$;
- кількість КТМ $N_{КТМ}=1$;
- кількість можливих ПТМ1 $N_{ПТМ1}=20$;
- кількість можливих ПТМ2 $N_{ПТМ2}=20$;
- “жадібність” методу $\beta=1$;

– вага феромонів $\alpha=2$;

– швидкість випаровування феромону $\rho=10^{-3}$;

– кількість ітерацій методу $N=400$;

– кількість агентів в ітерації $m=10$;

– початкова кількість феромону $F_0=10^{-2}$.

В результаті $N=400$ ітерацій маршрут руху агента наведений на рис. 2 жирною суцільною лінією. Переривчастою лінією показано оптимальний маршрут руху агентів.

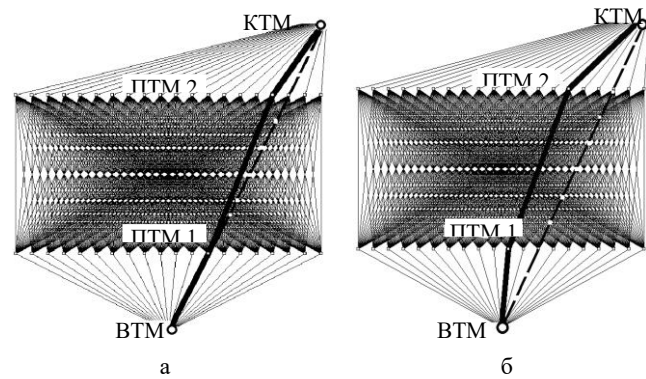


Рис. 2. Результати перевірки працездатності методу прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму:
а) маршрут руху агентів відрізняється від оптимального маршруту;
б) маршрут руху агентів істотно відрізняється від оптимального маршруту

З аналізу рис. 2, а видно, що отриманий маршрут руху агентів незначно відрізняється від оптимального маршруту. Проте, як показано на рис. 2, б, можливі і реалізації методу з відверто невдалими результатами, що вказує на необхідність проведення подальшого дослідження.

Висновки

Удосконалено метод прогнозування траєкторії польоту крилатих ракет на основі мурашиного алгоритму, в якому, на відміну від відомих, розрахунок траєкторії польоту крилатої ракети передбачає розрахунок цільової функції, визначення сукупності ділянок руху агентів, концентрації феромону на маршрутах руху агентів, переміщення агентів по визначених ділянках руху. Проведена перевірка працездатності удосконаленого методу прогнозування дій повітряного противника на основі мурашиного алгоритму.

Встановлено, що в усіх випадках метод прогнозування дій повітряного противника на основі мурашиного алгоритму забезпечує найменшу довжину маршруту від ВТМ до тих об'єктів ураження.

Напрямок подальших досліджень: побудова системи радіолокаційної розвідки за результатами моделювання варіантів дій повітряного противника.

СПИСОК ЛІТЕРАТУРИ

1. Amble J. Intelligence and the Future Battlefield, with Lt. Gen. Scott Berrier / J. Amble // Modern War Institute. – October 25, 2019. URL: <https://mwi.usma.edu/mwi-podcast-intelligence-future-battlefield-lt-gen-scott-berrier/>.
2. Air & Space Operations Review : A Journal of Strategic Airpower & Spacepower. – 2022. – Vol. 1. – No. 4. URL: <https://www.airuniversity.af.edu/ASOR/>.

3. Синтез адаптивних структур системи зенітного ракетно-артилерійського прикриття об'єктів і військ та оцінка її ефективності (теорія, практика, тенденції розвитку) : монографія / Торопчин А. Я. та ін. Харків : ХУ ПС, 2006. 349 с.
4. Городнов В. П., Ермошин М. О., Шулежко В. В. Методика оцінки статистичних параметрів удару повітряного противника при відновленні системи зенітного ракетно-артилерійського прикриття. Наука і техніка Повітряних Сил Збройних Сил України. Харків: ХУ ПС, 2013. № 1(10). С. 57–60.
5. Пуховий О. В. Удосконалена методика визначення варіанту бойового застосування угруповання радіотехнічних військ. Системи озброєння і військова техніка. 2012. № 3(31). С.66–70.
6. Моделювання бойових дій військ (сил) протиповітряної оборони та інформаційне забезпечення процесів управління ними (теорія, практика, історія розвитку) : монографія / Городнов В. П та ін. Харків : ХВУ, 2004. 410 с.
7. Ярош С. П., Рогуля О. В. Аналіз тактики бойового застосування крилатих ракет при нанесенні ударів по важливих державних об'єктах та угрупованнях військ. Збірник наукових праць Харківського національного університету Повітряних Сил. 2019. № 3(61). С. 35–44. <https://doi.org/10.30748/zhups.2019.61.06>.
8. Жарик О. М., Тристан А. В. Військова операція Азербайджану в Нагірному Карабасі в контексті розвитку Повітряних Сил Збройних Сил України. Наука і техніка Повітряних Сил Збройних Сил України. 2021. № 2(43). С. 19–24. <https://doi.org/10.30748/nitps.2021.43.02>.
9. Ярошенко Я., Герасименко В., Блискун О., Базіло С., Ікаєв Д. Досвід застосування безпілотної авіації у вірмено-азербайджанському конфлікті восени 2020 року. Уроки для України. Восно-історичний вісник. 2021. № 2(40). С. 53–71. <https://doi.org/10.33099/2707-1383-2021-40-2-53-71>.
10. Олексенко О. О., Ярошенко Я. В. Розробка методу оптимального маршруту польоту безпілотної літального апарату на основі мурашиного алгоритму. Сучасний стан проведення наукових досліджень у ІТ-технологіях, галузях електроніки, інженерії, нанотехнологіях та транспортній сфері : кол. наук. монографія. Вінниця, 2021. Вид. 1. С. 147–158. <https://doi.org/10.36074/csriteenat.ed-1.10>.
11. Олексенко О. О., Герасименко В. В. Метод визначення варіантів польоту безпілотної літального апарату на основі макс-мінного мурашиного алгоритму. Сучасний стан проведення наукових досліджень у ІТ-технологіях, галузях електроніки, інженерії, нанотехнологіях та транспортній сфері : кол. наук. монографія. Вінниця, 2021. Вид. 2. С. 2–10. <https://doi.org/10.36074/csriteenat.ed-2.01>.
12. Khudov, H., Oleksenko, O., Kuchuk, N., Yaroshenko, Y., Ishchenko, O., Ikaiev, D., Kireienko V. Determining the unmanned aerial vehicle optimal flight route based on the ant colony optimization. Turkish Online Journal of Qualitative Inquiry (TOJQI). 2021. Vol. 12, Issue 6, pp. 5173–5178.
13. Khudov, H., Oleksenko, O., Lukianchuk, V., Herasymenko, V., Yaroshenko, Y., Ishchenko, O., Ikaiev, D., Golovchenko O., Volobuiev, A., Drob, Y., Solomonenko, Y., Khizhnyak, I. The Determining the Flight Routes of Unmanned Aerial Vehicles Groups Based on Improved Ant Colony Algorithms. International Journal of Emerging Technology and Advanced Engineering. 2021. Vol. 11. Issue 09, pp. 23–32. https://doi.org/10.46338/ijetae0921_03.
14. Олексенко О. О., Худов Г. В., Гниря В. В., Неводничий А. О. Методика прогнозування дій повітряного противника на основі мурашиного алгоритму. Перспективи розвитку озброєння та військової техніки Сухопутних військ : тези доп. Міжн. наук.-техн. конф.(м. Львів, 14-15 трав. 2020 р.). Львів, 2020. С. 250.
15. M. Dorigo, K. Socha. An Introduction to Ant Colony Optimization. Technical Report Series. Iridia : Universite Libre de Bruxelles, 2006. 19 p.
16. M. Dorigo, M. Birattari, T. Stützle, Ant Colony Optimization. Artificial Ants as a Computational Intelligence Technique. IEEE Computational Intelligence Magazine, 2006. № TR/IRIDIA/2006-023.14 p. <https://iridia.ulb.ac.be/IridiaTrSeries/rev/IridiaTr2006-023r001.pdf>.
17. M. Dorigo, T. Stützle. Ant Colony Optimization: Overview and Recent Advances. M. Gendreau and Y. Potvin, editors, Handbook of Metaheuristics, 2nd edition. in International Series in Operations Research & Management Science, Springer, Verlag, New York, 2010. Vol. 146. P. 227–263.

Received (Надійшла) 23.04.2023

Accepted for publication (Прийнята до друку) 31.05.2023

An improved method for predicting the flight path of a cruise missile based on the ant colony algorithm

Hennadii Khudov, Igor Romanenko, Petro Mynko, Yuriy Solomonenko, Vitaliy Ivanyuk

Abstract. The subject matter of the article is the method for predicting the flight path of cruise missiles based on the ant colony algorithm. The goal is to obtain adequate results in predicting the trajectory of cruise missiles when planning the use of air defense forces and means and choosing their rational construction. Task: analysis of known methods for estimating the flight path of a cruise missile; modelling the flight path of cruise missiles as a search for a sequence of actions of a cruise missile that can move from one place to another, avoiding any obstacles on the flight path; improvement of the method for predicting the flight path of cruise missiles. The methods used are: methods of system analysis, probability theory, methods of swarm intelligence, mathematical statistics. The following results are obtained. The method of predicting the flight path of cruise missiles based on the ant colony algorithm has been improved, in which, unlike the known ones: the calculation of the flight path of a cruise missile provides for the calculation of the objective function; the determination of the set of agent movement areas; the pheromone concentration on the agent movement routes, and the movement of agents along certain movement areas. The operability of an improved method for predicting the actions of an air enemy based on the ant colony algorithm has been tested. Conclusions. It has been established that in all cases the method of predicting the actions of an air enemy based on the ant colony algorithm provides the shortest route length from the starting point of the route to the objects of destruction. The direction of further research is the construction of radar reconnaissance based on the results of modelling options for the actions of an air enemy.

Keywords: cruise missile, air enemy, flight path, target, ant algorithm, radar reconnaissance.

Р. Бевзюк, О. Ляшенко

Харківський національний університет радіоелектроніки, Харків, Україна

СИСТЕМА РОЗПІЗНАВАННЯ ОБ'ЄКТІВ ДЛЯ АВТОМАТИЗАЦІЇ СКЛАДІВ З ВИКОРИСТАННЯМ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Анотація. **Актуальність.** З збільшенням онлайн продаж в світі, зростає потреба в системах автоматизації складів. При великій кількості товарів постає питання розпізнавання схожих за виглядом товарів, але які мають різні характеристики. В зв'язку з цим використання елементів штучного інтелекту та побудові систем комп'ютерного зору для великих складських підприємств є необхідністю. **Метою роботи** є побудова системи розпізнавання різноманітних товарів за допомогою згорткових нейронних мереж. **Об'єктом дослідження** є процеси побудови та навчання системи розпізнавання товарів на складі. **Предметом дослідження** є методи розпізнавання об'єктів на базі нейронної мережі з використанням CPU та GPU. **Висновок.** Побудована система розпізнавання об'єктів на базі згорткових нейронних мереж в середовищі MatLab. Проведено експерименти з використанням CPU та GPU для навчання нейронної мережі. Отримані результати показали, що підвищення якості розпізнавання необхідно збільшувати тестову вибірку.

Ключові слова: нейрона мережа, машинне навчання, розпізнавання об'єктів, CPU, GPU.

Вступ

Однією з основних причин швидкого зростання глобального ринку автоматизації складів є значне зростання електронної комерції та вимоги до виконання замовлень, які створюються в результаті (швидка доставка, відстеження тощо). Існує підвищена потреба в системах, здатних обробляти великі обсяги замовлень із кількох позицій.

Найбільша проблема в реалізації повністю автономних систем виконання замовлень полягає в ідентифікації та обробці сторонніх предметів [1–3]. Сучасні методи сегментації зображення забезпечують достатню точність для обробки сторонніх об'єктів для операцій вибору та розміщення. Ці методи сегментації зображення дозволяють ідентифікувати та розрізнити геометричні форми, краї, поверхні та інші характеристики, достатні для надійної обробки сторонніх об'єктів. Однак цих ознак недостатньо для ідентифікації сторонніх предметів. Це обмеження технології зумовлює необхідність присутності фізичного працівника, який забезпечує ідентифікацію сторонніх предметів у процесі виконання замовлення. Для реалізації повної автоматизації система виконання замовлень повинна мати можливість автономно розпізнавати об'єкт, який обробляється, щоб об'єкт можна було транспортувати до відповідного місця призначення.

Постановка задачі

Розробити систему, яка може розпізнавати сторонні об'єкти та обробляти їх, важко через безліч візуальних явищ, які відбуваються в реальному світі, наприклад оклюзію об'єктів або світлові ефекти. Відрізнити два об'єкти, які майже однакові за формою та кольором, але є абсолютно різними, є однією з найскладніших проблем, яку можна вирішити за допомогою сучасних технологій комп'ютерного зору.

Приклад цієї ситуації наведено на рис. 1 та 2, де два майже однакові на вигляд продукти все ще мають різний вміст.



Рис. 1. Паприка копчена (мелена)

Рис. 2. Паприка (мелена)

Щоб розробити надійну систему обробки замовлень, багато з цих складних або незвичайних випадків потрібно правильно визначити. Щоб досягти цього, властивості, унікальні для кожного елемента, повинні бути знайдені та використані для правильної класифікації кожного елемента. У деяких із найскладніших випадків, коли майже не видно візуальної різниці, залишається унікальна властивість, яку можна використовувати для правильної класифікації: штрих-код. Кожен товар, який обробляється на складах, містить штрих-код, який у певний момент сканується фізичним працівником, щоб визначити його призначення під час процесу повторного збирання замовлення.

Основна увага в системі розпізнавання об'єктів буде зосереджена на використанні штрих-коду як унікальної властивості для правильної ідентифікації об'єктів. Якщо штрих-код кожного об'єкта можна знайти та надійно розпізнати, кожен об'єкт можна ідентифікувати та правильно обробити. Незважаючи на те, що існують стандартні вказівки щодо розміщення штрих-кодів, певно, не всі товари відповідають цим правилам. Дуже цікаво дослідити, чи існує корисний зв'язок між характеристиками об'єкта та розташуванням штрих-коду. Якщо такий зв'язок існує, його

можна використовувати для дуже ефективного підбору та сканування товарів на складі, усуваючи потребу людей у ручній ідентифікації об'єктів.

Машинне навчання

Машинне навчання – це галузь, яка зосереджена на розробці алгоритмів, які роблять прогнози на основі даних. Завдання машинного навчання має на меті ідентифікувати (вивчити) функцію $f: X \rightarrow Y$, що відображає вхідну область X (даних) на вихідну область Y (можливих прогнозів) [4]. Функції f вибираються з різних класів функцій, залежно від типу алгоритму навчання, який використовується. Для завдання класифікації точність системи зазвичай вибирається як міра ефективності, де точність визначається як пропорція, для якої система правильно створює вихідні дані. Досвід, якому піддаються алгоритми машинного навчання, це набори даних. Ці набори даних містять набір прикладів, які використовуються для навчання та тестування цих алгоритмів.

Алгоритми машинного навчання можна в основному класифікувати на три категорії за типом наборів даних, які використовуються як досвід. Ці категорії включають контрольоване навчання, неконтрольоване навчання та навчання з підкріпленням. Системи контрольованого навчання використовують мічені набори даних $(x, y) \in X \times Y$, де x представляє точку даних, а y відповідне істинне передбачення для x . Цей навчальний набір пар введення-виведення використовується для пошуку детермінованої функції, яка відображає будь-який вхід на вихід, прогнозуючи майбутні спостереження вводу-виводу, мінімізуючи помилки, наскільки це можливо. Системи неконтрольованого навчання використовують немарковані набори даних для навчання системи. Метою неконтрольованого навчання є отримання структури з немаркованих даних шляхом дослідження подібності між парами об'єктів і зазвичай пов'язане з оцінкою щільності або групуванням даних. Системи навчання з підкріпленням відчувають не фіксований набір даних, а петлю зворотного зв'язку між системою та її досвідом [5]. Розглядається динамічне середовище, в якому як дані спостерігаються трійки стан-дія-винагорода. Метою навчання з підкріпленням є відображення ситуацій у діях з метою максимізації винагороди. Існують інші навчальні системи, які є комбінацією двох категорій, наприклад напівкероване навчання, яке використовує як позначені, так і немарковані дані [6].

Існує велика різноманітність завдань, які можна вирішити за допомогою машинного навчання. Двома популярними завданнями машинного навчання є регресійний аналіз і класифікація. У регресійному аналізі зв'язок між змінними є наближеним для успішного прогнозування значення за певних вхідних даних. Ця задача розв'язується виведенням функції $f: R^n \rightarrow R$, що відповідає даним [7]. Регресійний аналіз можна використовувати, наприклад, для прогнозування майбутніх цін на акції у світі торгівлі. Під час класифікації машину просять визначити категорію n , до якої належить певний вхід. Задачу можна розв'язати, вивішивши функцію $f: R^n \rightarrow \{1, \dots, n\}$ [8]. Популярною проблемою класифікації є розпізнавання об'єктів для інтелектуальних

систем. Класифікація може бути використана, наприклад, для класифікації об'єктів на складі, щоб визначити правильне призначення кожного об'єкта, за допомогою сучасного розпізнавання об'єктів із використанням алгоритмів глибокого навчання [5].

Правила розміщення штрих-коду

Існує багато програм, які можуть знаходити та зчитувати штрих-коди зі складних сцен за допомогою камери та використовувати їх для ідентифікації об'єктів. Різниця в освітленні, позах і спотворенні перспективи може ускладнити ідентифікацію штрих-коду, але існують алгоритми, які пропонують адаптовані рішення для цих складних сцен. Найцікавішим є випадок, коли штрих-код приховано в результаті пози продукту, і його неможливо ідентифікувати в результаті цього без зміни положення об'єкта. Складський контейнер складається з предметів, на яких видно штрих-код, а також інших предметів зі штрих-кодами, прихованих за іншими поверхнями.

Відомо, що існує певна схема розміщення штрих-коду на предметах різного розміру та форми. Більшість товарів, які обробляються на складах, відповідають інструкціям щодо розміщення символів штрих-кодів. Ці настанови містять загальні принципи, які застосовуються, обов'язкові правила та рекомендації щодо розміщення символу штрих-коду на певних типах упаковок.

Ці вказівки можуть виявитися дуже зручними для прогнозування правильного розташування штрих-коду, коли штрих-код приховано. Однак у багатьох випадках розміщення штрих-коду на товарах може порушувати правила. Хоча не всі предмети відповідають цим правилам розміщення штрих-кодів, дуже цікаво дослідити, чи існує корисний зв'язок між такими характеристиками об'єкта, як розмір або форма, і розташуванням штрих-коду. Якщо цей зв'язок існує, його можна використовувати для дуже ефективного вибору та сканування товарів на складі, усуваючи потребу в ідентифікації об'єкта за допомогою людини. Машинне навчання буде впроваджено для отримання цих моделей розміщення штрих-кодів на об'єктах різних форм і розмірів. Аналізуючи форму, текстуру та набору інших можливих функцій пропонується нова структура розпізнавання об'єктів, яка повинна надійно визначати розташування штрих-коду об'єктів.

В ході досліджень було визначено структуру повної системи розпізнавання об'єктів. Було вирішено, що систему буде створено з використанням згорткової нейронної мережі (ЗНМ). Тоді для цього методу локалізації штрих-коду потрібно побудувати схему розміщення штрих-коду на продуктах за допомогою згорткової нейронної мережі, щоб визначити, на якій поверхні товару лежить штрих-код.

Щоб автоматизувати фазу повторного генерування системи виконання замовлень, нещодавно розроблена структура розпізнавання об'єктів має задовольнити двом важливим вимогам:

1. Система повинна досягати 40% точності передбачення локалізації штрих-коду: оскільки системи виконання замовлень обробляють величезну

кількість продуктів, нові системи бачення повинні прагнути до принаймні 40% точності передбачення. У подальшому можна покращити, додавши більше даних для навчання, або запровадивши більш оптимальні методи навчання.

2. Система повинна бути масштабованою до дуже великих наборів продуктів: спроектована система бачення повинна мати можливість обробляти поточні набори продуктів, які складаються в порядку 107 різних продуктів.

Визначення параметрів

Щоб побудувати структуру локалізації штрих-коду, спочатку визначаються вхід, вихід і визначення класу. Складські підприємства розробляють рішення автоматизації для різних ринків: посилок, складів і аеропортів, що означає, що об'єкти, з якими ці рішення обробляють, зазвичай мають велику різноманітність за формою, кольором, вагою та іншими фізичними властивостями. Важливо розуміти, що в цьому проєкті акцент робиться на складському сегменті. Крім того, продукти вибираються зі стандартизованих категорій, оскільки це дозволяє уникнути неоднозначності щодо визначення та тлумачення класу. Сфера дії цього дослідження обмежена елементами категорії "Коробки", для яких будується доказ концепції. Після створення робочої структури до неї можна додати інші категорії продуктів.

Створений набір даних складається із зображень елемента категорії коробки, взятих із вигляду зверху збоку. Приклади цих зображень з окремими предметами в різних позах можна знайти на рис. 3 та 4.



Рис. 3. Коробка пластівців із видимим штрих-кодом у верхній частині



Рис. 4. Коробка печива з прихованим штрих-кодом

І для коробки пластівців, і для коробки з печивом система локалізації штрих-коду повинна мати можливість визначати розташування штрих-коду.

Зображення будуть імпортовані у застосунок як дані RGB з роздільною здатністю 400x300. Таким



Вхідні зображення

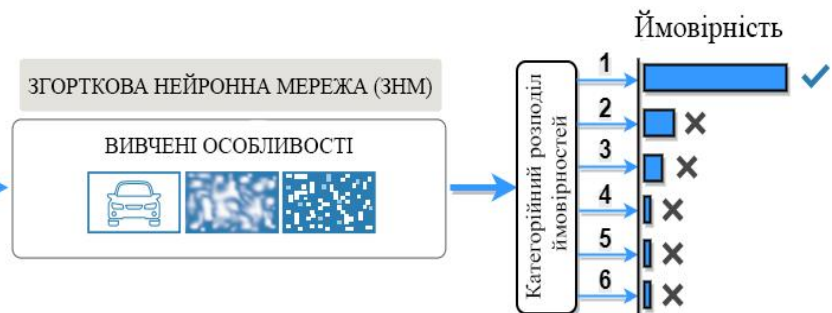


Рис. 6. Визначення класу для об'єкта у формі коробки

чином, вхідні дані складаються з матриці дизайну $X \in R^{N \times 400 \times 300 \times 3}$, де «N» представляє кількість зразків у наборі даних.

Оскільки згорткові нейронні мережі є типом контрольованого навчання, для етапу навчання рfспjсеуге потрібні дані з мітками. Для ефективного маркування всіх даних вибрано таке визначення класу:

$$c_n = \{1, 2, 3, 4, 5, 6\}$$

У цьому визначенні класу кожне з чисел представляє одну з шести граней об'єкта розміром з коробку. Це визначення класу завжди застосовується з точки зору камери, рис. 5.

Для всіх випадків, грань яка знаходиться найближче до камери, визначається як верхня грань (с=1), звідти визначаються інші обличчя.

Грань, протилежна верхній грані, є нижньою (с=3). Гранями, що прилягають до верхньої та нижньої граней, є передня (с=2), задня (с=4), права (с=5) і ліва грані (с=6).

Враховуючи перспективу камери, визначення тлумачиться та ілюструється як:

$$c_n = \{1, 2, 3, 4, 5, 6\} = \{\text{верх, спереду, знизу, ззаду, справа, зліва}\}$$

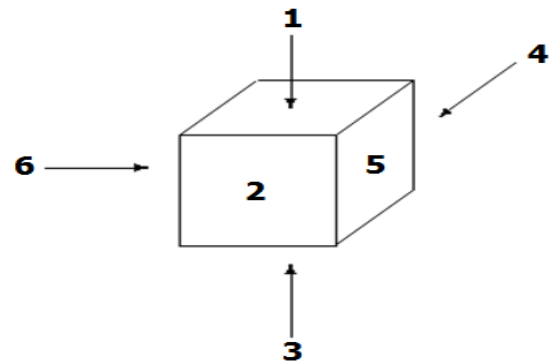


Рис. 5. Пропонована структура розпізнавання об'єктів із визначенням класу

Вихідні дані застосунку складаються з категоричного вектора розподілу ймовірностей, який визначає ймовірність наявності штрих-коду на певній грані. Можливий результат можна визначити за допомогою:

$$c_n = \{1, 2, 3, 4, 5, 6\} = \{0.8, 0.1, 0.05, 0.02, 0.02, 0.01\}$$

Цей результат означає, що з імовірністю 80% штрих-код буде розміщено на верхній стороні, 10% – на лицьовій стороні тощо. Повний каркас зображено на рис. 6.

Згорткова нейронна мережа

Згорткові нейронні мережі наразі є одним із найвідоміших алгоритмів глибокого навчання із зображеннями. У той час як для традиційного машинного навчання відповідні функції потрібно витягувати вручну, глибоке навчання використовує необроблені зображення як вхідні дані для вивчення певних функцій. ЗНМ складаються з вхідного та вихідного рівня та кількох проміжних рівнів між входом і виходом. Прикладами проміжних шарів є згорткові шари, шари максимального об'єднання та повністю зв'язані шари [9].

Архітектури ЗНМ відрізняються кількістю та типом шарів, реалізованих для конкретного застосування. Для безперервних відповідей мережа повинна включати рівень регресії в кінці мережі, тоді як для категоричних відповідей система повинна включати функцію класифікації та рівень. Нейрони в кожному шарі ЗНМ розташовані в тривимірному порядку та перетворюють тривимірний вихід із тривимірного вхідного. Для нашої конкретної програми вхідний рівень містить зображення як 3D-вхідні дані зі значеннями висоти, ширини та RGB як розмірів. Далі на згортковому шарі нейрони прикріплюються до областей зображення та перетворюються на тривимірний результат.

Конфігурації ЗНМ складаються з безлічі прихованих шарів. У кожному шарі обсяги активації змінюються за допомогою диференційованих функцій. Існують чотири основні типи шарів, які використовуються для побудови конфігурацій ЗНМ:

1. Згортковий рівень – згорткові фільтри використовуються для отримання карти активації з вхідних даних.

2. Rectified Linear Unit Layer (ReLU) – фільтрує від'ємні значення, щоб отримати лише додатні значення для набагато швидшого часу навчання.

3. Рівень об'єднання – виконує нелінійну низьку дискретизацію та скорочує кількість параметрів для простішого виведення.



Рис. 7. Конфігурація рівня згорткової нейронної мережі.

Розроблена мережа складається з 13 рівнів із різними розмірами фільтрів

Використання графічного процесора значно скорочує час навчання, і це можна зробити за допомогою Matlab без додаткового програмування. Однак для паралельних обчислень необхідний графічний процесор NVidia на основі CUDA, щонайменше 3.0. Існує також можливість використовувати кілька графічних процесорів, що ще більше зменшує час обробки. Нарешті, хмарні обчислення GPU розглядають використання хмарних ресурсів для обчислювальної потужності. Написаний код MATLAB можна покращити для цілей хмарних обчислень. У рамках цієї роботи обчислення, як на CPU, так і на GPU розглядаються для дослідження впливу використання GPU на зменшення часу обчислення.

4. Повністю зв'язаний рівень – обчислює оцінки ймовірності класу шляхом виведення вектора розмірів C , де C є кількістю класів. Усі нейрони підключені до цього шару.

Створення ЗНМ в Matlab

Структура, описана в цій роботі, розроблена в Matlab. Сам застосунок призначений для роботи на настільному комп'ютері. Використовуються два набори інструментів Matlab: Parallel Computing і Neural Network Toolboxes. Ці набори інструментів забезпечують алгоритми згорткового навчання, попередньо запрограмовані рівні та згорткові мережі, а також можливість паралельного обчислення для значного скорочення часу обчислення.

При визначенні конфігурації мережевого рівня важливо зазначити, що не існує точної формули для оптимальної конфігурації рівня. Замість цього найкращий підхід – це метод проб і помилок, коли кілька конфігурацій шарів досліджуються та порівнюються, щоб побачити, наскільки добре вони працюють. Попередньо навчені мережі, можна використовувати як початкову конфігурацію рівня. Однак у цьому випадку, переслідуються власні реалізації конфігурації шару. Вибрану конфігурацію шару показано на рис. 7. Навчання глибокої згорткової мережі, зображеної на рис. 7, може зайняти різну кількість часу, залежно від розміру набору даних і доступної потужності обробки. Для навчання ЗНМ доступні три варіанти обчислень, і вибір найбільш оптимального є вирішальним залежно від кількості доступного часу для вирішення конкретного завдання:

1. Обчислення на основі CPU.

2. Обчислення на основі GPU.

3. Хмарні обчислення GPU.

Варіанти обчислень на основі CPU є найбільш простими та доступними. Однак оскільки центральний процесор обчислює завдання в послідовній конфігурації, навчання мережі займає найдовше часу за допомогою цього методу.

Навчання ЗНМ

Під час імпорту позначеного набору даних 75%/25% розподіляються між даними навчання та даними перевірки. Це означає, що 75% даних використовується для навчання мережі, а 25% даних використовується для перевірки мережі. Після досягнення достатньої точності перевірки можна використати додатковий набір тестів, щоб побачити, наскільки добре працює мережа. Перша конфігурація виконується на основі обчислювальної моделі ЦП. Використовується процесор Intel Core i7 720QM з чотирма фізичними та чотирма віртуальними ядрами. Хід навчання та результат показані на рис. 8:



Рис. 8. ЗНМ навчальний запуск 1, з використанням центрального процесора. Точність перевірки 16% досягається за час навчання приблизно 431 хвилину

Незважаючи на те, що використовується CPU має чотири фізичні ядра, за час, що минув, можна відразу помітити, що навчання з CPU все одно займає надзвичайно багато часу, і його неможливо точно налаштувати в рамках цього проекту. Навіть якщо центральний процесор складається з кількох ядер і є найкращим у лінійці, велика тривалість обробки залишається проблемою. Дивлячись на результати тесту, точність перевірки в 16% є дуже недостатньою, і схоже, що ЗНМ не дізнався жодних цінних функцій, які можна було б використати для правильної класифікації. Змінюючи кількість функцій, які обчислюють згорткові шари, і розмір фільтра, який використовує згортковий шар, система додатково налаштована для підвищення точності перевірки. Однак з другого запуску на основі GPU беруться обчислення. Для наступних прогонів використовується графічний процесор NVidia GTX1060 і процесор Intel i7-6700HQ. Результати другого запуску показані на рис. 9.

Що відразу можна відзначити, так це те, що час обчислень зменшився з 431 хвилини до приблизно 1 хвилини. Це означає, що час обчислення зменшився приблизно в 400 разів. Це принаймні дозволяє ефективно налаштувати ЗНМ. Дивлячись на результати перевірки, досягнуто точності 40,16%. У той час як точність навчального набору збільшується приблизно до 100%, точність перевірки продовжує коливатися близько 40%.

Порівнюючи втрату під час навчання з втратою підтвердження, можна побачити, що хоча навчальний набір справді збігається до нуля, набір підтвердження залишається приблизно 1,75. Подальше тонке налаштування виконується для ЗНМ, щоб дослідити, чи може зміна швидкості навчання, розміру мінібатча чи епохи, або конфігурацій і налаштувань шару ще більше покращити ЗНМ. Це робиться кілька разів поспіль, показуючи деякі з цікавих результатів на рис. 10-12.

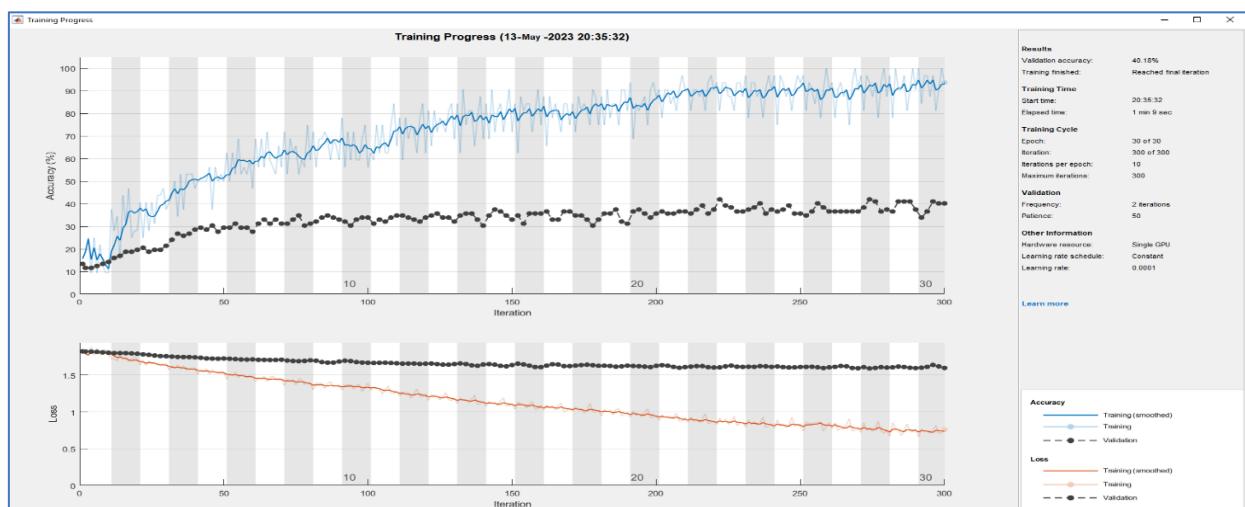


Рис. 9. Тренування ЗНМ 2 з використанням GPU. Параметри системи вже дещо налаштовані. За рахунок зменшення розміру фільтра та кількості фільтрів точність перевірки 40,18% досягається з часом навчання 69 секунд

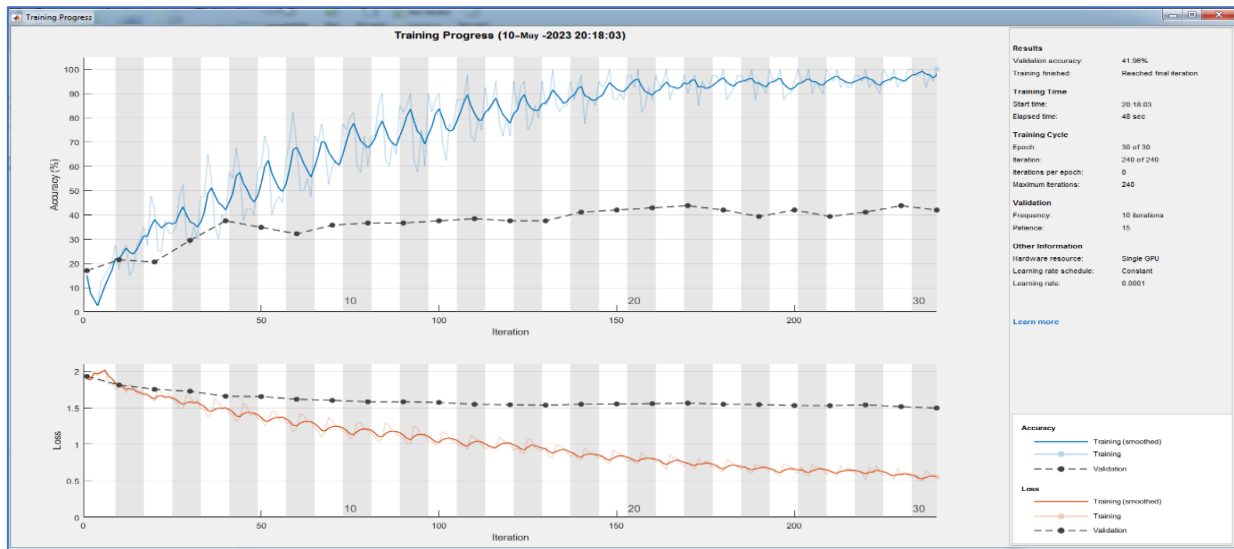


Рис. 10. Навчальний запуск ЗНМ (з GPU) зі зміненими темпами навчання та розмірами фільтрів шарів. Точність перевірки 41,96% досягається за час навчання 48 секунд

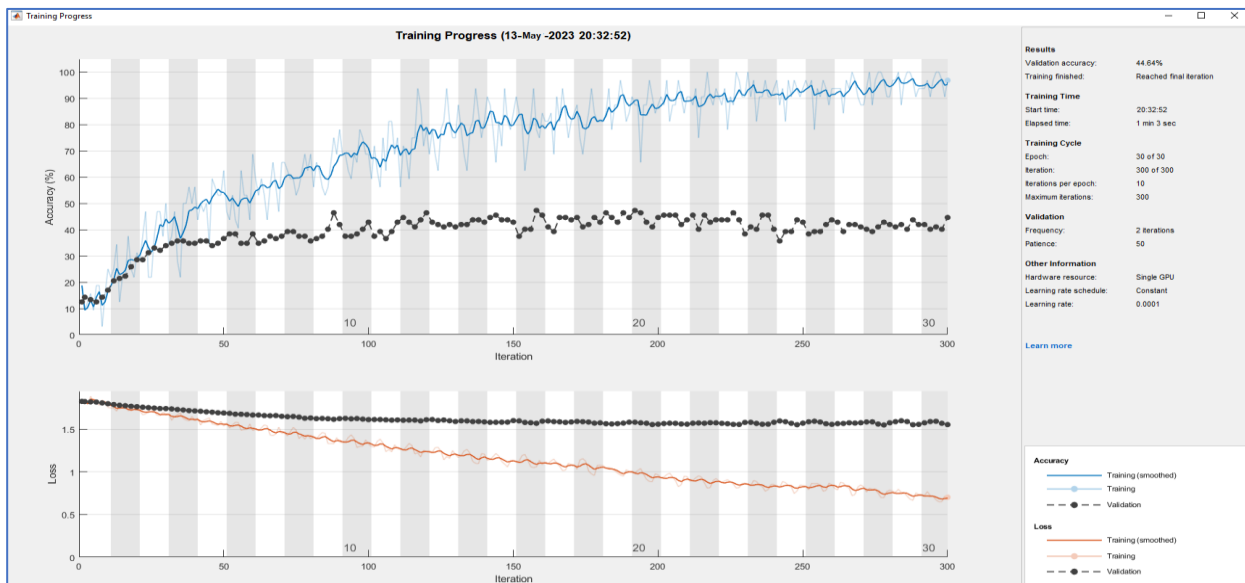


Рис. 11. Навчальний запуск ЗНМ (з GPU) зі зміненими темпами навчання та розмірами фільтрів шарів. Точність перевірки 44,64% досягається за час навчання 63 секунди

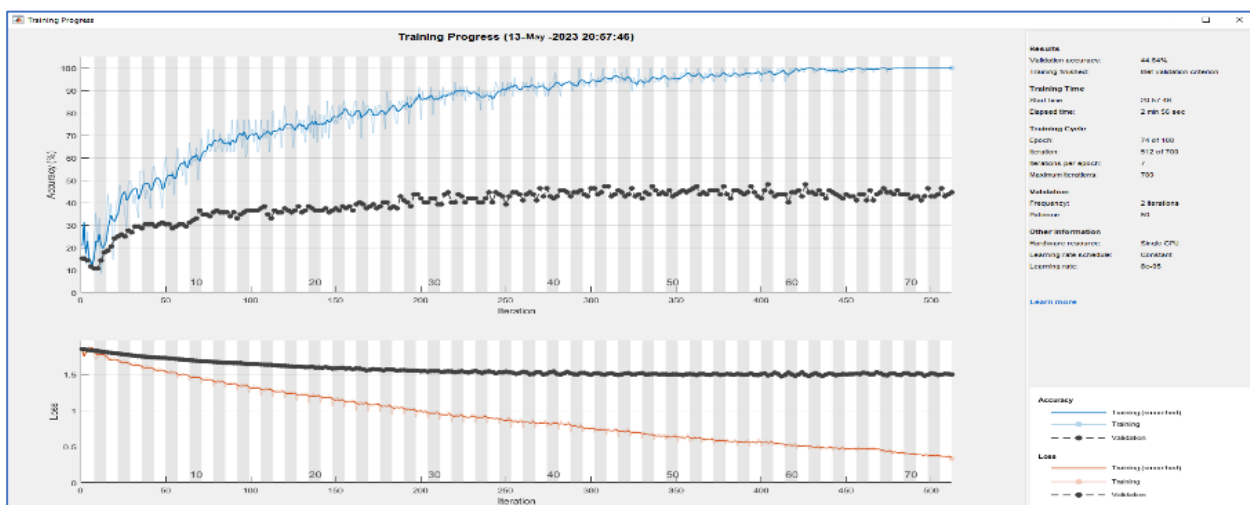


Рис. 12. Навчальний запуск ЗНМ (з GPU) зі зміненими темпами навчання та розмірами фільтрів шарів. Точність перевірки 44,64% досягається за час навчання 2 хвилини 56 секунд

Після проведення кількох десятків тестів було зроблено висновок, що точність перевірки набору даних становить приблизно 45% після точного налаштування системи. Точність навчального набору зростає до 100%, але набір перевірки продовжує коливатися на рівні 45%. З цих результатів можна зробити висновок, що з поточним розміром набору даних близько 1000 зображень ЗНМ не може достатньо розрізнити екземпляри даних. Ці результати є ідеальною ілюстрацією випадку, коли мережа не узагальнює достатньо добре, щоб не побачити екземпляри даних. ЗНМ переповнив дані, як видно зі 100% точності на навчальному наборі, а налаштування швидкості навчання та розмірів пакетів, здається, не мають жодного додаткового ефекту після точного налаштування, щоб досягти точності близько 45%. Висновок, який можна зробити, такий: для навчання ЗНМ потрібен більший набір даних, щоб зменшити переобладнання та досягти вищої точності.

Висновки

Розглядаючи запропоновану структуру розпізнавання об'єктів у світлі визначених вимог і обмежень, можна зробити висновок, що досягнуто достатньої точності 44,64%, що перевищує вимогу 40%. Логічний висновок, який можна зробити з результатів навчання, полягає в тому, що набір даних має бути різко збільшений, щоб запобігти переобладнанню даних. Крім того, набори даних усіх інших категорій продуктів повинні бути додані до ЗНМ, щоб

структура була реалізована на складі. Хоча генерація та збір цих даних може зайняти багато часу, з результатів, досягнутих на сьогоднішній день, можна побачити, що навіть з невеликими наборами даних показані багатообіцяючі результати.

Під час навчання ЗНМ необхідна реалізація GPU. Навчання ЗНМ лише за допомогою процесора може зайняти місяці для повного набору даних категорії продуктів. Логічним є висновок, що тонке налаштування системи під час цього процесу неможливе в цьому часовому режимі. Масштабованість, здається, не є проблемою при реалізації обчислень GPU, і, крім того, можна реалізувати хмарні обчислення, щоб ще більше скоротити час навчання під час навчання дуже великих наборів даних. Коли досягається достатня точність класифікації ЗНМ, можна застосувати розумні стратегії комплектування, коли робот-комплектатор намагається класифікувати продукти, показуючи грані продукту вздовж вектора ймовірності класифікації. Цей метод забезпечує негайний вибір, коли продукт класифіковано правильно, і вторинний вибір у разі неправильної класифікації, коли робот намагається сканувати обличчя відповідно до вектора ймовірності класифікації, доки не буде знайдено штрих-код. Це гарантує мінімальні або повну відсутність витрат на помилки, оскільки продукт завжди сканується. У цьому випадку точність класифікації ЗНМ корелює з тим, наскільки швидко сканер штрих-кодів класифікує продукти з першої спроби, де підвищення точності означає швидший вибір.

СПИСОК ЛІТЕРАТУРИ

1. Kuchuk, H., Kovalenko, A., Ibrahim, B.F. and Ruban, I. (2019), "Adaptive compression method for video information", *Int. Journal of Advanced Trends in Computer Science and Engineering*, 8(1), pp. 66-69, DOI: <http://dx.doi.org/10.30534/ijatcse/2019/1181.22019>
2. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: <https://doi.org/10.20998/2522-9052.2018.1.04>
3. Свиридов А. С., Коваленко А. А., Кучук Г. А. Метод перерозподілу пропускної здатності критичної ділянки мережі на основі удосконалення ON/OFF-моделі трафіку. *Сучасні інформаційні системи*. 2018. Т. 2, № 2. С. 139–144. DOI: <https://doi.org/10.20998/2522-9052.2018.2.24>
4. R. Bekkerman, M. Bilenko, J. Langford, *Scaling Up Machine Learning*, Cambridge University Press, January 2012
5. Mitchell, T. (1997). *Machine Learning*. McGraw Hill.
6. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge: MIT Press
7. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017), "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, No. 6
8. Alexander Toshev, Christian Szegedy (2014), "DeepPose: Human Pose Estimation via Deep Neural Networks", DOI: <https://doi.org/10.1109/CVPR.2014.214>
9. I. Mykhailichenko, H. Ivashchenko, O. Barkovska and O. Liashenko, "Application of Deep Neural Network for Real-Time Voice Command Recognition," *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek)*, Kharkiv, Ukraine, 2022, pp. 1-4, doi: 10.1109/KhPIWeek57572.2022.9916473.

Received (Надійшла) 17.03.2023

Accepted for publication (Прийнята до друку) 16.05.2023

Object recognition system for component automation using a convergular neural network

Roman Bevzyuk, Oleksii Lyashenko

Abstract. Topicality. With the increase in online sales in the world, the need for warehouse automation systems is growing. With a large number of products, the question arises of recognizing products that are similar in appearance, but have different characteristics. In this regard, the use of elements of artificial intelligence and the construction of computer vision systems for large warehouse enterprises is a necessity. **The purpose of the work** is to build a system for recognizing various goods using convolutional neural networks. **The object** of the study is the process of building and training a system for recognizing goods in a warehouse. **The subject** of the research is object recognition methods based on a neural network using CPU and GPU. **Conclusion.** An object recognition system was built based on convolutional neural networks in the MatLab environment. Experiments were conducted using CPU and GPU for neural network training. The obtained results showed that to increase the quality of recognition, it is necessary to increase the test sample.

Keywords: neural network, machine learning, object recognition, CPU, GPU.

Daniil Vyshnivskyi, Oleksii Liashenko, Nataliia Yeromina

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

HUMAN POSE ESTIMATION SYSTEM USING DEEP LEARNING ALGORITHMS

Abstract. The purpose of this work is the software implementation of neural network that can solve problem of Human Pose Estimation. With rapid improvements of neural network models and computing resources over last 10 years it's become possible to automate a lot of processes, carry out research and improve quality of life. One of the directions is Computer Vision: it allows to recognize objects, track motions, image segmentation, facial recognition etc. Human pose estimation is the part of Computer Vision area of research. It allows to capture human pose from a video or an image and have many uses in medicine, sport, augmented reality, video games etc. Therefore, the goal of this work is to find and optimize algorithm, that is relatively accurate, for identifying and classifying the joints in the human body. To achieve the goal, the following tasks were solved: current methods and technologies that is commonly used to solve problem of human pose estimation were reviewed and analyzed, artificial neural networks were used as a mathematical apparatus for the model, software implementation for human pose estimation was developed and tested, outputs from model were analyzed and evaluated, results and conclusion were formulated.

Keywords: human pose estimation; classification of objects; object detection; convolutional neural networks.

Introduction

Human Pose Estimation (HPE) is a fundamental and rapidly evolving area of research within the field of computer vision. It aims to estimate the positions and orientations of human body joints in images or videos, enabling machines to understand and analyze human body poses. The ability to perceive and interpret human poses has far-reaching implications for numerous applications, such as:

Human-computer interaction: Pose estimation can be used to create natural user interfaces for controlling devices, applications, or games using body gestures.

Activity recognition: By analyzing the body poses of people in a scene, you can determine the activities they are engaged in, such as walking, running, or sitting.

Animation and gaming: Pose estimation can be used to drive the animation of virtual characters, enabling realistic motion and interaction in video games or computer-generated films.

Sports analysis: Coaches and athletes can use pose estimation to analyze and improve their techniques, postures, and movements during training or competitions.

Healthcare and rehabilitation: Pose estimation can be employed in monitoring patients during physiotherapy, tracking their progress, and providing feedback on their exercises.

Fitness and wellness: Applications can leverage pose estimation to offer real-time guidance and feedback on workout routines, yoga postures, or dance moves.

Surveillance and safety: Pose estimation can aid in detecting unusual or dangerous activities in public places, such as detecting falls, fights, or accidents.

Robotics: Robots can use HPE to understand human activities and interact with people more effectively, enabling better collaboration between humans and robots.

Fashion and retail: Pose estimation can be used to create virtual fitting rooms, allowing customers to see how clothes fit on a digital representation of themselves.

Augmented reality: By estimating human poses, augmented reality applications can overlay virtual content on real-world scenes in a more context-aware and interactive manner.

Over the past decade, HPE has witnessed significant advancements, primarily driven by the proliferation of deep learning techniques and the availability of large-scale annotated datasets. Convolutional Neural Networks (CNNs) [1] have emerged as a powerful approach to model the complex spatial relationships between body joints and achieve state-of-the-art performance on various HPE benchmarks.

There are several popular deep learning architectures [2] for HPE, such as Stacked Hourglass Networks [3], and Simple Baselines [4]. These architectures utilize CNNs to estimate human poses. Recent advancements also involve using Transformer-based models, such as Vision Transformers [5] (ViT), for pose estimation tasks.

Analysis of recent research and publications. Vision Transformers have emerged as a powerful architecture for computer vision tasks, originally proposed by Dosovitskiy et al. for image classification. They have shown remarkable performance, often surpassing traditional CNNs on various tasks. The core idea behind Vision Transformers is to leverage the self-attention mechanism from the Transformer architecture, initially designed for natural language processing, and apply it to images by treating them as sequences of tokens.

Recently, researchers have started to explore the potential of ViT for pose estimation tasks. To adapt the ViT architecture for human pose estimation, several modifications and approaches can be considered:

Tokenization: divide the input image into non-overlapping patches and linearly embed them as tokens. Include an additional learnable token, known as the class token, to aggregate global information across the image.

Position embeddings: add learnable position embeddings to the patch tokens to capture the spatial information, which is essential for pose estimation tasks.

Multi-scale feature extraction: integrate multi-scale feature extraction into the ViT architecture to capture both local and global context. This can be achieved by incorporating multi-scale patch sizes, multi-head self-attention, or pyramid-like architectures.

Heatmap prediction: modify the output layer of the ViT architecture to predict heatmaps for each keypoint.

Heatmaps are 2D probability maps that represent the likelihood of a keypoint being present at each pixel location. The final keypoint locations can be obtained by finding the maximum value in each heatmap.

Multi-stage refinement: incorporate a multi-stage architecture to refine the pose estimation iteratively. Each stage can consist of a Vision Transformer module followed by an output prediction layer. Intermediate supervision can be applied to encourage the model to learn better intermediate representations.

Multi-person pose estimation: for multi-person pose estimation tasks, incorporate an additional output layer to predict the presence of individuals in the image, along with their corresponding keypoints.

Several recent works have successfully applied ViT to HPE tasks, achieving competitive performance compared to traditional CNN-based methods. Some notable examples include Swin Transformer and VisTR which demonstrate the potential of Transformer-based architectures for pose estimation.

While Vision Transformers have shown promise in pose estimation tasks, there remain challenges to address, such as computational complexity and the need for large amounts of training data. Ongoing research in this area aims to overcome these challenges and further improve the performance of Vision Transformers for HPE and other computer vision task

The Stacked Hourglass Network, proposed by Newell et al. in their paper "Stacked Hourglass Networks for Human Pose Estimation," is a deep learning architecture specifically designed for HPE tasks. This model has been particularly influential in the field, as it introduced a novel approach to capturing and combining multi-scale contextual information in a deep learning framework.

The Stacked Hourglass Network consists of multiple stacked hourglass modules, each of which is responsible for predicting the heatmaps for each keypoint. The overall architecture is symmetric, with a series of downsampling layers followed by upsampling layers, resembling an hourglass shape. Here's an overview of the key components Stacked Hourglass Network:

Hourglass module: The hourglass module is designed to capture and process multi-scale contextual information within an image. It consists of a series of convolutional and pooling layers that progressively downsample the input, followed by a series of upsampling and convolutional layers that gradually restore the original resolution. The downsampling and upsampling stages are connected through skip connections, which allow the model to learn both local and global context.

Intermediate supervision: The Stacked Hourglass Network employs intermediate supervision between the hourglass modules. After each hourglass module, the model predicts a heatmap for each keypoint. The predicted heatmaps are compared to the ground truth heatmaps, and the loss is backpropagated to refine the model's predictions. This intermediate supervision encourages the model to learn better intermediate representations.

Stacking hourglass modules: multiple hourglass modules are stacked together to form the complete Stacked Hourglass Network. The output of each module is passed as input to the next module, allowing the model to refine its predictions iteratively. The final pose estimation is obtained by combining the predictions from all the stacked modules.

Heatmap prediction: The Stacked Hourglass Network predicts a heatmap for each keypoint, which represents the likelihood of the keypoint being present at each pixel location in the image. The final keypoint locations are obtained by finding the maximum value in each heatmap.

The Stacked Hourglass Network has been highly successful in HPE tasks, achieving state-of-the-art performance on several benchmarks when it was introduced. Its design and principles have influenced many subsequent pose estimation models. While newer architectures, such as those based on Vision Transformers, have emerged, the Stacked Hourglass Network remains a valuable reference point for understanding and developing HPE models.

The Simple Baselines model for HPE, proposed by Xiao et al. in their paper "Simple Baselines for Human Pose Estimation and Tracking," is a deep learning-based approach that focuses on simplicity and efficiency while achieving competitive performance compared to more complex models. The Simple Baselines model uses a ResNet backbone for feature extraction, followed by a few deconvolutional layers to predict heatmaps for each keypoint.

The model employs a ResNet backbone (e.g., ResNet-50, ResNet-101, or ResNet-152) to extract high-level features from the input image. ResNet architectures have proven to be effective for various computer vision tasks due to their ability to learn deep representations while mitigating the vanishing gradient problem through residual connections.

"The Simple Baselines for Human Pose Estimation and Tracking" model stands out for its simplicity and computational efficiency while maintaining competitive performance in HPE tasks.

Its straightforward design makes it easier to implement, understand, and extend compared to more complex models, making it an attractive choice for various pose estimation applications.

Materials and methods

HPE involves predicting (Fig. 1) the spatial locations of human body joints or keypoints from input data such as images or videos [6]. It typically focuses on 2D or 3D pose estimation:

2D Pose Estimation [7]: this estimates the 2D coordinates (x, y) of the body keypoints in an image or video frame. It is widely used due to its computational efficiency and applicability to various applications.

3D Pose Estimation: this estimates the 3D coordinates (x, y, z) of the body keypoints. It provides more information about the pose but requires additional information or assumptions, such as camera parameters, depth information, or temporal information across video frames.



Fig. 1. Examples of HPE model output

Here are the main approaches to HPE:

- Top-Down Approach: this involves two main steps: human detection and keypoint localization. First, a human detector (e.g., Faster R-CNN, YOLO, or SSD) is used to locate people in the input image. Then, a pose estimation model is applied to each detected bounding box to estimate the keypoints.

- Bottom-Up Approach: this approach first predicts the individual body keypoints for all the people in the input image. Then, a grouping algorithm (e.g., greedy inference, bipartite matching) is used to associate the detected keypoints with each person. This approach is computationally efficient for scenes with multiple people.

HPE is one of the problems that are solved using image classification. Image classification is a fundamental computer vision task that aims to assign a predefined class label to an input image based on the objects or features present within it. A typical algorithm for performing image classification consists of several steps, such as preprocessing, feature extraction, and classification. In recent years, deep learning techniques, particularly (CNNs), have become the go-to approach for image classification tasks due to their superior performance [8, 9].

Here is a high-level overview of the algorithm for performing image classification using a CNN:

- Data preprocessing: prepare the input data by performing operations such as resizing, normalization, and data augmentation. These steps help ensure that the input images are in a consistent format and improve the robustness of the model to variations in the data.

- Feature extraction: a CNN consists of several convolutional, pooling, and activation layers that learn to extract meaningful features from the input image. These layers capture both low-level features, such as edges and textures, and high-level features, such as object parts and semantic information. The output of the feature extraction stage is a high-dimensional feature map that represents the input image in a more compact and informative manner.

Fully connected layers: after feature extraction, one or more fully connected layers, also known as dense layers, are used to process the high-dimensional feature maps. These layers help the model learn non-linear combinations of the extracted features, which can improve the discriminative power of the model.

Output layer: the final layer of the CNN is a fully connected layer with as many neurons as there are class labels. This layer is typically followed by a softmax activation function, which converts the output of the neurons into class probabilities.

Loss function and optimization: the CNN is trained using a suitable loss function, such as cross-entropy loss, which measures the difference between the predicted class probabilities and the ground truth class labels. The model parameters are optimized using gradient-based optimization techniques, such as stochastic gradient descent (SGD) or adaptive optimizers like Adam, to minimize the loss function.

Evaluation: once the model is trained, it can be evaluated on a test set to measure its performance using metrics such as accuracy, precision, recall, or F1-score.

Inference: for a new input image, the trained CNN processes the image through the same series of layers, and the softmax activation function outputs a probability distribution over the class labels. The class label with the highest probability is assigned to the input image as the final classification result.

For this paper Simple Baselines model was chosen because of its simplicity and resource efficiency while still achieving state-of-the-art performance.

This model takes a ResNet as basis and several changes that will be described below.

ResNet is a deep learning architecture proposed by He et al. in their paper "Deep Residual Learning for Image Recognition." ResNet introduces residual connections, or skip connections, to mitigate the vanishing gradient problem that arises in deep neural networks. This allows the model to learn deep representations more effectively and achieve state-of-the-art performance on various computer vision tasks.

Basic building block: The basic building block of a ResNet model consists of two or three convolutional layers followed by batch normalization [10], [11] and ReLU (Rectified Linear Unit) activation functions. Let's denote the input to a basic block as x , and the output of the convolutional layers, batch normalization, and ReLU as $F(x)$. In a simple case, where the input and output have the same dimensions, the residual connection is applied as follows:

$$y = F(x) + x$$

Here, y represents the output of the basic block, and the addition operation is performed element-wise.

Bottleneck block: For deeper ResNet models, a bottleneck block is often used to reduce the number of parameters and computational complexity.

Projection shortcut: When the input and output dimensions of a basic or bottleneck block are different, a projection shortcut is used to match the dimensions.

Stacking layers: The ResNet architecture is constructed by stacking basic or bottleneck blocks, forming a deep neural network. It begins with an initial convolutional layer and pooling layer, followed by a series of blocks grouped into stages, and finally, an average pooling layer and a fully connected layer for classification.

Loss function: ResNet is trained using the cross-entropy loss function, which is commonly used for multi-class classification problems.

In summary, the ResNet model leverages residual connections to enable the training of deep neural networks, allowing it to learn complex hierarchical features effectively. The simplicity and effectiveness of the ResNet architecture have made it widely adopted in various computer vision tasks, including image classification, object detection, and semantic segmentation. Simple Baselines (Fig. 2) propose a few changes to the original model:

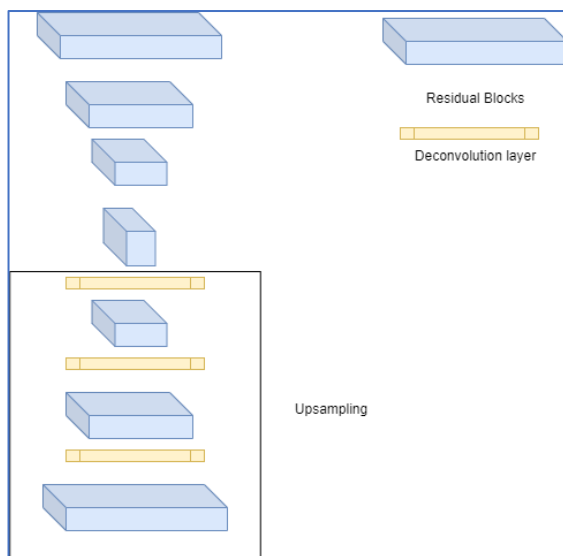


Fig. 2 Simple Baselines architecture

Deconvolutional layers: After feature extraction, the model utilizes a series of deconvolutional layers to upsample the feature maps to a higher resolution.

Deconvolutional layers, also known as transposed convolution or fractionally-strided convolution, are used to increase the spatial dimensions of the feature maps while preserving the spatial information. The purpose of these layers is to generate heatmaps for each keypoint at a suitable resolution, enabling accurate keypoint localization.

Batch normalization and ReLU: Between the deconvolutional layers, batch normalization and ReLU activation functions are applied to normalize the feature distribution and introduce nonlinearity, respectively.

Heatmap prediction: The final layer of the model is a 1×1 convolutional layer that predicts a heatmap for each keypoint. Heatmaps represent the likelihood of a keypoint being present at each pixel location in the image. The final keypoint locations are obtained by finding the maximum value in each heatmap.

Loss function: the model is trained using the Mean Squared Error (MSE) loss between the predicted heatmaps and the ground truth heatmaps. This loss encourages the model to learn accurate keypoint localization.

So, to build solution using this model following steps were performed:

For the training a dataset was chosen and collected – COCO (Common Objects in Context) [12] 2017 training dataset. It's set of data with already classified objects in it. Next step is building a neural network mode. As the backbone the ResNet50, pretrained on ImageNet is used. After that deconvolution layers are added. Next stage is training on previously prepared COCO 2017 dataset, using following params: Mean Square error as loss function, Adam optimization, accuracy metric. After this training steps quality of model is evaluated using default metrics for COCO Keypoint Detection.

In the context of HPE, Average Precision (AP) and Average Recall (AR) are two evaluation metrics commonly used to quantify the performance of a model. They are based on precision and recall metrics that are widely used in various computer vision tasks (Table 1).

Table 1 – Metric results using ResNet 50 and 256x192 image size

Metric Name	Value	Metric Name	Value
AP	70.4	AR	76.2
AP .5	88.6	AR .5	93
AP .75	77.8	AR .75	83
AP (M)	67	AR (M)	71.9
AP (L)	76.9	AR (L)	82.4

Precision: Precision measures the proportion of true positive predictions (correct keypoint detections) out of all positive predictions made by the model. In HPE, it assesses how accurate the model is in detecting keypoints.

Recall: Recall measures the proportion of true positive predictions (correct keypoint detections) out of all actual positive instances (ground truth keypoints) in the dataset. In HPE, it assesses how well the model detects all keypoints present in the image.

AP and AR for HPE are calculated as follows:

Average Precision (AP): to compute the AP for HPE, we first calculate the precision at different levels of

keypoint localization error (thresholds). For example, the model's predictions are considered correct if the distance between the predicted and ground truth keypoints is below a certain threshold (e.g., a percentage of the head segment length). The AP is then calculated as the average of precision values across different error thresholds. The AP provides a single value that summarizes the model's performance in terms of both accuracy and robustness to keypoint localization errors.

Average Recall (AR): similarly, the AR for HPE is calculated by computing the recall at different levels of keypoint localization error (thresholds). The AR is then obtained as the average of recall values across different error thresholds. The AR offers a single value that summarizes the model's performance in terms of both the ability to detect keypoints and its robustness to localization errors.

Both AP and AR can be computed for individual keypoints or averaged across all keypoints to obtain a single value that represents the overall performance of

the HPE model. High AP and AR values indicate that the model is both accurate and robust in detecting and localizing keypoints in the input images.

Conclusions

Problem of Human pose estimation was reviewed in this paper. Methods and models for human joints classification were explored and of them was implemented and evaluated. It has showed results that meet current expectations of accuracy for neural networks for HPE. It's important to note that this is only one of the models that can be used for HPE and can't be considered as best and only solution. As main object of HPE is human it will only increase its actuality in future. Results can be further used in activity recognition, animation, gaming sports analysis, healthcare and rehabilitation, fitness and wellness surveillance and safety, robotics, fashion and retail, augmented reality modern approaches what results in further growth of demand of such solutions in future.

REFERENCES

1. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017), "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, No. 6
2. Alexander Toshev, Christian Szegedy (2014), "DeepPose: Human Pose Estimation via Deep Neural Networks", DOI: <https://doi.org/10.1109/CVPR.2014.214>
3. Alejandro Newell, Kaiyu Yang, Jia Deng (2016), "Stacked Hourglass Networks for Human Pose Estimation", DOI: <https://doi.org/10.48550/arXiv.1603.06937>
4. Bin Xiao, Haiping Wu, Yichen Wei (2018), "Simple Baselines for Human Pose Estimation and Tracking", DOI: <https://doi.org/10.48550/arXiv.1804.06208>
5. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby (2020), "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", DOI: <https://doi.org/10.48550/arXiv.2010.11929>
6. Girdhar, R., Gkioxari, G., Torresani, L., Paluri, M., Tran, D., "Detect-and-track: Efficient pose estimation in videos. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*"
7. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B. 20(14) "2d human pose estimation: New benchmark and state of the art analysis. In: *IEEE Conference on Computer Vision and Pattern Recognition*"
8. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2015), "Deep Residual Learning for Image Recognition", DOI: <https://doi.org/10.48550/arXiv.1512.03385>
9. Huang, Gao; Liu, Zhuang; Van Der Maaten, Laurens; Weinberger, Kilian Q. (2017). *Densely Connected Convolutional Networks*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), doi:10.1109/CVPR.2017.243
10. Ioffe, S., Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning*."
11. Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik, (2018) "End-to-end Recovery of Human Shape and Pose", DOI: <https://doi.org/10.48550/arXiv.1712.06584>
12. COCO2017 dataset, available at <https://cocodataset.org/?ref=blog.roboflow.com#download>

Received (Надійшла) 12.03.2023

Accepted for publication (Прийнята до друку) 16.05.2023

Система оцінки пози людини з використанням алгоритмів глибокого навчання

Д. Вишнівський, О. Ляшенко, Н. Єрємона

Анотація. Метою даної роботи є програмна реалізація нейронної мережі, яка може вирішити задачу оцінки пози людини. Завдяки швидкому вдосконаленню моделей нейронної мережі та обчислювальних ресурсів за останні 10 років стало можливим автоматизувати багато процесів, проводити дослідження та покращувати якість життя. Одним із напрямків є комп'ютерний зір: він дозволяє розпізнавати об'єкти, відстежувати рухи, сегментувати зображення, розпізнавати обличчя тощо. Оцінка пози людини є частиною напряму досліджень комп'ютерного зору. Це дозволяє захопити людську позу з відео або зображення та має багато застосувань у медицині, спорті, доповненій реальності, відеоіграх тощо. Таким чином, ціль цієї роботи полягає в тому, щоб знайти та оптимізувати алгоритм, який є відносно точним, для ідентифікації та класифікації суглобів в тілі людини. Для досягнення поставленої мети були вирішені наступні завдання: розглянуто та проаналізовано сучасні методи та технології, які зазвичай використовуються для вирішення задачі оцінки пози людини, використано штучні нейронні мережі як математичний апарат для моделі, програмна реалізація для оцінки пози людини. Було розроблено та протестовано, результати моделі проаналізовано та оцінено, результати та висновки сформульовано.

Ключові слова: оцінка пози людини; класифікація об'єктів; виявлення об'єктів; згорткові нейронні мережі.

М. О. Волк, М. В. Гора, В. Г. Лабазов, А. В. Міщенко, А. І. Барсуков, В. В. Голець

Харківський національний університет радіоелектроніки, Харків, Україна

ЖУРНАЛІЗАЦІЯ СТАНУ ПРОГРАМ ДЛЯ САМОВІДНОВЛЕННЯ ПАРАЛЕЛЬНИХ ПРОГРАМНИХ СИСТЕМ

Анотація. У статті розглянута система самовідновлення паралельного програмного забезпечення з використанням журналізації точок відновлення. Самовідновлення є необхідною властивістю сучасних програмних засобів, яка надає можливість автоматичного виявлення, діагностики та відновлення працездатності систем. Основними етапами відновлення є зберігання стану програм (журналізація) в точках відновлення, моніторинг стану програм для виявлення помилок, створення патчів, відновлення стану програм до відповідної точки відновлення. У роботі запропоновано структуру системи, описано алгоритм її функціонування; обговорюються питання призначення та віртуалізація точок відновлення; наведено опис експериментальної програмної системи та її застосування для відновлення поширених програмних систем. Результати можуть бути чималого поширення та застосовуватись у роботі більшості програмних, інформаційних систем з метою автоматизації відновлення, налагодження та експлуатації сучасних комп'ютерних та хмарних систем.

Ключові слова: програмна система, самовідновлення, точка відновлення, стан програми, точка відновлення.

Вступ

Самовідновлення є необхідною властивістю сучасних інформаційних систем, що надає здатність автоматично виявляти, діагностувати та вирішувати проблеми, які виникають у системі, без втручання людини. Ці системи розроблені таким чином, щоб бути відмовостійкими, тобто вони здатні продовжувати функціонувати, навіть якщо певні компоненти відмовляють. Окрім автоматичного вирішення проблем, системи самовідновлення також можуть вживати профілактичних заходів, щоб уникнути можливих проблем у майбутньому. Це може включати передбачення потенційних проблем на основі шаблонів або тенденцій у системних даних і вжиття профілактичних заходів для запобігання виникненню цих проблем.

Інформаційні системи, що самовідновлюються, стають все більш важливими, оскільки підприємства та організації все більше покладаються на технології для роботи. Зводячи до мінімуму час простою та зменшуючи потребу у втручанні людини у вирішення системних проблем, системи самовідновлення можуть підвищити доступність, надійність і продуктивність системи.

Системи самовідновлення використовують різні методи, такі як резервування, реплікація та моніторинг, щоб гарантувати, що система залишається в робочому стані та може швидко відновлюватися після будь-яких збоїв.

Наприклад, якщо сервер у системі самовідновлення виходить з ладу, система може автоматично перенаправляти трафік на інші сервери в мережі, щоб користувачі не постраждали.

В роботі [1] було наведено класифікаційну базу щодо відмов, процесів зберігання стану компонентів системи та задач забезпечення самовідновлення комп'ютерних систем. Зокрема сформульовані основні складові вирішення проблеми самовідновлення: створення механізму самовідновлення, моніторинг стану системи (діагностика та прогнозування збоїв), виявлення порушення нормальної роботи,

корегування стану системи, створення умов для запобігання повторення помилки.

Стосовно самовідновлення програмних засобів, як складової інформаційних систем, одним з ефективних методів самовідновлення є створення бекапів – точок зберігання стану програм в фіксований момент часу.

Аналіз існуючих підходів щодо самовідновлення програмних систем

Існує багато підходів для відновлення програмного забезпечення після збоїв. Найбільш поширені з них використовують методи[^]

перезавантаження, включаючи перезапуск усієї програми [2],

оновлення програмного забезпечення [3] і мікроперезавантаження [4], які повертають систему до початкового стану до або після виявлення несправності.

Повний перезапуск програми може зайняти багато часу, що призведе до значного простою програмної системи.

Мікроперезавантаження може бути швидшим, якщо перезапустити лише частину програм системи, але вимагає повного аналізу стану програм для розуміння причин невдачі. Жоден із цих методів неефективно справляється з детермінованими помилками, оскільки вони можуть повторюватися після перезапуску.

Методи відкату до контрольної точки [5] можна використовувати у спосіб, подібний до перезапуску всієї програми, але може забезпечити швидший час перезапуску, оскільки перезапуски виконуються з контрольної точки. При такому використанні ці методи все ще не обробляють детерміновані помилки, оскільки ці помилки все одно виникатимуть після перезапуску.

Також були запропоновані інші варіанти використання контрольної точки-перезапуску в поєднанні з запуском кількох версій програми [6], які можуть пережити детерміновані помилки, якщо збої відбуваються незалежно. Однак вони несуть непо-

мірні витрати для більшості програм з точки зору розробки, підтримки та запуску кількох версій програми одночасно.

Автоматична генерація сигнатур для систем виявлення вторгнень у мережу [7, 8] захищає від вразливостей, фільтруючи вхідні дані для відсівання атак. Ключова проблема полягає в тому, що такі сигнатури досить сприйнятливі до помилкових спрацювань, особливо для поліморфних атак.

Крім того, було показано, що поліморфна поведінка надто різноманітна, щоб її можна було ефективно моделювати за допомогою сигнатур [9].

У проєкті Vigilante [10] покращено мережеву фільтрацію вхідних даних завдяки автоматичному створенню вхідних фільтрів на базі хоста. Фільтри на основі хоста пропонують покращену точність для виявлення семантично еквівалентних вхідних даних. На жаль, вони потребують специфічних для протоколу парсерів і не можуть працювати зі складними правилами, шифруванням і конкретним станом програми.

Проєкт Rx [11] використовує механізм перезапуску контрольної точки в поєднанні з механізмами для зміни середовища виконання з метою відновлення після помилок.

Однак попередні роботи [12] виявили, що понад 86% помилок програм не залежать від операційного середовища, є повністю детермінованими та повторюваними, і що відновлення, ймовірно, буде успішним лише за допомогою методів, пов'язаних із програмою.

У той час, як Rx розглядає ширші можливості для зміни середовища, включаючи відкидання шкідливих запитів на введення, на практиці виявилось неефективним через поліморфну поведінку [13]. Rx намагається замаскувати прояв несправностей для клієнта, але потребує використання проксі-сервера додатка з підтримкою протоколу, який має бути здатний відфільтрувати інформацію, таку як мітки часу, яка могла б заплутати клієнтську програму.

Використання проксі-сервера ускладнює використання все більшої кількості програм, які використовують шифрування. Rx вимагає змін ядра операційної системи, що є ще однією перешкодою для розгортання. Нарешті, Rx не вирішує проблеми узгодженості під час встановлення контрольних точок і перезапуску програм, що включають кілька процесів.

"Прибиральник" (sweeper) [14] поєднує механізм перезапуску контрольної точки Rx і проксі з VSEF. Якщо виникає помилка, Sweeper використовує аналіз "забруднення", щоб визначити вхідні дані, які призвели до збою, генерує вхідний фільтр для видалення цього та подібних майбутніх запитів, а потім повертається до попередньої контрольної точки та повторно відтворює вхідні дані. Оскільки Sweeper зменшує VSEF до використання для генерації вхідного підпису, він страждає від тих самих обмежень фільтрації введення, які описані раніше (поліморфізм і зашифрований трафік).

Обчислення, орієнтовані на "прийнятність" (acceptability) [15] просуває ідею про те, що поточні

зусилля з розробки програмного забезпечення можуть бути неправильно спрямованими, ґрунтуючись на спостереженні про те, що можна знехтувати певними регіонами програми без негативного впливу на загальну доступність системи.

Обчислення без збоїв [16] — це спекулятивна техніка відновлення, яка базується на компіляторі і вимагає вставку додаткового коду для роботи із записами в нерозподілену пам'ять шляхом віртуального розширення цільового буфера. Така можливість має на меті забезпечити більш надійну реакцію на помилку, ніж просто збій, хоча й із значними накладними витратами на продуктивність, які варіюються від 80% до 500% для різноманітних програм.

Вибіркова транзакційна емуляція (STEM), яка використовується в реактивній імунній системі [17] — це спекулятивна техніка відновлення, розроблена двома авторами, яка визначає функцію, у якій сталася помилка, а потім вибірково емулює цю функцію та, можливо, інші в більшому діапазоні, щоб повернути значення помилок у спробі відновлення після помилки. STEM використовує поняття віртуалізації помилок, щоб означати евристичне значення помилки від функції, у якій сталася помилка. Це дуже відрізняється від поняття віртуалізації помилок точки відновлення, яке використовується в системі, яке повторно використовує існуючий код обробки помилок у програмах і повертає значення на основі профілювання цих функцій для імітації поведінки системи в умовах контрольованих і очікуваних помилок. На відміну від STEM, наші пропозиції не потребують вихідного коду, працюють з багатопроцесорними та багатопоточними програмами, забезпечують суттєві покращення продуктивності системи.

Найбільш формалізований підхід до відновлення розподіленого програмного забезпечення присутній у роботах [18, 19], де на прикладі програмних моделей зорганізується управління станом програми у часі на основі дамів пам'яті. Якщо використати цей підхід до нашої задачі, можлива організація самовідновлення на низькому рівні (операційної системи, віртуальної машини тощо).

Тож **метою** статті є опис механізму самовідновлення паралельного програмного забезпечення з використанням бекапів стану програм на основі журналізації точок відновлення.

Забезпечення життєвого циклу програмних систем з самовідновленням

Система забезпечує архітектурну підтримку для самовідновлення програми за наявності непередбачених збоїв у повністю автоматичному режимі. Система постійно стежить за програмою на наявність збоїв і визначає стратегії, використовуючи точки відновлення для реагування на майбутні випадки таких самих або подібних збоїв.

Після вибору стратегії, система динамічно модифікує програму, використовуючи динамічну бінарну ін'єкцію, щоб вона могла виявити та оминувати ту саму помилку в майбутньому. Метою нашої системи є автоматичне створення тимчасового ви-

правлення певної проблеми, доки не стане доступним рішення користувача.

Рис. 1 ілюструє роботу системи на високому рівні. Перед розгортанням, програма профілюється для виявлення потенційних точок відновлення. Після завершення профілювання, програма розгортається у своєму робочому середовищі.

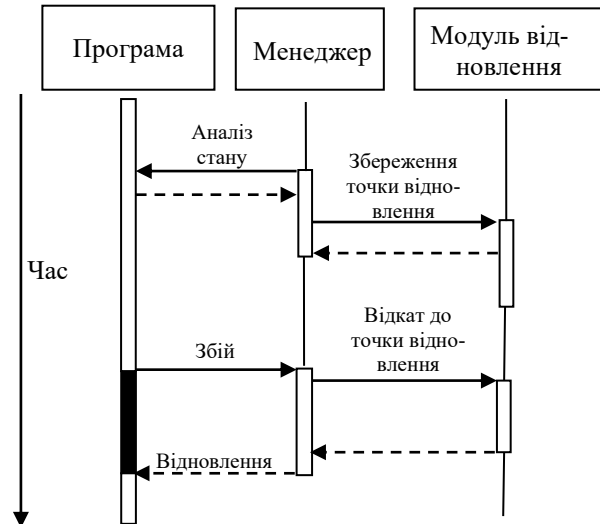


Рис. 1. Журналізація точки відновлення та відновлення системи після збою

Під час нормального виконання, система контролює програму за допомогою інструментальних засобів, які виявляють та звітують про неправильну поведінку програми та системи.

Крім того, система періодично перевіряє стан програми та веде журнал виконання (включаючи мережевий трафік), де періодично фіксує контрольну точку зі збереженням стану програмної системи.

Коли, під час виконання, виявляється помилка, стан останньої контрольної точки програми разом із журналом усіх вхідних даних, починаючи з цієї контрольної точки, передається в систему сортування, де аналізується помилка.

Потім система виконує автоматичний процес, метою якого є визначення відповідної точки відновлення, до якої програма може відновити виконання, при умові, що ця конкретна помилка не виявиться повторно. Протягом цього часу робоча система залишається вразливою до повторного виникнення помилки, що призводить до вікна вразливості, у якому програмі може знадобитися вдатися до повного перезапуску програми для відновлення нормального процесу виконання. Хоча нашій системі може знадобитися деякий час простою для фази аналізу, досвід показує, що це зазвичай декілька секунд, і вартість амортизується, оскільки вона виникає один раз за нову помилку.

Після вибору точки відновлення, система підтверджує, що вона придатна для розгортання, перевіряючи, що вона задовольняє трьом критеріям: живучість, правильність і продуктивність. Вибрана точка відновлення забезпечує живучість, якщо віртуалізація помилок у цей момент дозволяє програмі

вижити після повторення помилки.

Точка відновлення є дійсною, якщо вона не вносить семантичних помилок і якщо програма може правильно обслуговувати майбутні запити. Точка відновлення є справжньою, якщо продуктивність захисту не спричиняє значних витрат часу виконання.

Живучість перевіряється шляхом відтворення послідовності подій, які, очевидно, спричинили несправність. Правильність перевіряється за допомогою ретельного тестування, яке адаптовано для конкретної роботи програми.

Ефективність оптимізується за рахунок використання продуктивності як метрики при виборі точки відновлення. Щойно відповідну точку відновлення буде перевірено, система створює виправлення, яке динамічно застосовується до програмного забезпечення, поки програма виконується в робочій системі.

Патч створює точку відновлення всередині програми, щоб захистити програму від повторення конкретної помилки. Змінена програма запускати точку відновлення кожного разу, коли виконання досягне точки відновлення, і повертатиме свій стан до цієї точки, якщо помилка повториться.

Після відкату виконання до точки відновлення, віртуалізація помилок використовується для використання існуючих можливостей обробки помилок програми для ефективного усунення помилок.

Замість фільтрації певних вхідних даних, які можуть спричинити збої, патч захищає програму від збоїв, які можуть виникнути в певному місці програми.

Отриманий механізм відновлення не залежить від вхідних даних і, отже, захищений від ризиків, пов'язаних із поліморфізмом помилок/вхідних даних.

Структура системи з самовідновленням

Система забезпечує загальний механізм, який програми можуть використовувати для відновлення виконання за наявності помилок. Запропонований підхід можна розділити на наступні етапи офлайнових і онлайнових дій.

В автономному режимі додатки профілюються під час «хибних запусків», щоб побудувати модель поведінки додатків. Ці хибні прогони генеруються регресійними тестами, якщо вони доступні, або за допомогою методів фазингу [3,10], який підкреслює можливості обробки помилок програм. Дії полягають в тому, що існує набір тестованих програмістами точок програми, які регулярно використовуються для поширення «очікуваних» помилок.

У свою чергу, ці прикладні точки можна використовувати для відновлення після збоїв і таким чином підтримувати доступність системи.

Використовуючи цю модель, ми виділяємо місце розташування програми, які можна використовувати як потенційні точки відновлення. Наша архітектура в режимі он-лайн дозволяє використовувати різноманітні монітори несправностей. Після виявлення помилки, стан програми повертається до по-

передньо визначеного (точки відновлення), де програма змушена повертати помилку, імітуючи поведінку, що спостерігається під час помилкових запусків. Модулі системи, представлені на рис. 2:

набір засобів (зондів), який постійно відстежує програму на наявність помилок і бере на себе контроль щоразу, коли вони виявляються;

помилка віртуалізації активує компонент, який відповідає за визначення значень для введення в разі несправності;

виявлення точки відновлення активує компонент, який використовується для визначення потенційних точок відновлення за допомогою статистики та аналізу;

активатор, який використовує механізм перезапуску контрольної точки для фіксації стану програми та відкату до збереженого стану;

генератор патчів, який створює патчі з підтримкою точки відновлення для аплікації;

середовище тестування, в якому запропоновані патчі оцінюються згодом;

вставка засобів, які полегшують інсталяцію затверджених патчів у запуснені бінарні файли.

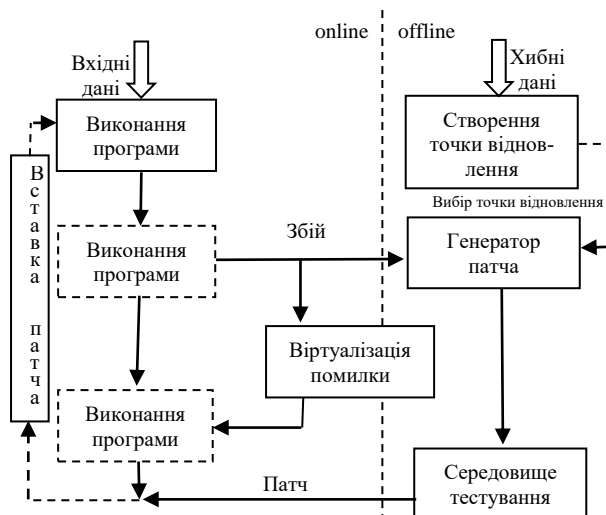


Рис. 2. Робота системи з самовідновленням

Усі компоненти розроблені для роботи в автоматичному режимі, щоб мінімізувати час реакції.

Призначення та віртуалізація точок відновлення

Визначення прийнятної точки відновлення має ключове значення для віртуалізації помилок. Це багато в чому визначає ступінь, ймовірність того, що програма продовжить роботу при несправності. Для виявлення відновлення використовуються два механізми: один статичний і один динамічний.

Динамічний аналіз є кращим механізмом для виявлення відповідних точок відновлення, оскільки він надає однозначне розуміння поведінки програми. Зокрема, метою є дізнатися, як реагує програма на «хибні вхідні дані» у контрольованих умовах і використовувати це знання, щоб зробити так, щоб раніше невидимі дефекти були прогнозованими. Існує набір перевірених програмістами точок про-

грам, які є регулярними та використовуються для обробки «очікуваних» помилок. Наприклад, можна побачити, як зазвичай поширюється програма помилки під час стрес-тестування тестами забезпечення якості.

Навчання на основі помилок використовувалося в машинному навчанні і згодом системи виявлення помилок, було використано для відновлення програмного забезпечення при помилках. Зокрема, ми інструментуємо програми шляхом вставки коду моніторингу в точках входу та виходу кожної функції, які можливі впровадження під час виконання [2]. Засоби записують обидва параметри функції - типи та значення, що повертаються програмою при помилці. З цих записів можна побудувати графіки викликів функцій разом із інформацією про тип повернення для кожної точки на графіку ("графіки відновлення"). Rescue-графи використовуються як програмні фрагменти на функціональному рівні деталізації, які, у свою чергу, можна використовувати щоб ізолювати контрольний потік несправностей та визначити можливе відновлення.

Статичний аналіз використовується для посилення результатів описаних методів динамічного аналізу. Зокрема, ми використовуємо статичний аналіз для полегшення віртуалізації помилок і виявлення точки відновлення. Для віртуалізації помилок статичний аналіз може допомогти визначити відповідні значення повернення помилок через перевірку коду та зворотні фрагментації програми. Детально досліджуються шляхи, що виникають при прояві несправності, де вразлива функція знаходиться під час виклику. На цьому етапі перевіряється, як значення функції, що повертається, використовується в коді обробки. Це забезпечує розуміння того, як ми можемо використовувати під час помилки механізм віртуалізації.

Наприклад, коли функція повертає значення оператора керування, за яким слідує оператор виходу – це створює умови для використання цього значення як відповідного відновленого значення під час віртуалізації помилок. Протягом цього процесу ми також приділяємо увагу фрагментам програми для будь-яких проблемних випадків віртуалізації помилок, включаючи введення-виведення, яке виконується разом із використанням глобальних змінних та існування коду обробки сигналів.

Опишемо основну концепцію віртуалізації помилок за допомогою точок відновлення. Віртуалізацію помилок можна зробити такими кроками:

- зберегти стан програми в точці відновлення;
- контролювати систему на наявність помилок;
- коли виникає помилка, скасувати зміни стану, зроблені функцією, аж до точки відновлення;
- після відновлення виконання до точки відновлення.

Компоненти виявлення несправностей можна розглядати як чорний ящик, якому потрібно лише повідомити про виникнення несправності. На додаток до стандартної обробки помилок операційної системи, можна використовувати додаткові механізми для виявлення помилок пам'яті.

Існує ряд доступних компонентів виявлення несправностей, які можуть виявляти помилки пам'яті (наприклад, ProPolice [9] і TaintCheck [12]) і деякі, які виявляють порушення основних політик безпеки [1,11].

Проведення експериментів з прототипами системи самовідновлення

Експериментальну симуляцію роботи системи було реалізовано для операційної системи Linux. Було використано стандартні утиліти та модулі ядра ядра Linux, які забезпечують віртуальне середовище виконання з точками перезавантаження та відтворення журналу, а також Dyninst 10.1 для аналізу паралельного коду під час виконання.

Використовуючи цей прототип, було оцінено ефективність системи щодо реальних помилок і стандартних робочих навантажень для низки популярних багатопроцесних і багатопоточних серверних програм. Для всіх експериментів процес був повністю автоматизованим, за винятком генерування інформації профілю та ініціювання помилки. Процес профілювання має відбуватися один раз для кожної програми (або надаватися як частина тестового пакету). Усі експерименти проводилися на машинах з процесорами Intel Core i7, 8/16, 2,9-4,8 ГГц і 16 ГБ оперативної пам'яті, з підключенням через з'єднання Ethernet 1 Гбіт/с.

Сервери та клієнти працювали на окремих машинах. Оцінка функціонування системи у вирішенні помилок відбувалася за трьома напрямками: живучість, правильність і продуктивність. Живучість перевіряла здатність системи підтримувати доступність служби за наявності збою програмного забезпечення, спричиненого помилкою.

Система виявляла збої та автоматично починала процес відновлення. Після відновлення, сервер перевіряв збої, які могли бути спричинені нашим механізмом, та продовжував правильно обслуговувати запити. Оскільки можливо, що механізм відновлення вносить побічні ефекти, перевірка на правильність вихідних даних сервера проходила і після відновлення.

Потім проводили оцінку продуктивності системи з точки зору як повних системних накладних витрат на робочому сервері, так і часткової перевірки компонентів системи.

Було використано п'ять помилок для трьох популярних програм: Apache (версія 2.2), MySQL (8.0), і PostgreSQL (12.14). Серед помилок були: порушення прав доступу до пам'яті, переповнення буфера, некоректні вхідні дані. Також, для порівняння ми заклали зовнішні помилки для Linux. Дослідження виконано для програм із відкритим вихідним кодом.

Помилки генерувалися під час виконання тесту, паралельно з цим програма відстежувалася, щоб перевірити її здатність успішно пройти контрольний тест. Якщо тест завершувався, виконувалася перевірка живучості та оцінка продуктивності. Програма перевірялася на правильність або шляхом перевірки результатів порівняльного тесту, або через додаткові тести, які перевіряли і порівнювали вихідні дані з очікуваним набором результатів. Для кожної помилки фіксувалася глибина відновлення та значення відновлення: відстань між несправністю і точкою відновлення та значення віртуалізації помилки, яке потім використовувалася для поширення помилок.

Результати часової оцінки різних етапів процесу підтримки відновлення представлено на рис. 3.

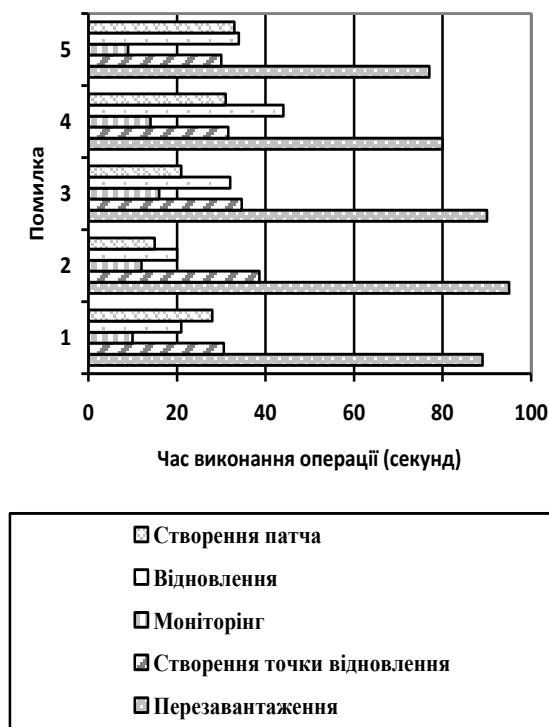


Рис. 3. Оцінка часу виконання операцій при виявленні помилки

Було реалізовано наступні помилки:
(1) переповнення буферу (Apache);
(2) невірні вхідні дані (Apache);
(3) порушення прав доступу до даних (MySQL);
(4) переповнення буферу (MySQL);
(5) невірні вхідні дані (PostgreSQL).

Щоб оцінити швидкість реагування системи на створення виправлення для нововиявленої помилки, вимірювався загальний час, необхідний для переходу від помилки до виправлення. Іншими словами, від моменту першого виявлення несправності у виробничій системі до динамічного застосування виправлення. Однак, треба відмітити, що такі процеси, як моніторинг та створення точки відновлення може виконуватись паралельно з роботою системи. Створення патча для зафіксованої помилки виконується один раз та зберігається у системі. Таким чином, коли помилка трапляється наступного разу, система виконує тільки відновлення.

Рис. 3 показує середній час у секундах для створення робочого, перевіреного патча для кожної з помилок. Вказується загальний час, необхідний для створення, тестування та застосування патча. Загальний час розбивається на дві частини: час, необхідний для підключення до запущеного процесу та введення точки відновлення за допомогою інструментів системи, і тестовий час, необхідний для виконання тесту живучості, правильності та продуктивності в середньому для успішної точки відновлення. Зауважимо, що час, необхідний для створення графіка відновлення для конкретної помилки, незначний, враховуючи, що траса стека була глибиною менше 5 функцій у всіх випадках.

Для кожної помилки ми оцінювали ефективність відновлення після несправності. Зокрема, час відновлення стану програми до точки відновлення після виявлення несправності. Помилка виникала, коли програма була зайнята виконанням заданих тестів для вимірювання відновлення під навантаженням.

Ми порівняли час відновлення системи із часом перезапуску всієї програми після збою, у якому ми виміряли час, що минув від запуску програми до моменту, коли вона стане працездатною та готовою обслуговувати запити. Повний перезапуск програми не обов'язково дозволяє відновлення, але це може скинути сервер, щоб він міг обслуговувати майбутні запити, навіть якщо він не дозволяє завершити роботу навантаження, яке виконувалося на момент помилки. Хоча він не забезпечує такого ж рівня живучості, як наша система, він забезпечує відновлення. Як видно з рис. 3, перезавантаження призводить к значним витратам часу в порівнянні з етапами запропонованого механізму самовідновлення. Щоб виміряти реалістичний час перезапуску додатка, ми тестували перезапуск програми за допомогою реальних навантажень. Наприклад, для PostgreSQL ми вимірювали час, необхідний для перезапуску програми, коли в ній попередньо завантажено набір даних Wisconsin, для Apache визначили середній час, необхідний для відновлення виконання до точ-

ки відновлення, або, іншими словами, відкат до точки колишнього виконання.

Висновки

Розроблена система надає можливість створення нового програмного забезпечення для самовідновлення програмних систем, зокрема для відновлення від програмних збоїв у серверних програмах. Точки відновлення – це місця, які визначені в існуючому коді програми, де обробка помилки виконується щодо заданого набору передбачених або виявлених збоїв.

Існуючі методи перевірки достовірності можна використовувати для створення відомих хибних вхідних даних програм, щоб визначити потенційні точки відновлення.

При першому виявленні несправності, система використовує репліку (стан програми у точці відновлення) програми, щоб визначити, які точки відновлення можна використовувати найбільш ефективно для відновлення майбутнього виконання програми.

Після того, як система перевірить, чи створено виправлення, яке усуває несправність, вона динамічно виконує відновлення. Якщо помилка виникає знову, система повертає програму до точки відновлення та використовує власний вбудований програмний код обробки помилок для відновлення після несправності та правильного виправлення внутрішнього і зовнішнього стану.

У роботі показана ефективність на кількох серверних програмах, включаючи веб сервер та бази даних.

Експериментальні результати, як з реальними помилками, так і з ін'єкціями штучних помилок показали, що запропонований механізм може бути використано для відновлення виконання у більшості розглянутих випадків з невеликими операційними витратами. Використання неоптимізованого прототипу, який повністю автоматизує процес відновлення програмного забезпечення, займає всього десятки секунд, а замовлення на відновлення швидше, ніж поточне розгортання усунення помилок людиною. Крім того, не вимагається наявність вихідного коду програми. Кінцевим результатом є автоматичне відновлення програмних служб з невідомих і непередбачених програмних збоїв.

Було окреслено межі віртуалізації помилок за допомогою точок відновлення, запропоновано нову техніку самовідновлення програмного забезпечення для відновлення після помилок використано цільові системи для виявлення помилок програмного забезпечення в програмах, викликаних втручанням з метою використання вразливостей програмного забезпечення, і отримання кінцевого стека викликів. Це узгоджується з потенційним набором точок відновлення шляхом відкату та повторення виконання з помилкою, щоб визначити, яку точку відновлення можна використовувати для відновлення після помилки.

Система динамічно виправляє запущену роботу програму до точки самоперевірки в точці відновлення та, якщо виникає помилка, повертається до

точки відновлення і повертає відоме значення, якщо використовується для відповіді на неправильний вхід, який використовується власними вбудованими програмами у механізмах обробки помилок для відновлення після несправності.

Результати роботи можуть набути поширення та застосовуватись у розробці більшості програмних, інформаційних систем з метою автоматизації відновлення, налагодження та експлуатації сучасних комп'ютерних та хмарних систем.

СПИСОК ЛІТЕРАТУРИ

1. Волк М.О., Лунічкін О.Г. Комп'ютерні системи з самовідновленням. Системи управління, навігації та зв'язку, 2022, випуск 1(67), с. 48-51
2. Sullivan, M., Chillarege, R.. Software Defects and Their Impact on System Availability—A Study of Field Failures In Operating Systems. In Proceedings of the 21st International Symposium on Fault-Tolerant Computing (FTCS-21), June 2021. pages 2–9,
3. Kolettis, N., Fulton, N. D. Software Rejuvenation: Analysis, Module and Applications. In Proceedings of the 25th International Symposium on Fault-Tolerant Computing (FTCS- 25), pages 381–395, June 2019.
4. Candea, G., Fox, A.. Crash-Only Software. In Proceedings of the 9th Workshop on Hot Topics in Operating Systems (HotOS IX), pages 12–20, May 2013.
5. King, S. T., Dunlap, G. W., Chen, P. M.. Debugging Operating Systems With Time-Traveling Virtual Machines. In Proceedings of the 2015 USENIX Annual Technical Conference (USENIX 2015), pages 1–15, Apr. 2015.
6. Bressoud, T. C., Schneider F. B. Hypervisor-Based Fault Tolerance. In Proceedings of the 15th ACM Symposium on Operating Systems Principles (SOSP 1995), pages 1–11, Dec. 1995.
7. Paxson, V.A System For Detecting Network Intruders In Real-Time. Computer Networks, 31(23-24):2435–2463, Dec. 2019.
8. Norton, M., Roelker, D. Snort 2.0 Protocol Flow Analyzer. Sourcefire White Paper, Apr. 2014.
9. Song, Y., Locasto, M. E., Stavrou, A., Keromytis, A. D., Stolfo, S. J. On the Infeasibility of Modeling Polymorphic Shellcode. In Proceedings of the 24th ACM Conference on Computer and Communications Security (CCS 2017), pages 541–551, Oct. 2017.
10. Costa, M., Crowcroft, J., Castro, M., Rowstron, A., Zhou, L., Zhang, L., Barham., P. Vigilante: End-To-End Containment of Internet Worms. In Proceedings of the 20th ACM Symposium on Operating Systems Principles (SOSP 2005), pages 133–147, Dec. 2015.
11. Qin, F., Tucek, J., Sundaresan, J., Zhou, Y., Rx: Treating Bugs As Allergies—A Safe Method To Survive Software Failures. In Proceedings of the 20th ACM Symposium on Operating Systems Principles (SOSP 2015), pages 235–248, Oct. 2015.
12. Chandra, S. An Evaluation of the Recovery-Related Properties of Software Faults. PhD thesis, University of Michigan, Sept. 2000.
13. Song, Y., Locasto, M. E., Stavrou, A., Keromytis, A. D., Stolfo S. J. On the Infeasibility of Modeling Polymorphic Shellcode. In Proceedings of the 24th ACM Conference on Computer and Communications Security (CCS 2017), pages 541–551, Oct. 2017.
14. Tucek, J., Newsome, J., Lu, S., Huang, C., Xanthos, S., Brumley, D., Zhou, Y., Song., D. Sweeper: A Lightweight End-To-End System For Defending Against Fast Worms. In Proceedings of the 2nd European Conference on Computer Systems (EuroSys 2017), pages 115–128, Mar. 2017.
15. Demsky, B., Rinard., M. Automatic Detection and Repair of Errors In Data Structures. In Proceedings of the 18th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA 2003), pages 78–95, Oct. 2021.
16. Rinard, M., Cadar, C., Dumitran, D., Roy, D. M., Leu, T., William J., Beebe, S. Enhancing Server Availability and Security Through Failure-Oblivious Computing. In Proceedings of the 6th Symposium on Operating Systems Design and Implementation (OSDI 2014), pages 303–316, Dec. 2014.
17. Sidiroglou, S., Locasto, M. E., Boyd, S. W., Keromytis, A. D. Building A Reactive Immune System For Software Services. In Proceedings of the 2015 USENIX Annual Technical Conference (USENIX 2015), pages 149–161, Apr. 2015.
18. Волк М.О. Журналізація станів програмних розподілених моделей та її використання в оптимістичних алгоритмах синхронізації. Збірник наукових праць Харківського університету Повітряних Сил. 2010, випуск 1 (23). С.104–107.
19. Рубан І.В., Волк М.О., Рісукін М.В. Метод самовідновлення розподіленого програмного забезпечення в гетерогенних комп'ютерних системах. Телекомунікаційні та інформаційні технології. 2019. № 3 (64), с. 17-23

Received (Надійшла) 14.03.2023

Accepted for publication (Прийнята до друку) 02.05.2023

Self-healing computer systems

Maksym Volk, Maksym Hora, Vladyslav Labazov, Andriy Mishchenko, Anton Barsukov, Vladyslav Holetz

Abstract. The article deals with the self-healing system of parallel software using journaling of recovery points. Self-healing is a necessary feature of modern software, which provides the possibility of automatic detection, diagnosis and restoration of system performance. The main stages of recovery are storing the state of programs (journaling) in recovery points, monitoring the state of programs to detect errors, creating patches, restoring the state of programs to the corresponding recovery point. The structure of the system is proposed in the work, the algorithm of its functioning is described; issues of destination and virtualization of restore points are discussed; a description of the experimental software system and its application to recovery of common software systems is given. The results can become widespread and be used in the development of most software and information systems for the purpose of automating the recovery, debugging and operation of modern computer and cloud systems.

Keywords: computer system, self-healing, fault, backup, recovery point.

С. Ю. Гавриленко¹, В. Д. Зозуля¹, В. В. Омельченко²

¹ Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна

² Харківський національний університет радіоелектроніки, Харків, Україна

ДОСЛІДЖЕННЯ МЕТОДІВ ПІДВИЩЕННЯ ЯКОСТІ КЛАСИФІКАЦІЇ НА НЕЗБАЛАНСОВАНИХ ДАНИХ

Анотація. Предметом дослідження є методи балансування вихідних даних. Метою статті є підвищення якості систем виявлення вторгнень у комп'ютерні мережі за рахунок використання методів балансування класів. Завдання: дослідити методи балансування класів та розробити метод класифікації на незбалансованих даних для підвищення рівня мережевої безпеки. Використовуваними методами є: методи штучного інтелекту, машинного навчання. Отримано такі результати: Досліджено методи балансування класів, які базуються на технології Undersampling, Oversampling та їх комбінації. Для подальшого дослідження обрано такі методи: SMOTEENN, SVM SMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE. У якості вихідних даних використано набір UNSW-NB 15, який містить інформацію про нормальне функціонування мережі та під час вторгнень. У якості базового класифікатора використано дерево рішень на основі CART (Classification And Regression Tree) алгоритму. За результатами досліджень отримано, що використання методу SMOTEENN надає можливість підвищити якість виявлення вторгнень у функціонування мережі. **Висновки.** Наукова новизна отриманих результатів полягає у комплексному використанні методів балансування даних та методу класифікації даних на основі дерева рішень для виявлення вторгнень у комп'ютерні мережі, що дозволило зменшити кількість помилок II роду.

Ключові слова: машинне навчання, мережева безпека, балансування даних, системи виявлення вторгнень, Undersampling, Oversampling, SMOTEENN, SVM SMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE.

Вступ

Системи виявлення вторгнень (СВВ) є одним із механізмів аналізу поведінки комп'ютерної мережі та інструментом моніторингу та спостереження підозрілої активності, аналізу ресурсів мережі, ідентифікації аномальних подій у мережі.

Функціонування мережі характеризується великим обсягом показників, що призводить до наявності труднощів з вибором найбільш інформативних показників та розробки моделей виявлення вторгнень.

Моделі виявлення вторгнень, зазвичай, базуються на процесі класифікації – групуванню об'єктів за певними ознаками та встановленні ієрархії між ними на основі спільних характеристик. Завдання класифікації найчастіше спирається на методи машинного навчання з учителем, та використовує розмічені дані для навчання. Це означає, що якість моделі багато в чому залежить від вихідних даних, на основі яких вона «навчається», в тому числі від збалансованості даних [1].

Проблема незбалансованих даних відноситься до ситуації, коли класи в цільовій змінній не однаково представлені в наборі даних і є однією з найскладніших проблем алгоритмів машинного навчання.

Дисбаланс може бути властивістю проблемної галузі коли присутність одного класу може домінувати над іншими класами. Це може бути пов'язано з тим, що часто неможливо або важко просто зібрати більше зразків міноритарного класу, щоб покращити розподіл класів. При цьому, потрібна модель, щоб класифікувати дані. Прикладами областей, які, зазвичай, оперують незбалансованими класами є: виявлення шахрайства, спаму, аномалій, викидів, відтоку клієнтів, виявлення рідкісних захворювань та ін. Ситуації з незбалансованими даними також досить поширені в кібербезпеці. Так, при виявленні мережних вторгнень

кількість доброякісного трафіку на порядки перевищує кількість шкідливого. Іншим прикладом є боротьба з інсайдерськими загрозами, коли кількість даних про звичайну поведінку на порядок перевищує кількість даних про зловмисну поведінку.

Об'єктом дослідження є процес балансування вихідних даних з метою підвищення рівня мережевої безпеки. **Предметом дослідження** є методи балансування (семплінгу) вихідних даних.

Постановка проблеми та огляд наукових публікацій

Незбалансовані класифікації створюють проблеми для прогнозного моделювання тому, що більшість класифікаційних алгоритмів машинного навчання були розроблені на основі припущення про кількість прикладів для кожного класу, тобто вони не враховують класового розподілу/пропорції або балансу класів. Це призводить до побудови упереджених моделей із поганою прогностичною ефективністю, особливо для класу меншості. Модель повністю ігнорує клас меншості і маркує всі об'єкти класифікації мітками мажоритарного класу [2].

Крім того, як правило, клас меншості важливіший, і тому проблема більш чутлива до помилок класифікації для меншості, ніж для більшості. Це призводить до різної ціни помилок першого та другого роду при класифікації даних. Особливо це актуально для систем виявлення вторгнень, які, зазвичай, містять невелику кількість прикладів виявлення вторгнень і для яких помилково-негативна класифікація (помилка II роду) може мати фатальні наслідки [3].

Основна мета врівноваження класів полягає в тому, щоб збільшити частоту класу меншості, або зменшити частоту класу більшості. Це робиться для того, щоб отримати приблизно однакову кількість примірників для обох класів.

Один із підходів до вирішення зазначеної проблеми – застосування різних стратегій балансування класів. Виділяють наступні підходи балансування даних: Undersampling, Oversampling та їх поєднання (Over+Undersampling) [4].

У випадку Undersampling видаляють частину прикладів мажоритарного класу. При цьому зберігаються пропорції між класами, але зменшується кількість зразків у класі, який має більшість екземплярів. Використовують такі методи Undersampling: Random undersampling, Condensed nearest neighbors (CNN) undersampling, Tomek Links method, Edited nearest neighbors (ENN), One-sided selection (OSS, combines Tomek Links and CNN), Neighborhood cleaning rule (NCR, combines CNN and ENN) [5,6]

Random undersampling це техніка, яка балансує дані шляхом випадкового видалення вибірок із основного класу. Недоліком даного методу є можливість втрати важливої інформації в процесі видалення вибірок.

Методи Condensed nearest neighbors (CNN) та Edited nearest neighbors (ENN) засновані на алгоритмі найближчого сусіда, який використовує відстань між точками для визначення того, які об'єкти повинні бути видалені. Алгоритм CNN ітераційно додає до нового набору даних ті об'єкти, які є найближчими сусідами до об'єкту іншого класу. ENN працює, видаляючи з набору даних будь-які зразки, які не узгоджуються з їхніми найближчими сусідами. Цей процес допомагає усунути шум і покращити якість даних шляхом видалення будь-яких викидів або неправильно позначених зразків. ENN часто використовується в поєднанні з іншими методами попередньої обробки даних для покращення загальної продуктивності алгоритмів машинного навчання [7]. Основна відмінність між CNN і ENN полягає в тому, що CNN працює шляхом конденсації класу більшості в репрезентативну підмножину даних, тоді як ENN видаляє приклади з класу більшості, які знаходяться поблизу межі прийняття рішення. Ще одна відмінність полягає в тому, що CNN аналізує дані один раз, тоді як ENN використовує кілька проходів.

Головним недоліком цих алгоритмів, як і більшості методів балансування, є можливість надмірної підгонки навчальних даних, що призводить до пере-навчання моделі

Tomek Links (TL) method базується на пошуку пар екземплярів протилежних класів, які є своїми найближчими сусідами та видаляє їх. Це не тільки вирівнює баланс даних, а й робить межі класів більш чіткими і вираженими, що підвищує якість класифікації. Метод зв'язків Томека також використовується для видалення точок даних із шумом.

One-sided selection (OSS) поєднує метод CNN та метод Томека. На першому кроці застосовується правило Condensed nearest neighbors, а на другому видаляються всі мажоритарні спостереження, які беруть участь у зв'язках Томека. Таким чином, видаляються великі згустки мажоритарних спостережень, а потім область простору зі згущенням міноритарних очищається від мажоритарних, які створюють ефект шуму на межах класів і заважають їх розпізнаванню.

Neighborhood cleaning rule (NCR) класифікує усі спостереження за правилом трьох найближчих сусідів (3-NN). Потім видаляються такі приклади мажоритарного класу, які правильно розпізнані або є сусідами міноритарних прикладів або були неправильно класифіковані. Перевага даного підходу в тому, що збільшення області сусідства дає змогу краще «очищати» дані від шумів.

У випадку Oversampling збільшують число екземплярів міноритарного класу. Найбільш поширеними є такі методи: Random Oversampling, алгоритм SMOTE, ASMO, ADASYN[8].

Алгоритм Random Oversampling балансує дані за рахунок дублювання даних міноритарного класу. Це робить розподіл класів більш збалансованим, але все не відображає розподіл класів у реальному світі. Це може негативно вплинути на якість прогнозів моделі, особливо для рідкісних або важливих подій [9].

В основі алгоритму SMOTE лежить ідея генерації деякої кількості штучних спостережень, які були б схожі на спостереження, що є в міноритарному класі, але при цьому не дублювали їх. Для створення нових спостережень вибирають об'єкт міноритарного класу b та використовуючи метод найближчого сусіда знаходять k його найближчих сусідів a . Надалі алгоритм формує нові спостереження, вибираючи випадково місце на прямий між об'єктом b та кожним із його сусідів a . Завдяки такому підходу штучні спостереження завжди формуватимуться поблизу існуючих об'єктів, але не будуть збігатися з ними. Алгоритм SMOTE дозволяє задавати кількість спостережень, яку необхідно штучно згенерувати. При цьому ступінь подібності прикладів a і b можна регулювати шляхом зміни числа k його найближчих сусідів: чим менше k , тим вище буде ступінь подібності. Недоліком цього підходу є те, що алгоритм просто збільшує щільність спостережень у областях згущення об'єктів. У випадку, коли приклади міноритарного класу розташовані рівномірно, це призводить до перемішування класів, що ускладнює класифікацію [10-12].

Алгоритм ASMO (Adaptive Synthetic Minority Oversampling) є модифікацією алгоритму SMOTE і надає можливість генерувати дані у всіх областях незалежно від згущеності об'єктів.

Ще одним недоліком алгоритму SMOTE є те, що він для кожного прикладу міноритарного класу створює одну і ту ж кількість штучних прикладів. Це не зовсім оптимально, оскільки не всі приклади однаково прості в навчанні. ADASYN є адаптивним алгоритмом, який створює синтетичні зразки для класів, які мають недостатньо представницьких зразків у навчальних даних. Як наслідок, підхід ADASYN покращує навчання щодо розподілу даних двома способами: зменшуючи зміщення, викликане дисбалансом класів, і адаптивно зміщуючи межі класифікаційних рішень у бік складних прикладів.

Алгоритм Borderline SMOTE (Synthetic Minority Oversampling Technique) є також модифікацією алгоритму SMOTE. Він фокусується саме на прикладах класу меншості, які розташовані на лінії розмежування класів. Для таких прикладів логічно генеру-

вати більше штучних спостережень, щоб зробити межу класу більш чіткою. BorderlineSMOTE генерує синтетичні приклади для цих граничних випадків замість надмірної вибірки всіх вибірок класу меншин, тим самим покращуючи продуктивність узагальнення класифікатора [13-14].

Алгоритм KMeansSMOTE спрямований на зменшення шумових синтетичних точок, які генерують інші методи балансування. Спочатку дані кластеризуються та визначаються кластери з високою частотою більше п'ятдесяти відсотків або визначену користувачем) вибірок міноритарного класу. Потім до кожного із обраних кластерів застосовується алгоритм SMOTE та генеруються нові об'єкти. Кількість цих згенерованих балів залежить від розрідженості класу меншості в кластері; чим більше розрідженість, тим більше нових об'єктів.

Алгоритм SVMSMOTE фокусується також на збільшенні об'єктів міноритарного класу вздовж межі розмежування класів. Він контрастує з методом K-Means, який ми бачили раніше, але узгоджується з варіантом Borderline). При цьому, для кожного з опорних векторів ми знаходимо його K-найближчих сусідів і створюємо вибірки вздовж лінії, що з'єднує опорний вектор і найближчих сусідів, використовуючи або інтерполяцію, або екстраполяцію. Якщо менше половини найближчих сусідів належать до класу більшості, ми виконуємо екстраполяцію. Це допомагає розширити територію класу меншості до території більшості. Якщо ні, виконуємо інтерполяцію. Ідея полягає в тому, що оскільки більшість сусідів належать до класу більшості, ми натомість консолідуємо поточну територію класу меншості.

SMOTEENN – це алгоритм, який поєднує в собі алгоритми SMOTE (Synthetic Minority Oversampling Technique) і Edited Nearest Neighbor (ENN) для генерації екземплярів міноритарного класу, а потім для видалення зашумлених даних.

Процес алгоритму SMOTEENN можна описати так:

1. Вибрати випадковим чином об'єкт x_j даних із класу меншості.
2. Обчислити відстань d_i між об'єктом x_j та його k найближчими сусідами.
3. Помножити отримане значення відстані на випадкове число $a \in [0-1]$ та отримати координати нового об'єкту. Додати об'єкт до класу меншості як синтетично-згенерований екземпляр даних.
4. Повтори кроки 1–3, доки не буде досягнуто бажаної пропорції класу меншості.
5. Для видалення зашумлених даних, встановити параметр k – кількість найближчих сусідів (за замовченням $k=3$).
6. Для кожного об'єкту даних x_j , знайти k -найближчих сусідів серед інших спостережень у наборі даних.
7. Визначити клас більшості для знайдених сусідів m .
8. Якщо клас спостереження x_j та клас більшості m знайдених сусідів спостереження відрізняються, то спостереження x_j та його k -найближчих сусідів видаляються з набору даних.

9. Повторити кроки 6-8 доки не буде досягнуто бажаної пропорції кожного класу.

Перевагою даного алгоритму є можливість збалансувати розподіл класів, одночасно зменшуючи ризик перенавчання.

За результатами аналізу методів балансування класів для подальшого дослідження вибрано такі методи: SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE.

Дослідження методів балансування класів

У цій роботі для дослідження використано набір даних UNSW-NB 15, який був створений інструментом IXIA PerfectStorm у лабораторії Cyber Range Австралійського центру кібербезпеки (ACCS) та містить інформацію про нормальне функціонування мережі та під час синтетичних вторгнень. Для захоплення 100 ГБ необробленого трафіку (наприклад, pcap файлів) використовується інструмент Tcpdump [15].

Модифікований набір даних містить 39 атрибутів, які включають потік між хостами і перевірку мережевих пакетів для розрізнення наявності або відсутності дев'яти типів атак: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms. Кількість записів у навчальному наборі складає 175 341 запис, а в тестовому наборі – 82 332 записи різних типів, атаки та нормального режиму [15].

Для дослідження ефективності використання обраних методів балансування класів, у середовищі Collab Python розроблено їх програмні моделі.

Для оцінки якості моделі до та після виконання балансування класів використано модель дерев рішень на основі CART (Classification And Regression Tree) алгоритму. Дерево рішень – це метод представлення вирішальних правил у визначеній ієрархії, що включає в себе елементи двох типів: вузли (node) і листя (leaf). Вузли містять в собі вирішальні правила і виконують перевірку прикладів відповідно до обраного атрибуту. Вибір атрибуту відбувається на підставі критерію Джині:

$$I = \sum_{k=1}^C p_k (1 - p_k),$$

де p_k – це частка зразків, що належить до класу C для конкретного вузла.

Основний принцип методу полягає в послідовному, рекурсивному розбитті навчальної множини на підмножини із застосуванням вирішальних правил у вузлах. Для оцінки якості моделі було використано такі характеристики якості моделі: Точність (Accuracy), Влучність (Precision), повнота (Recall) та міра F_1 (F_1 score) [16].

Влучність і повнота характеризують різні сторони якості класифікатора:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN},$$

де TP – кількість вірно класифікованих позитивних подій, FP (помилка I роду, false alarm) – кількість невірно класифікованих позитивних подій, FN

(помилка II роду, miss target) – кількість невірно класифікованих негативних подій.

Чим вище влучність, тим менше помилкових спрацьовувань (помилка I роду). Чим вище повнота, тим менше помилкових пропусків (помилка II роду). Повнота демонструє здатність алгоритму виявляти клас взагалі, а Влучність – здатність відрізнити цей клас від інших класів.

Accuracy, на відміну від *Precision* і *Recall*, залежить від співвідношення класів і тому рідко застосовується за умови незбалансованих вибірок.

Міра F_1 – поєднує *Precision* і *Recall* в агрегований критерій якості і є їх середнім гармонійним:

$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2TP}{2TP + FP + FN}$$

Оскільки нашим завданням є виявлення усіх позитивних прикладів (прикладів вторгнення в систему), то формальна постановка дослідження це максимізація характеристик якості моделі: *Recall* та F_1 score.

За результатами дослідження було проаналізовано показники якості моделі на основі дерев рішень тільки для класу шкідливого програмного забезпечення, яке моделює типи атак: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms.

Отримано середнє значення показників якості: *Recall* та F_1 score за умови використання різних методів балансування класів. Результати досліджень наведено на рис.1-2. Як видно із результатів *Recall* та F_1 score на незбалансованих даних є низькою і дорівнює приблизно 61%.

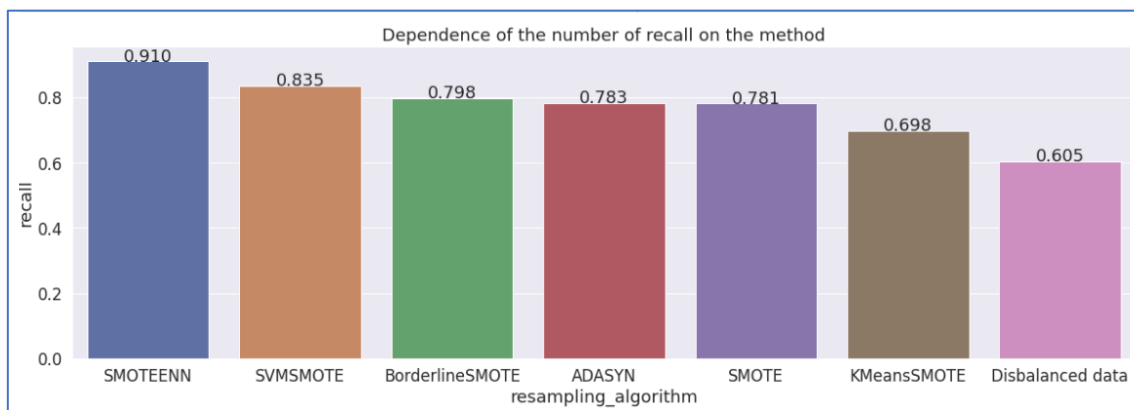


Рис. 1. Залежність повноти класифікації від методу балансування класів

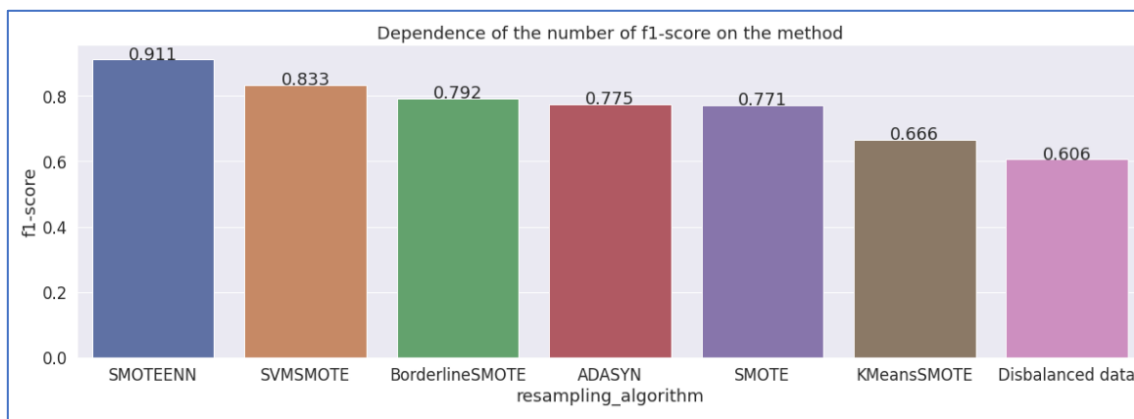


Рис. 2. Залежність міри F_1 класифікації від методу балансування класів

У всіх випадках застосування методів балансування дозволило отримати більш високу якість класифікації, ніж на незбалансованих даних. Найкращі результати отримано для методу SMOTEENN. Використання цього методу на етапі попередньої обробки даних для балансування класів дозволило покращити якість на 30 % при вирішенні завдання ідентифікації вторгнень в функціонування комп'ютерних мереж.

Висновки

Таким чином, у даній статті вирішено завдання підвищення рівня мережевої безпеки за рахунок

балансирування вхідних даних моделі виявлення вторгнень.

Наукова новизна отриманих у даній статті результатів полягає у комплексному використанні методу балансування даних SMOTEENN та методу класифікації даних на основі дерев рішень з метою підвищення якості виявлення вторгнень у комп'ютерні мережі, що дозволило зменшити кількість помилок II роду.

Досліджено методи балансування класів, які базуються на технології Undersampling, Oversampling та їх комбінацій.

Для подальшого дослідження обрано такі методи: SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeans-SMOTE.

У якості вихідних даних використано набір UNSW-NB 15, який містить інформацію про нормальне функціонування комп'ютерної мережі та під час вторгнень.

Набір даних містить інформацію про потік між хостами і перевірку мережевих пакетів для розрізнення наявності або відсутності дев'яти типів атак: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms.

У якості базової моделі класифікації даних використано дерево рішень. Розроблено програмне забезпечення мовою Python у середовищі Collab.

Результати досліджень методів балансування класів показали, що найбільш якісним є використання методів SMOTEENN, SVMSMOTE.

За результатами досліджень отримано, що комплексне використання методу балансування даних SMOTEENN та методу класифікації даних на основі дерев рішень надає можливість підвищити на 30% кількість виявлених вторгнень у функціонування мережі.

СПИСОК ЛІТЕРАТУРИ

1. S. Gavrylenko, V. Chelak, S. Semenov. Development of Method for Identification the Computer System State based on the Decision Tree with Multi-Dimensional Nodes. *Radio Electronics, Computer Science, Control (RECSC)*. 2022, V.4, pp.113-121.
2. Krawczyk, Bartosz. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 2016, V.5, pp.221-232.
3. C. Wheelus, E. Bou-Harb and X. Zhu. Tackling Class Imbalance in Cyber Security Datasets. *2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA*. 2018, pp.229-232.
4. Abdi L, Sattar H. To combat multi-class imbalanced problems by means of over-sampling techniques. *IEEE Trans Knowl Data Eng*. 2016, V.28, pp.238–251.
5. Will Badr. Having an Imbalanced Dataset? Here Is How You Can Fix It. [Електронний ресурс] – Режим доступу: <https://towardsdatascience.com/having-an-imbalanced-dataset-here-is-how-you-can-solve-it-1640568947eb>.
6. Jason Brownlee. Cost-Sensitive Learning for Imbalanced Classification. [Електронний ресурс] – Режим доступу: <https://machinelearningmastery.com/cost-sensitive-learning-for-imbalanced-classification/>.
7. D. L. Wilson. Asymptotic Properties of Nearest Neighbor Rules Using Edited Data. *IEEE Transactions on Systems, Man, and Cybernetics*. 1972, V.3, pp.408-421.
8. Luque A, Carrasco A, Martin A, Heras de las A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recogn*. 2019, pp.216–231.
9. Batista, Gustavo EAPA, Ronaldo C. Prati, and Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*. 2004,V.6, pp.20-29.
10. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over sampling technique. *J Artif Intellig Res*. 2002, pp.321–357.
11. Douzas G, Bacao F, Last F. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Inf Sci*. 2018, V.465, pp.1–20.
12. Blagus R, Lusa L. SMOTE for High-dimensional class-imbalanced data. *BMC Bioinf*. 2013, V.14, pp.14-106.
13. Fu G.H., Xu F., Zhang B.Y., Yi L.Zh. Stable variable selection of class-imbalanced data with precision-recall criterion. *Chemo-metrics and Intelligent Laboratory Systems*. 2017, V.171, pp.241-250.
14. Haixiang G., Yijing L., Shang J., Mingyun G., Yuanue H., Bing G. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*. 2017, V.73, pp.220-239.
15. Nour Moustafa and Jill Slay. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). *Military Communications and Information Systems Conference (MilCIS)*. 2015, pp.1-6.
16. Douzas Georgios, Fernando Bacao, and Felix Last. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Information Sciences*. 2018, V. 465, pp.1-20.

Надійшла (received) 27.03.2023

Прийнято до друку (accepted for publication) 23.04.2023

Research of methods for improving the quality of classification on imbalanced data

Svitlana Gavrylenko, Vladislav Zozulia, Viktoriia Omelchenko

Abstract. The subject of the study is methods of balancing raw data. The purpose of the article is to improve the quality of intrusion detection in computer networks by using class balancing methods. Task: to investigate methods of balancing classes and to develop a classification method on imbalanced data to increase the level of network security. The methods used are: methods of artificial intelligence, machine learning. The following results were obtained: Class balancing methods based on Undersampling, Oversampling and their combinations were studied. The following methods were chosen for further research: SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE. The UNSW-NB 15 set was used as the source data, which contains information about the normal functioning of the network and during intrusions. A decision tree based on the CART (Classification And Regression Tree) algorithm was used as the basic classifier. According to the research results, it was found that the use of the SMOTEENN method provides an opportunity to improve the quality of detection of intrusions in the functioning of the network. Conclusions. The scientific novelty of the obtained results lies in the complex use of data balancing methods and the method of data classification based on decision trees to detect intrusions into computer networks, which made it possible to reduce the number of Type II errors.

Keywords: classification, imbalanced data, data balancing, Undersampling, Oversampling, SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE.

А. В. Гаптельманов, О. П. Міхаль, О. О. Щепка

Харківський національний технічний університет радіоелектроніки, Харків, Україна

ПІДВИЩЕННЯ ЕНЕРГОЗБЕРЕЖЕННЯ БЕЗДРОТОВИХ СЕНСОРНИХ МЕРЕЖ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Анотація. Актуальність. Бездротові сенсорні мережі (БСМ) є перспективною галуззю розвитку комп'ютерних мереж. Ключова ідея БСМ – автоматизація збору інформації про навколишнє середовище та керування об'єктами. БСМ особливо корисні там, де присутність людини в контрольованій зоні неможлива або збирання інформації необхідно проводити тривалий час. БСМ отримали широке поширення після активного розвитку сучасної мікроелектроніки, технологій бездротової взаємодії та відповідного апаратного забезпечення. **Метою даної роботи** є розробка аналіз алгоритмів підвищення енергозбереження в бездротових сенсорних мережах з використанням методів машинного навчання. **Об'єктом дослідження** є тривалість функціонування вузла в бездротовій сенсорній мережі. **Предметом дослідження** є алгоритми підвищення енергозбереження. **Результати.** Запропоновано адаптивний енергозберігаючий алгоритм передачі даних, який може застосовуватися в бездротових сенсорних мережах, де необхідно приймати рішення про стан сенсорної мережі (завдання класифікації) на основі безлічі показань із сенсорів. Для функціонування алгоритму необхідно навчити дерево рішень та отримати предикати для сенсорів. **Висновок.** Розроблено алгоритми роботи сенсора БСМ та побудови бінарного дерева рішень, які можуть бути використані для підвищення енергозбереження функціонування вузлів в бездротових сенсорних мережах. Розглянуто методи машинного навчання. Вони дозволяють оперувати з безліччю структурованих даних, отримуючи з них знання в вигляді моделі, яку можна використовувати в майбутньому для ухвалення рішення. Для реалізації запропонованого алгоритму обрано метод дерева рішень. Вибір пов'язаний з тим, що даний метод має високу інтерпретацію та схожий із процесом прийняття рішення оператором. Крім цього, дерева рішень дозволяють проводити класифікацію з пропусками у вхідних даних, що може досить часто відбуватися у сенсорних мережах. Грунтуючись на ідемпотентності предикатів дерева рішень, пропонується скоротити кількість сеансів зв'язку за рахунок динамічного визначення частоти передачі. Таким чином, цей підхід дозволить позбутися постійної передачі даних бездротовим каналом зв'язку, тим самим буде зекономлено енергоресурси вузлів мережі.

Ключові слова: енергозбереження, бездротова сенсорна мережа, вузол, машинне навчання, дерево рішень, класифікація.

Вступ

Структура БСМ часто залежить від її цільового призначення та навколишнього середовища інфраструктури. У класичному варіанті БСМ є безліч взаємопов'язаних по бездротовій мережі вузлів (сенсорів), розташованих на певній території та виконуючих збір характеристик довкілля. Як правило, вузол мережі – це малогабаритний пристрій, до складу якого входять датчик, блок обробки даних (мікропроцесор і зовнішня пам'ять), приймач і автономне джерело живлення. Сенсор реєструє зміни конкретних параметрів довкілля (освітлення, температура, вологість, тиск, рух тощо), потім обробляє їх та передає на базову станцію. Вузол мережі також може виступати в ролі ретранслятора повідомлень, за рахунок чого досягається значна зона покриття мережею за низьких фінансових витрат.

БСМ використовують для виконання таких функцій, як автоматичний збір даних, стеження параметрами середовища [1]. Серед основних типів БСМ виділяють: наземні, підземні, підводні, мультимедійні та мобільні мережі. Їх відмінні риси полягають у різних топологіях, технічній складовій сенсора, періодичності та швидкості передачі даних, особливості каналу зв'язку.

За рахунок низької вартості, швидкості розгортання та високої ефективності БСМ широко застосовуються практично у всіх сферах життєдіяльності, особливо популярні в оборонній галузі, промисловості, сільському господарстві та охороні здоров'я.

За великим потенціалом використання БСМ стоїть чимало наукових та технічних проблем, що починаються ще з проектування мережі [2]. Ці проблеми мають концептуально новий характер. Важливість їх рішення полягає в тому, що невраховані деталі можуть суттєво вплинути на функціонування всієї мережі у майбутньому. На відміну від інформаційних мереж, БСМ містять набір відмінних характеристик, які сприяють створенню нових способів та методик проектування. В силу особливостей використання БСМ та обмеженого функціоналу сенсорів, що формують мережу, виникають нові параметри системи, такі як розгортання та конфігурація маршрутизації, зона покриття, енергоспоживання та термін автономної роботи вузла мережі [3]. Оскільки все частіше БСМ є важливим елементом у процесах зовнішніх систем, то необхідно приділяти велику увагу надійності та відмовостійкості мережі.

Проведений аналіз публікацій [4-7] показав актуальність питання енергозбереження в БСМ. Значна частина досліджень спрямована на рішення вузькоспеціалізованих проблем, що належать до конкретних областей застосування БСМ, і лише невелика частина досліджень сфокусована на загальних задачах, властивих різним типам БСМ [8, 9]. При проектуванні БСМ за заданими вимогами часто стикаються з безліччю обмежень, зумовлених зовнішнім середовищем, рівнем розвитку техніки, доступними ресурсами [10, 11].

Серед розв'язуваних завдань виділяють наступні значні напрямки: проектування апаратного

забезпечення, способи і алгоритми управління та обробки інформації, питання експлуатації.

Серед перерахованих вище завдань приділимо увагу розробці методів і алгоритмів управління та обробки інформації у мережі. Ключовими параметрами при оцінці запропонованих методів та алгоритмів комунікації в мережі є швидкість передачі даних та термін експлуатації БСМ.

Термін експлуатації мережі безпосередньо залежить від тривалості функціонування вузлів мережі. У свою чергу, вузол мережі може припинити функціонувати з причин внутрішньої несправності, агресивного зовнішнього середовища, виходу із зони дії бездротової мережі або вичерпання енергії джерела живлення. Вузол мережі споживає електроенергію для збору, обробки та передачі інформації, також при цьому повинна бути вирішена задача маршрутизації при топології, відмінної від топології «зірка». Як правило, вузол мережі містить автономне джерело живлення з лімітованим енергоресурсом та відсутні способи поповнення енергії із зовнішнього середовища. Час автономної роботи вузла мережі пов'язано безпосередньо з часом служби автономного джерела живлення.

Незважаючи на те, що для вузлів БСМ можна передбачити механізми отримання енергоресурсів із навколишнього середовища (сонячні батареї, механічна, електромагнітна енергія), не завжди існує прямий і постійний доступ до перерахованих ресурсів [12]. Також варто враховувати, що елементи перетворення енергії з довкілля мають високу вартість, яка доходить до собівартості самого датчика; в такому випадку використання перетворювачів стає безглуздом, оскільки часто однією з базових вимог до БСМ є низька вартість датчиків. У дослідженні [13] було підкреслено, що питома енергетична щільність сучасних хімічних джерел живлення зростає повільніше, ніж обсяг споживаної енергії. Ще важливим фактом є те, що джерела енергії з великою енергоємністю становлять підвищену небезпеку і вимагають особливих умов експлуатації, тому що в іншому випадку можуть статися загоряння та вибухи.

Отже, вибір енергоефективних методів і алгоритмів збору та передачі інформації між вузлами мережі є однією з основних актуальних наукових проблем під час проектування БСМ, що підтверджує актуальність даної роботи.

Метою роботи є розробка аналіз алгоритмів підвищення енергозбереження в бездротових сенсорних мережах з використанням методів машинного навчання.

Основна частина

Сенсорні мережі дозволяють автоматизувати збирання даних про навколишнє середовище чи контрольований об'єкт. Бездротова взаємодія між вузлами мережі значно розширює сферу застосування, але несе безліч нових задач, якими представлено ряд досліджень. Однією з важливих завдань при проектуванні БСМ є збільшення терміну експлуатації мережі, яку, у свою чергу, пов'язано з терміном експлуатації вузлів мережі. Функціонування сенсора може бути порушено з кількох причин, і в цій роботі розглядаються проблеми обмеженого енергоресурсу біля вузла мережі та неенергоефективних алгоритмів обробки даних.

Аналіз споживання електроенергії вузлом БСМ ґрунтується на експериментальних даних, оскільки дослідження теоретичної моделі енергоспоживання вузла БСМ досить трудомістке і має низьку точність через наявність безлічі факторів, таких як: тип сенсора, дальність передачі даних, що виконуються обчислення. Результатом аналізу енергоспоживання є залежність витраченої енергії від виконуваних сенсором функцій. Перед аналізом енергоспоживання вузла мережі розглянемо його узагальнену структуру у складі БСМ. До складу вузла мережі (сенсора) входять (рис. 1):

- датчики, що вимірюють параметри навколишнього середовища;
- блок обробки даних, що складається з мікроконтролера та зовнішньої пам'яті;
- приймач;
- автономне джерело живлення.

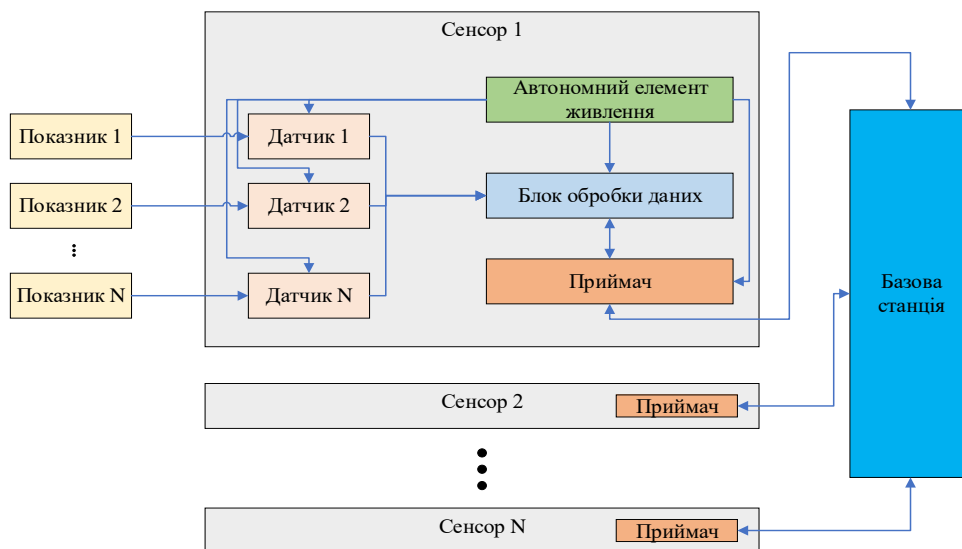


Рис. 1. Узагальнена структура вузла у складі БСМ

Вузол мережі вимірює параметри навколишнього середовища, обробляє їх, а потім посилає базову станцію, куди надходять дані з усіх сенсорів. З представленої структури видно, що енергія автономного елемента живлення витрачається на енергопостачання датчиків, мікроконтролера з пам'яттю та на приймач. На рис. 2 представлено запропонований алгоритм роботи сенсора БСМ.

Як правило, сенсор мережі функціонує циклічно, доки не закінчиться енергоресурс, тому представлений алгоритм не має завершення. Більшість досліджень у галузі підвищення енергоефективності вузла



Рис. 2. Алгоритм роботи сенсора бездротової сенсорної мережі

Збір даних – на зовнішній запит від базової станції сенсори зчитують показання з датчиків про параметри зовнішнього середовища або контрольованого об'єкта та відправляють на базову станцію. Моніторинг являє собою збір даних із певної періодичністю. Чергова ітерація збору даних може бути ініційована не тільки на запит від базової станції, а й за заданим розкладом, який зберігається локально у пам'яті сенсора. Як правило, на основі моніторингу в наступних приймається рішення про стан мережі або події подіях. Під функцією контролю мають на увазі моніторинг параметрів навколишнього середовища та подальшу їх перевірку на спрацювання заздалегідь певних умов. Спостереження – передача показань на головний модуль, якщо сенсор виявив конкретний об'єкт чи поведінку за заданим шаблоном. Формування керуючих команд – на основі отриманих із сенсорів показань приймається рішення про те, яку керуючу команду необхідно сформулювати для керованої системи. Для функцій моніторингу, контролю, стеження та формування керуючих команд потрібно на основі даних з вузлів БСМ приймати рішення про стан мережі або необхідну керуючу команду.

Тенденція нових розробок іде до того, щоб зберегти енергоресурси за рахунок зменшення кількості сеансів зв'язку між вузлами мережі. З іншого боку, завдання БСМ – своєчасно збирати дані про параметри навколишнього середовища та на їх основі приймати наступні рішення. У результаті, рішення, спрямоване на зменшення кількості сеансів зв'язку, створює технічну суперечність: «Зі збільшенням числа

БСМ мають характерну рису: запропоноване рішення є результатом пошуку компромісу між підвищенням енергоефективності мережі та зниженням якості інших параметрів мережі, таких як: своєчасність доставки повідомлень, простота реалізації комунікації та обробки інформації. Таким чином, досі залишається відкритим питання побудови алгоритмів передачі даних та обробки інформації в БСМ, що забезпечують одночасне підвищення енергозбереження, продуктивності та адаптивності мережі до довкілля.

Розглянемо п'ять основних напрямів функціонування БСМ (рис. 3).

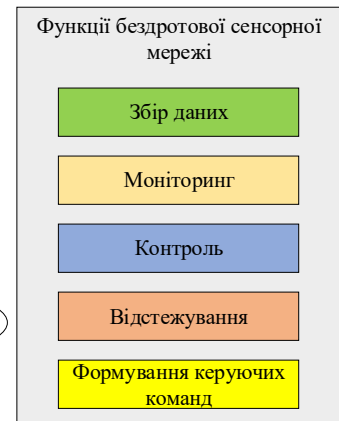


Рис. 3. Функції бездротової сенсорної мережі

передач даних збільшується енергоспоживання вузла мережі, що призводить до зменшення часу автономного термін служби сенсора, зі зменшенням кількості передач даних – збільшується ймовірність пропуску важливих подій, що може призвести до невірної ухвалення рішення».

Для вирішення даної технічної суперечності пропонується розробити алгоритм обробки інформації у БСМ, який з одного боку скоротить кількість сеансів зв'язку, тим самим підвищить енергоефективність вузла БСМ, а з іншого боку гарантуватиме, щоб головний модуль завжди володів необхідною інформацією про параметри довкілля для коректного прийняття рішення.

Для динамічного визначення частоти передач даних пропонується використовувати інформацію з бази знань, щоб сенсори самостійно ухвалювали рішення, коли необхідно передавати дані на базову станцію для загальної класифікації системи. Для реалізації цього алгоритму придатними є методи машинного навчання [14]. Дані методи спрямовані на створення математичної моделі процесу прийняття рішення на основі знань, отриманих з навчальної вибірки. Реалізація процесу ухвалення рішення про необхідність передачі даних на головний модуль має розташовуватися в логіці роботи вузла мережі, оскільки першоджерелом інформації є дані з датчиків, що знаходяться безпосередньо у складі вузла мережі. З іншого боку, для ухвалення рішення про передачу даних, можливо, знадобляться знання про процес класифікації, тобто дані про математичну модель прийняття рішення про стан мережі.

Аналіз існуючих застосувань [15,16] з використанням методів машинного навчання у роботі БСМ виявив, що часто такий підхід пов'язаний з тим, що ручне створення моделі ухвалення рішення є складним процесом, іноді навіть нездійсненним за рахунок різноманіття зовнішніх факторів, що впливають на результат, а також їх непередбачуваність у майбутньому. Отже, створення методів прийняття рішення в БСМ є трудомісткою операцією, вимагає адаптивної настройки параметрів алгоритму для виконуваної задачі. Можлива ситуація, коли модель має надто багато вхідних даних, в результаті процес прийняття рішення може бути занадто довгим і

задовольняти поставленим вимогам. Тому важливо генерувати модель, яка з одного боку надаватиме коректні результати прийняття рішення, а з іншого – має бути не надто складним, щоб час прийняття рішення було у допустимому інтервалі. У машинному навчанні для спрощення моделі використовують ймовірні підходи, які дозволяють знайти компроміс між складністю моделі та точністю прийнятих рішень. Тому найчастіше під час використання методів машинного навчання виконують попередній етап аналізу вхідних даних. Серед основних завдань машинного навчання (рис. 4) виділяють завдання регресії, класифікації, кластеризації, пошуку аномалій.



Рис. 4. Узагальнена класифікація методів машинного навчання

Для кожного завдання використовуються різні методи, ефективність яких залежить від типу вхідних даних та бажаного результату. Представлені методи дозволяють генерувати модель прийняття рішення шляхом навчання на раніше відомій вибірці даних та результатах їх обробки. У випадку з БСМ вхідними даними будуть параметри зовнішнього середовища, отримані з сенсорів, та результати ухвалення рішення про стан мережі. Отже, у цій роботі для завдання ухвалення рішення пропонується використати методи машинного навчання. Для цього на основі раніше відомих прецедентів (бази знань) буде отримано математичну модель, за допомогою якої надалі автоматично прийматиметься рішення, імітуючи роботу оператора. Завдання визначення стану в сенсорній мережі за даними із сенсорів відноситься до задачі класифікації. Найбільш популярними методами класифікації на даний момент є: а) k-найближчих сусідів; б) найкращий байєсівський класифікатор; в) метод опорних векторів; г) лінійна регресія; г) індукція правил (дерево рішень, випадковий ліс). Вибір методу класифікації залежить від показників вхідних даних (кількість параметрів, пропуски в даних), точності та повноти класифікації, часу навчання та перенавчання.

Велика кількість досліджень із завдання класифікації об'єктів представлено у галузі машинного

навчання, штучного інтелекту. Проведено аналіз методів машинного навчання з метою пошуку найбільш відповідної реалізації запропонованого алгоритму. Для ухвалення рішення про стан об'єкта використовуються різні моделі класифікації об'єктів, які широко представлені в галузі машинного навчання. Безліч розроблених методів класифікації об'єктів дозволяє обрати найбільш відповідний спосіб залежно від специфіки даних, їх кількості, необхідної точності та часу класифікації. Для вирішення поставленої задачі пропонується використовувати метод дерева рішень, оскільки він має ряд переваг: інтерпретованість моделі, робота з перепустками в даних.

У цьому алгоритмі процес класифікації вхідних об'єктів побудований на кон'юнктивних правилах, що складаються з безлічі предикатів за вхідними даними. Для цього на основі навчальної вибірки виконується побудова логічної схеми, що реалізує організовану систему ієрархічно послідовних питань. Процес класифікації полягає у почерговій відповіді на питання моделі, причому наступне питання залежить від попередньої відповіді. В результаті після відповіді на всі питання визначається єдиний клас, який і є відповіддю класифікації. Також є можливість для кожного предикату моделі зберігати ймовірність відповіді, тоді за неможливості відповісти питанням

(наприклад, за відсутності даних), обирається найбільш популярна відповідь, тим самим метод дозволяє оперувати з перепустками у вхідних даних. До недоліків даного методу відноситься неможливість перенавчання моделі, так як навчання використовується вся навчальна вибірка.

На алгоритми обробки інформації в БСМ накладають низку характерних обмежень, таких як розмір програмного коду через обмежений ресурс внутрішньої пам'яті, кількість виконуваних операцій для економії енергоресурсу.

Далі зробимо опис моделі дерева рішень для завдання класифікації БСМ.

Структура даних «дерево» є звичайним пов'язаним графом (без наявності циклів), у якого виділено кореневу вершину. Листком дерева називають вершину, яка не має нащадків. Шлях у графі "дерево" проходить від кореня до листа без повторення вершин. Дерево рішень представлено кореневим деревом, у вершинах якого містяться логічні правила, а листі – результати класифікації. Логічні правила містять умову у форматі «якщо... то... інакше...». Дерева рішень відмінно справляються із завданням багатокласової класифікації, тобто дозволяють визначити належність вхідного об'єкта до одного із заздалегідь відомих класів. Цільова змінна може мати категоріальні, дискретні та бінарні значення. Алгоритм класифікації стану мережі за допомогою дерева рішень являє собою процес перетворення вхідного вектора даних на результат класифікації. Для кожної вершини дерева використовуються дані вхідного вектору, щоб відповідати на поставлені запитання. У листі дерева розташовуються результати класифікації, і навіть розмір класу, щоб у разі пропуску даних обирати найбільш популярний клас.

Запитання, що знаходяться у вершинах дерева рішень, використовують виключно один параметр вхідного об'єкта та мають уявлення "значення параметра x_{nk} більше значення ...". При позитивній відповіді перехід здійснюється до правого нащадка, при негативному – до лівого. Якщо нащадок містить правило, знову виконується процес відповіді питання. Внаслідок відповідей на запитання досягається один із листів дерева, де міститься інформація про цільовий параметр. За рахунок прозорої інтерпретації роботи алгоритму цей метод має високу наочність.

Алгоритм побудови бінарного дерева рішень представлений на рис. 5. На першій ітерації алгоритму розглядається вся безліч вхідних даних, далі алгоритм повторюється рекурсивно для підмножин, отриманих при розбитті:

Крок 1. Перевірка умови виходу з алгоритму. При виконанні умови слід вибрати для поточного вузла (аркуша дерева) значення з множини Y (результат класифікації вхідного об'єкта), інакше переходимо до виконання кроку 2.

Крок 2. Множина даних S розбивається на 2 непересічні підмножини S_1 і S_2 за раніше встановленим критерієм мінімізації ентропії. Цей критерій виражений як функція $Qt(x)$, на вхід якої подаються різні розбиття множини S і виконується пошук максимальної якості розбиття (максимальне зменшення ентропії).

Крок 3. Для підмножин S_1 і S_2 алгоритм продовжує виконуватись рекурсивно, для цього кожна підмножина окремо передається до кроку 1.

Для функції розбиття зазвичай використовують формулу розрахунку ентропії, і на кожній ітерації алгоритму прагнуть її мінімізувати. Таким чином, досягається найменший розмір дерева рішень. Для зменшення розміру дерева застосовують спеціальні методи раннього завершення алгоритму, такі як обмеження по глибині дерева, мінімальну кількість об'єктів у множині S , мінімально-допустиму зміну значення ентропії.

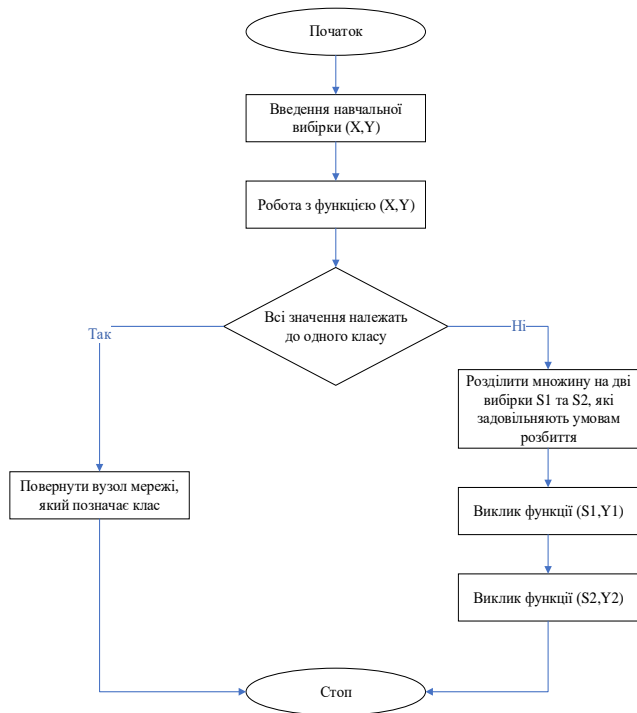


Рис. 5. Алгоритм побудови бінарного дерева рішень

В результаті використання даних методик зменшується ймовірність перенавчання моделі дерева рішень, що позитивно позначається на точності класифікації моделі, і навіть зменшується загальний розмір моделі. Спосіб розрахунку абсолютного значення енергоефективності пропонованого алгоритму досить складно висловити, оскільки підсумковий результат в основному залежить від динаміки параметрів довкілля. Для більш точної оцінки енергоефективності запропонованого адаптивного енергозберігаючого алгоритму необхідно мати дані щодо поведінки параметрів зовнішнього середовища.

Висновки

Розроблено алгоритми роботи сенсора БСМ та побудови бінарного дерева рішень, які можуть бути використані для підвищення енергозбереження функціонування вузлів в бездротових сенсорних мережах. Розглянуто методи машинного навчання. Вони дозволяють оперувати з безліччю структурованих даних, отримуючи з них знання в вигляді моделі, яку можна використовувати в майбутньому для ухвалення рішення. Для реалізації запропонованого алгоритму обрано метод дерева рішень. Вибір пов'язаний

з тим, що даний метод має високу інтерпретацію та схожий із процесом прийняття рішення оператором. Крім цього, дерева рішень дозволяють проводити класифікацію з пропусками у вхідних даних, що може досить часто відбуватися у сенсорних мережах. Грунтуючись на ідемпотентності предикатів дерева

рішень, пропонується скоротити кількість сеансів зв'язку за рахунок динамічного визначення частоти передачі. Таким чином, цей підхід дозволить позбутися постійної передачі даних бездротовим каналом зв'язку, тим самим буде зекономлено енергоресурси вузлів мережі.

СПИСОК ЛІТЕРАТУРИ

1. Akyildiz, I. F., Weilian, S. A survey on sensor networks // *IEEE Communications magazine*. 2002. 40(8). P. 102–114. doi: 10.1109/MCOM.2002.1024422.
2. Abbas, I., Liu, J., Faheem, M., Noor, R. S. Different sensor based intelligent spraying systems in Agriculture // *Sensors and Actuators A: Physical*. 2020. Volume 316. P. 1–16.
3. Alias, F., Carrie, J. C., Alsina-Pages, R. WASN-Based day–night characterization of urban anomalous noise events in narrow and wide streets // *Sensors*. 2020. 20(17) 26 p.
4. Chen, Y., Nasser, N. Energy-balancing multipath routing protocol for wireless sensor networks // *Quality of service in heterogeneous wired/wireless networks*. New York: Qshine, 2006. Vol. 21. – P. 245–249. doi: 10.1145/1185373.1185401.
5. Das, K., Behera, R. N. A survey on machine learning: concept, algorithms and applications // *International Journal of Innovative Research in Computer and Communication Engineering*. 2017. 5(2). P. 1301–1309.
6. Gao, T., Greenspan, D., Welsh, M., Juang, R. R., Alm, A. Vital signs monitoring and patient tracking over a wireless network // *Proceedings of the 27th IEEE EMBS Annual International Conference*. 2005. P. 66–74.
7. Huang, J.H., Amjad, S., Mishra, S. CenWits: A sensor-based loosely coupled search and rescue system using witnesses // *Proceedings of the Third Int. Conf. on Embedded Networked Sensor Systems (Sensys)*, San Diego, CA. 2005. P. 180–191.
8. Monitoring volcanic eruptions with a wireless sensor network / Werner-Allen, G., Johnson, J., Ruiz, M., Lees, J., Welsh, M. // *Proceedings of the Second European Workshop on Wireless Sensor Networks, IEEE*. Istanbul, 2005. 13 p.
9. Hardware design experiences in ZebraNet / Zhang, P., Sadler, C. M., Lyon, S. A., Martonosi, M. // *Proceedings of the SenSys'04*, Baltimore, MD. 2004. P. 227–238. Li Q.-L. (2016), “Nonlinear Markov processes in large networks”, *Special Matrices*, Vol. 4(1), doi: <https://doi.org/10.1515/spma-2016-0019>
10. Yick, J., Mukherjee, B., Ghosal, D. Analysis of a Prediction-based Mobility Adaptive Tracking Algorithm // *Proceedings of the IEEE Second International Conference on Broadband Networks (BROADNETS)*, Boston. 2005. Vol. 1. P. 753–760
11. Yu, Z., Tsai J. J., Weigert T. An automatically tuning intrusion detection system // *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 2007. 37(2). P. 373–384.
12. Design considerations for solar energy harvesting wireless embedded systems / Raghunathan, V., Kansai, A., Hse, J., Friedman, J., Srivastava, M. // *Proceedings of the IPSN*. 2005. P. 457–462.
13. Second Generation System-on-Chip Solution for 2.4 GHz IEEE 802.15.4 / RF4CE ZigBee [Electronic resource] // *Texas Instruments*. Available at: <http://www.ti.com/lit/ds/symlink/cc2530.pdf>. 05.05.2023. 36 p
14. Pelckmans, K., Jos, D. B., Johan, A. S., Bart, D. M. Handling missing values in support vector machine classifiers // *Neural Networks* 18. 2005. №5-6. P. 684–692
15. Venkatesan, R., Meng, J. E. A novel progressive learning technique for multi-class classification // *Neurocomputing*. 2016. 207. P. 310–321.
16. Yu, Z., Tsai, J. J. P. A Framework of Machine Learning Based Intrusion Detection for Wireless Sensor Networks // *2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*. 2008. P. 272–279.

Received (Надійшла) 01.03.2023

Accepted for publication (Прийнята до друку) 24.05.2023

Enhancement of energy saving of wireless sensor networks using machine learning methods

Artem Haptelmanov, Oleg Mikhal, Oleksii Schepka

Abstract. Topicality. Wireless sensor networks (WSN) are a promising branch of computer network development. The key idea of BSM is to automate the collection of information about the environment and controlled objects. WSN is especially useful where the presence of a person in the controlled area is impossible or the collection of information must be carried out for a long time. BSMs became widespread after the active development of modern microelectronics, wireless communication technologies and corresponding hardware. **The goal of this work** is the development and analysis of algorithms for increasing energy saving in wireless sensor networks using machine learning methods. **The object of research** is the duration of the operation of a node in a wireless sensor network. The subject of research is algorithms for increasing energy saving. **The subject of research** is algorithms for increasing energy saving. **Results.** In this paper, a method of simulation modeling of computer network traffic with fractal properties is proposed. The theory of fractal analysis and the theory of Markov processes were used to generate traffic. This method can be part of a software simulation model of a computer network, which in turn can be used for testing network algorithms and protocols. **Conclusions.** Algorithms for the operation of the WSN sensor and the construction of a binary decision tree have been developed, which can be used to increase the energy saving of the functioning of nodes in wireless sensor networks. Methods of machine learning are considered. They allow you to operate with a lot of structured data, obtaining knowledge from them in the form of a model that can be used in the future to make a decision. The decision tree method was chosen to implement the proposed algorithm. The choice is due to the fact that this method has a high interpretation and is similar to the process of decision-making by the operator. In addition, decision trees allow classification with gaps in the input data, which can happen quite often in sensor networks. Based on the idempotency of the predicates of the decision tree, it is proposed to reduce the number of communication sessions due to the dynamic determination of the transmission frequency. Thus, this approach will make it possible to get rid of the constant transmission of data through a wireless communication channel, thereby saving energy resources of network nodes.

Keywords: energy saving, wireless sensor network, node, machine learning, decision tree, classification.

G. Golovko, D. Iievliev

National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine

ENHANCED AUTHORIZATION FOR SECURE MANAGEMENT OF SENSITIVE DATA IN HYBRID APPLICATIONS

Abstract: Sensitive data is often managed by cloud-based applications, which can be vulnerable to attackers who seek unauthorized access to this data. Traditional approaches to authorization may not be sufficient to protect sensitive data from such attacks. In this article, we propose an enhanced authorization approach that uses a combination of symmetric and asymmetric cryptography to secure sensitive data. Specifically, we propose generating a unique encryption key per file and a set of public and private keys per user, which are used to encrypt and decrypt the data. We demonstrate the feasibility of our approach with examples in Node.js, showing how to generate public and private keys, encrypt and decrypt files, and store encrypted data on a drive. Our approach provides an effective solution to the problem of managing sensitive data in hybrid applications, while preserving user and developer convenience.

Keywords: Cryptography, Public-key cryptography, Authorization, Node.js.

Introduction

Cloud-based applications are commonly used to manage sensitive data, such as customer information and messages. However, such data can be vulnerable to attacks by unauthorized users seeking to gain access to this data. Traditional approaches to authorization, such as role-based authorization and granular permissions systems, may not be sufficient to protect sensitive data from such attacks. This is because even if an application is designed with the strongest authentication and authorization mechanisms, it can still be vulnerable to security issues in the libraries used to build it. For example, in Node.js, one of the most commonly exploited vulnerabilities is Prototype Pollution [1], which can allow attackers to gain administrative privileges and access sensitive data. We need to embrace the fact that attackers could potentially get access to storage with sensitive data.

Proposal

We propose an enhanced authorization approach that uses a combination of symmetric and asymmetric cryptography [2] to secure sensitive data. Specifically, we propose:

1. Do not store unencrypted data. We accept that the attackers could access the data, but they will not be able to read it without the encryption key.
2. Use a unique encryption key per file. We accept that attackers might get much faster hardware in the future and find an encryption key. Using a unique encryption key per file dramatically increases attackers' effort to decrypt data and make it much harder to use the Harvest now, decrypt later [3] attack.
3. Use a unique set of public and private encryption keys per user. Even if one user's key is compromised, the attacker will not be able to decrypt files from other users.
4. Always encrypt a file encryption key with a user public encryption key. This ensures that only authorized users can decrypt it and read file data.

Implementation

We demonstrate the feasibility of our approach with examples in Node.js, a popular platform for building cloud-based applications. Node.js has a built-in standard

crypto [4] package which provides routines for symmetric and asymmetric cryptography. Each user should have their own public and private keys. In order to generate a key pair we should use the generateKeyPair() [5] (Listing 1).

```
const { generateKeyPair } = require('node:crypto');
generateKeyPair('rsa', {
  modulusLength: 4096,
  publicKeyEncoding: {
    type: 'spki',
    format: 'pem',
  },
  privateKeyEncoding: {
    type: 'pkcs8',
    format: 'pem',
    cipher: 'aes-256-cbc',
    passphrase: 'top secret',
  },
}, (err, publicKey, privateKey) => {
  // Handle errors and use the generated key pair.
});
```

Listing 1. RSA key pair generation

The following 4 scenarios provide implementation guidelines that could help to understand the main idea.

Scenario 1: create a file with a sensitive information

1. Prior to storing of any secret, the owner needs to generate a unique symmetric key.
2. Owner stores encrypted file on a drive.
3. Owner stores encrypted symmetric key on a drive (Fig. 1).

The randomBytes() [6] routine allows to generate cryptographically strong pseudorandom data which is then used as a symmetric encryption key (Listing 2).

The createCipheriv() [7] is being used to create an encryption object which is then used to encrypt a file stream (Listing 3).

```
const { randomBytes } = require('node:crypto');
const symmetricEncryptionKey = randomBytes(32);
...
```

Listing 2. Symmetric key generation

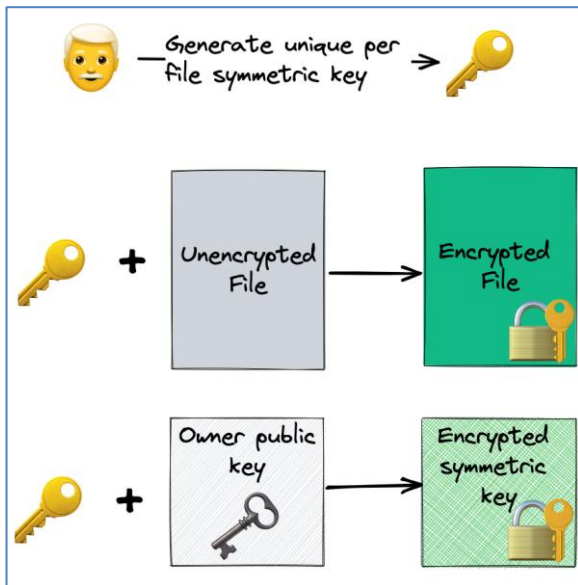


Fig. 1. Encrypted file creation flow

```
const { createCipheriv } = require('node:crypto');
const { createWriteStream } = require('node:fs');
const busboy = require('busboy');
const
INITIALIZATION_VECTOR="23ca408b8ff898bb" //
16 bytes
const bb = busboy({ headers: req.headers });
bb.on("file", async (_, file, { filename }) => {
  const symmetricEncryptionKey = randomBytes(32);
  const cipher = createCipheriv(
    'aes-256-cbc',
    symmetricEncryptionKey,
    INITIALIZATION_VECTOR
  );
  const stream = createWriteStream(filePath);
  file.pipe(cipher).pipe(stream);
  ...
}
```

Listing 3. Create encrypted files stream

The publicEncrypt() [8] is being used to encrypt file's symmetric encryption key (Listing 4).

```
const { publicEncrypt } = require('node:crypto');
const { writeFile } = require('node:fs');

const encryptedKey = publicEncrypt(publicKey,
fileSymEncryptionKey)
writeFile(pathToEncryptedKey,
encryptedKey.toString('hex'))
...
```

Listing 4. symmetric key encryption

Scenario 2: decrypt file content

1. User needs a private key to decrypt symmetric encryption key.
2. With help of symmetric encryption key user could decrypt encrypted file content (Fig. 2).

The privateDecrypt() [9] is being used to get file's symmetric encryption key (Listing 5). The createDecipheriv() [10] is being used to create a decryption object which then sends decrypted data to an output stream (Listing 6).

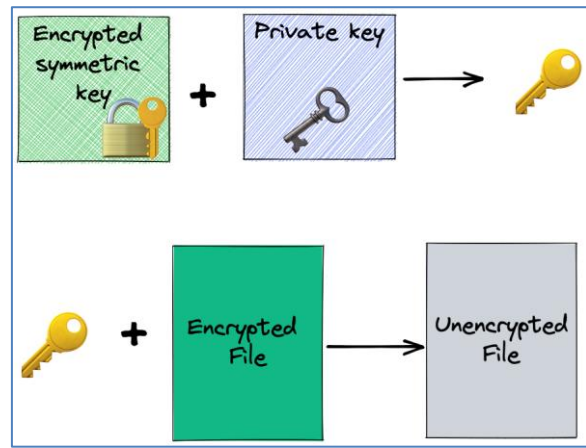


Fig. 2. File decryption flow

```
const { privateDecrypt } = require('node:crypto');
const encryptedKey = Buffer.from(rawFileData, 'hex')
const fileSymEncryptionKey = privateDecrypt({
  privateKey,
  Passphrase,
},
encryptedKey);
...
```

Listing 5. Symmetric key decryption

```
const { createDecipheriv } = require('node:crypto');
const { pipeline, createReadStream } =
require('node:fs');
const
INITIALIZATION_VECTOR="23ca408b8ff898bb" //
16 bytes
const encryptedFileStream =
createReadStream(filePath);
const decipher = createDecipheriv(
'aes-256-cbc',
symmetricEncryptionKey,
INITIALIZATION_VECTOR);
pipeline(encryptedFileStream, decipher,
outputStream, (error) => {
  ...
})
```

Listing 6. Decrypt file stream

Scenario 3: share a file with a sensitive information

1. File owner decrypts symmetric encryption key.
2. File owner encrypts symmetric encryption key with user's public key.
3. File owner adds to the storage a new file with encrypted symmetric key.
4. Only user could decrypt symmetric key with its private key and then decrypt file content (Fig. 3).

Important to mention that the encrypted file remains intact. The file storage just got a tiny extra file with encrypted symmetric key

Scenario 4: revoke access to a file with a sensitive information

1. Owner removes encrypted symmetric key of the user.
2. User without a key cannot decrypt file content.
3. Encrypted file remains intact (Fig. 4).

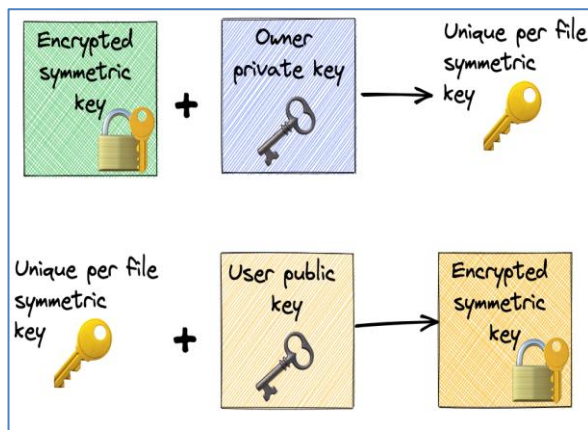


Fig. 3. Encrypted file sharing flow

Conclusion

We have proposed an enhanced authorization approach. Idea behind it is not a new one, it has already been used for decades in EFS [11] in Microsoft Windows. It has proven to be reliable over the years. Since the file encryption is based on the symmetric cryptography, it allows us to encrypt/decrypt files at a high speed. Moreover, modern CPUs for more than a decade have had a hardware acceleration [12] for the AES [13] algorithm. The biggest drawback of the approach is the recovery process.

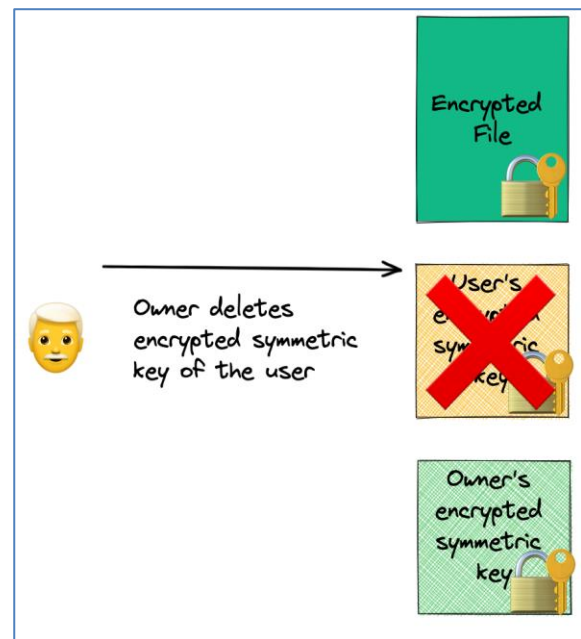


Fig. 4. Revoking file access flow

As a developer of an application, you need to take care of the recovery key option and the organization should define security policy for creation and storing of the recovery keys.

REFERENCES

1. *Prototype Pollution*. Snyk. <https://learn.snyk.io/lessons/prototype-pollution/javascript/>
2. Г. В. Головка. *Конспект лекцій з дисципліни "Захист інформації в комп'ютерних системах і Кібербезпека"* (2021). Національний університет «Полтавська політехніка імені Юрія Кондратюка». <https://dist.nupp.edu.ua/mod/resource/view.php?id=122282>
3. *Harvest now, decrypt later*. Wikipedia. https://en.wikipedia.org/wiki/Harvest_now,_decrypt_later
4. *Crypto package*. Node.js. <https://nodejs.org/api/crypto.html#crypto>
5. *crypto.generateKeyPair()* Node.js. <https://nodejs.org/api/crypto.html#cryptogeneratekeypairtype-options-callback>
6. *crypto.randomBytes()*. Node.js. <https://nodejs.org/api/crypto.html#cryptorandombytesize-callback>
7. *crypto.createCipheriv()*. Node.js. *Crypto package*. Node.js. <https://nodejs.org/api/crypto.html#cryptocreatecipheralgorithm-key-iv-options>
8. *crypto.publicEncrypt()*. Node.js. <https://nodejs.org/api/crypto.html#cryptopublicencryptkey-buffer>
9. *crypto.privateDecrypt()*. Node.js. <https://nodejs.org/api/crypto.html#cryptoprivatedecryptprivatekey-buffer>
10. *crypto.createDecipheriv()*. Node.js. <https://nodejs.org/api/crypto.html#cryptocreatedecipheralgorithm-key-iv-options>
11. *Encrypting File System*. Wikipedia. https://en.wikipedia.org/wiki/Encrypting_File_System
12. *AES instruction set*. Wikipedia. https://en.wikipedia.org/wiki/AES_instruction_set
13. *Advanced Encryption Standard*. Wikipedia. https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

Надійшла (received) 22.02.2023

Прийнято до друку (accepted for publication) 16.05.2023

Розширена авторизація для безпечного керування конфіденційними даними в гібридних програмах

Г. Головка, Д. Ієвлев

Анотація. Конфіденційними даними часто керують хмарні додатки, які можуть бути вразливими для зломисників, які прагнуть несанкціонованого доступу до цих даних. Традиційних підходів до авторизації може бути недостатньо для захисту конфіденційних даних від таких атак. У цій статті ми пропонуємо розширений підхід авторизації, який використовує комбінацію симетричної та асиметричної криптографії для захисту конфіденційних даних. Зокрема, ми пропонуємо створити унікальний ключ шифрування для кожного файлу та набір відкритих і закритих ключів для кожного користувача, які використовуються для шифрування та дешифрування даних. Ми демонструємо здійсненність нашого підходу на прикладах у Node.js, показуючи, як генерувати відкритий і закритий ключі, шифрувати та розшифровувати файли та зберігати зашифровані дані на диску. Наш підхід забезпечує ефективне вирішення проблеми управління конфіденційними даними в гібридних програмах, зберігаючи при цьому зручність для користувачів і розробників.

Ключові слова: криптографія, криптографія з відкритим ключем, авторизація, Node.js.

M. Demydenko, M. Kalnytskyi

National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine

ONLINE PLATFORM PROTOTYPE USING MICROSERVICE ARCHITECTURE AND CONTAINERIZATION FOR DIGITALIZATION OF THE EDUCATIONAL PROCESS

Abstract. The COVID pandemic and full-scale military operations in our territory affected the implementation of digital and information technologies in the educational process. Digitalization has covered all spheres of education, and has especially affected distance learning technologies. But the problem of automating the process of managing the educational process remains unexplained. Ensuring quality management of the educational process requires a review of approaches to architecture, development processes, and implementation of appropriate software. Most of the modern information systems for the management of the educational process (ISMEP) are built according to the "monolithic" architecture. This limits the possibility of modernization without rebuilding the core of the system. The consequence of this is the problem of scaling and efficient use of cloud technologies. The purpose of the work is to develop a prototype of an online ISMEP platform using microservice architecture and containerization.

Keywords: digitalization of the educational process, microservice architecture, containerization.

Introduction

The COVID pandemic and full-scale military operations in our territory affected the implementation of digital and information technologies in the educational process. Digitalization has covered all spheres of education, and has especially affected distance learning technologies. But the problem of automating the process of managing the educational process remains unexplained.

Actuality of theme. Ensuring quality management of the educational process requires a review of approaches to architecture, development processes, and implementation of appropriate software. Most of the modern information systems for the management of the educational process (ISMEP) are built according to the "monolithic" architecture:

1. MKR [1].
2. Politek-soft [2].
3. ASU VNZ [3].

The listed *ISMEPs* are built on client/server technology and consist of several modules that automate the activities of the main structural divisions. This limits the possibility of modernization without rebuilding the core of the system. The consequence of this is the problem of scaling and efficient use of cloud technologies.

Complex solutions of educational tools from Google and Microsoft are mostly focused on e-learning. That is, there are no ISMEP on the market that are built on a microservice architecture.

Analysis of the latest research. Most of the publications devoted to microservice architecture in the educational process consider its use in e-learning. Thus, in the work [4], the main principles of the organization of microservice architectures (MSA), their parameters, functioning, and the possibility of application in electronic educational systems (EES) were considered. A comparative analysis of construction options for the MSA, their components and methods was carried out. A comprehensive method of building electronic educational systems based on MSA is proposed. In

work [5], a study was conducted on increasing the productivity of EES when switching to MSA.

The objective of the work – to develop a prototype of an online ISMEP platform using microservice architecture and containerization.

Main part

Features of application of monolithic and MAS architecture in ISMEP. The parameters of ISMEP are determined not only by functionality and software, but also by architecture. Most ISMEPs with a developed core use monolithic structure. This gives them such advantages as good dynamics of development and support. An alternative to monolithic systems is the division of ISMEP into separate microservices. MSA provides the following advantages:

1. Modularity. Separate parts of the application are responsible for specific operations.
2. Reliability. When one of the microservices fails, only those functions for which it is responsible are disrupted. It will not be able to damage the server completely.
3. Changeability. You can change the microservice independently of other system services.
4. Scalability. You can add or remove a module at any time. If the microservice reaches its maximum load, new instances of the service can be deployed in a neighboring cluster.
5. Testability. New features can be tested immediately in production. It is enough to implement them for some segments of users to check how they will work [5].

Disadvantages using microservice architecture:

1. Organization of interaction of microservices in the entire system. This depends on the number of processes involved and requires additional efforts in planning, control, monitoring, testing and deployment.
2. Expenses. Microservices require the use of a larger total amount of memory than monolithic architectures, this is due to the need to organize isolated environments, as well as the costs of organizing communications between them (interfaces, protocols, communication channels, etc.).

3. Security issues. Microservices require security solutions due to the presence of internal service communication.

4. Decrease in productivity compared to monolithic systems. The reason is the occurrence of additional delays in data exchange between microservices. Therefore, it is necessary to solve

problems with the correct design of queues, asynchronous processing and balancing.

5. Problems of inconsistency. Unlike monolithic systems, in MSA, errors when updating individual microservices can lead to out of sync violations [4].

For the realization of ISMEP, there are chosen the following functions (Fig. 1):

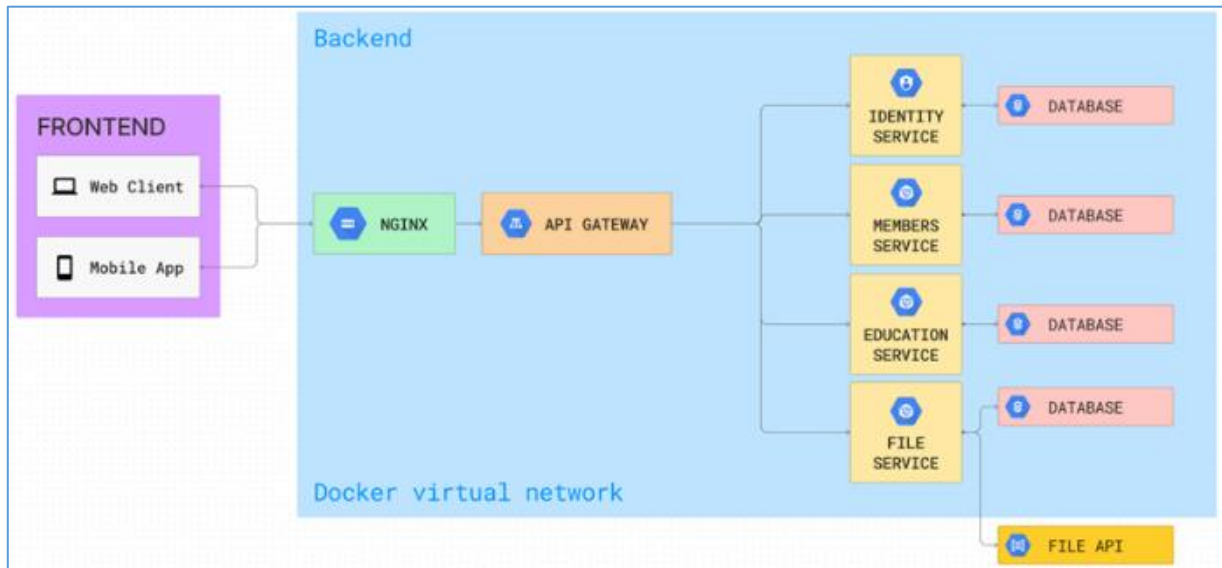


Fig. 1. ISMEP prototype architecture

The implementation of the platform prototype was divided into two parts - the creation of the server (backend) and client (frontend) parts. The server part is a REST API system built on a microservice architecture and deployed using containerization technologies, the task of which is to receive HTTP requests from the client (web application), process them and return data in JSON format.

1. API Gateway is a request distributor and aggregator, provides only permitted functionality for users with different access levels, and collects data from multiple microservices as needed. The only point of interaction between the client and the server is the API GATEWAY. This microservice does not contain any database, and performs two functions: a role limiter and a request aggregator.

The first role of Api Gateway is to provide a certain function only to authenticated users with the required role (student/teacher/administrator). Authorization starts through guest access points that are directed to the identity microservice.

The second role of Api Gateway is aggregation. Some requests require data and processing in multiple microservices, so gateway's job is to send requests to all required microservices, collect all responses separately, filter, and return the response to the client. Specific microservices are closed from the user, so there is no role checking. One big problem in the aggregation process is the lack of atomicity of operations. Ideally, when sending requests to 2-3 microservices, if one of them does not perform the operation and gives an error, then all changes in other microservices should be canceled.

Identity Service is created for authorization, registration and storing of personal data of users. Application authorization and authentication is built on a modern approach that uses JWT tokens.

JWT (JSON Web Token) is a way of securely transferring information between a client and a server in the form of a special string token. The token contains information required for authorization and authentication (such as an ID and role) and can be used to access certain resources or functions in the application. The JWT is transmitted with each request, accordingly, the validity of the token and the possibility of access of the given user to the requested sections of the functionality at the moment will be checked each time.

When entering the correct login and password on the site, the Identity microservice generates a token that stores the unique user ID and role (student/teacher/administrator) and returns it to the client for storage. The token itself is signed with a secret key, making counterfeiting by a third party impossible.

The lifetime of the token, which determines its validity, is recorded in the body of the token and is determined by the server that issues it. The recommended time is up to 30 minutes.

In order for the user not to re-enter his login and password every ten minutes to log in to the account, a second key, called Refresh Token, is issued during authorization. This is a special code that can be used by the client to update the JWT without passing the login and password.

The Refresh Token is stored in the browser's Http-only cookie, which guarantees the impossibility of receiving the token by third-party scripts and programs.

The single access point to the server functionality (API Gateway) contains four routes through which requests can pass. Each of the three roles (student/teacher/administrator) has its own route, which checks the existence of the required role in the

transferred JWT token before executing the commands. The fourth route is made for non-authorized users, it does not require any tokens and is used for user registration or login.

The microservice data scheme is shown in Fig. 2.

2. Members Service is designed to store and process data about students, teachers, groups and reviews. The data scheme is shown in Fig. 3.

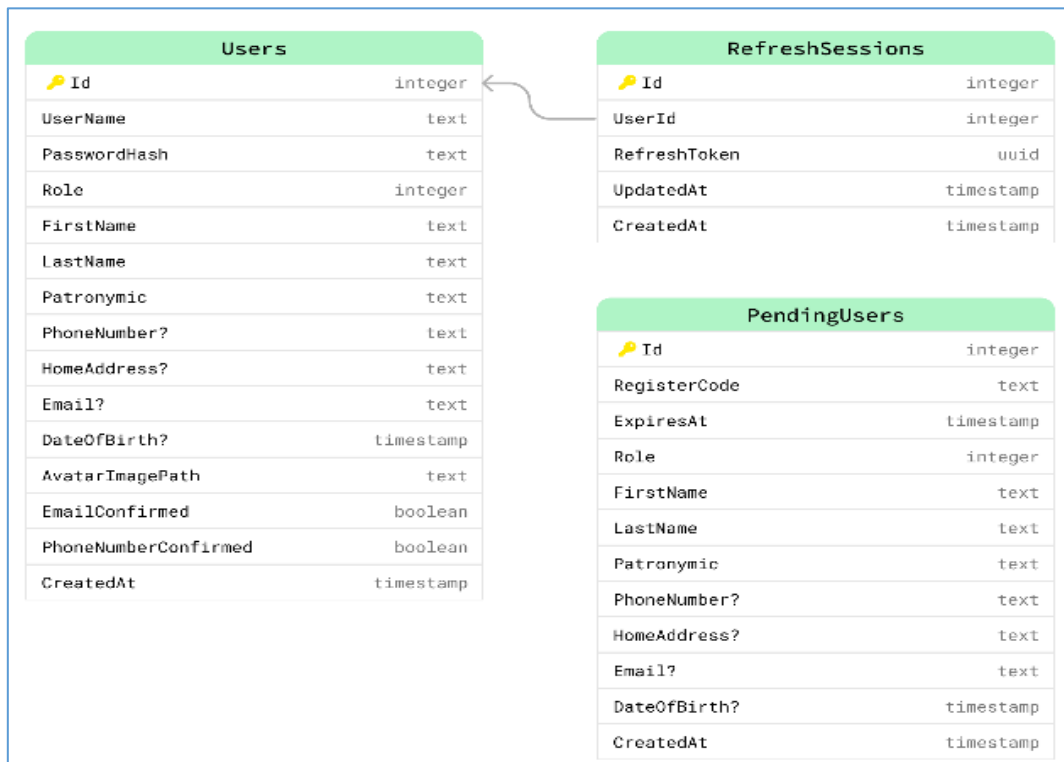


Fig. 2. Identity microservice data scheme

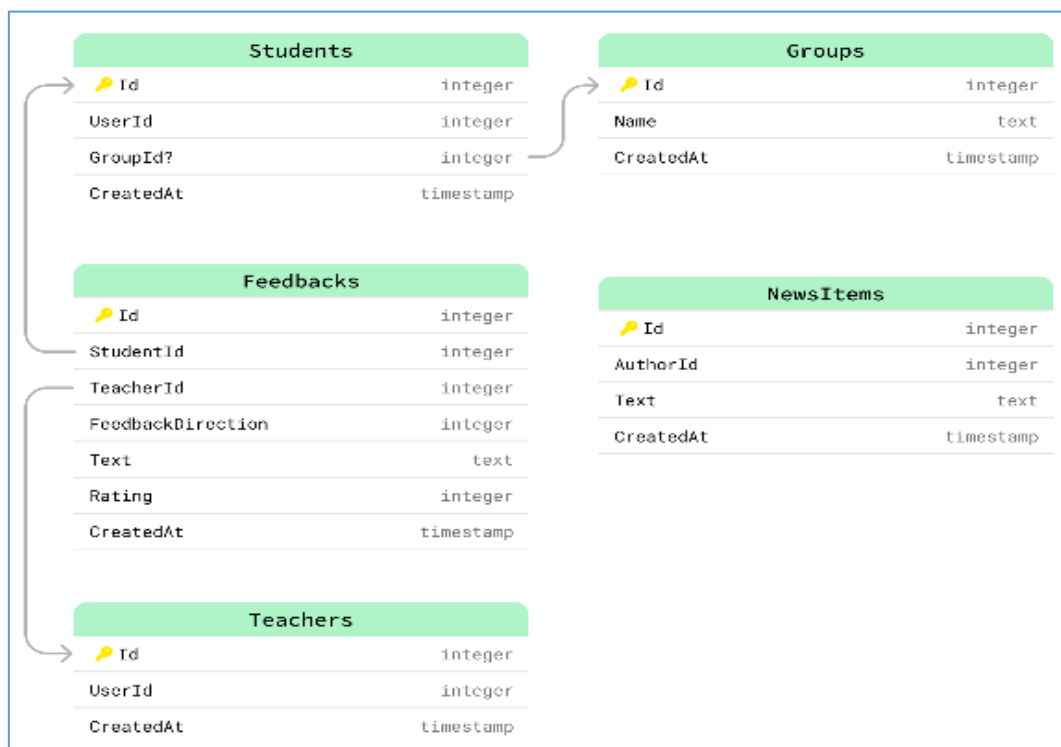


Fig. 3. Members microservice data scheme

3. Education service is designed to store and process everything related to the educational process: assignments, schedules, grades, assessments, exams. The data scheme is shown in Fig. 4.

4. File Service is designed to work with external applications for working with files and images. Amazon S3 is used to store and download files. The data scheme is shown in Fig. 5.

Core functions of the prototype of the online platform ISMEP.

There are three user groups defined:

- student;
- teacher;
- administrator.

Diagram of the possible actions is shown in Fig. 6. To ensure containerization, Docker containers are used,

which are a kind of virtual machines that allow developers to package and run programs in an isolated environment.

They contain everything the application needs to run, including dependencies and settings.

Conclusions

The analysis of recent researches showed that the majority of publications devoted to microservice architecture in the educational process consider its use in e-learning.

Most of the modern information systems for the management of the educational process (ISMEP) are built according to the "monolithic" architecture. Based on the results of the research, a prototype of the online platform was developed.

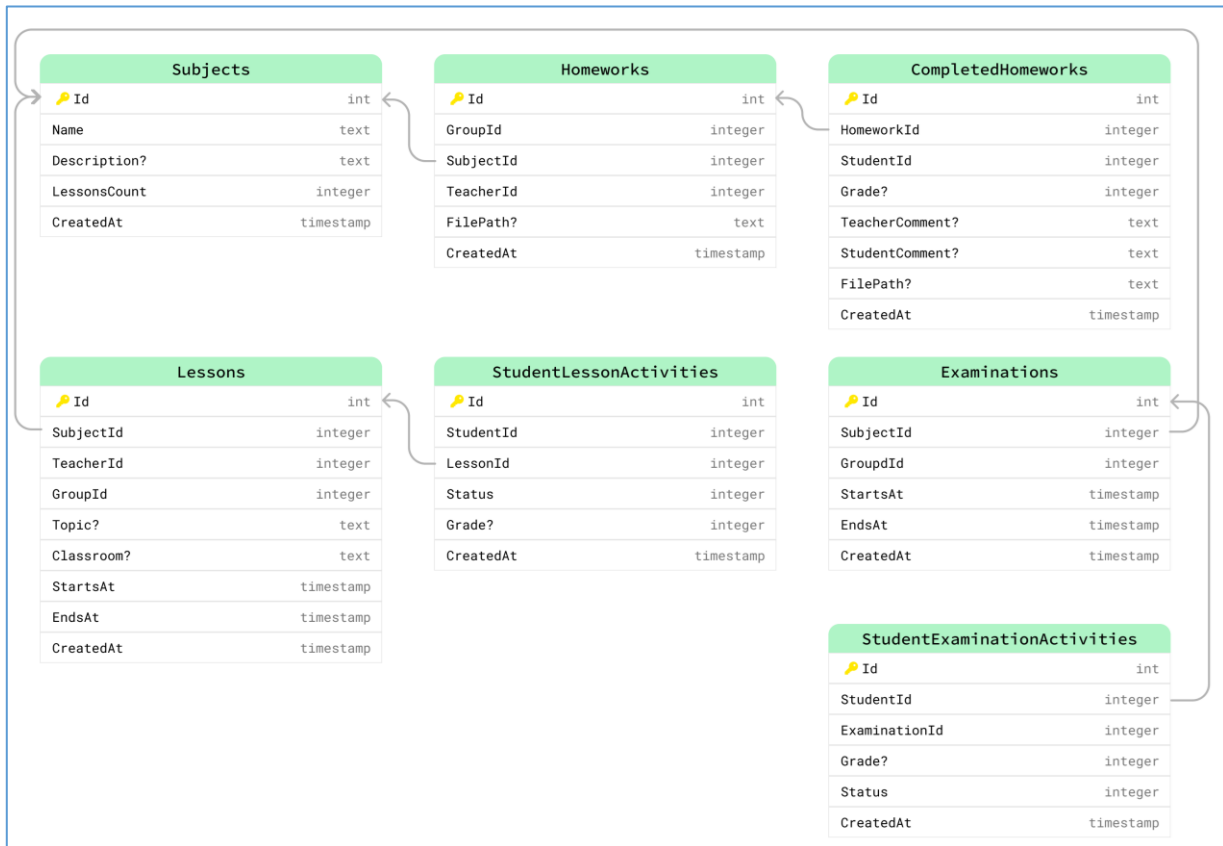


Fig. 4. Education microservice data scheme

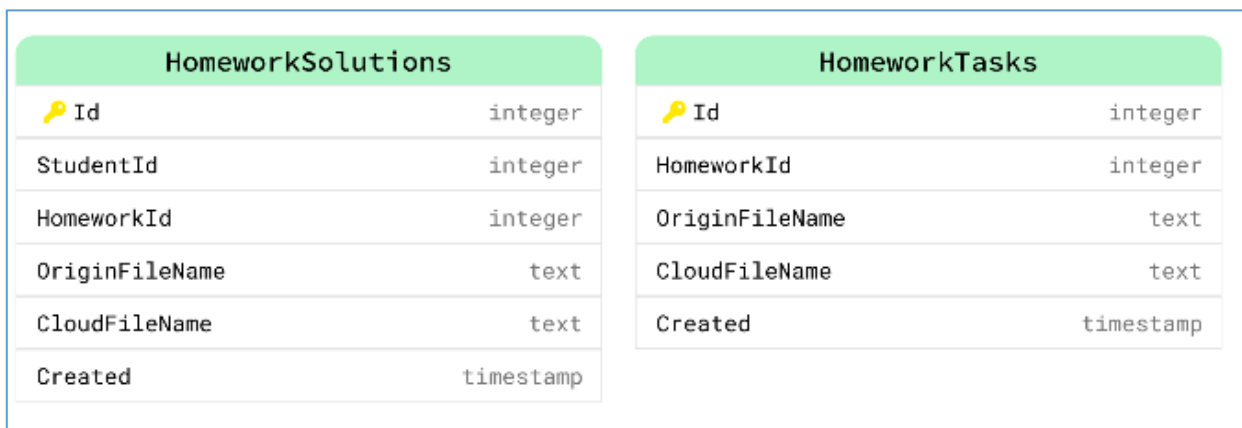


Fig. 5. File microservice data scheme

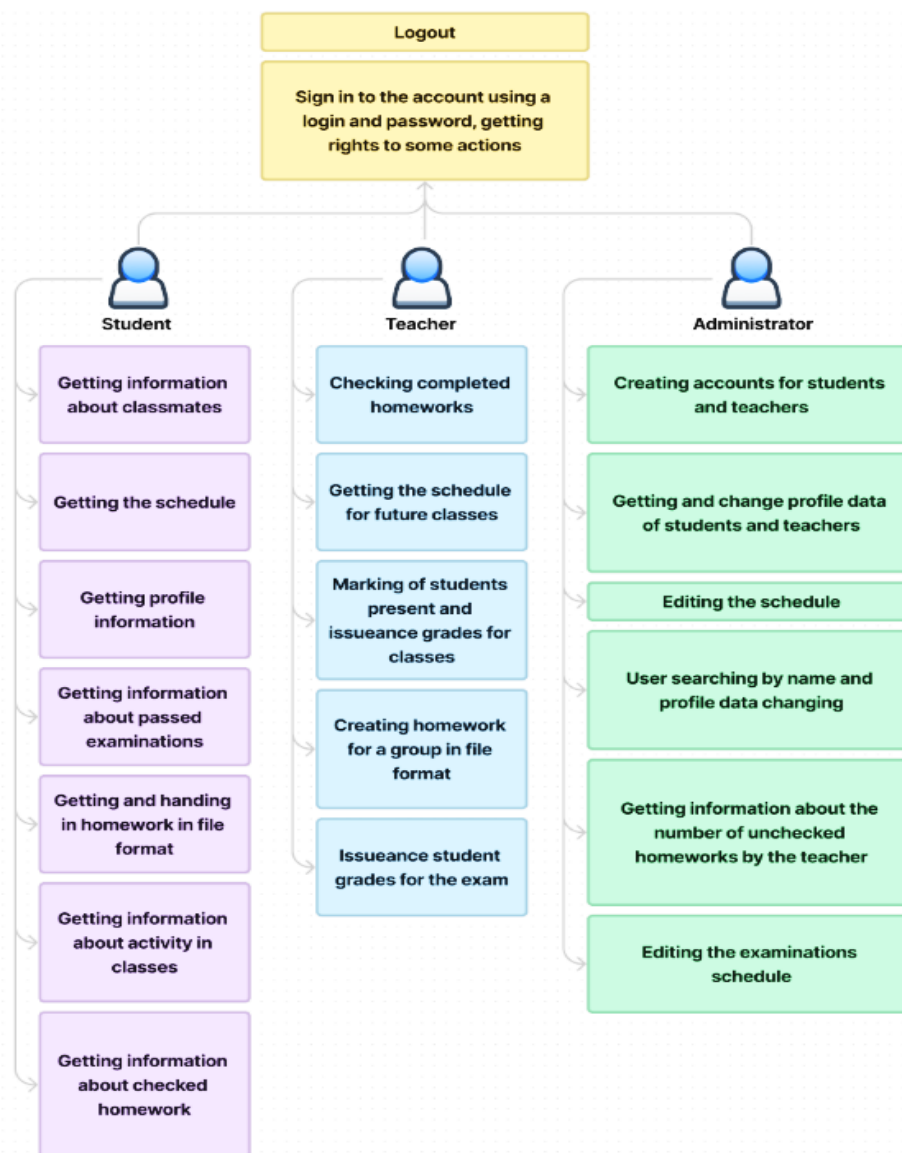


Fig. 6. ISMEP core functions

REFERENCES

1. MKR [Electronic resource] –Access mode:<https://mkr.org.ua/>
2. Politek-soft [Electronic resource] –Access mode: <http://www.politek-soft.kiev.ua/ru/>
3. ASU VNZ [Electronic resource] –Access mode: <http://ndipit.com.ua/ua/rozrobky/asu-vnz>
4. Є. Паламарчук, «МЕТОДИ ПОБУДОВИ МІКРОСЕРВІСНОЇ АРХІТЕКТУРИ СИСТЕМ ЕЛЕКТРОННОГО НАВЧАННЯ», ІТКІ, вип. 53, вип. 1, с. 43–54, Лют 2022. <https://doi.org/10.31649/1999-9941-2022-53-1-43-54>
5. A. Kravchuk Мікросервісна архітектура: плюси та мінуси. Access mode <https://blog.iteducenter.ua/articles/microservices-architecture-advantages-and-disadvantages/>

Надійшла (received) 22.03.2023

Прийнято до друку (accepted for publication) 23.05.2023

Прототип онлайн платформи з використанням мікросервісної архітектури та контейнеризації для діджиталізації освітнього процесу

М. І. Демиденко, М. А. Кальницький

Анотація. Пандемія COVID, та повномасштабні військові дії на нашій території вплинули на впровадження цифрових та інформаційних технологій в освітній процес. Діджиталізація охопила всі сфери освіти, а особливо вплинула на технології дистанційного навчання. Але проблема автоматизації процесу управління освітнім процесом залишається не висвітленою. Забезпечення якісного менеджменту освітнього процесу вимагає перегляду підходів до архітектури, процесів розроблення, та впровадження відповідного програмного забезпечення. Більшість сучасних інформаційних систем управління освітнім процесом (ІСУОП) побудовані по «монолітній» архітектурі. Це обмежує можливість модернізації без перебудови ядра системи. Наслідком цього є проблема масштабування та ефективного використання хмарних технологій. Мета роботи – розробити прототип онлайн платформи ІСУОС з використанням мікросервісної архітектури та контейнеризації.

Ключові слова: діджиталізація освітній процес, мікросервісна архітектура, контейнеризація.

О. А. Ільяшов, К. В. Покора, В. О. Дяченко, А. А. Коваленко

Харківський національний технічний університет радіоелектроніки, Харків, Україна

КЛАСИФІКАЦІЯ ДАНИХ АПАРАТНИМИ ПРИСКОРЮВАЧАМИ FPGA У ЦЕНТРАХ ОБРОБКИ ДАНИХ ТА ХМАРАХ

Анотація. **Актуальність.** Аналіз даних, зокрема їх класифікація) часто виконується за допомогою методів машинного навчання. Часто задіяні алгоритми потрібні мати справу з великими наборами даних, що призводить до тривалого часу виконання. Таким чином, актуальним є дослідження апаратних прискорювачів, на базі програмованої вентиляльної матриці (FPGA) для покращення продуктивності. FPGA є перспективним рішенням для апаратного прискорення, конфігурації після виробництва та можливості перепрограмування. **Метою даної роботи** є дослідження та аналіз тенденцій у існуючих хмарних архітектурах FPGA, які підкреслюють складний зв'язок між архітектурою та системними вимогами та призначені для вирішення завдань класифікації даних методами машинного навчання. Це дозволяє нам ідентифікувати нові архітектури, які, ймовірно, запропонують значні переваги для хмарних робочих навантажень. **Об'єктом дослідження** є процес еволюції FPGA-прискорювачів для обчислень у центрах обробки даних (ЦОД) та хмарах. **Предметом дослідження** є методи та алгоритми дослідження хмарних архітектур FPGA на основі таксономічних категорій. **Результати.** У роботі обговорюється майбутнє використання FPGA у центрах обробки даних та хмарах. Також у роботі досліджуються поточні архітектури та обговорюється масштабованість і абстракції, які підтримуються операційними системами, проміжним програмним забезпеченням і віртуалізацією. Висновки. Розроблено класифікацію щодо дослідження хмарних архітектур FPGA на основі таксономічних категорій. Розглянута та запропонована архітектурна організація щодо розгортання додатків FPGA, що використовуються в хмарних середовищах і середовищах центрів обробки даних. Запропоновані дві моделі проектування додатків машинного навчання для класифікації даних з використанням апаратних FPGA-прискорювачів.

Ключові слова: реконфігурована логіка, FPGA прискорення, хмарні обчислення, центр обробки даних, віртуалізація, таксономічні категорії, класифікація даних, машинне навчання, програмне забезпечення.

Вступ

Майбутнє штучного інтелекту (AI), зокрема машинного навчання (ML), наближається, і воно залежить від нього. Штучний інтелект (AI) став однією з найпопулярніших обчислювальних технологій у світі, і компанії, такі як Intel, Google, Microsoft, IBM та інші, всі вони спираються на FPGAs. Масштабно-програмовані вентиляльні матриці (FPGAs) стали більш, ніж екзотичною та нішевою технологією, яку можуть досягнути тільки магіки апаратного забезпечення. Гравці такого рівня, як Alibaba, Amazon, Baidu, Huawei та Tencent, тепер використовують FPGAs для розробки додатків у своїх інфраструктурах дата-центрів. Інші використовують FPGAs для надання додатків як сервіс (Microsoft [2] та Nimbix) або для внутрішньо розроблених додатків. Крім того, по всьому світу здійснюється ряд проєктів в академічних та інших дослідницьких організаціях з метою надання користувачам віддаленої можливості прискорення та гнучкості. Хоча FPGAs стають все більш доступними у дата-центрах, їх довгострокове прийняття в цих місцях не гарантоване. Архітектура та інтеграція на рівні плати та системи повинні забезпечити, щоб застосунки використовували переваги переконфігурації та пом'якшували його недоліки. Надання апаратних ресурсів вимагає відповідного програмного забезпечення, апаратної віртуалізації та механізмів розділення доменів.

З усього світу, в академічних та інших наукових організаціях, все більше запускається проєктів, щоб надавати користувачам віддалену можливість прискорення та гнучкості. Хоча FPGAs стають все більш доступними у датацентрах, їх довгострокове використання у таких місцях не гарантується. Архітектура та

інтеграція на рівні плати та системи повинна забезпечити те, що застосунки використовують переваги переконфігурування та зменшують його недоліки. Постачання апаратних ресурсів потребує відповідної проміжної програми, апаратної віртуалізації та механізмів розділення доменів. Ефективне та гнучке постачання апаратних ресурсів збільшує можливість безперешкодної інтеграції, ідеально без переконструювання, в існуючу інфраструктуру керування хмарою. Просторовий спільний доступ до FPGA збільшує складність планувальників та менеджерів ресурсів хмари. На щастя, великий обсяг робіт, розроблених за останнє десятиліття, може бути використаний у системах планування завдань хмари, які в даний час використовуються в датацентрах, таких як спискове планування, що використовується в Amazon Cloud. Просторовий спільний доступ до FPGA збільшує складність планування та менеджменту ресурсів хмари.

Метою цієї роботи є дослідження та аналіз тенденцій у існуючих хмарних архітектурах FPGA, які підкреслюють складний зв'язок між архітектурою та системними вимогами та призначені для вирішення завдань класифікації даних методами машинного навчання. Це дозволяє нам ідентифікувати нові архітектури, які, ймовірно, запропонують значні переваги для хмарних робочих навантажень.

Основна частина

На зміну універсальним обчисленням приходять гетерогенні: звичайні ЦП доповнюються спеціалізованими процесорами або прискорювачами, що підвищують продуктивність та енергоефективність серверів при ресурсомістких робочих навантаженнях, які стають звичайним явищем у сучасному дата-центрі. В якості таких прискорювачів обчислень

використовуються графічні процесори (GPU), програмовані користувачем вентиляльні матриці (FPGA) та інтегральні схеми спеціального призначення (ASIC). Нові робочі навантаження визначають нові системні вимоги та функції, сприяють розвитку та впровадженню таких технологій, як NVMe, FPGA та бази даних у оперативній пам'яті (in-memory). За даними аналітиків Forrester, спеціалізовані прискорювачі, такі як FPGA та GPU, будуть використовуватися

все ширше. Застосування графічних процесорів може приблизно в 25 разів прискорити виконання окремих видів обчислень, відзначають у 451 Research [6]. У ієрархії процесорів - від ЦП загального призначення до ASIC - існує компроміс між гнучкістю та ефективністю. При цьому, як стверджують у Accenture, коли якась конкретна програма реалізується за допомогою спеціалізованих рішень, продуктивність може збільшуватися на порядки (табл. 1).

Таблиця 1 – Відносна ієрархія апаратних прискорювачів

Апаратне рішення	CPU	GPU	FPGA	ASIC
Призначення	Для додатків загального призначення	Для обчислень пов'язаних з графікою	Масив програмованих блоків з програмованим міжз'єднанням	Розробляються для конкретної функціональності
Відносна продуктивність	1	100	1000	10000-100000
Гнучкість	Універсальний	Спеціалізований	Програмований	Для спеціального застосування
Ринок	Широкий	Обмежений	Обмежений	Конкретні ніші ринку
Простота програмування	Широко доступні навички програмування	Потребує спеціальних знань	Потребує спеціальних знань	Тільки інтерфейс
Ключові гравці	Intel, AMD, ARM	NVIDIA, AMD, Intel	Xilinx, Intel (Altera), Actel	NEC, LSI, Samsung

GPU, FPGA та ASIC можуть використовуватися для прискорення завдань віртуалізації та хмарних обчислень, 3D/2D-графіки, високопродуктивних обчислень (HPC), штучного інтелекту та машинного навчання, аналізу великих даних, віртуалізації мережевих функцій (NFV) тощо. Наприклад, їх можна успішно застосовувати в застосунках машинного навчання, призначених для розпізнавання голосу та зображень або аналізу тексту.

Комбінування універсальних процесорних ядер зі спеціалізованими блоками обчислень за нейромережевими алгоритмами – актуальна тенденція. Спеціалізовані мікросхеми для систем штучного інтелекту – ще один новий ринок, що стрімко розвивається.

Свій ASIC – Edge TPU – для задач III випустила й Google. Проте зараз у сфері обробки даних з застосуванням III найбільш відомі продукти NVIDIA з її графічними процесорами та рішення Intel, такі як Nervana Neural Network Processor (NNP-I), заявлена продуктивність якого щодо штучного інтелекту в 10 разів вища, ніж у графічних карт, а також лінійка FPGA Agilex для задач III.

Кожен тип процесора, універсального (CPU), графічного (GPU) або FPGA, має свої переваги, інакше їх просто не виробляли б. CPU забезпечують хорошу продуктивність при найвищій універсальності та застосовності. Приблизно 99% всіх існуючих програм написані для виконання під CPU. GPU мають більшу кількість ядер та векторну архітектуру, високу швидкість обміну з пам'яттю та введення-виведення. FPGA мають найбільшу продуктивність на Ватт споживаної потужності завдяки властивостям апаратури, але потребують дуже ретельного та складного програмування.

Нижче про ці відмінності трохи детальніше:

Універсальні процесори CPU - це в суті, працюючі нині IT-індустрії. їх можна використовувати для найрізноманітніших завдань, але через свою

архітектуру CPU не настільки ефективні для паралельних обчислень. Останнім часом ця проблема частково вирішується за рахунок імплементації в чипі процесора багатьох ядер. Однак навіть у найпродуктивніших CPU кількість ядер поки що вимірюється декількома десятками.

Графічні процесори (GPU) довгі роки працювали тільки в ніші відображення інформації на екрані. І лише відносно недавно GPU стали застосовуватися для завдань високопродуктивних обчислень, в тому числі майнінгу криптовалют. Робота з графікою як векторними завданнями призвела до такого розвитку архітектури GPU, яка стала пристосованою для цілей паралельних обчислень. Як результат, сучасна архітектура графічного процесора дозволяє прискорити прохід векторизованих даних через свої конвейери, які в іншому випадку довелося б запускати через багато інших логічних блоків у ЦП з відповідною втратою продуктивності. Сучасні GPU містять у чипі кілька тисяч процесорних ядер.

FPGA, на відміну від універсальних і графічних процесорів, можна перепрограмувати згідно з особливостями вирішуваної на них обчислювальної задачі. В результаті отримується синтез спеціалізованого процесора для конкретної задачі. Іншими важливими відмінностями FPGA є знижене споживання енергії на одиницю обчислювальної потужності, а також архітектура з паралельним виконанням багатьох векторних операцій одночасно - так звана масивно-паралельна дрібнозерниста архітектура. Кількість ядер у чипі FPGA може досягати одного мільйона і більше.

FPGA-прискорювач представляє собою адаптер, який крім чипа FPGA містить на платі локальну оперативну пам'ять типу HBM (пам'ять DRAM з високою пропускну здатністю) та високошвидкісні інтерфейси введення-виведення, такі як 10/40 GE та PCI Express. FPGA-прискорювачі також випуска-

ються у форм-факторі SOM (як одномодульні системи). Кожен чіп FPGA містить понад мільйон комірок, що можуть бути перепрограмовані для різних функцій - кеш-пам'ять, сигнальні процесори, блоки цифрової обробки, блоки інтерфейсів.

На сьогодні FPGA перевершує ASIC у продуктивності завдяки більш сучасним технологічним процесам створення кристалів. Для FPGA використовуються техпроцеси рівня 20 нм і навіть 14 нм. Тоді як для створення кристалів ASIC використовуються більш "старі" техпроцеси рівня 60 нм. Відповідно, на тій же площі кристала у FPGA можна розмістити у декілька разів більше логічних комірок, ніж у ASIC, що забезпечує перевагу в продуктивності.

У хмарних обчисленнях FPGA застосовуються для швидкого розрахунку, прискорення мережевого трафіку та доступу до масивів даних. Сюди також можна віднести використання FPGA для високочастотної торгівлі на біржах. У сервери вставляють плати FPGA з PCI Express та оптичним мережевим інтерфейсом виробництва Intel (Altera) або Xilinx.

На FPGA чудово лягають криптографічні алгоритми, порівняння послідовностей ДНК та наукові задачі, такі як молекулярна динаміка. У Microsoft давно використовують FPGA для прискорення пошукового сервісу Bing, а також для організації Software Defined Networking всередині хмари Azure.

Бум машинного навчання також не обійшов FPGA стороною. Компанії Xilinx та Intel пропонують засоби на основі FPGA для роботи з глибокими нейронними мережами. Вони дозволяють отримувати прошивки FPGA, які реалізують ту чи іншу мережу безпосередньо з фреймворків, таких як Caffe та TensorFlow.

До того ж все це можна спробувати, не виходячи з дому та використовуючи хмарні сервіси. Наприклад, в Amazon можна орендувати віртуальну машину з доступом до плати FPGA та будь-яких засобів розробки, в тому числі й машинного навчання.

На сьогоднішній день більшість хмарних провайдерів зосередилися на платформах більш загального призначення, здаючи в оренду стандартизовану масштабовану інфраструктуру. Пропозиції в основному відрізняються кількістю (віртуальних) ядер, обсягом пам'яті та пропускною спроможністю мережі на ЦП і сховище. Пропозиції з більш спеціалізованим апаратним забезпеченням, насамперед із графічним процесором, є переважно окремими спеціалізованими пропозиціями, наприклад, для високопродуктивних обчислень (HPC), а не елементами, які забезпечують широку основу архітектури хмарної платформи.

Деякі хмарні провайдери почали використовувати спеціалізоване обладнання (включаючи FPGA) [2] і навіть почали розробляти власні SoC [5] замість того, щоб покладатися на кремнієвий процесор від постачальника, який відкриває двері для більшої диференціації платформ на основі апаратного забезпечення. Це передбачається як хмарна інфраструктура прогресує від інфраструктури як послуги (IaaS) до віртуалізованих платформ, а потім до мікросервісів і функцій як послуги (FaaS), хмарні обчислення також можуть надати екосистеми програмного забезпечення,

які сприятимуть зростанню впровадження спеціалізованого апаратного забезпечення.

Хмарні архітектури FPGA.

Попит на послуги дата-центрів має шанс продовжувати експоненційне зростання [1]. Оскільки закон Мура сповільнився, а обчислювальні витрати та складність хмарних завдань продовжують зростати та еволюціонувати, ПЛІС пропонує перспективний механізм для вирішення парадоксу, пов'язаного з поєднанням продуктивності, енергоефективності та гнучкості. Вони можуть бути встановлені майже будь-де в дата-центрі для прискорення обчислень, зберігання та мереж, як показано на рис. 1. Завдяки неочіненним перевагам ПЛІС їх використання в дата-центрах очікується зростати з гігантською річною зростанням в 48% між 2020 та 2027 роками [2,10].

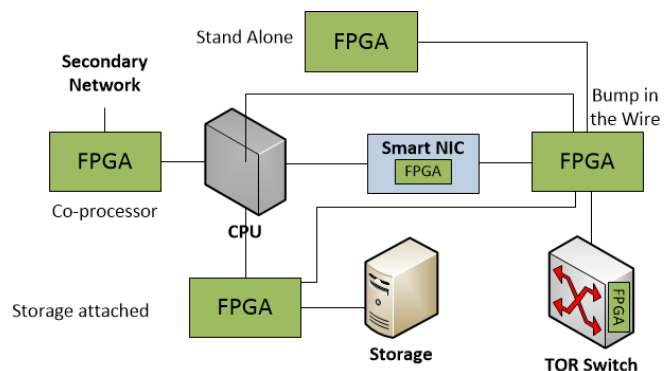


Рис. 1. Розгортання FPGAs в центрі обробки даних

Незважаючи на те, що ПЛІС пропонують низку переваг для дата-центрів, вибір певної архітектури для використання є складним та непростим; ми не можемо просто розміщувати будь-яку кількість та розмір цих пристроїв в будь-якому місці у хмарі. Екстремальна потреба в ефективності витрат приводить до акценту на розмірі, енергоспоживанні, охолодженні, сумісності, автоматизації та можливості оновлення на місці, забезпечуючи при цьому задоволення певних потреб хмарних завдань у термінах продуктивності, пам'яті, можливості сервера, надійності, безпеки та підключення до мережі.

Cloud FPGA - це інфраструктура пристроїв FPGA або програмних засобів проектування, які доступні в хмарі. Щоб використовувати переваги "розміщення FPGA в хмарі", ми пропонуємо класифікувати cloud FPGA за трьома рівнями сервісів:

1) FPGA програмні засоби як сервіс: з моделі хмарного SaaS засоби FPGA стали дематеріалізовані з 2010 року [4]. Ця модель отримала користь від масштабної обчислювальної потужності хмари без потреби в процесі ускладненого налаштування інструментів. Ця модель надає легкий доступ до хмари для апаратних проектувальників без складності інфраструктури та програмного забезпечення, але не дає доступу до фізичного FPGA.

2) FPGA платформи як сервіс: з моделі хмарного PaaS платформи FPGA були дематеріалізовані до 2010 року [5, 6]. Немає потреби в придбанні FPGA платформ. Ця модель дозволяє дизайнерам додатків отримувати доступ до однієї або кількох FPGA плат-

форм. Дизайнер може розробляти та впроваджувати свої хмарні додатки на старих або нових платформах FPGA.

3) FPGA ресурси як сервіс: з моделі хмарного IaaS FPGA та його ресурси стали віртуалізованими з 2014 року [7]. У цій моделі FPGA розділено на кілька незалежних віртуальних FPGA областей. Ці області можуть надаватися кільком дизайнерам додатків в багатокористувацькому середовищі з віртуальним доступом до фізичного FPGA.

Ця класифікація може змінюватись з часом в залежності від різних критеріїв через появу нових застосувань або зміни вимог. Залежно від сервісу, запропонованого для хмари FPGA, можна вирішити раніше поставлені виклики.

Використання FPGA в хмарі полягає у віддаленій оренді набору конкретних програмних засобів, платформ або ресурсів FPGA у вартісно-ефективний

спосіб. Така FPGA-забезпечена хмара зберігає переваги FPGA (наприклад, низьке споживання енергії та програмованість) і забезпечує масштабованість, гнучкість та багатокористувацьку обслуговування.

Надання ресурсів FPGA подібно до надання традиційних хмар на основі центральних процесорів (CPU) та графічних процесорів (GPU). Щодо категорій сервісів у традиційному обчислюванні в хмарі, провайдери хмари FPGA пропонують FPGA як інфраструктуру як сервіс (IaaS) або програмне забезпечення як сервіс (SaaS) [6]. Рисунок 2 наводить приклад ієрархічного відображення в FPGA-хмарі. Немає стандартного визначення або класифікації для FPGA-хмар, і ієрархічне відображення може змінюватись з часом. "Vendor manage (optional)" та "User manage (optional)" вказують на те, що ця ієрархія не завжди існує в хмарі FPGA, і вона настроюється кожним виробником або користувачем хмари FPGA.

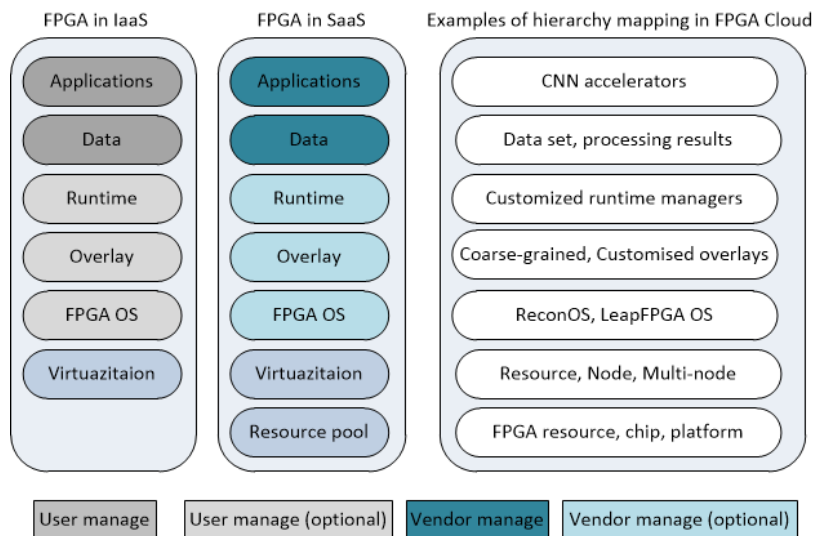


Рис. 2. IaaS та SaaS хмара FPGA

Фреймворк FPGA у хмарі відноситься до інфраструктури та програмного забезпечення, необхідного для використання FPGA в середовищі хмари. Фреймворк повинен забезпечувати спосіб відкриття FPGA як пулу ресурсів, які можуть бути виділені та звільнені арендарями. Це потребує відстеження використання FPGA для спрощення розрахунків.

Крім того, фреймворк повинен дозволяти арендарям програмувати FPGA. Традиційні системні стеки вважають FPGA нерухомими прискорювальними пристроями, не враховуючи їх програмованість. Для вирішення цих проблем потрібно додати нові модулі до обчислювальних вузлів, а планувальник в контролюючому вузлі повинен бути покращений. Системний стек обчислювальних вузлів можна розділити на чотири рівні: апаратний, гіпервізор, бібліотека та застосунок. Модулі FPGA повинні бути додані на рівень апаратного забезпечення та зроблені доступними як пул ресурсів для арендарів.

Рівень гіпервізора повинен бути змінений для підтримки віртуалізації FPGA, що дозволяє кільком арендарям ділитися однією FPGA без взаємодії. Рівень бібліотеки повинен забезпечувати програмні

абстракції та інструменти, які дозволяють арендарям програмувати FPGA.

Нарешті, рівень застосунку повинен забезпечувати застосунки та сервіси, які можуть використовувати прискорення FPGA. Фреймворк також повинен вирішувати проблеми безпеки, такі як забезпечення того, що арендарі не можуть отримати доступ до FPGA.

FPGA мають бути представлені у стеку хмарних послуг як басейн ресурсів, який може бути активно керований, тобто його можна запитувати, виділяти та звільняти користувачами. Його використання повинно бути відстежене, щоб сприяти розрахунку, пов'язаному зі збором платежів за публічну хмарну модель. Крім того, якщо FPGA було надано арендареві, його повинно бути можливо програмувати орендарем, аналогічно до інших ресурсів, таких як ЦП та ГПУ. Однак, традиційні стеки системного програмного забезпечення, тобто операційна система та гіпервізор, вважають FPGA лише функціональними прискорювальними пристроями, ігноруючи їхню природу програмованості. Крім модулів, які зображені сірим кольором, на рис. 3 показані типові компоненти хмари на базі OpenStack.

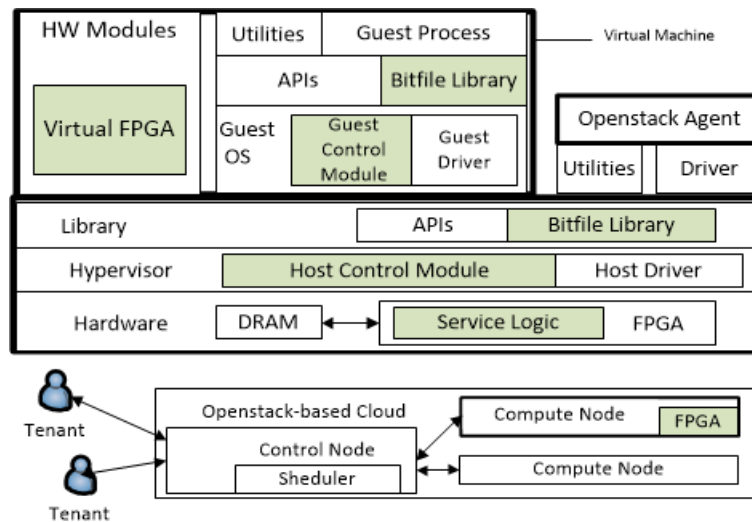


Рис. 3. Фреймворк FPGA у хмарі

Кілька вузлів обчислень, кожен з яких є фізичною машиною, надають фізичні ресурси, включаючи ЦП, пам'ять, диски та мережу. Керуючий вузол обробляє запити від користувачів, розподіляє ресурси та створює віртуальні машини (ВМ) на вибраних фізичних машинах. Після цього користувачі отримують доступ до своїх віртуальних машин та розгортають на них додатки.

Для введення FPGA в хмару ми надаємо нові модулі для вузлів обчислень, як показано сірим на рис. 2. Крім того, планувальник у керуючому вузлі OpenStack покращений. Щоб полегшити опис, ми розділили стек системи вузлів обчислень на 4 рівні: апаратний, гіпервізор, бібліотека та додаток.

Таксономія FPGA відноситься до системи класифікації, яка використовується для категоризації різних типів FGAs на основі їх властивостей та можливостей. Існують кілька способів класифікації FGAs, включаючи їх архітектуру, розмір, продуктивність та функціональність.

Один зі способів класифікації FPGA базується на їх архітектурі, яка може бути розподілена на три категорії: SRAM-базовані, anti-fuse-базовані та flash-базовані. SRAM-базовані FGAs використовують лічильники пам'яті для зберігання конфігураційних даних, тоді як anti-fuse-базовані FGAs використовують невідновлювальні плавки, які програмується один раз і не можуть бути перепрограмовані. Flash-базовані FGAs використовують клітини пам'яті flash для зберігання конфігураційних даних і можуть бути перепрограмовані багато разів.

FGAs також можна класифікувати за їх розміром та продуктивністю, які зазвичай вимірюються за кількістю логічних елементів, тактовою частотою та споживанням енергії. Малі FGAs зазвичай мають менше ніж 10 000 логічних елементів і використовуються для простих застосувань, тоді як великі FGAs можуть мати понад 10 мільйонів логічних елементів і використовуються для складних застосувань, які вимагають високої продуктивності. Інший спосіб класифікації FGAs базується на їх функціональності, яка може бути розподілена на загального призначення FGAs та спеціалізовані FPGA (ASPs) [8-9].

Іншими словами таксономія базується на критичних аспектах будь-якої хмарної системи FPGA: тип плат FPGA, розміщення FPGA в системі, підключення до мережі, підключення до вузла та випадки використання. Ці таксономічні категорії не є ні взаємовиключними, ні комплексними: системи можуть мати кілька підкатегорій, а нові підкатегорії додаватимуться пізніше для інкапсуляції майбутніх інновацій.

Далі розглянемо аспект використання FPGA-прискорювачів методів машинного навчання при роботі з ЦОД та хмарами.

Однією з проблем реалізації машинного навчання є не тільки поділ на апаратне та програмне забезпечення. Як критична складова будь-якого рішення у галузі машинного навчання, проектування апаратного забезпечення становить серйозну проблему перед постачальниками FPGA-прискорювачів, які повинні мати можливість реалізовувати свою продукцію клієнтам за рахунок інструментальних програмних засобів, що дозволяють абстрагуватися від апаратного забезпечення. Таким чином, для виведення нових FPGA-прискорювачів на ринок їх розробникам доводиться залучати і фахівців у галузі створення програмного забезпечення (ПЗ). Апаратні структури, які критично важливі для обчислень машинного навчання, можуть створюватися розробниками апаратного забезпечення до того моменту, коли подальше вдосконалення моделі ще можна здійснити без проектування апаратної частини. Один з варіантів реалізації досягається просто шляхом створення екземпляра складного функціонального блоку, за допомогою якого буде керуватися інструментальним засобом вищого рівня. Деякі фірми-постачальники FPGA-прискорювачів поділяють процес проектування таким чином, щоб проектування апаратного забезпечення було значною мірою відокремлено від проектування моделі. Завдяки цьому FPGA стають доступнішими для розробників, які не є фахівцями у даній галузі розробки апаратного забезпечення. Ще одне питання стосується того, які інструментальні засоби слід використовувати, і яким чином проектування апаратного забезпечення було значною мірою відокремлено від проектування моделі.

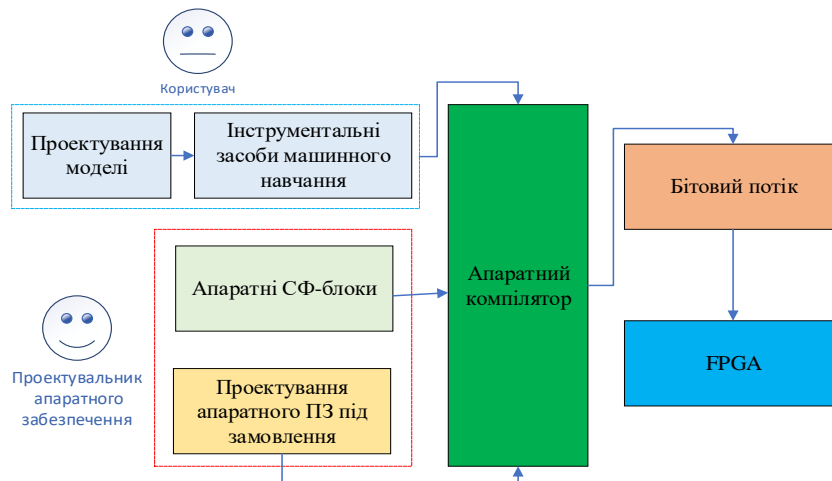


Рис. 4. Перша модель проектування додатків машинного навчання на FPGA

Детальна інформація про конструкцію FPGA зазвичай вказується в низькорівневому бітовому потоці, що завантажується в прилад. Але коли справа доходить до машинного навчання, деякі фірми реалізують у бітовому потоці всю модель, тоді як інші використовують його тільки для апаратної частини моделі, використовуючи для інформації про особливості моделі машинного

навчання програмований двійковий файл. На рис. 4 всі аспекти конструкції відображені в одному бітовому потоці. Розробник може бути здатним здійснювати проектування на високому рівні, але інструментальні засоби все одно пропускатимуть інформацію через апаратний компілятор. На рис. 5 дані моделі високого рівня виділені в окремий двійковий файл.

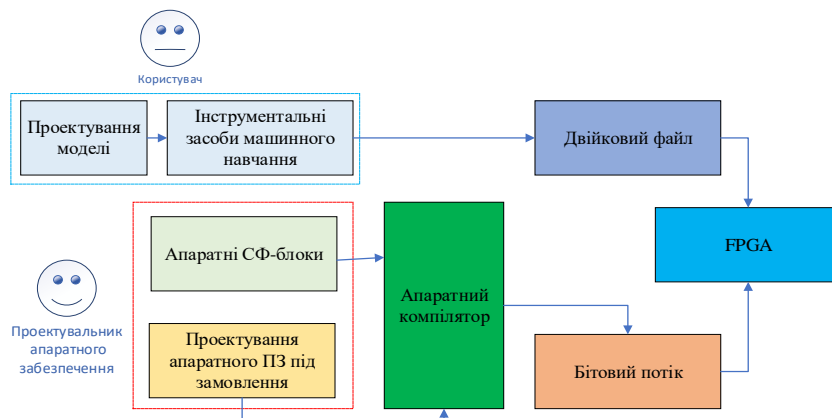


Рис. 5. Друга модель проектування додатків машинного навчання на FPGA

Сам момент при зміни моделі, якщо потрібна зміна бітового потоку апаратного забезпечення, визначається постачальником FPGA. Модель проектування має значення як для поточних модифікацій конструкції, та для оновлення приладів, які вже використовуються замовниками. У випадках, коли всі дані конструкції були включені в бітовий потік, можливі зміни в майбутньому для створення оновленого бітового потоку, який, в свою чергу, вимагатиме використання інструментальних апаратних засобів FPGA низького рівня, навіть якщо інструментальні засоби високого рівня дозволяють уникнути необхідності будь-якої явної зміни конструкції апаратного забезпечення. В інших випадках зміни в моделі призводять до змін лише у програмному двійковому файлі, який компілюється набагато швидше, ніж здійснюється повна рекомпіляція апаратного забезпечення. У цьому випадку основні апаратні засоби залишаються незмінними поряд із зміною аспектів моделі. Точна межа між змінами лише програмного забезпечення та

змінами апаратного забезпечення залежатиме від конкретної архітектури.

Висновки

Поточні інфраструктури обчислення в хмарі на базі FPGA ефективно використовують наявні плати FPGA, встановлюють їх у сервери та надають доступ до них кінцевим користувачам. Це дозволило багатьом провайдерам хмарних послуг швидко розгорнути ці плати FPGA у своїх датацентрах і надати доступ до них кінцевим користувачам. Однак, на всіх рівнях архітектури апаратного та програмного забезпечення є багато можливостей для покращення. Від нових апаратних конструкцій (наприклад, нової інтеграції FPGA у ЦП), до конструкцій системи (наприклад, вибору типу бампера-на-проводі в порівнянні з типом прискорювача) та програмного забезпечення та програмування (наприклад, використання HLS або нових абстракцій). Крім того, безпека залишається відкритим завданням, оскільки низькорівневий доступ

до апаратного забезпечення FPGA дозволяє зловмисникам створювати нові датчикові схеми, які можуть сприяти витоку інформації.

Зокрема, доступність FPGA в хмарних датацентрах відкрила безпрецедентні можливості щодо гнучкості та продуктивності застосувань. Але FPGA в хмарі також надають унікальні шляхи для зловживань з боку зловмисників на інших користувачів хмари або інфраструктуру, що не є можливим з CPU або GPU, базованих на обчисленнях в хмарі. Безпека залишається викликом, де архітектури повинні балансувати низькорівневий доступ, який дозволяє користувачам створювати власні потужні прискорювачі.

Основною перешкодою для широкого застосування FPGA в центрах обробки даних є відсутність стека програмного забезпечення, яке дозволяє легко розгортати, керувати та масштабувати FPGA в хмарі.

Надання стека програмного забезпечення, подібного до того, що привело до успіху центральних/графічних процесорів, збільшить диверсифікацію архітектури в хмарі, коли FPGA працюватимуть як рівноправні партнери на боці центральних і графічних процесорів.

У роботі був зроблений огляд архітектур хмарних FPGA, який показує складні та нетривіальні взаємозв'язки між вимогами системи та конфігураціями розгортання, і визначає можливі напрямки майбутньої інновації в цьому просторі. Для цього, було використано поняття таксономії, яке абстрагує низькорівневі деталі реалізації, при цьому виділяючи переваги та обмеження даної архітектури. Використовуючи цю таксономію, ми класифікували як виробничі, так і дослідницькі системи, що демонструє основні тенденції в архітектурі хмарних FPGA.

СПИСОК ЛІТЕРАТУРИ

1. A. Putnam, A. M. Caulfield, E. S. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmailzadeh, J. Fowers, G. P. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Y. Xiao, and D. Burger. 2014. A reconfigurable fabric for accelerating large-scale data center services. In ACM/IEEE 41st International Symposium on Computer Architecture (ISCA). 13–24. DOI:<https://doi.org/10.1109/ISCA.2014.6853195>.
2. A. M. Caulfield, E. S. Chung, A. Putnam, H. Angepat, Jeremy Fowers, Michael Haselman, Stephen Heil, Matt Humphrey, Puneet Kaur, Joo-Young Kim, Daniel Lo, Todd Massengill, Kalin Ovtcharov, Lisa Woods, Sitaram Lanka, Derek Chiou, and Doug Burger. 2016. A cloud-scale acceleration architecture. In 49th IEEE/ACM Int. Symp. on Microarchitecture. 1–13.
3. J. Fowers, G. Brown, P. Cooke, and G. Stitt, "A performance and energy comparison of FPGAs, GPUs, and multicores for sliding-window applications," in Proceedings of the ACM/ SIGDA International Symposium on Field Programmable Gate Arrays, pp. 47–56, ACM, Monterey, CA, USA, February 2012.
4. Chethan Ramesh, Shivukumar B. Patil, Siva Nishok Dhanuskodi, George Provelengios, Sébastien Pillement, Daniel Holcomb, and Russell Tessier. 2018. FPGA side channel attacks without physical access. In IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). 45–52.
5. Kasper Rasmussen, Ilias Giechaskiel, and Ken Eguro. 2019. Leakier wires: Exploiting FPGA long wires for covert-and side-channel attacks. ACM Trans. Reconfig. Technol. Syst. 12, 3 (2019).
6. Ilias Giechaskiel, Kasper Rasmussen, and Jakub Szefer. 2020. CAPSULE: Cross-FPGA covert-channel attacks through power supply unit leakage. In IEEE Symposium on Security and Privacy. 1728–1741.
7. George Provelengios, Daniel Holcomb, and Russell Tessier. 2019. Characterizing power distribution attacks in multi-user FPGA environments. In International Conference on Field Programmable Logic and Applications (FPL). 194–201.
8. Xilinx Case Study. [n.d.]. Xilinx Powers Alibaba Cloud FaaS with AI Acceleration Solution for E-Commerce Business. Retrieved from <https://www.xilinx.com/publications/powered-by-xilinx/xilinx-alibaba-case-study.pdf>.
9. Norihisa Fujita, Ryohei Kobayashi, Yoshiki Yamaguchi, and Taisuke Boku. 2019. Parallel processing on FPGA combining computation and communication in OpenCL programming. In IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). IEEE, 479–488.
10. Mohamed S Abdelfattah, David Han, Andrew Bitar, Roberto DiCecco, Shane O'Connell, Nitika Shanker, Joseph Chu, Ian Prins, Joshua Fender, Andrew C Ling, et al. DLA: Compiler and fpga overlay for neural network inference acceleration. In 2018 28th International Conference on Field Programmable Logic and Applications (FPL), pages 411–4117. IEEE, 2018.

Received (Надійшла) 31.01.2023

Accepted for publication (Прийнята до друку) 19.04.2023

Future of FPGA - accelerating computations in data processing centers and clouds

Oleksandr Ilyashov, Kostiantyn Pokora, Vladislav Diachenko, Andriy Kovalenko

Abstract. Relevance. Data analysis is often performed using machine learning methods. Often, the algorithms involved need to deal with large datasets, which leads to long execution times. Therefore, research into hardware accelerators based on field-programmable gate arrays (FPGAs) to improve performance is relevant. FPGAs are a promising solution for hardware acceleration, post-production configuration, and reprogramming capabilities. **The purpose** of this study is to investigate and analyze trends in existing cloud FPGA architectures, which highlight the complex relationship between architectures and system requirements. This allows us to identify new architectures that are likely to offer significant benefits for cloud workloads. **The object** of the study is the evolution of FPGA accelerators for data center (DC) and cloud computing. **The subject** of the study is methods and algorithms for researching cloud FPGA architectures based on taxonomic categories. **Results.** The paper discusses the future use of FPGAs in data centers and clouds. Current architectures are also investigated and scalability and abstractions supported by operating systems, middleware, and virtualization are discussed. **Conclusion.** A classification of cloud FPGA architectures based on taxonomic categories has been developed. An architectural organization for deploying FPGA applications used in cloud environments and data center environments is considered and proposed.

Keywords: reconfigurable logic, FPGA acceleration, cloud computing, data center, virtualization, taxonomic categories, data classification, machine learning, software.

А. О. Каргін^{1,2}, С. В. Жуков¹, Д. А. Сергєєв¹, Є. Л. Сілін²

¹Харківський національний технічний університет радіоелектроніки, Харків, Україна

²Український державний університет залізничного транспорту, Харків, Україна

МОДЕЛЬ БАЗОВОГО РІВНЯ ШТУЧНОГО ОТОЧЕННЯ АВТОНОМНИХ ІНТЕЛЕКТУАЛЬНИХ БЕЗЛЮДНИХ СИСТЕМ НА ПРИКЛАДІ МОБІЛЬНОГО РОБОТУ ЩО ОБСЛУГОВУЄ

Анотація. Мотивація. Незважаючи на значний прогрес у галузі створення безлюдних систем, забезпечення необхідного рівня їх автономності залишається актуальним завданням. Важливу роль у його вирішенні відіграє штучний інтелект. Особливості безлюдних систем породили нову модель «Штучний інтелект що відчуває» (ШІВ), що підтримує автономність. **Метою даної роботи** є створення штучного оточення для експериментів з безлюдними системами, підтримуваними ШІВ, на прикладі складу контейнерів, що обслуговується колесним роботом. **Предметом дослідження** є методи та моделі управління автономними мобільними роботами на підставі даних від різноманітних сенсорів. **Результати.** Описується архітектура штучного оточення, базові компоненти системи що мають багатошарову організацію. На прикладі штучного оточення «Склад контейнерів», як прототипу автономної безлюдної системи що обслуговується колесним роботом, показано набір контролерів, приклад схем підключення сенсорів та актуаторів, методи та алгоритми управління, які потрібні на базовому рівні. Наведено обґрунтування методу ідентифікації номерів контейнерів що задовольняє вимогам, що накладаються автономними системами. **Висновок.** Доведено що архітектура базового рівня штучного оточення має бути універсальною щодо підтримки різних методів й алгоритмів управління на базі різноманітних даних від сенсорів. Універсальність досягається завдяки багатошаровій організації контролерів за для підтримки функцій ШІВ. Експерименти зі контролерами базового рівня штучного оточення «Склад контейнерів» продемонстрували можливість підвищення рівня автономності безлюдної системи за рахунок розширених можливостей використання в управляючій програмі робота різноманітних методів управління на різних етапах плану із урахуванням поточної ситуації, що оцінюється на множині даних від різноманітних сенсорів. **Напрямок подальших досліджень.** Майбутні кроки можуть бути присвячені розвитку моделі та інструментарію для інших класів безлюдних систем.

Ключові слова: автономні інтелектуальні безлюдні системи, штучний інтелект що відчуває, штучне оточення, програмне управління, робот, сенсор, розпізнавання номерів контейнерів.

Вступ

В сучасному світі транспортування товарів є важливою складовою бізнесу в багатьох галузях промисловості. Товари що транспортуються упаковуються у контейнери. Технологія транспортування передбачає проміжні етапи тимчасового зберігання контейнерів на складах, обслуговування яких автоматизовано. Більшість сучасних складів відносяться до безлюдних систем (БС) [1, 2]. Автоматичні склади, як БС, передбачають автономне функціонування у тому числі й в непередбачених ситуаціях, коли системи керування різноманітним обладнанням складу повинні приймати рішення без втручання людини. Такі склади в залежності від спеціалізації обладнанні мобільними роботами, які виконують, як завдання по транспортуванню контейнерів, так й допоміжні функції, наприклад, додатковий облік й контроль наявності певних контейнерів на певних місцях й таке інше [2]. Мобільні роботи (МР) мають переміщуватися до вказаної позиції складу по певному маршруту й використовуючи зорову камеру ідентифікувати місце, наявність контейнеру певного типу й його номер для подальшої обробки системою обліку. За для того, щоб не порушувати цикли основних технологічних процесів МР повинні орієнтуватися в поточній ситуації, уникати аврійних ситуацій й виконувати своє завдання в автономному режимі.

Незважаючи на гігантський прогрес у галузі створення БС, таких як безпілотні автомобілі [3], апарати військового призначення [4] та інші системи, забезпечення необхідного рівня автономності БС

залишається актуальним завданням [5]. Важливу роль у вирішенні цього завдання відіграє штучний інтелект (ШІ). Автономні системи з підтримкою штучного інтелекту пройшли у своєму розвитку етапи програмування автономних систем та інтелектуальних БС. Сьогодні є актуальними автономні інтелектуальні БС (АІБС) [5, 6]. На ШІ в АІБС покладають рішення багатьох завдань з обробки інформації: отримання та первинна обробка даних від сенсорів, формування моделі оточення, планування дій у цьому оточенні та прийняття рішень, що призводять до досягнення мети, реалізація реакцій, що управляють, навчання та адаптація до змін в оточенні [7, 8]. Ця специфіка знайшла відображення у класифікації ШІ, який пропонується для АІБС. В роботах [5, 9] вводиться три типи штучного інтелекту – механічний, мислячий і той що відчуває. Механічний ШІ використовується для простих, стандартизованих, повторюваних і рутинних завдань. Мислячий ШІ використовується для складних, систематичних, заснованих на правилах і чітко визначених завдань. ШІ що відчуває, зокрема у військовій сфері, обговорюється у зв'язку з необхідністю АІБС вирішувати соціальні, емоційні, комунікативні та інтерактивні завдання. В роботі [10] пропонується модель і проект архітектури ШІВ не з точки зору прийняття рішень, які відповідають соціальним людським цінностям, а з технічної точки зору для розширення можливостей ШІВ щодо забезпечення необхідного рівня автономії АІБС.

Для тестування моделей ШІВ потрібні різного роду полігони [11], де створюється штучне оточення АІБС що моделює різноманітне застосування БС.

Метою цієї роботи є розробка штучного оточення АІБС на прикладі автоматизованого складу контейнерів. МР, як складова компонента автоматизованого складу контейнерів, розглядається у штучному оточенні як прототип АІБС. На базі колесного робота потрібно створити фізичну й програмну модель штучного оточення «Склад контейнерів» для подальших експериментів з АІБС що вирішує завдання аналізу ситуацій, ситуаційного планування маршруту й управління переміщеннями в умовах перешкод, розпізнаванню номерів контейнерів у різних умовах експлуатації.

Основна частина

Архітектура штучного оточення «Склад контейнерів». Створюється штучне оточення на прикладі автоматизованого сховища з різними секціями для вантажних контейнерів. Загальні вимоги до штучного оточення наступні.

У секціях сховища встановлені маркери відповідних номерів вантажних місць де зберігаються контейнери (ідентифікатори місця секції) та системи освітлення для забезпечення якісного зображення номеру контейнера у будь який час доби [1]. Для забезпечення безперебійної роботи МР, необхідно мати доступ до електричної мережі (позиції зарядки акумуляторних батарей, зарядної станції), а також бездротову мережу зв'язку, щоб передавати отримані дані для подальшої обробки та аналізу. Важливо також щоб МР мав права доступу до системи управління сховищем, щоб забезпечити швидкий й безпечний доступ к хмарним сховищам даних.

Для виконання основних функцій МР створюється штучна секція сховища для вантажних контейнерів. Приміщення секції поділяється на 3 основні зони з маршрутами «А-а», «А-б» та «А-с» та нумеровані вантажні місця від нуля до шести. Робот може курсувати між рядами, до будь-якого з контейнерів за маршрутом зображеним на рис. 1.

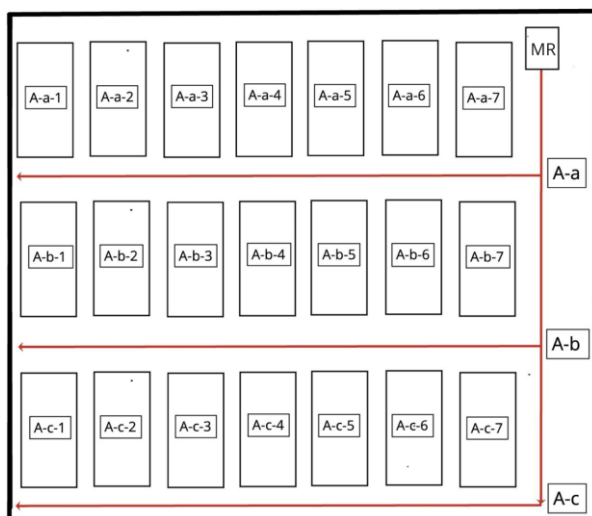


Рис. 1. Схема секції складу вантажних контейнерів

Підлога має розмітку для орієнтування МР. Робот, позначений на рисунку літерою «MR», має доступ до кожного вантажного місця. Вони позначені пунк-

тирною лінією – це є місце де, як вважається, контейнери мають стояти. Надписом «СН» позначена зарядна станція робота, яка винесена у окрему зону у приміщенні. Секція має одну зарядну станцію. МР може обслуговувати декілька секцій упродовж дня, періодично заряджаючись між патрулюванням та перевірки.

Для експериментів обрано два типу МР з різним апаратним забезпеченням. Перший тип МР на базі чотирьох колісного повнопривідного шасі Multi Chassis-4WD Robot Kit ATV, мікроконтролеру Arduino Mega з набором Motor shield, модулем wi-fi esp8266, матрицею із 10-ти інфрачервоних датчиків відображення ku-033 для управління переміщеннями вздовж розмітки на підлозі, ультразвуковим датчиком HC-SR04, встановленим на поворотній платформі з сервопривідом SG90, та датчиком одометром H206.

Другий тип МР реалізований на основі Raspberry Pi навчальний набір (PiCar-V для побудови роботів з відкритим вихідним кодом, оснащений ширококутною USB веб-камерою та укомплектований трьома новими платами мікроконтролеру Raspberry Pi, яка має вбудований порт для камери з високою роздільною здатністю що дозволяє працювати з відео та фото без додаткових інструментів або пакетного програмного забезпечення). Плата має достатньо велику оперативну пам'ять для складних обчислень, такі як зчитування інформації з зображення у режимі реального часу. Операційна система Linux прискорює процес розробки окремих компонентів системи на більш релевантних для типу задачі мовах.

Компоненти системи управління переміщеннями мобільного робота. ШІВ має шарову архітектуру [11, 12]. До нижнього першого шару належать програмні інтерфейси з апаратним забезпеченням АІБС, тобто драйвери сенсорів й актуаторів. Для штучного оточення що розглядається в данній роботі це драйвери датчику одометра, ультразвукового датчику відстані та матриці датчиків відображення що влаштовані на днищі шасі робота, а також драйвери сервопривіду поворотної платформи та приводів двигунів Motor Shield. До другого шару належать контролери руху. Це контролер руху вздовж розмітки на підлозі (K1), контролер руху вздовж штучної огорожі (K2), контролер руху до візуального маркера по вільній траєкторії (K3) та контролери відстані (S) й швидкості (V) руху. До третього шару належить контролер програмного управління рухом що реалізує план (програму в термінах управління роботами) досягнення кінцяї цільової точки. І на верхньому четвертому шарі знаходяться інтелектуальні компоненти прийняття рішень в умовах неповної інформації [12] що застосовують когнітивні моделі сприйняття даних від сенсорів та емоційної оцінки ситуації для прийняття рішень [10].

Перелічені вище контролери розроблені на мові C++ в середовищі Arduino IDE. Роль драйверів відіграють вбудовані функції та бібліотеки Arduino IDE. Розглянемо організацію драйверів й контролерів на прикладі одометра. При переміщеннях колесного робота одометр збирає дані про обертання колес. Дані обробляються мікроконтролером для визначення пройденої відстані та швидкості робота.

Серед трьох типів одометрів (механічні, електро-механічні та електронні) обрано електронний що представляє собою інфрачервоний датчик H206 та диск з 20 отворами, який кріпиться на вісь двигуна (рис. 2).

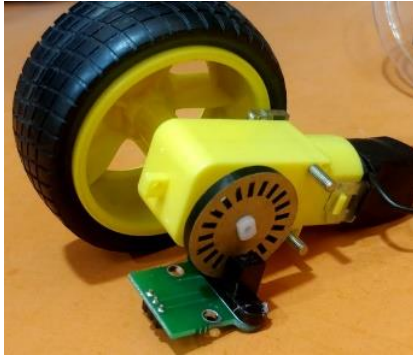


Рис. 2. Апаратна реалізація датчика одометра

Диск розміщується між інфрачервоним випромінювачем та приймачем (датчик H206) і при кожному проходженні отворів диска датчик реагує на зміну його сигналу. Оскільки отворів у диска 20, а датчик реагує на оби дві зміни сигналу, то за час повного обороту колеса МР мікроконтролер нарахує 40 переривань. Відстань, пройдена роботом за один оберт колеса, дорівнюватиме довжині його кола. Для радіуса колеса r пройдена відстань обчислюється за формулою:

$$S = 2\pi r * (n/40), \quad (1)$$

де n – кількість переривань.

Датчик H206 підключається до піну переривання (цифрові піни 2 та 3 на рис. 3) і його показання фіксує вбудована функція *attachInterrupt()*.

Таким чином роль драйвера датчика одометр виконує ця функція.

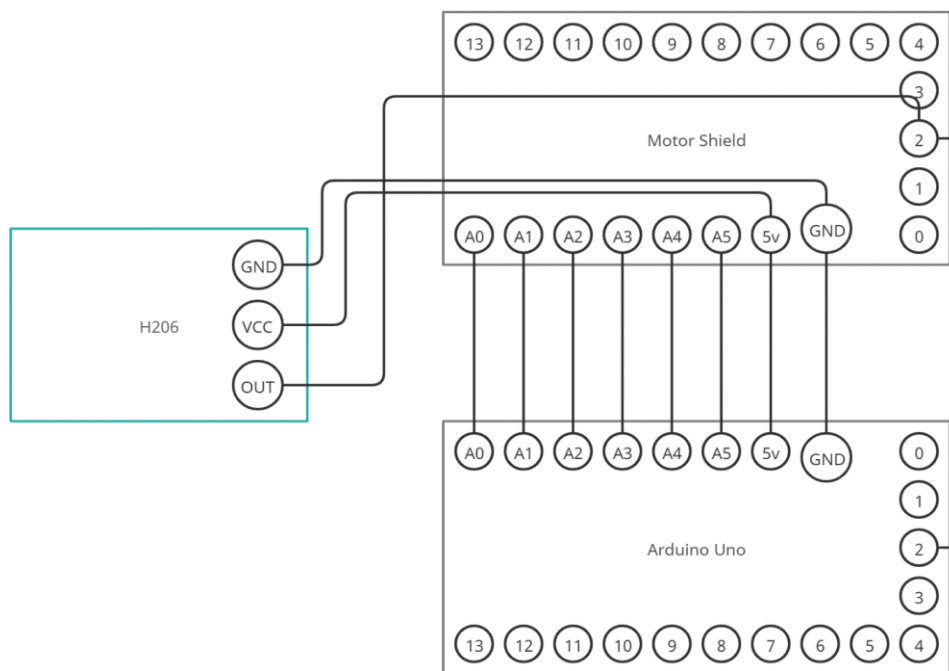


Рис. 3. Схема з'єднання компонентів на базі датчика H206

Контролер відстані видає по запиті поточне значення пройденної відстані (1) в метрах, реалізує функції обнулення $n=0$ в (1) та переривання при досягненні значення відстані S^* , яке задається.

Контролер швидкості підтримує поточну лінійну швидкість руху роботу рівною заданому значенню v^* . Використовується функція *millis()*, значення якої є кількість мілісекунду починаючи з часу подачі живлення на плату Arduino. Підраховується кількість мілісекунд t на кожний повний оберт колеса й обчислюється миттєва швидкість v в одиницях м/сек:

$$v = 2\pi r * 1000/t. \quad (2)$$

Знаходиться похибка управління

$$e = v * -v \quad (3)$$

й застосовуючи один з алгоритмів управління зі зворотними зв'язками корегуються оберти колес.

Контролери руху реалізують управління переміщеннями роботу окремих етапів плану переміщення. Контролери руху вздовж розмітки на підлозі використовує дані від драйверу матриці датчиків відображення на їх підставі обчислюється відхилення робота від лінії розмітки, тобто похибка управління, й на її основі знаходить значення управляючого сигналу. Аналогічно два інших контролерів руху обчислюються похибки управління на основі даних від ультразвукового датчику й відео камери, відповідно. Контролери реалізують управління зі зворотними зв'язками застосовуючи налаштовані ПІД алгоритм чи нечіткі алгоритми, такі як нечіткий лінійний статичний із зоною нечутливості й насиченням, нечіткий трьохпозиційний, нечіткий нелінійний із гістерезисом й насиченням, нечіткий ПІД.

Контролер програмного управління виконує функцію менеджера окремих етапів плану досягнення

цільової позиції. Маршрут від поточної точки місцезнаходження робота до цільової розбивається на окремі етапи, послідовність яких уявляє управляючу програму роботи. Контролер програмного управління активізує етапи управляючої програми шляхом активізації команди що вказана, контролює завершення виконання етапу (успішне чи аврійне), та приймає рішення щодо активізації наступного етапу. В управляючій програмі (УП) можуть бути етапи з різними методами управління, на відміну від класичного програмного управління. Етапи що реалізують жорстке програмне управління, коли умовою завершення етапу є час (в опису етапу УП задається інтервал часу) можуть перемішуватися з етапами що реалізують гнучке програмне управління, коли умовою завершення етапу є пройдена відстань, чи відстань до перешкоди, чи наявність певного маркера що ідентифікується камерою зору. Контролер цього рівня виконує моніторинг даних від відповідних датчиків (одометр, ультразвуковий та відео) й порівнює поточні дані з заданими в УП. При виконанні умов, вказаних в описі етапу, ініціюється завершення етапу УП. Жорстке програмне управління порівняно з гнучким менш ресурсоємне але може застосовуватися в умовах відсутності перешкод за підтримки контролерів нижчого шару що мають високі значення критерію якості управління.

Обов'язковою компонентою що належить до четвертого шару компонент зі штучним інтелектом є планувальник завдань, який в залежності від стану (ситуації) штучного оточення генерує УП, як послідовність етапів досягнення цілі. Планувальник завдань реалізує метод безперервного планування за прецедентами [10, 12]. Безперервне планування передбачає що після завершення кожного етапу УП (успішно чи аварійно) планувальник виходячи з поточної ситуації по новому генерує план досягнення цілі. Такий підхід забезпечує високий рівень автономності прийняття рішень в умовах збурень та неповної інформації.

Взаємозв'язки різного рівня контролерів при реалізації МР функції контролю в штучному оточенні «Склад контейнерів» розглянемо на наступному

прикладі. Нехай при виконанні УП після завершення поточного етапу МР знаходиться у позиції A-a-3 (рис.4).

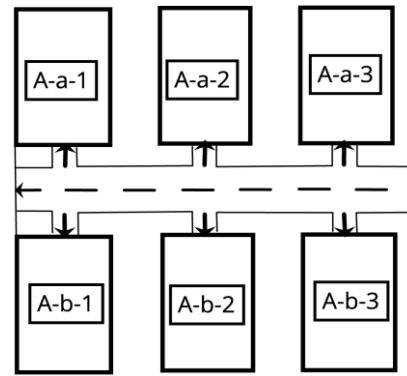


Рис. 4. Приклад маршруту переміщення

Планувальник сгенерував УП переміщення МР до вантажної позиції A-b-1. Приклад трьох типів УП наведено в табл. 1. Методи програмного управління роботом демонструються на прикладі обмеженого набору команд управління: вперед, праворуч, ліворуч. В методі жорсткого програмного управління видача команди викликає початок відповідного руху. Цей рух триває до тих пір, поки не буде видана наступна команда вимкнути попередньо увімкнутий рух. Для методу жорсткого програмного управління вважається, що переміщення здійснюються з постійними швидкостями й тривають константний час. Наприклад, t_{A-a-3} секунд потрібно часу для переміщення МР від місця моніторингу номеру контейнера на позиції A-a-3 (рис.4) до лінії розмітки A-a (рис. 1). До складу команд методу жорсткого програмного управління (інтерпретація програми управління роботом - жорстка послідовність виконання команд: i-a по порядку команда в програмі не може почати виконуватися, доки не буде завершено виконання i-1-ї команди) обов'язково входить затримки, наприклад, *затримка* t_{A-a-3} з параметром часу. Інтерпретатор команд контролеру при обробці цієї команди затримує на t_{A-a-3} одиниць часу початок виконання наступної команди.

Таблиця 1 – Приклади управляючих програм

Жорстке програмне управління		Гнучке програмне управління	Гнучке ситуаційне програмне управління
<i>вперед</i>	<i>затримка</i> t_{1-3}	<i>вперед</i> K1 $S=120$	<i>вперед</i> K1 $R=f, S=50, V=1$
<i>затримка</i> t_{A-a-3}	<i>вимк_вперед</i>	<i>праворуч</i> $\gamma=90^\circ$	<i>праворуч</i> $\gamma=90^\circ$
<i>вимк_вперед</i>	<i>ліворуч</i>	<i>вперед</i> K1 $S=500$	<i>вперед</i> K2 $R=h, S=50, V=2$
<i>праворуч</i>	<i>затримка</i> t_{90°	<i>ліворуч</i> $\gamma=90^\circ$	<i>ліворуч</i> $\gamma=90^\circ$
<i>затримка</i> t_{90°	<i>вперед</i>	<i>вперед</i> K1 $S=120$	<i>вперед</i> K1 $R=k, S=50, V=1$
<i>вимк_праворуч</i>	<i>затримка</i> t_{A-b-1}	<i>стоп</i>	<i>стоп</i>
<i>вперед</i>	<i>вимк_вперед</i>		
	<i>стоп</i>		

Метод жорсткого програмного управління вимогливі до попереднього впорядкування середовища: обов'язкові жорсткі вимоги до швидкості переміщення і миттєві режими розгону/гальмування робота. Ці вимоги для колесних роботів практично ніколи не можливо виконати. У методі програмного управління із зворотними зв'язками (гнучкого

програмного управління) для ідентифікації події виходу робота на певні позиції використані дані від датчиків, а не інтервали часу. В табл. 1 наведено приклад УП, в якій використовуються дані від одометрів (колес роботу ($S=120\text{см}$) і сервоприводу розвороту ($\gamma=90^\circ$). В УП указано що використовується контролер управління K1 (рух вздовж розмітки на підлозі на

підставі даних від датчиків відображення). Умова вимикання команди задається конкретним значенням пройдені МР відстані, наприклад, $S=120\text{см}$ у першій команді *вперед* в Таблиці 1. Контролер К1, отримуючи у реальному часі дані від драйверу одометра, розраховує поточну відстань (1) й видає сигнал вимкнути_вперед, коли значення відстані дорівнює чи більш вказаного, тобто, $S \geq 120$.

В методі гнучкого ситуаційного програмного управління є можливість задавати будь яку ситуацію й будь який контролер. В прикладі (Таблиця 1) на першому етапі УП використовується контролер К1, третій – контролер К2, п'ятий - знову К1. При реалізації першого етапу, крім контролера К1, паралельно працюють контролери швидкості ($V=I$) й відстані ($S=50$) та контролер руху вздовж смуги, який додатково відстежує наявність маркеру f на підлозі, що позначає перехрестя лінії розмітки від контейнера $A-a$



Рис. 5. Приклади контейнерів й область з їх номером

Перший метод на основі бібліотек OpenALPR (Automatic License Plate Recognition), алгоритму Haar Cascade та бібліотеки OpenCV. Бібліотека OpenALPR використовується в системі розпізнавання різного типу номерів, тому числі контейнерів. OpenALPR використовує навчальні дані, що містять інформацію про різні шрифти та символи знаків з різних країн. Бібліотека використовує алгоритми комп'ютерного зору для локалізації номерних знаків на зображенні та витягування з них текстової інформації. Вона налаштована для використання у режимі реального часу, що дозволяє розпізнавати номерні знаки на відео з використанням веб-камер. OpenCV - бібліотека відкритого коду для комп'ютерного зору та обробки зображень. Використовується для розв'язання різноманітних задач, пов'язаних з обробкою зображень, таких як детекція облич, розпізнавання тексту, відслідковування об'єктів та інші [13].

Запропоновано на отриманому зображенні з камери, використовуючи OpenCV та Haar Cascade Classifier, знаходити зображення номеру контейнера на відео та вирізати цей фрагмент, зберігаючи як окремий файл зображення. Якщо область успішно визначена, фрагмент передається до OpenALPR для розпізнавання. OpenALPR використовує неймерережеві алгоритми, щоб розпізнати цифри номеру контейнера та перетворити фрагмент у текстовий формат. Після цього результат розпізнавання номеру контейнера передається до програмного забезпечення автоматизованого складу.

Другий метод на основі Tesseract (двигун розроблений Google, що може бути використаний на

3 (рис. 4) та лінії розмітки $A-a$ (рис.1). Ця умова укавана в команді у вигляді $R=f$.

Методи ідентифікації номеру контейнера.

Позиціонування МР на вантажному місті секції виконується за розміткою на підлозі. Номер місця вказан кодом розмітки пілоги, який зчитується матрицею датчиків відображення. Ідентифікація номеру контейнера розбита на дві фази. Перша - позиціонування поворотної платформи на якій встановлена веб камера на контейнер та зчитування зображення. Друга фаза вже є безпосередньо фаза обробки зображення. Це локалізація контейнера на зображенні, локалізація області на контейнері та розпізнавання номеру в локалізованій області.

Були проаналізовані три підходи обробки даних від камери, під'єднаної до мікрокомп'ютера Raspberry Pi 3B+, для вирішення завдання розпізнавання номеру контейнеру, що зображено на рис. 5.

Raspberry Pi для розпізнавання тексту, включаючи номери контейнерів). Можна використовувати різні алгоритми машинного навчання, такі як згорткові або рекурентні нейронні мережі для навчання моделей Raspberry Pi розпізнаванню номерів контейнерів на зображеннях.

Розглянута наступна схема обробки даних. Препроцесинг зображень (операції попередньої обробки зображень, такі як бінаризація, видалення шуму, розпізнавання ліній тощо). Детекція тексту (виявлення регіону з текстом на зображенні, визначення координат розташування текстових блоків, рядків та окремих символів). Розпізнавання тексту (рекурентні нейронні мережі для розпізнавання тексту шляхом аналізу текстових блоків, рядків та символів, що дозволяє розпізнавати текст з варіаціями шрифту, розміру, кутів та інших атрибутів).

Третій метод на основі Tesseract TensorFlow та Keras. Ці популярні бібліотеки машинного навчання можуть бути використані для розпізнавання номерів контейнерів на зображеннях. Вони забезпечують широкий спектр можливостей для розробки та тренування власних моделей розпізнавання номерів контейнерів, включаючи використання камери Raspberry Pi як джерела зображень [14].

Висновки

Для дослідження можливостей нового напрямку штучного інтелекту, а саме штучний інтелект що відчуває, призначеного для підтримки автономності безлюдних систем, створюється штучне оточення. АІБС взаємодіє зі своїм оточенням на базовому рівні

шляхом отримання даних від сенсорів про стан оточення й впливу на стан оточення через свої актуатори. Доведено що архітектура базового рівня штучного оточення має бути універсальною щодо підтримки різних методів й алгоритмів управління на базі різноманітних даних від сенсорів. Універсальність досягається завдяки багатопаровій організації контролерів. На прикладі штучного оточення «Склад контейнерів», як прототипу АІБС показно набір контролерів, приклад схем підключення сенсорів та

актуаторів, методи та алгоритми управління, які потрібні на базовому рівні щоб підтримувати функції ШІВ. Експерименти зі контролерами базового рівня штучного оточення «Склад контейнерів» продемонстрували можливість підвищення рівня автономності безлюдної системи за рахунок розширених можливостей використання в управляючій програмі робота різноманітних методів управління на різних етапах плану із урахуванням поточної ситуації, що оцінюється на множині даних від різноманітних сенсорів.

СПИСОК ЛІТЕРАТУРИ

1. Функціональне призначення та класифікація складів в логістиці // WareTeKa News, 13.11.2020, <https://wareteka.com.ua/uk/blog/klassifikaciya-skladiv-v-logistyki/>
2. Rasmussen S, Kingston D, Humphrey L (2018) Brief Introduction to Unmanned Systems Autonomy Services (UxAS). Int. Conf. on Unmanned Aircraft Sys. (ICUAS). <https://doi.org/10.1109/ICUAS.2018.8453287>
3. Dimitrakos TD, Kyriakidis EG (2008), "A semi-Markov decision algorithm for the maintenance of a production system with buffer capacity and continuous repair times", International Journal of Production Economics, Vol. 111(2), pp. 752-762, doi: <https://doi.org/10.1016/j.ijpe.2007.03.010>
4. T. A. Litman, "Autonomous Vehicle Implementation Predictions: Implications for Transport Planning," Victoria Transport Policy Inst., Rep., Aug. 2022.
5. H. Shakhathreh et al., "Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges," J. IEEE Access, vol. 7, pp. 48572–48634, Apr. 2019, doi: 10.1109/ACCESS.2019.2909530.
6. T. Zhang et al., "Current trends in the development of intelligent unmanned autonomous systems," Frontiers Inf. Technol. Electron. Eng., vol. 18, Feb. 2017, pp. 68–85, doi: 10.1631/FITEE.1601650.
7. J. Reis, Y. Cohen, N. Melao, J. Costa, and D. Jorge, "High-Tech Defense Industries: Developing Autonomous Intelligent Systems," Appl. Sci., vol. 11, 4920, 2021, doi: 10.3390/app11114920.
8. M.Huang and R.Rust, "Artificial Intelligence in Service," J. of Service Res., vol. 21(2), Feb. 2018, pp.155-172, doi: 10.1177/1094670517752459.
9. M. Czerwinski, J. Hernandez, D. McDuff, "Building an AI that feels," Appl. Sci., vol.11, 4920, Apr. 2021, doi: 10.3390/app11114920.
10. A. Kargin, T. Petrenko, "Feeling Artificial Intelligence for AI-Enabled Autonomous Systems" in Conf. Proc. of 2022 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT) Alamein New City, Egypt, 18-21 December 2022, P.88-93.
11. A. Kargin and T. Petrenko, "Spatio-Temporal Data Interpretation Based on Perceptual Model," in Advances in Spatio-Temporal Segmentation of Visual Data. Studies in Computational Intelligence, V. Mashtalir, I. Ruban, V. Levashenko, Eds., vol. 876, Springer, Cham, 2020, pp. 101-159.
12. A. Kargin and T. Petrenko, "Multi-level Computing With Words Model to Autonomous Systems Control," in Proc. 9th Int. Conf. Inf. Control Sys.&Tech (ICST-2020), A. Pakštas, T. Hovorushchenko, H. Yin, N. Rudnichenko. Eds. Odessa, Ukraine, Sep. 24–26, 2020, CEUR Workshop Proceedings, vol. 2711, pp. 16-30. [Online]. Available: <http://ceur-ws.org/Vol-2711/>
13. Келер А., Брэдски Г. Изучаем OpenCV 3: навч. посіб. ДМК Пресс, 2017. 826 с.
14. Keras. Simple. Flexible. Powerful. [Online]. Available: <https://keras.io/>

Received (Надійшла) 15.03.2023

Accepted for publication (Прийнята до друку) 17.05.2023

A basic-level model of the artificial environment of autonomous intelligent unmanned systems as an example of a mobile robot that serves

Anatolii Kargin, Stanislav Zhukov, Danilo Sergeev, Yevhenii Silin

Abstract. Motivation. Despite significant progress in the field of creating unmanned systems, ensuring the necessary level of their autonomy remains an topicality task. Artificial intelligence plays an important role in its solution. Features of unmanned systems gave rise to a new model of Feeling Artificial Intelligence (FAI) that supports autonomy. **The goal of this work** is to create an artificial environment model for experiments with unmanned systems supported by FAI using the example of a container warehouse serviced by a wheeled robot. **The object of research** is the methods and models of controlling autonomous mobile robots based on data from various sensors. **Results.** The architecture of the artificial environment, the basic components of the system with a multi-layered organization are described. Using the example of the artificial environment "Container Warehouse" as a prototype of an autonomous unmanned system serviced by a wheeled robot, a set of controllers, an example of sensor and actuator connection schemes, control methods and algorithms that are required at the basic level are shown. The justification of the method of identification of container numbers, which satisfies the requirements imposed by autonomous systems, is presented. **Conclusions.** It is proven that the architecture of the basic level of the artificial environment should be universal in terms of supporting various control methods and algorithms based on various data from sensors. Versatility is achieved due to the multi-layered organization of controllers to support the functions of FAI. Experiments with controllers of the basic level of the artificial environment "Container Warehouse" demonstrated the possibility of increasing the level of autonomy of the unmanned system due to the expanded possibilities of using various control methods in the robot's control program at different stages of the plan, taking into account the current situation, which is evaluated on the basis of a set of data from various sensors.

Keywords: autonomous intelligent unmanned systems, artificial environment, programming control, robot, sensor, recognition of container numbers.

Nina Kuchuk¹, Yurii Zdorenko², Wang Shengquan¹

¹ National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine

² National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine

A LIGHTWEIGHT TARGET DETECTION ALGORITHM BASED ON MOBILENET CONVOLUTION

Abstract. Target detection algorithm based on deep learning needs high computer GPU configuration, even need to use high performance deep learning workstation, this not only makes the cost increase, also greatly limits the realizability of the ground, this paper introduces a kind of lightweight algorithm for target detection under the condition of the balance accuracy and computational efficiency, MobileNet as Backbone performs parameter The processing speed is 30fps on the RTX2060 card for images with the CNN separator layer. The processing speed is 30fps on the RTX2060 card for images with a resolution of 320×320.

Keywords: deep learning; MobileNet; global Average pooling layer; GPU; target detection.

Introduction

Target detection is important for understanding image content and finding target objects. This technique is essential in applications such as autonomous driving and augmented reality. In addition, real-time operation is also necessary, so a target detection model with fewer parameters is essential.

Since most devices have very limited computational power, it is important to build lightweight target detection models. Since MobileNet convolution is small, the main approach to increase the perceptual field of the convolutional neural network is to replace Backbone with MobileNet [1] using global pooling layers, and to reduce the size of the model by appropriate parameter reduction operations. The recognition speed of a single frame is improved.

1 Related study

Since 1943, there has been a growing trend to use machines to achieve human-like perception, learning, memory and recognition. In 1985, Geoffrey-Hinton replaced the original single feature layer with multiple hidden layers and used the BP algorithm (back-propagation algorithm, proposed in 1969) to compute network parameters. In 1988, Kohonen et al. introduced the concept of neural networks. Subsequently, the basic theory of perceptrons and multilayer perceptrons was proposed, giving neural networks a certain basic model. Hubel and Wiesel found that the information processing process of biological vision is hierarchical through an in-depth study of the neural structure in biological neuroscience theory. By continuously extracting information between levels, the initial simple surface information is abstracted into higher-level feature information. Recognition of handwritten characters in postal codes using deep neural networks. Between 1987 and 1989, researchers proposed convolutional neural networks. Around 1995, researchers used physiology and computer technology to study vision problems. They proposed a sparse coding algorithm that iterated over 400 image segments to select the best segment weighting factor. However, the weights chosen were essentially the edges of images of different objects, which were similar in shape but different in orientation. In 2018, Redmon. J et al. proposed YOLOv3 [5], which is a little larger than

YOLO9000[4] and YOLOv2, but more accurate. Also, the inference is faster. In 320x320 image detection, YOLOv3[5] has 28.2mAP in 22ms, which is as accurate as SSD, but three times faster.

2 Algorithms

2.1 Overall thinking

The first step is to analyse the role of features in the process in conjunction with deep learning image processing algorithms, as feature extraction plays an important role in deep learning.

Suppose a matrix

$$I = [I_1 \quad I_2 \quad I_3 \quad I_4 \quad I_5],$$

where

$$I_{\Theta j} = [\beta_{11} + \alpha_{11}X \quad \dots \quad \beta_{\Theta j} + \alpha_{\Theta j}X],$$

$$\Theta \in \{0, 1, 3, \dots\}, \quad j = \{1, 2, 3, \dots\}, \quad I_j = \begin{bmatrix} I_{\Theta 1} \\ \vdots \\ I_{\Theta j} \end{bmatrix}.$$

Since the convolution operation is more important in convolutional neural networks. Let the input matrix be the following:

$$I = \begin{pmatrix} I_{1,1} & I_{1,2} & I_{1,3} & I_{1,4} & I_{1,5} \\ I_{2,1} & I_{2,2} & I_{2,3} & I_{2,4} & I_{2,5} \\ I_{3,1} & I_{3,2} & I_{3,3} & I_{3,4} & I_{3,5} \\ I_{4,1} & I_{4,2} & I_{4,3} & I_{4,4} & I_{4,5} \\ I_{5,1} & I_{5,2} & I_{5,3} & I_{5,4} & I_{5,5} \end{pmatrix}. \quad (1)$$

Set a convolution kernel of 3 × 3:

$$k = \begin{pmatrix} kernal_{1,1} & kernal_{1,2} & kernal_{1,3} \\ kernal_{2,1} & kernal_{2,2} & kernal_{2,3} \\ kernal_{3,1} & kernal_{3,2} & kernal_{3,3} \end{pmatrix}. \quad (2)$$

Convolution operations are performed with the input matrix using a convolution kernel, as follows:

$$c = conv2(I, k). \quad (3)$$

The pooling operation is then carried out:

$$p = globalaveragepool(c). \quad (4)$$

From Eqs. (1 - 3), which is a common procedure for convolutional neural networks, Eq. (4) indicates that the pooling operation is set to global mean pooling.

A typical convolutional neural network structure is shown in Fig. 1, but the main purpose of this structure is to extract the features better, and to increase the sampling rate while extracting the feature regions separately for subsequent global pooling.

The purpose of this is actually to distinguish between finer target features such as riders and pedestrians in road scenes where it is not enough to extract the appearance of

the person, but the environment they are in is equally important. Experiments have shown that it is better to use classical convolutional neural networks.

However, downsampling has a great advantage. Filters operating on downsampled images have a larger receptive field, which allows them to collect more information. This is particularly important when differentiating between categories.

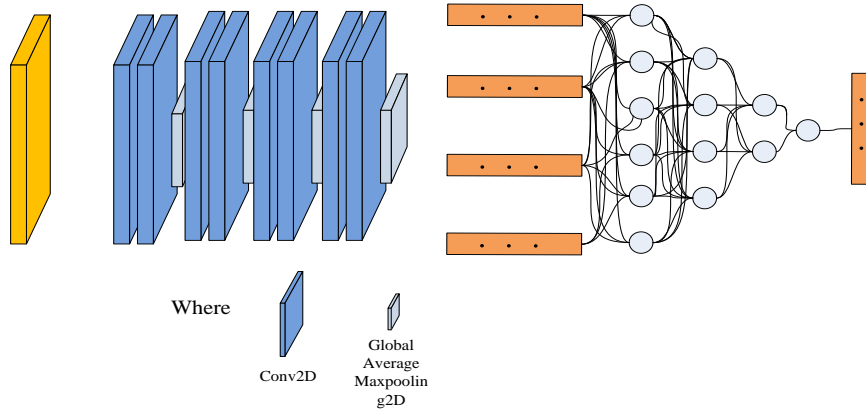


Fig. 1. Convolutional neural network [1-6]

For example, in road scenes, it is not enough for the target networks of riders and pedestrians to understand the appearance of people, but their environment is equally important. Finally, it was found to be better to use Mobilenet convolution.

As mentioned above, it is important that the network has a wide receptive domain, so that it can perform classification by taking into account a richer set of relevant data and features. The study wanted to avoid excessive downsampling of the feature maps and decided to use Mobilenet convolutions to improve the model in the study. They replace the main convolutional layers in several bottleneck modules in a stage that operates at minimal resolution. These greatly improved the accuracy of the operations, by about 4 percentage points, with no additional loss.

As shown in Fig. 2, 3, the structure of Darknet53 is a combination of DarknetConv2D, a deep network with 147 layers, which is effective in detecting small targets with high accuracy and speed, and is used in well-known target detection algorithms such as YOLOv3.

Each network was trained with the same settings, and all were tested with images from 256×256 , and all were single precision. Fig. 4 shows the interpretation of DarknetConvolutional2D_BN_Leaky in Fig. 2, 3 and Fig. 5 shows the performance of some of the current mainstream target detection algorithms on the open source dataset.

The model mentioned in this paper is improved by replacing the backbone with Mobilenet, which reduces the overall computational effort of the network.

This is done by balancing accuracy and computational efficiency.

In terms of targeted training and optimization, the following work has been done in this paper:

MobileNet as Backbone, and parameter optimization of Backbone:

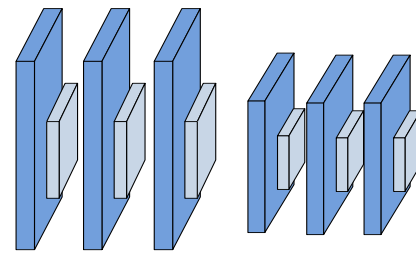


Fig. 2 Simplified Darknet diagram

Layer	Filter size	Repeat	Output size
Image			416*416
Conv	32 3*3/1	1	416*416
Conv	64 3*3/2	1	208*208
Conv	32 1*1/2		208*208
Conv	64 3*3/1	1	208*208
Residual			208*208
Conv	128 3*3/2	1	104*104
Conv	64 1*1/1		104*104
Conv	128 3*3/1	1	104*104
Residual			104*104
Conv	256 1*1/2	1	52*52
Conv	128 1*1/1		52*52
Conv	256 3*3/1	1	52*52
Residual			52*52
Conv	512 3*3/2	1	26*26
Conv	256 1*1/1		26*26
Conv	512 3*3/1	1	26*26
Residual			26*26
Conv	1024 3*3/2	1	13*13
Conv	512 1*1/1		13*13
Conv	1024 3*3/1	1	13*13
Residual			13*13

Fig. 3. Darknet structure diagram

increase sampling on Mobilenet to improve the recognition rate of small features;
 increasing the perceptual wildness of the convolutional neural network (CNN) partial separator layer by using a global pooling layer;
 enhanced for small sample data.

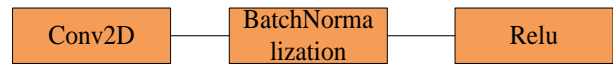


Fig. 4. DarknetConvolutional2D-BN-Leaky
 As a result of these operations, the structure in Fig. 6 is obtained.

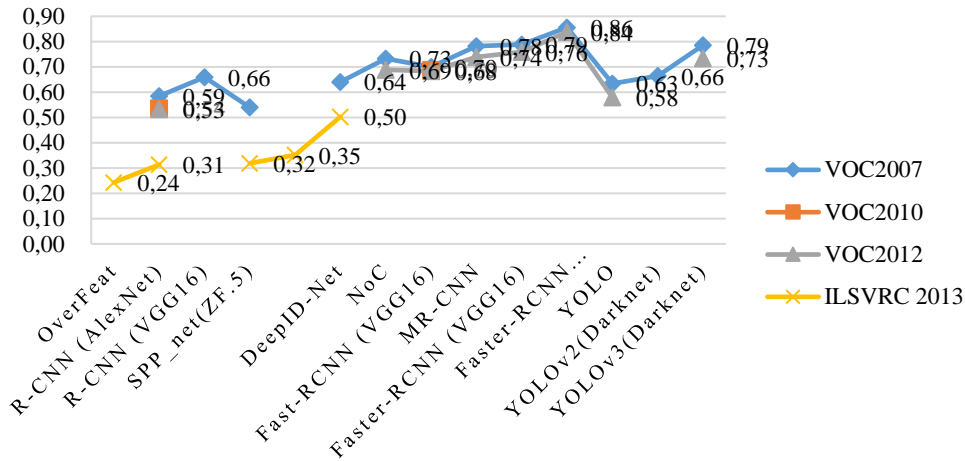


Fig. 5. The performance of the current mainstream target detection algorithms on VOC and ILSVRC, where the ordinate refers to the accuracy

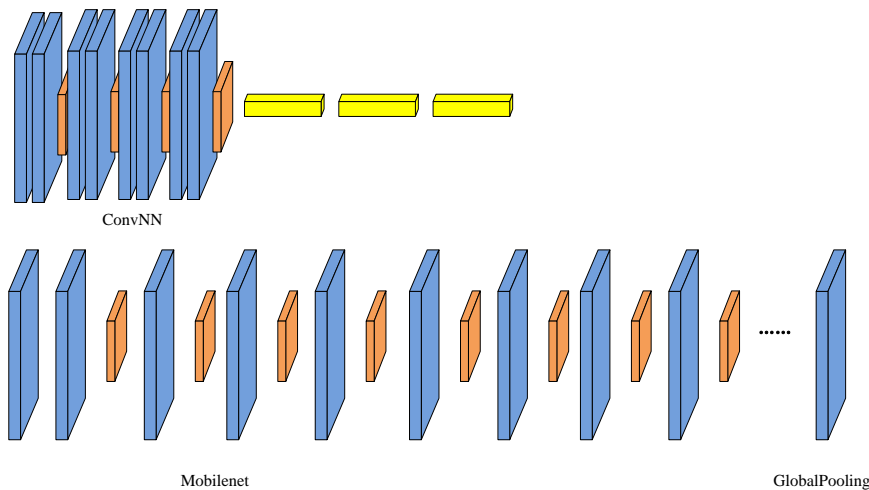


Fig. 6. Convolutional neural network + Mobilenet

2.2 Relevant explanations

2.2.1 Using global mean pooling to improve upsampling. There is a project for YOLOv3 on GITHUB, which uses a lot of downsampling for the tuning. Since it is only tested on the PASCAL VOC dataset, which has large targets with distinctive features, the overall improvement is made by adding an upsampling layer using a global pooling layer, which replaces the global average pooling layer with a global average pooling layer, while keeping the basic accuracy unchanged. This is to enable the classification methods with different labels to generate different feature maps.

2.2.2 Mobilenet. Mobilenet is an extremely simple network structure, with small network depth and fast inference. It is also the first of its kind for migration networks. MobileNet is a depthwise separable convolution [7] model, which uses depthwise separable

convolution to convolve its sub-channels and uses only the convolution kernel for each convolution, which can greatly reduce the number of parameters in the model. The structure of Mobilenet is shown in Fig. 7.

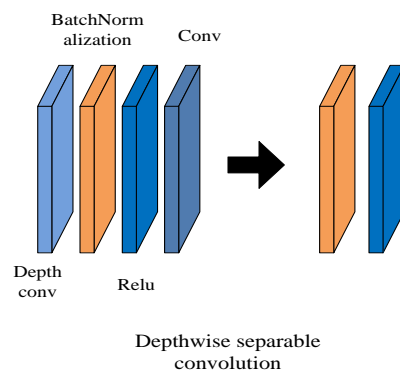


Fig. 7. The Mobilenet Simplified Structure

2.2.3 Reduction of parameters. The number of parameters determines to a large extent the size and learning ability of the model, but of course it is not a matter of increasing, let alone decreasing, the number of parameters. This limit should be determined by balancing accuracy and model size. In this paper, the parameters are reduced by 1/4 of the size, but the depth of the algorithm does not change as a result. In fact, since the algorithm was originally designed to understand typical targets such as pedestrian recognition, they do not necessarily need a large number of parameters to improve the learning ability, but only a certain number of parameters to extract the features, which is not only better for the task of target detection. This is not only better for target detection, but also for reducing the size of the model.

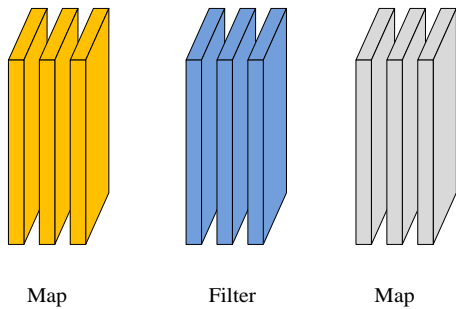


Fig. 8. Depthwise separable convolution

4 Training

Due to the significant disproportion between small and large samples in the dataset, the data were normalized by removing the very sparse labels and augmented with a generative adversarial network [8]

(GAN) and denoised with a wavelet algorithm. The cost function used is Eq. (5), where S is the number of all grids, B is the number of predicted borders, and x, y, w, h, c are the five calibration values.

$$\begin{aligned}
 L(x, y, w, h, c) = & \\
 = & \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^B l_{i,j}^{obj} \left[(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2 \right] + \\
 + & \lambda_{toord} \sum_{i=0}^{s^2} \sum_{j=0}^B l_{i,j}^{obj} \left[(\sqrt{w_i} - \sqrt{\bar{w}_i})^2 + (\sqrt{h_i} - \sqrt{\bar{h}_i})^2 \right] + \quad (5) \\
 + & \sum_{i=0}^{s^2} \sum_{j=0}^B \left[l_{i,j}^{bb} (C_i - \bar{C}_i)^2 + \lambda_{hoob} l_{i,j}^{noobj} (C_i - \bar{C}_i)^2 \right] + \\
 + & \sum_{i=0}^{s^2} l_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2.
 \end{aligned}$$

The overall structure is shown in Fig. 9. Since the computer configuration used for the experiments was far from the level of a standard deep learning lab's arithmetic workstation, the size of the input images was reduced from the original design to a uniform size. The training data was obtained from PASCAL VOC images on a computer with an I5 processor, 16 GB of RAM, and an RTX2080 graphics card, while the weights from the training tests were obtained using an RTX2060 graphics card. Tests were also conducted on a GTX1050TI laptop, but as there is still a gap between the graphics cards of laptops and desktops, it is not recommended to use a laptop platform. It is not recommended to use a laptop platform for testing.

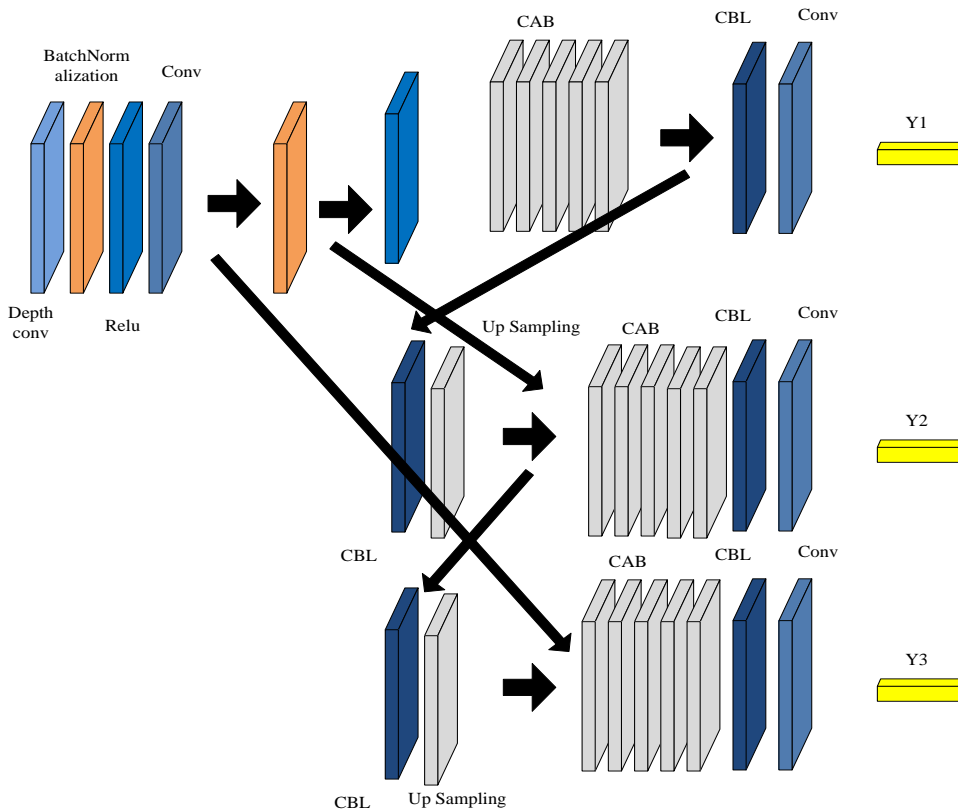


Fig. 9. Overall Structure

As shown in Fig. 10, Mobilenet is used as the Backbone of the grafting network, which contains a combination of modules as shown in Figure 4, including a number of positive feedback connections, the data set is fed into the network and trained 1000 times, including Training, Trainingval, Test and other processes, and finally the corresponding weight file is obtained The test is ready for testing.

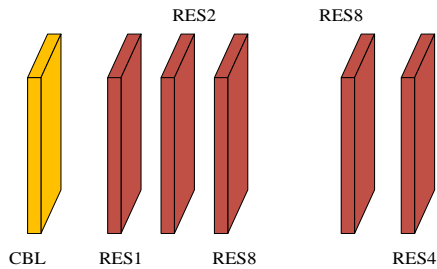


Fig. 10. The Darknet53 replaced by Mobilenet

As can be seen, the structure has an upsampling component, which is actually necessary for the current mainstream YOLO target detection model, to improve the quality of the detection by increasing the upsampling to obtain the desired target frame. Although the experimental metrics produced by the method are comparable to YOLOv3, and even slightly behind in

AP50 and AP, the weight size is significantly reduced, due to the above-mentioned reduction in the number of parameters and the use of a mean pooling layer in Mobilenet with increased upsampling to reduce the model size.

Fig. 10 shows a simplified structure of part of Fig. 3, mainly to explain the Darknet 53. that Mobilenet has replaced. Res is the residual block, CBL is introduced in Fig. 4.

Darknet53 has a very large number of participants, while Mobilenet is very small.

In addition, the smaller GPU memory can also be used to set a larger Batchsize during training, thus improving the training efficiency.

Since Batchsize determines the number of input images in a single training batch, it would make sense to increase its size with limited memory.

The performance of this target detection model is not very different from the rest of the models. On the contrary, its size is smaller than most models. And its inference speed is 30 FPS.

The results for the smaller samples are worse than the others, but the rest of the results are better and do not differ much, which is reasonable due to the small number of parameters and the simple structure, and will be improved if the number of training sessions increases [9-16], Fig. 11.

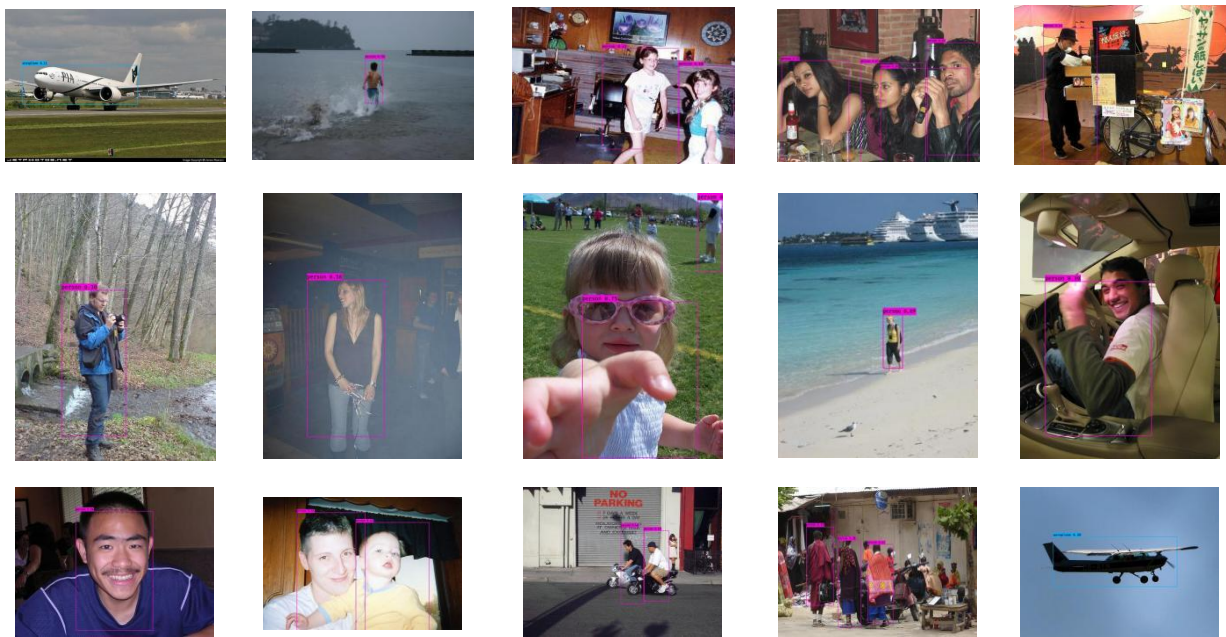


Fig. 11. In the PASCAL VOC test set test, in the PASCAL VOC test effect, it can be very intuitive to see, in the large sample identification effect is very good

Fig. 12 shows a simplified diagram of the structure of a discrete graphics card, with hundreds of stream processors in green and the memory module in yellow. The memory usage during training is in accordance with the following equation.

$$RAM_for_training = Modelsize + batchsize \times N. \quad (6)$$

Obviously, the 28.7mb weight file reference is smaller than 230mb and 150mb, so the model has a significantly

lower footprint in the video memory and thus has some advantages for training devices with low video memory.

5 Conclusion

The efficiency of feature extraction has been improved and the chance of overfitting has been reduced by using parameter reduction. With these improvements, the network is able to improve the speed of target extraction and target recognition in images.

Table 1 – Relevant indicators

	AP / %	AP50 / %	AP75 / %	APs / %	APM / %	APL / %	FPS / %	Size / mb
Two-stage model: Faster R-CNN ^[2]	34.9	55.7	37.4	15.6	38.7	50.9	5-17fps	≈ 150mb
Faster R-CNN+ FPN ^[2]	36.2	59.1	39.0	18.2	39.0	48.2		
Faster R-CNN +G-RMI ^[2]	34.7	55.5	36.7	13.5	38.1	52.0		
Faster R-CNN +TDM ^[2]	36.8	57.7	39.2	16.2	39.8	52.1		
Single stage model: YOLOv3	21.6	44.0	19.2	5.0	22.4	35.5	28-30fps	≈ 230mb
Ours	20.5	40.3	30.6	6.9	24.6	35.9	30fps	≈ 28.7mb

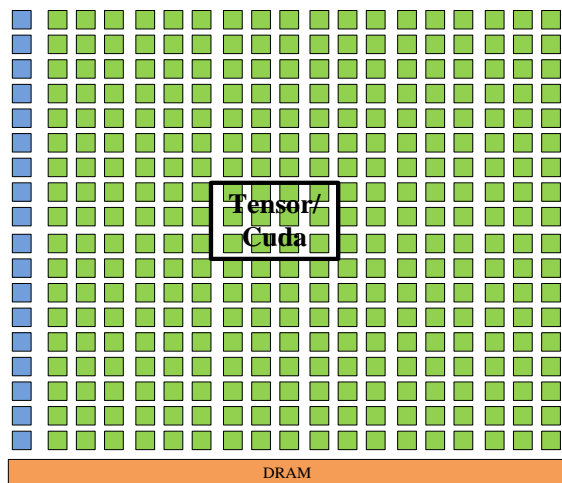


Fig. 12. Structural simplified diagram of the independent

Of course, the accuracy of this algorithm is somewhat affected by the fact that it is not as deep as traditional deep learning algorithms, with a size of only 28MB, but this can be solved by increasing the number of iterations and the dataset. When training the classical dataset, we did not have time to investigate how the classical network was augmented due to the time and equipment impact. The classical dataset VOC PASCAL has a large number of small samples of images, so some augmentation is necessary to improve the results.^[17] The test case achieved good results as the number of labels were all collected in a near 1:1 ratio. Finally, although the target detection algorithm described in this paper is not very advantageous to a certain extent, the idea of migrating the training and solving the overfitting problem by reducing the number of parameters is of great importance in future research, and many algorithms are currently validated by such methods.

REFERENCES

- Howard A G , Zhu M , Chen B , et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Appl. [J]. 2017.
- Girshick R . Fast R-CNN[J]. Computer Science, 2015.
- He K , Gkioxari G , P Dollár, et al. Mask R-CNN [C]// IEEE. IEEE, 2017.
- Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[C] // IEEE Conf. on Computer Vision & Pattern Recognition. IEEE, 2017:6517-6525.
- Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
- He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. IEEE, 2016.
- Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.
- Goodfellow I J. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
- Technicolor T , Related S , Technicolor T , et al. ImageNet Classification with Deep Convolutional Neural Networks [50].
- Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- Pang Y , Sun M , Jiang X , et al. Convolution in Convolution for Network in Network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, PP(99):1587-1597.
- GoogLeNet [J]. Journal of Thesis of Korean Cultural Information Society, 2018, 18.
- He K , Zhang X , Ren S , et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[C]// CVPR. IEEE Computer Society, 2015.
- Li H , Xiong P , Fan H , et al. DFANet: Deep Feature Aggregation for Real-Time Semantic Segmentation[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- Wang H , Jiang X , Ren H , et al. SwiftNet: Real-time Video Object Segmentation[J]. 2021.
- Li Wei,Liu Kai. Confidence-Aware Object Detection Based on MobileNetv2 for Autonomous Driving[J]. Sensors,2021,21(7).
- Batra Varun,Kumar Vijay. Real-Time Object Detection and Localization for Vision-Based Robot Manipulator[J]. SN Computer Science,2021,2(3).

Надійшла (received) 11.04.2023

Прийнято до друку (accepted for publication) 31.05.2023

Полегшений алгоритм виявлення цілей на основі Mobilenet Convolution

Н. Г. Кучук, Ю. М. Здоренко, Ван Шенцюань

Анотація. Алгоритм виявлення цілі, заснований на глибокому навчанні, потребує високої конфігурації графічного процесора комп'ютера, навіть потребує використання високопродуктивної робочої станції глибокого навчання, це не тільки підвищує вартість, а також значно обмежує реалізованість землі, ця стаття представляє свого роду полегшений алгоритм для виявлення цілей за умови балансової точності та обчислювальної ефективності MobileNet як Backbone виконує параметр Швидкість обробки становить 30 кадрів/с на карті RTX2060 для зображень із шаром розділення CNN. Швидкість обробки становить 30 кадрів в секунду на карті RTX2060 для зображень з роздільною здатністю 320×320.

Ключові слова: глибоке навчання; Mobilenet; шар глобального середнього об'єднання; GPU; виявлення цілі.

Є. В. Маловічко, В. С. Харченко

Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна

ІНТЕРАКТИВНЕ МИСТЕЦТВО: МУЛЬТИСЦЕНАРНІ AR-ТЕХНОЛОГІЇ НА ПІДСТАВІ АНАЛІЗУ ЕМОЦІЙНОГО СТАНУ

Анотація. Мотивація. Розвиток технологій доповненої реальності (AR) створює нові можливості для мистецтва, які можуть змінити сприйняття та взаємодію глядачів. Це стає актуальним у сучасному світі, де цифрові технології стають все більш поширеними та доступними. Інтерактивне мистецтво є важливим та перспективним напрямком сучасності, оскільки дозволяє глядачам стати активними учасниками творчого процесу. У кількості інтерактивних мистецьких виставок значною мірою переважають картини, які мають абстрактний характер. Проблематика полягає в тому, як допомогти створити мистецтво, яке здатне змінювати образи та адаптуватись до емоційного стану користувача, створюючи більш особистісні зв'язки між мистецтвом та глядачем, а також залучаючи широку аудиторію. Результати дослідження можуть стати корисними для художників, які створюють інтерактивні виставки, а також для дослідників, які цікавляться взаємодією людини та мистецтва. **Метою дослідження** є створення та аналіз сценаріїв взаємодії глядачів з абстрактними картинами за допомогою доповненої реальності на підставі аналізу емоційного стану людини, а також дослідження того, які елементи картини сприймаються глядачами як найбільш важливі. **Етапи дослідження.** В статті аналізуються різні сценарії взаємодії глядачів з картинами, пропонуються методи для визначення емоційного стану людини, аналізу отриманих даних та висновки. **Висновки.** В статті представлено та досліджено сценарії з використанням доповненої реальності AR на підставі аналізу емоційного стану людини. Обговорено оптимальні методи визначення емоційного стану людини. Проведено дослідження серед учасників та представлено узагальнені оцінки по обраним критеріям. Структура застосунку складається з фреймворку Vue.js та бібліотеки MindAR. Серверна частина написана на Node.js, в якості бази даних обрано об'єктно-реляційну систему PostgreSQL. **Напрямок подальших досліджень.** Майбутні кроки можуть бути присвячені аналізу та збору даних, що допоможе розробникам створити більш ефективні сценарії, які будуть здатні викликати очікувані емоційні реакції у глядачів в інтерактивному мистецтві.

Ключові слова: доповнена реальність, мистецтво, сценарії, емоційний стан, абстрактні картини, технології.

Вступ

Мотивація. У світі сучасного мистецтва все більше використовується доповнена реальність (AR). AR дає можливість створювати інтерактивні віртуальні об'єкти та взаємодіяти з ними у реальному часі. Це відкриває нові можливості для мистецтва.

Одним з найбільш інноваційних напрямків AR є інтерактивне мистецтво. Це мистецтво дає можливість глядачам стати активними учасниками творчого процесу та взаємодіяти з творами у новий спосіб. AR – це технологія, яка дозволяє додати віртуальний контент до реального світу за допомогою камери смартфона або іншого пристрою. Інтерактивне мистецтво з використанням AR вже знайшло застосування в багатьох галузях, таких як музика, театр, кіно, музеї та галереї мистецтва. Із зростанням популярності AR-технологій можна очікувати більшого розширення їх використання в мистецтві та культурі.

Застосування AR в мистецтві може мати і значний комерційний потенціал. Наприклад, виробники мистецьких картин можуть створювати віртуальні тури по своїх виставках, що дозволяє глядачам переглядати твори мистецтва з будь-якої точки світу.

AR-технології задають нові горизонти для мистецтва та культури, і слід очікувати більшого їх використання в майбутньому.

Аналіз публікацій. У своїх публікаціях багато мистецьких критиків обговорюють можливості AR в мистецтві. Вони відзначають, що AR дозволяє створювати інтерактивні твори, які взаємодіють з глядачем і змінюються в залежності від його дій. В одній з останніх публікацій [1] автори досліджували

взаємодію глядачів з інтерактивними картинами, особливо звертаючи увагу на тіло глядача як важливий елемент у взаємодії з мистецтвом. Також не менш захоплююче є стаття, де проводилось дослідження, як взаємодія з інтерактивним мистецтвом впливає на емоційний стан глядачів та їх поведінку [2, 3].

Одним з цікавих аспектів взаємодії глядачів з абстрактними картинами є дослідження сприйняття кольорів та форм [4]. Та як вибір кольору та його яскравість впливає на емоційну реакцію глядачів на інтерактивне мистецтво.

У своїх публікаціях автори статей [5, 6] розробляли та досліджували сценарії використання доповненої реальності, а також застосовували метод семантичного аналізу.

Метою дослідження є проаналізувати та розробити сценарії використання доповненої реальності (AR) за емоційним станом глядачів та розглянути можливі напрямки розвитку інтерактивного мистецтва з використанням AR.

Стаття структурується наступним чином. У першому розділі пропонуються різноманітні формалізовані сценарії взаємодії глядачів з картиною, у другому – описуються методи для визначення емоційного стану людини. У третьому розділі розглядаються результати застосування методу. У четвертому розділі підсумовуються результати дослідження та описуються їх майбутні напрями. Дане дослідження є актуальним у зв'язку з розвитком інтерактивних технологій та збільшенням інтересу до абстрактного мистецтва. Результати дослідження будуть корисними для художників, дизайнерів та IT-фахівців, які

удосконалюють технології доповненої та віртуальної реальності.

1 Сценарій застосування за допомогою доповненої реальності для абстрактних картин

Існують можливі сценарії візуалізації на підставі емоційного стану людини:

- опитування відвідувачів для визначення настрою;
- за допомогою голосу або словом відвідувач описує свої враження стосовно картини;
- розпізнавання обличчя та настрою користувача;
- спілкування з картиною, за допомогою штучного інтелекту.

1.1 Сценарій 1. Опитування відвідувачів для визначення настрою. Сценарій базується (рис. 1) на тому, щоб краще зрозуміти емоційний стан людини та визначити наблизений варіант доповненої реальності із існуючих. Цього можна досягти за допомогою опитування. Для використання цього підходу художник повинен завантажити картину на сервер, а також декілька варіантів доповненої реальності. Після входу користувача в застосунок і використання камери для сканування картини, система розпізнає картину та на екрані з'являється опитування. Після того як візитор відповість, йому слід за допомогою камери навести на картину і система підбере один з варіантів доповненої реальності виходячи з відповідей користувача.

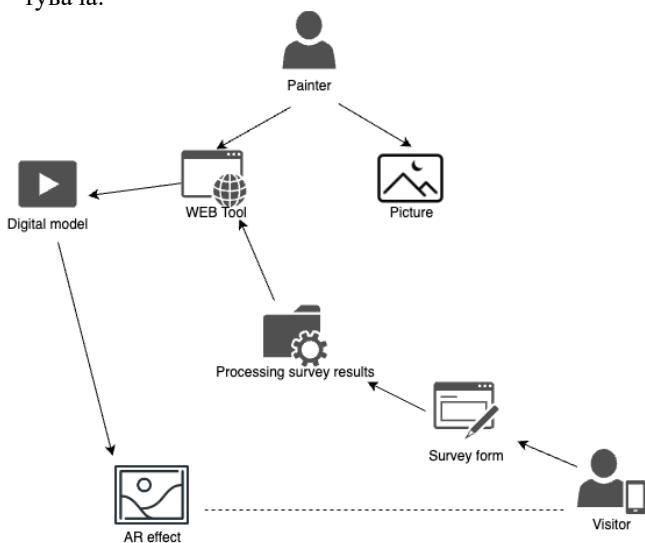


Рис. 1. Сценарій 1

Приклад застосунку, який використовує опитування для визначення настрою користувачів, - Moodnotes – Mood Tracker [7] який дозволяє пройти тест, щоб визначити емоційний стан, відстежувати свій настрій та покращувати його. Технології, які використовуються в додатку Moodnotes – Mood Tracker, включають аналіз даних та розпізнавання обличчя. Застосунок використовує ці технології для надання індивідуальних порад, які можуть підтримувати користувачів у керуванні їх емоційним станом та його покращенні. Крім того, він використовує аналіз

даних, щоб зрозуміти, які фактори можуть впливати на емоційний стан користувачів, та надає поради, що ґрунтуються на цих даних.

1.2 Сценарій 2. За допомогою голосу або слів відвідувач описує враження стосовно картини. Сценарій базується на тому, що за допомогою голосу або слів (рис. 2) відвідувач описує свої враження стосовно картини та асоціації, пов'язані абстракцією. Система має обробити запит і за допомогою алгоритмів пошуку знайти необхідний об'єкт доповненої реальності. Таким чином, система може збирати необхідну інформацію про емоційний стан людини і обрати найбільш наблизений варіант доповненої реальності до картини, яка має сподобатись користувачеві. Художник відповідає за створення картини та завантаження зображення картини у додаток для застосування технологій AR. Що стосується користувача, після входу в застосунок і наводячи камерою на картину, він вербально описує свої враження, пов'язані з абстракцією, де система за ключовими словами шукає збіги і підбирає візуальний ефект AR.

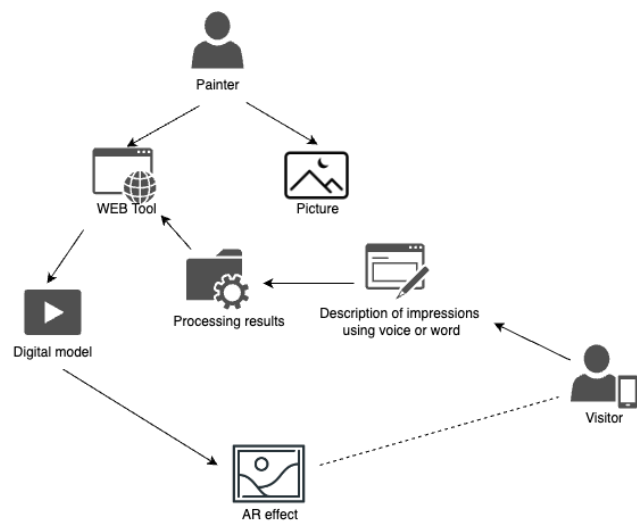


Рис. 2. Сценарій 2

Для порівняння даного сценарію можна обрати застосунок EmoVoice [8]. EmoVoice являє собою всеохоплюючу систему для розпізнавання емоцій. Може працювати у режимі реального часу та відштовхується лише від акустичних властивостей мовлення (не використовуючи інформацію про слова). EmoVoice містить два модулі: один для незалежного аналізу емоційної бази мови; інший - для відстеження онлайн впливу мовлення, під час розмови.

1.3 Сценарій 3. Розпізнавання обличчя та настрою користувача. Для цього сценарію (рис. 3) можна використовувати нейронні мережі, які здатні аналізувати відомості з відеопотоку з камери мобільного пристрою та визначити параметри обличчя і настрою користувача, такі як наявність посмішки, гніву, суму, задоволення та інші емоції. Таким чином система може підібрати найбільш наблизений варіант доповненої реальності до картини та до настрою користувача. Художник відповідає за створення картини та завантаження зображення картини у додаток для застосування технологій AR. Що стосується

користувача, після того як він увійшов в застосунок, слід дозволити використання камери, для того щоб система розпізнала обличчя, а також настроїв відвідувача. Після цього слід навести камерою на картину і спостерігати візуальний ефект AR.

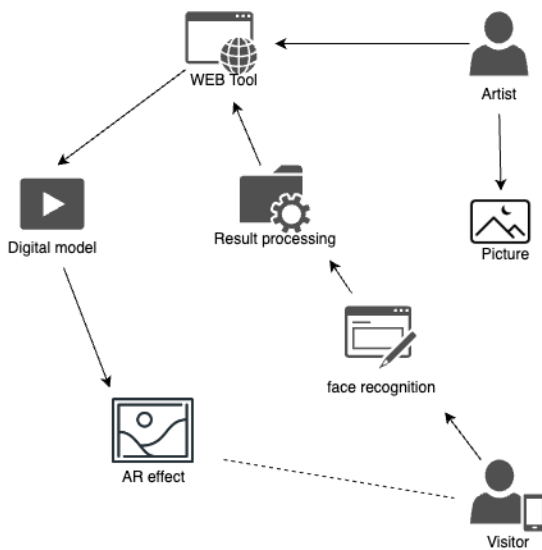


Рис. 3. Сценарій 3

Для порівняння даного сценарію можна обрати застосунок Affectiva [9]. Серед комерційних застосунків ця технологія використовується, щоб допомогти брендам вдосконалити свої рекламні та маркетингові повідомлення. Особливість застосунку полягає в тому, що він використовує глибоке навчання (deep learning) та комп'ютерний зір (computer vision) для аналізу облич та розпізнавання емоцій користувачів. Крім того, Affectiva використовує машинне навчання (machine learning) та інші алгоритми для підвищення точності розпізнавання емоцій.

Незважаючи на переваги, застосунок має деякі недоліки: точність розпізнавання емоцій може залежати від якості освітлення та якості зображення. Крім того, деякі користувачі можуть бути менш експресивними та мати менш видимі ознаки емоцій, що може призвести до менш точного аналізу їх емоційного стану.

1.4 Сценарій 4. Спілкування з картиною, за допомогою штучного інтелекту. Сценарій базується (рис. 4) на тому, що відвідувач може спілкуватися з картиною, як з чат-ботом, наприклад, ChatGPT, і отримувати відповідь. Використання AR та інтерактивної взаємодії з абстрактною картиною може бути цікавим для відвідувачів мистецьких виставок і галерей, оскільки воно може допомогти розширити їхнє сприйняття творів мистецтва та забезпечити новий рівень взаємодії між користувачами та картиною. Крім того, цей сценарій може бути використаний як засіб освіти та навчання, де користувачі можуть отримати інформацію про художника, його твори та історію їх створення. Такий досвід допомагає запам'ятовувати та зберегти інформацію про творців та їхні роботи в більш приємному та ефективному способі. Застосування цього сценарію також може допомогти розширити можливості мистецтва та

зробити його більш інтерактивним та доступним для широкої аудиторії.

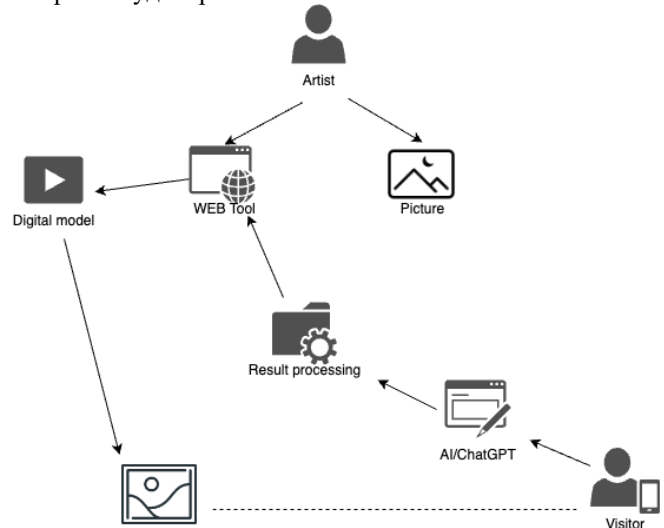


Рис. 4. Сценарій 4

Художник відповідає за створення картини та завантаження зображення картини у додаток для застосування технологій AR. Що стосується користувача, після того як він увійшов в застосунок, слід навести камерою на абстрактну картину. Картина знаходиться в центрі екрану, і користувач може продовжити спостерігати за доповненою реальністю або об'єктом "розмова", яка розпочинає інтерактивну сесію з картиною. Користувач починає розмову, задаючи питання або команди, наприклад, "Привіт, як справи?", "Розкажи мені про свою історію". Картина використовує інтелектуальний чат-бот, наприклад, ChatGPT, для аналізу запитів користувача та його емоційного стану і надає відповіді, які відображаються на екрані. Картина також може відображати анімацію та змінювати свій вигляд, щоб краще передати відчуття та настрої, пов'язаний з темою розмови.

Один з прикладів додатків AR зі штучним інтелектом на основі ChatGPT - це застосунок "Replika" [10]. Він використовує технологію ChatGPT для створення віртуального друга, з яким користувач може спілкуватися у режимі реального часу. Застосунок збирає дані про користувача, його інтереси та поведінку, щоб створити персоналізовану взаємодію та відповіді на повідомлення. Він навчається на основі даних, які користувач надає під час спілкування, та з кожним днем стає все більш інтуїтивною та особистісною. Replika відрізняється тим, що створена спеціально для підтримки психічного здоров'я, а не для розваг.

2 Методи визначення емоційного стану

Емоційний стан людини - це стан, в якому перебуває людина в даний момент, відображаючи її емоції та почуття. Емоції можуть бути позитивними, такими як радість, любов, захоплення, або негативними, такими як гнів, розчарування, страх.

Емоційний стан може впливати на різні аспекти життя людини, такі як поведінка, сприйняття світу навколо, фізичний стан, рівень стресу та багато

іншого. Часто емоційний стан є реакцією на певну ситуацію, тому може змінюватися впродовж дня.

Незважаючи на те, що методів для визначення емоційного стану дуже багато, можна виділити деякі з них:

- вимірювання фізіологічних параметрів;
- розпізнавання обличчя;
- аналіз голосу;
- вибір кольору;
- опитування.

Один з методів - вимірювання фізіологічних параметрів, таких як швидкість серцевого ритму, рівень потовиділення, температура тіла. Ці показники можуть змінюватися залежно від емоційного стану людини. Для вимірювання фізіологічних параметрів можна використовувати різноманітні сенсори, які надають дані в реальному часі для подальшого аналізу.

Інший метод - розпізнавання емоцій через обличчя. Для цього можна використовувати камери, які реєструють вираз обличчя та дозволяють визначити емоційний стан людини за допомогою алгоритмів обробки зображень.

Третій метод - аналіз голосу. За допомогою програмного забезпечення можна аналізувати голос та визначити емоційний стан людини, враховуючи такі характеристики як тон, інтонація та швидкість мовлення.

Напевно, найбільш відомою методикою визначення емоційного стану особистості крізь призму кольору є тест М. Люшера [13]. Метод передбачає розташування 8 кольорових карт: червоний, жовтий, зелений – розглядаються як позитивний стан; синій, фіолетовий – як нейтральний стан; сірий, коричневий та чорний – розглядаються в негативному сенсі.

Однак одним з найбільш поширених методів визначення емоційного стану є опитування, які можуть містити запитання про настрій, рівень тривоги, ступінь стресу тощо. Зазвичай вони базуються на шкалах оцінок, наприклад, від 1 до 10, де 1 означає найнижчий рівень емоції, а 10 - найвищий. Це досить економічний та доступний спосіб дослідження емоційного стану. Опитування можна проводити дистанційно, що дозволяє досліджувати емоційний стан учасників навіть у віддаленому режимі. Також, за допомогою цього методу можна отримати велику кількість даних за короткий проміжок часу, що дозволяє швидко проводити аналіз та порівняння результатів.

Крім того, можна отримати корисну інформацію про те, які елементи картин привертають увагу відвідувачів, які емоції вони викликають, які елементи можна покращити або змінити, щоб зробити виставку більш цікавою та захоплюючою.

Окрім технологій, важливо також пам'ятати про звичайні методи спілкування з відвідувачами, такі як проведення екскурсій, організація лекцій та майстер-класів. Це дозволить створити додаткові можливості для взаємодії з картинами та забезпечити більш глибоке розуміння творів мистецтва.

На жаль, не існує єдиної формули для визначення емоційного стану людини. Емоційний стан залежить від багатьох факторів, таких як настрій, досвід, ситуація, фізичний стан та багато іншого.

Існують психологічні тести, що дозволяють оцінити емоційний стан людини, такі як "Шкала депресії Бека" [14] або "Тест на оцінювання стану тривоги" [15]. Однак, ці тести можуть бути недостатньо точними та повними, оскільки людський емоційний стан є складним та індивідуальним.

Тому для визначення емоційного стану людини краще застосовувати комплексний підхід, що включає в себе спостереження, бесіду та опитування. Такий підхід дозволить отримати більш повну та точну картину емоційного стану людини.

Оцінка за результатами опитування

Дослідження проведене на базі одного з методу визначення емоційного стану людини за допомогою опитування. Сукупність людей була представлена, які переважно ніяк не пов'язані з живописом. В експерименті брали участь 10 людей, а також була обрана одна абстрактна картина (рис. 5). Питання були пов'язані з їх емоційним станом та уявлень про картину.



Рис. 5. Приклад абстрактної картини

У результаті опитування було виявлено, що більшість користувачів перебувають у позитивному емоційному стані, але також були ті, які відчували тривогу або стомленість. Також було встановлено, що більшість людей уявляють картину у яскравих кольорах та бажають побачити на ній природні елементи, такі як річка, вода або квіти.

На основі отриманих даних було створено прототип застосунку (рис. 6), який мав за мету підібрати віртуальне доповнення до картини для кожного відвідувача в залежності від його емоційного стану та уявлень. Він надавав можливість показувати картину в різних кольорах та з різними візуальними ефектами, щоб відобразити бажання користувача.

Застосунок має такі характеристики:

1. Веб застосунок може працювати на будь-якій операційній системі, яка підтримує сучасні веб-браузери.

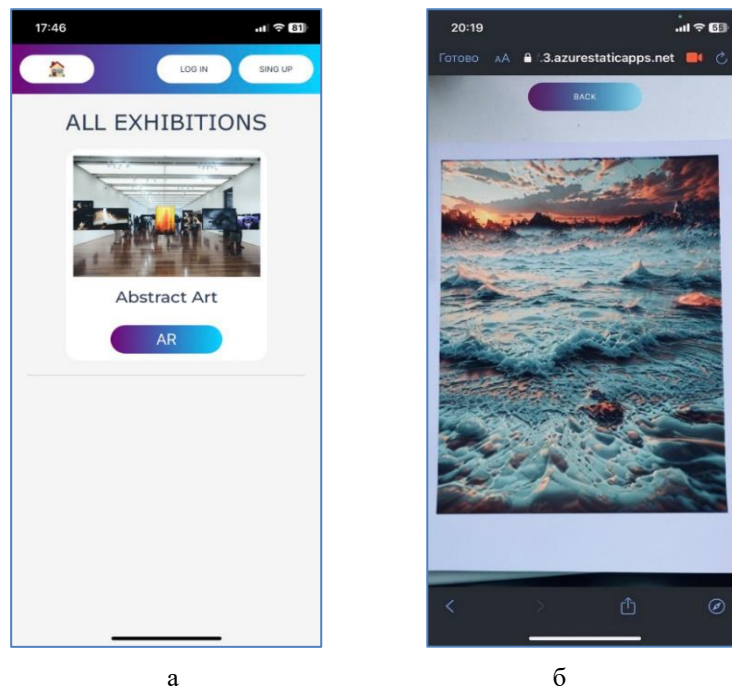


Рис.6. Застосунок: а – інтерфейс головної сторінки; б – AR відео

2. Наявність камери: для використання доповненої реальності потрібно мати камеру, яка забезпечує якість зображення та може працювати з веб додатком;

3. Сенсорний екран: для відображення картин та доповненої реальності;

4. Інтернет-з'єднання: Для завантаження картин та контенту з сервера потрібне стабільне Інтернет-з'єднання зі швидкістю не менше 50 КБ/сек.

У подальшому, було проведено ще одне опитування, де було запитано про подобу візуалізації картини з використанням прототипу застосунку. Загалом, більшість користувачів відзначили, що підібрані елементи AR їм сподобалися. Однак, деякі люди відзначили, що візуальні ефекти були надто насиченими, тому вони бажали більш природний вигляд картини.

Після тестування учасники повинні були оцінити відповідні параметри, де:

1-3 бали – найнижча оцінка;

4-6 балів – середня оцінка;

7-10 балів найвища оцінка.

Результати оцінок для всієї сукупності учасників представлені в табл.1.

Таблиця 1 – Результати оцінок учасників за вибраними критеріями

Критерії	Бали		
	1-3	4-6	7-10
AR ефект для зображення	1	3	6
Якість відео та ефекту	-	3	7
Швидкість завантаження	1	-	9
Візуальний ефект заемоційним станом	-	2	8
Підбір кольорів	-	1	9
Загальні враження	-	-	10

В результаті дослідження зазначимо, що емоційний стан людини можна визначити за допомогою опитування та спостереження. Багатолюдей були готові поділитись своїми емоціями та настроєм, і це дало змогу зрозуміти, яким чином можна краще підібрати віртуальне доповнення до картини для кожного окремого користувача.

Також важливою була інформація про те, які кольори та елементи візуалізації подобаються людям. За допомогою цих даних вдалось створити прототип застосунку, який містив елементи та кольори, які були більш прийнятними для цільової аудиторії.

Висновки

Результати. Дослідження стосовно інтерактивного мистецтва з використанням AR допомагають показати, які нові можливості надаються митцям, щоб створювати твори, які розширюють та поглиблюють досвід глядачів. Також дослідження демонструють, як AR може допомогти зберігати та відновлювати культурну спадщину, забезпечуючи доступність до неї для широкої аудиторії.

Крім того, дослідження про AR в мистецтві допомагають виявляти нові можливості для інтерактивного досвіду та експериментування зновими формами взаємодії між глядачами та творами мистецтва.

В статті авторами представлено та проаналізовано сценарії з використанням доповненої реальності AR на підставі аналізу емоційного стану людини. Обговорено раціональні методи визначення емоційного стану людини, проведене їх порівняння, дослідження варіантів і оцінювання з допомогою розробленого застосунку представлено та проаналізовано їх результати.

Майбутні кроки можуть бути спрямовані на подальший розвиток технологій та створення більш складних інтерактивних мистецьких інсталяцій з

використанням AR. Також можливими напрямками розвитку можуть бути створення інтерактивних мистецьких проєктів, які будуть спрямовані на підвищення інформованості про проблеми соціуму,

наприклад, певні соціальні проблеми, які можна було б вирішити через використання відповідних картин і доповненої реальності для покращення емоційного і психологічного стану людини.

СПИСОК ЛІТЕРАТУРИ

1. Interactive Art and Embodiment: The Implicit Body as Performance (Arts Future Book) Paperback [Text] // Nathaniel Stern. –2013.
2. Experiencing affective interactive art [Text] / Leticia S.S. Bialoskorski, Joyce Westerink Philips, Egon L.// van den Broek Utrecht University. - 2010 International Journal of Arts and Technology 3(4). DOI:10.1504/IJART.2010.035826
3. Aesthetic Experience and Creativity in Interactive Art // Esma Betul Savas. -2021 .Art & Perception 9(2):167-198. DOI:10.1163/22134913-bja10024
4. A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People // Jun Dong Cho. – 2021. Electronics 10(4):470. DOI:10.3390/electronics10040470
5. Interactive Abstract Painting by Augmented Reality: Scenarios and Architecture Solutions/ Yaroslav Bershadskiy, Ilona Zaika, Vyacheslav Kharchenko, Olena Golembovska .-2019 [Electronic resource]. – Available at: <https://ceur-ws.org/Vol-2533/paper29.pdf>
6. Augmented Reality for the Abstract Paintings: Application Scenarios, Semantic Similarity Analysis and Case Study / Mariia Danova, Igor Shostak, Olena Feoktystova, Vladyslav Plietnov, Vyacheslav Kharchenko, Olena Golembovska // 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) . DOI: 10.1109/IDAACS.2019.8924411
7. Moodnotes - Mood Tracker. -2015-2021, [Electronic resource]. – Available at: <https://overlyapp.com/blog/augmented-reality-art-opportunities-and-examples-for-artists-and-creatives/https://apps.apple.com/us/app/moodnotes%20-%D1%89%D0%BE%D0%B4%D0%B5%D0%BD%D0%BD%D0%B8%D0%BA-%D0%BD%D0%B0%D1%81%D1%82%D1%80%D0%BE%D1%8E/id1019230398?l=uk> - 23.03.2023
8. EmoVoice [Electronic resource]. – Available at: <https://github.com/hcmlab/emovoice> - 23.03.2023
9. Affectiva a smart eye company [Electronic resource]. – Available at: <https://www.affectiva.com/> - 23.03.2023
10. Replica. The AI companion who cares. Always here to listen and talk. Always on your side [Electronic resource]. – Available at: <https://replika.com/> 24.03.2023
11. Augmented reality art: opportunities and examples for artists and creatives//Ainars Klavins . -2022, [Electronic resource]. – Available at: <https://overlyapp.com/blog/augmented-reality-art-opportunities-and-examples-for-artists-and-creatives/> - 25.03.2023
12. Augmented reality in art: The rising trend. -2022, [Electronic resource]. – Available at: <https://www.ienhance.co/insights/augmented-reality-in-art> -27.03.2023
13. Офіційний сайт Lüscher Diagnostics [Electronic resource]. – Available at: <https://de.luscher-color.com/> -28.03.2023
14. Шкала депресії Бека [Electronic resource]. – Available at: <https://cbt-therapist.net/BDI> -28.03.2023
15. Шкала Гамільтона для оцінки тривоги (HAM-A) [Electronic resource]. – Available at: <https://www.darnitsa.ua/ptr/shkala-gamiltona-dlya-otsinki-trivogi-ham-a> -29.03.2023
16. Augmented Dreams / Samantha Culp // the article appears in the January. -2021 issue, pp. 40– 47. [Electronic resource]. – Available at: <https://www.artnews.com/art-in-america/features/augmented-reality-art-1234581120/> -1.04.2023.
17. Van Krevelen, D., and Poelman, R., / A survey of augmented reality technologies, applications and limitations [Text] // International Journal of VirtualReality. -vol. 9 (2). -pp. 1. -2010.

Received (Надійшла) 12.04.2023

Accepted for publication (Прийнята до друку) 17.05.2023

AR interactive art: scenarios by emotional state

Yelyzaveta Malovichko, Vyacheslav Kharchenko

Abstract. Motivation. The development of Augmented Reality (AR) technologies creates new possibilities for art that can change the perception and interaction of viewers. This becomes relevant in today's world, where digital technologies are becoming more and more common and accessible. Interactive art is an important and promising direction of modernity, as it allows viewers to participate actively in the creative process. The number of interactive art exhibitions is largely dominated by paintings of an abstract nature. The challenge is how to help create art that can change and adapt to the emotional state of the user, thereby creating a more personal connection between the art and the viewer, as well as engaging a wider audience. The results of the study can be useful for artists who create interactive exhibitions, as well as for researchers who are interested in the interaction of people and art. **The goal of the research** is to create and analyze scenarios of viewers' interaction with abstract paintings using augmented reality based on the emotional state of a person and consider which elements of the picture are perceived by the audience as the most important. **Stages of research.** The article analyzes various scenarios of audience interaction with paintings, offers methods for determining the emotional state of a person, analyzes the obtained data, and draws conclusions. **Conclusions.** The article presents and analyzes scenarios using AR augmented reality based on a person's emotional state. The most optimal methods of determining a person's emotional state are discussed. Conducted research among participants and presented generalized evaluations according to the selected criteria. The framework of the application consists of the Vue.js framework and the MindAR library. The server part is written in Node.js. But the object-relational system PostgreSQL was chosen as the database. The direction of further research. Future steps can be devoted to analysis and data collection, which will help art developers to create more effective scenarios that will be able to evoke the desired emotional reactions from viewers in interactive art.

Keywords: augmented reality, art, scenarios, emotional state, abstract paintings, technologies.

М. І. Матвеев, Г. А. Кучук

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

ОПТИМІЗАЦІЯ МОБІЛЬНОЇ ГРИ В СЕРЕДОВИЩІ UNITY

Анотація. Актуальність. В наш час мобільні ігри – найшвидше зростаючий сегмент ігрового ринку. Ігри впевнено займають перше місце за виручкою та завантаженням у *Google Play* та *App Store*. Однак вже сьогодні розробники все частіше створюють ігри не тільки під мобільні платформи, а й під персональні комп'ютери одночасно, адже спостерігається збільшення інтересу користувачів до мобільних ігор, які здатні працювати не на одній, а на декількох платформах. Такі ігри дозволяють розгорнути той самий код на кількох платформах, отже у таку гру люди можуть грати на будь-якому пристрої (смартфон або комп'ютер). **Метою даної роботи** є оптимізація роботи власної розробленої мобільної гри в середовищі *Unity* та розширення її доступності для персональних комп'ютерів на системі *Windows*. **Об'єктом дослідження** є 2D мобільна гра *AirMan* в середовищі *Unity*. **Предметом дослідження** є методи діагностики та оптимізації гри. **Результати.** У результаті оптимізації гри навантаження на оперативну пам'ять при старті ігрового процесу зменшено з 445 до 336 мегабайт, а середній показник використання пам'яті становив 310,2 проти 363,5 мегабайт при початковому тестуванні гри. Гру *AirMan* зроблено кросплатформною. **Висновок.** Покращено роботу гри *AirMan* в середовищі *Unity* та розширено її доступність для персональних комп'ютерів на системі *Windows* і як наслідок – збільшено аудиторію гравців.

Ключові слова: мобільна гра, оптимізація, кросплатформність, unity, android, windows.

Вступ

Індустрія комп'ютерних ігор (розробка, видання та просування ігор) – одна з галузей комп'ютерних технологій і одночасно глобального сектору розваг, яка на сьогодні найшвидше розвивається.

Світовий ринок комп'ютерних ігор став більш глибоким та масштабним. За останні п'ять років, за оцінками *Newzoo* [1], він зростав у середньому на 11% у рік і у 2021 році досяг позначки 192,7 млрд дол., обігнав низку інших конкурентних контентних ринків, зокрема, кіно- та музичної індустрії. Ключовими драйверами зростання стали широка доступність інтернету та комп'ютеризованих пристроїв, насамперед смартфонів, а також розширення дистрибутивних каналів.

У 2020 році, за попередніми оцінками аналітиків, обсяг світового ринку комп'ютерних ігор міг зрости на 18%, до 160 млрд дол. Але пандемія COVID-19 внесла свої корективи, виступив своєрідним драйвером і ринок ігор зріс до 179,18 млрд дол. Сума доходів, що отримано з 2020 по 2022 рік, майже на 43 мільярди доларів вища, ніж прогнозувалось аналітиками до пандемії.

За прогнозами *Newzoo* [2], ринок ігор і далі зростатиме до 2025 року, збільшуючись у середньому на 3,4%, майже до 211,2 млрд. доларів.

Найбільшим сегментом глобального ринку комп'ютерних ігор у 2022 році, як і в попередні роки, залишились мобільні ігри (50%, або 92,2 млрд дол. проти 45%, або 68,5 млрд дол. у 2019 році).

За прогнозами аналітиків [2], цей напрямок якщо не зростатиме, то своїх позицій у найближчі роки не здаватиме, адже кількість мобільних пристроїв значно більша, ніж ПК та консолі.

На другому місці за вагою знаходився сегмент ігор для консолей (28%, або 51,8 млрд дол.). Десктопні рішення сумарно займали 22% ринку (40,5 млрд дол.), з яких 1% – це браузерні онлайн-ігри, а решта 21% – скачані або коробкові версії.

Однак у 2022 році ринок ігор приніс 184,4 мільярда доларів, що на 4,3% менше, ніж у минулому 2021 році, тому що 2022 рік став роком корекції для ігрового ринку після двох виняткових років зростання світового ігрового ринку, що було викликано карантинном.

В наш час мобільні ігри – ігрові програми для смартфонів і планшетів – сегмент ігрового ринку, що найшвидше зростає.

Це пояснюється доступністю мобільних пристроїв, мобільного трафіку і навіть легкістю освоєння мобільних ігор.

Сучасні мобільні пристрої можуть працювати під управлінням різних операційних систем, але в більшості це *Android* або *iOS* [3]. Кожна з операційних систем має свій майданчик з додатками.

У *Android* це *PlayMarket*, а в *iOS* – *Appstore*, де щодня користувачі завантажують мільйони різних додатків, зокрема й ігор.

На сьогодні у всьому світі налічується близько 14 мільйонів розробників мобільних програм, і майже половина з них займається розробкою програм для *Android* [4].

Але щоб додаток набув більшої популярності, має сенс зробити його кросплатформним [5], тобто здатним запускатися на іншій платформі.

Вже сьогодні можна спостерігати як розробники створюють ігри одночасно під мобільні платформи та персональні комп'ютери або під персональні комп'ютери та консолі, адже останнім притаманна простота ігрового процесу.

Ситуацію, що склалася, називають «танцем платформ» (*platform dance*).

Можливо кросплатформність у найближчій перспективі стане новим трендом [6].

Але такі розробки додатків для більш ніж однієї операційної системи звісно можуть стати більш складними, тому що розробник стикається з купою проблем, пов'язаних з технічними характеристиками пристроїв.

І як наслідок – вартість такої розробки стає вищою [3].

Однак, при створенні кросплатформної гри потрібно враховувати те, що люди зможуть грати в неї на будь-якому пристрої.

У наші дні багато людей можуть дозволити собі відразу кілька пристроїв, тому вони нерідко грають на кількох платформах або перемикаються між ними. Гравці, які використовують дві або більше платформи, залучені до ігрового процесу набагато більше, ніж інші [7-15].

Отже, спираючись на вищевикладене, дослідження та оптимізація мобільних ігор з урахуванням різних платформ на сьогодні є актуальним та доцільним.

Метою цієї роботи є дослідження та оптимізація роботи власно розробленої 2D мобільної гри у жанрі «Платформер» *AirMan* в середовищі програмування *Unity* [16, 17] з урахуванням різних платформ задля розширення її доступності.

Основна частина

Для досягнення зазначеної мети роботи необхідно виконати наступні задачі:

- обрати програмний інструментарій для дослідження та проведення оптимізації мобільної гри;
- зменшити використання оперативної пам'яті девайсу 2D мобільною грою у жанрі «Платформер» *AirMan* в середовищі програмування *Unity*;
- розширити доступність гри для персональних комп'ютерів на системі *Windows*.

На першому етапі роботи обґрунтовано обрано програмний інструментарій для дослідження та оптимізації мобільної гри.

Для дослідження, а саме тестування, обрано десктопний додаток *Apptim*. За допомогою даного додатку можна виявляти критичні помилки, проблеми продуктивності, встановити використання оперативної пам'яті та потужності процесора, які залучені під час дослідження, та фіксувати збої.

Додаток дозволяє побачити відеозапис, зроблений під час тестування.

До того ж він має можливість надавати інформацію у вигляді залежностей навантаження *CPU*, *Memory*, *Network*, *Storage* від часу, завдяки яким можна побачити зміни у використанні потужності зазначених ресурсів на досліджуваному девайсі і дізнатися на який саме момент припадає максимальне навантаження на систему девайсу.

Найбільш ключовими є залежності навантаження *CPU* та *Memory*.

Оптимізацію гри виконано у кілька кроків за допомогою інструментарію, який вбудований в *Unity*.

Загальну оптимізацію, без глибокого втручання в функціонал гри реалізовано за допомогою вікна *Player Settings*, що містить загальні налаштування проєкту, які скореговані під власну 2D гру.

Оптимізацію асетів проведено за допомогою *Profiler* – інструменту, який збирає дані про продуктивність гри в таких областях, як центральний процесор, пам'ять, рендерінг та аудіо, та виводить їх у графічному та текстовому вигляді.

Оптимізацію коду задля покращення роботи 2D мобільної гри у жанрі «Платформер» *AirMan* виконано за допомогою інструменту – *Visual Studio Code*.

На другому етапі роботи задля встановлення вихідної продуктивності мобільної гри *AirMan* за допомогою додатку *Apptim* проведено її тестування. Встановлено, що під час завантаження ігрового рівня, а це двадцять друга секунда, спостерігається зростання навантаження на оперативну пам'ять.

При старті ігрового процесу навантаження досягає більше ніж 445 мегабайт, а піком є 448 мегабайт.

В середньому під час тестування використання оперативної пам'яті становило 363,5 мегабайта.

Першим кроком оптимізації стало відключення тих налаштувань, які непотрібні для 2D гри.

Для оптимізації замість *API OpenGL4*:

використано *OpenGL2*,

відключено налаштування згладжування тіні,

виключено *Reflection Probes*,

виключено *Soft Particles*

перемкнено тип рендерінгу з *Linear* на *Gamma* [18].

В результаті тестування після першого кроку оптимізації одержано результат у вигляді зменшення використання максимального піка оперативної пам'яті на 3 мегабайти (445 мегабайта проти 448 мегабайта), а середнє значення впало майже на 4 мегабайти (359,7 МВ проти 363,5 МВ).

Другим кроком оптимізації стало зменшення використання асетами гри оперативної пам'яті. За допомогою *Unity Profiler* з'ясовано, що текстури займають найбільше всього пам'яті – 327 мегабайтів, а найбільше місця в пам'яті займають картинки для фону.

Зменшенням якості картинок досягнуто зменшення використання оперативної пам'яті. Так, наприклад, в результаті оптимізації картинка *bgr-level1*, яка займала 24,9 мегабайтів пам'яті, після оптимізації стала важити 12,5 мегабайтів.

Слід зазначити, що наявна якість картинок від самого початку була надто високою, зокрема, розмір *bgr-level1* становив 2048 пікселів. Тому її розмір зменшено до 64 пікселів. Такі дії не вплинули на вигляд картинок.

Щоб побачити наскільки вплинуло зменшення розмірів картинки на обсяг використання оперативної пам'яті, проведено тестування ігрового рівня в *Profiler* і з'ясовано, що зменшенням якості однієї картини *bgr-level1* досягнуто зменшення використання оперативної пам'яті з 24,9 мегабайтів до 25,5 кілобайтів.

Отже, усі картини ігрового рівня, які можна зменшити – зменшено. При цьому без змін залишені ті спрайти, на вигляд яких зменшення розміру впливає суттєво, наприклад, *bgr-town* залишено без змін. В

результаті оптимізації асетів за допомогою *Profiler* маємо зменшення використання текстурами оперативної пам'яті з 250 мегабайт до 71,3 мегабайт.

За даними тестування в *Apptim* після оптимізації за допомогою *Profiler* маємо результат: при старті

ігрового процесу пік навантаження на оперативну пам'ять досягає 353 мегабайтів, це менше майже на 100 мегабайт ніж було після першого кроку оптимізації *Player Settings* (445 мегабайт) та на 28,9 мегабайт менше (330,3 проти 359,2 мегабайт) в середньому під час тестування використання оперативної пам'яті.

Третім кроком оптимізації була оптимізація коду проекту, а саме одного з найголовніших та найбільших за кількістю коду скриптів – *Player*, який містить у собі функціональність пов'язану з персонажем на ігровому рівні, яким керує користувач. В

скриптах *OnClickRight()*, *GetComponent()*, з яким пов'язані переміщення головного героя, здійснено перенесення коду *Find()*, до функції *Start()*, щоб дія виконувалась лише один раз – на самому початку роботи скрипта.

За даними тестування в *Apptim* маємо результат у вигляді використання 336 мегабайтів оперативної пам'яті у піку, це менше на 17 мегабайт ніж було після попередньої оптимізації (353 мегабайт) та на 20,1 мегабайт менше (310,2 проти 330,3 мегабайт) середній показник використання пам'яті. Покрокові зміни процесу оптимізації надано на рис. 1.

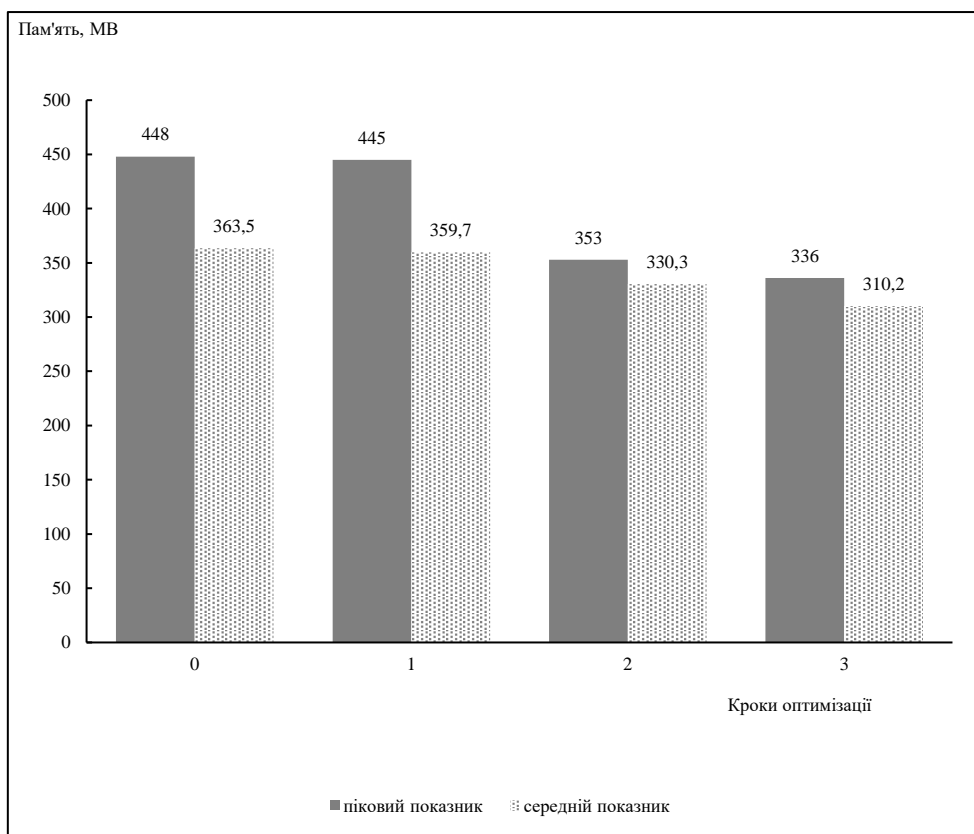


Рис. 1. Зміни використання оперативної пам'яті девайсу грою *AirMan* в процесі оптимізації

Спочатку доступність мобільної гри *AirMan* була обмежена тільки *Android* платформою, тобто на всіх інших платформах гра була недоступною. Але з метою охоплення більшої аудиторії користувачів ігри вирішено було зробити її кросплатформною та розширити доступність для персональних комп'ютерів на системі *Windows*. Через те, що розмір вікна гри від самого початку був розрахований тільки під мобільні екрани, то багато елементів гри стали маленькими за розмірами, через що з'явився порожній простір, який юзер не повинен бачити.

В процесі роботи над доступністю гри під *Windows*, були домальовані деякі елементи гри, підігнані інтерфейс та бекграунд під розміри екрана комп'ютера.

В результаті гра *AirMan* стала доступною і на платформі *Windows*.

Висновки

В результаті проведених досліджень навантаження на оперативну пам'ять девайсу при старті ігрового процесу зменшено з 445 до 336 мегабайт, а середній показник використання пам'яті становив 310,2 проти 363,5 мегабайт при початковому тестуванні гри.

Розширено доступність гри для персональних комп'ютерів на системі *Windows*, що надалі стане поштовхом для збільшення аудиторії гравців, адже у таку гру люди зможуть грати на будь-якому пристрої – смартфоні або комп'ютері.

СПИСОК ЛІТЕРАТУРИ

1. Електронна стаття Tom Wijman «The Games Market Will Decline -4.3% to \$184.4 Billion in 2022; Long-Term Outlook Remains Positive» [Електронний ресурс] – Режим доступу: <https://newzoo.com/insights/articles/the-games-market-will-decline-4-3-to-184-4-billion-in-2022> (дата звернення 25.11.2022).

2. Електронна стаття Emma McDonald «Newzoo's Games Market Estimates and Forecasts» [Електронний ресурс] – Режим доступу: <https://newzoo.com/insights/articles/the-latest-games-market-size-estimates-and-forecasts> (дата звернення 25.11.2022).
3. Hui, Ng & Chieng, Liu & Ting, Wen & Mohamed, Hasimah & Mohd Arshad, Muhammad. (2013), «Cross-platform mobile applications for android and iOS», *6th Joint IFIP Wireless and Mobile Networking Conference (WMNC)*, Pp. 1-4. doi: <https://doi.org/10.1109/WMNC.2013.6548969>.
4. П. Дейтел, Х. Дейтел, Э. Дейтел, М. Моргано (2013), «Android для программистов: создаём приложения», 560 с.
5. Yan, Q., Hu, GY., Ni, GQ., Jiang, JS., Long, JX. (2016), «Research on the Development Technology of Cross Platform Hybrid Mobile Application Based on HTML5», *Proceedings of the 2016 International conference on communications, information management and network security*, 47, Pp.168-171, WOS:000390854400042, ISBN 978-94-6252-247-3, ISSN 2352-538X
6. Седых И.А. (2020), «Индустрия компьютерных игр», 74 с.
7. Cheon, Y., Chavez, C. (2021), «Converting Android Native Apps to Flutter Cross-Platform Apps», *IEEE: 2021 International conference on computational science and computational intelligence (CSCI 2021)*, Pp. 1898-1904. doi: <https://doi.org/10.1109/CSCI54926.2021.00355>, WOS:000832229300341, ISBN: 978-1-6654-5841-2.
8. Bennis, L., Amali, S., El Faddouli, NE. (2020), «Generating Adaptive Learning Games to Multiple Mobile Platforms», *Advanced intelligent systems for sustainable development (AI2SD'2019)*, V. 1 – Advanced intelligent systems for education and intelligent learning system. *Advances in Intelligent Systems and Computing*, 1102, Pp.129-135. doi: https://doi.org/10.1007/978-3-030-36653-7_13, WOS:000656953100013, ISSN: 2194-5357, eISSN: 2194-5365. ISBN: 978-3-030-36653-7; 978-3-030-36652-0.
9. Vaquero-Melchor, D., Bernardos, AM., Bergesio, L. (2020), «SARA: A Microservice-Based Architecture for Cross-Platform Collaborative Augmented Reality», *Applied sciences-basel*. 10, 6, 2074, doi: <https://doi.org/10.3390/app10062074> WOS:000529252800175.
10. Smilovitch, M., Lachman, R. (2019), «BirdQuestVR: A Cross-Platform Asymmetric Communication Game», *CHI PLAY'19: Extended abstracts of the annual symposium on computer-human interaction in play*, Pp. 307-313, doi: <https://doi.org/10.1145/3341215.3358246/>, ISBN: 978-1-4503-6871-1.
11. Sarinho, VT., de Azevedo, GS., Boaventura, FMB. (2018), «Providing an IM Cross-Platform Game Engine for Text-Messaging Games», *17th brazilian symposium on computer games and digital entertainment (SBGAMES 2018): Brazilian Symposium on Games and Digital Entertainment SBGAMES*, Pp.212-220, doi: <https://doi.org/10.1109/SBGAMES.2018.00033> WOS:000467962600024, ISSN: 2159-6654, eISSN: 2159-6662 ISBN: 978-1-5386-7769-8.
12. Schlueter, J., Baiotto, H., Hoover, M., Kalivarapu, V., Evans, G., Winer, E. (2017), «Best practices for cross-platform virtual reality development», *Degraded environments: sensing, processing, and display 2017*, V. 10197, UNSP 1019709. doi: <https://doi.org/10.1117/12.2262718>, WOS:000406803100007.
13. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: <https://doi.org/10.20998/2522-9052.2018.1.04>.
14. Свиридов А. С., Коваленко А. А., Кучук Г. А. Метод перерозподілу пропускної здатності критичної ділянки мережі на основі удосконалення ON/OFF-моделі трафіку. *Сучасні інформаційні системи*. 2018. Т. 2, № 2. С. 139–144. DOI: <https://doi.org/10.20998/2522-9052.2018.2.24>.
15. Левашенко В. Г., Ляшенко О. С., Кучук Г. А. Побудова системи підтримки прийняття рішень на основі нечітких даних. *Сучасні інформаційні системи*. 2020. Том 4, № 4. С. 48-56. DOI: <https://10.20998/2522-9052.2020.4.07>
16. Кучук Г.А., Матвеев М.І (2022), «Мобільна гра в середовищі Unity проблеми інформатизації», *Тези доповідей десятої міжнародної науково-технічної конференції: «Проблеми інформатизації», 24 – 25 листопада 2022 року*, Т. 2, С.116, available at: <https://er.chdntu.edu.ua/handle/ChSTU/4231>.
17. Кучук Г.А., Матвеев М.І (2022), «Розробка двовимірного ігрового застосування», *Матеріали IV Міжнарод. наук.-техн. конфер.: «Автоматизація, електроніка, інформаційно-вимірні технології: освіта, наука, практика», 01- 02 грудня 2022 р.*, С. 19-20, available at: <https://repository.kpi.kharkov.ua/handle/KhPI-Press/60095>.
18. Електронна стаття «Гайд по оптимізації мобільних ігор в Unity» [Електронний ресурс] – Режим доступу: <https://dtf.ru/s/unity/293563-gayd-po-optimizacii-mobilnyh-igr-v-unity> (дата звернення 19.03.2023).

Received (Надійшла) 11.03.2023

Accepted for publication (Прийнята до друку) 26.04.2023

Mobile game optimization in the unity environment

Mykyta Matvieiev, Heorhii Kuchuk

Abstract. Topicality. Currently, mobile games are the fastest growing segment of the gaming market. Games confidently rank first in terms of revenue and downloads on *Google Play* and the *App Store*. However, even today, developers are increasingly switching from mobile platforms to personal computers, because there is an increase in user interest in mobile games that can work not on one, but on several platforms. These games allow the same code to be deployed across multiple platforms so that people can play the game on any device (smartphone or computer). **The goal of this work** is to optimize the work of our own developed mobile game in the *Unity* environment and expand its availability for *Windows* personal computers. **The object of research** is the 2D mobile game *AirMan* in the *Unity* environment. **The subject of research** is the methods of diagnostics and optimization of the game. **Results.** As a result of game optimization, the load on RAM at the start of the game process was reduced from 445 to 336 megabytes, and the average memory usage was 310.2 megabytes compared to 363.5 megabytes during the initial testing of the game. *AirMan* game is made cross-platform. **Conclusions.** The work of the *AirMan* game in the *Unity* environment has been improved and its availability for personal computers on the *Windows* system has been expanded, and as a result, the audience of players has been increased.

Keywords: mobile game, optimization, cross-platform, unity, android, windows.

А. О. Рибальченко

Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна

АЛГОРИТМИ РІШЕННЯ ЗАДАЧІ ОПТИМАЛЬНОГО РОЗМІЩЕННЯ ДАНИХ В БІЛІНГОВИХ OLTP-СИСТЕМАХ НА ОСНОВІ РЕАЛІЗАЦІЇ РАНГОВОГО ПІДХОДУ

Анотація. У статті приведено результати розробки наближених та точних алгоритмів рішення задачі оптимального розміщення даних у білінгових OLTP-системах на основі реалізації рангового підходу. Даний тип задач відноситься до класу цілочисельного лінійного програмування (ЦЛП) з булевими змінними (БЗ). Домінуюче місце у методах рішення таких задач у даний час займають комбінаторні методи та еволюційні алгоритми. Практичне застосування даних методів ускладнено при рішенні задач великої розмірності. Для усунення даної проблеми пропонується використовувати ідею рангового підходу. Наведено аналіз підходів до оптимального розміщення даних у білінгових OLTP-системах, модель рангового підходу, а також наближені та точні алгоритми. **Об'єктом дослідження** є алгоритми функціонування і розміщення інформаційних ресурсів у хмарному середовищі, концепція хмарних обчислень та багаторівневих інформаційних систем. **Предметом дослідження** є принципи розміщення даних, що зберігаються у розподілених базах даних (РБД) та циркулюючих у хмарній мережі, а також специфіка процесів обслуговування абонентів у сучасних реалізаціях OLTP-систем. **Метою наукової роботи** є розробка наближених та точних алгоритмів оптимізації розміщення фрагментів РБД по вузлах мережі хмарної структури, які дозволять збільшити продуктивність інформаційної системи за рахунок раціонального розподілу даних. **Висновки.** Запропоновано стратегії відсікання безперспективних шляхів у множинах, що призводять до наближених і точних рішень задачі ЦЛП з БЗ та побудовано ефективні точні і наближені алгоритми. Показано, що важливою перевагою розроблених алгоритмів на основі рангового підходу є той факт, що збільшення числа обмежень практично не впливає на погіршеність рішень алгоритмів, тоді як для методів рішення задач дискретної оптимізації, що засновані на ідеях методу гілок та кордонів, зростання числа обмежень до декількох сотень приводить фактично до неможливості їхнього практичного застосування.

Ключові слова: ранговий підхід, цілочисельне лінійне програмування, булеві змінні, оптимізація за напрямком, наближені та точні алгоритми.

Вступ

Одним із етапів проектування розподілених баз даних (РБД) є планування фрагментації та розміщення, тобто розбиття баз даних (БД) на фрагменти та ухвалення рішення про те, де зберігатимуться ці фрагменти. Однак, проектування схем фрагментації та розміщення відношень ґрунтується на інформації про способи та методи використання РБД. Методи використання залежать від схеми фрагментації та розміщення.

Отже, задачу проектування РБД слід формулювати наступним чином: для даної конфігурації обчислювальної системи (ОС) необхідно описати схему розміщення фрагментів таким чином, щоб оптимізувати цільову функцію.

Для досягнення високої продуктивності розподілених OLTP-систем необхідні ефективні методи управління даними та грамотний вибір стратегії зберігання інформації у РБД. Особливо це стосується транзакційних систем, які побудовані на базі хмарної платформи, оскільки концепція плати у міру використання ресурсів змушує шукати компроміс між продуктивністю та витратами.

Тому, одним із актуальних завдань є оптимальне розміщення даних у розподіленому середовищі (хмарі) за такими критеріями, як загальна мінімальна вартість трафіку, який породжений функціонуванням ОС протягом одиниці часу, загальна мінімальна вартість оренди дискового простору та ін. Їх рішення дозволить досягти значного скорочення витрат та підвищення швидкості роботи системи.

Даний тип задач відноситься до рішення задачі цілочисельного лінійного програмування (ЦЛП) з булевими змінними (БЗ).

Домінуюче місце у методах рішення таких задач у даний час займають комбінаторні методи. До них, у першу чергу, можна віднести методи повного перебору, гілок та кордонів, динамічного програмування, а також локальні алгоритми. Практичне застосування даних методів ускладнено при рішенні задач великої розмірності.

Спроби зменшення часу рішення задач ЦЛП з БЗ за рахунок розпаралелення зштовхуються із іншою проблемою теорії паралельних обчислень, яка полягає у тому, що з точки зору паралельних алгоритмів даний тип задач відноситься до класу сильнов'язаних задач. Тому, погано піддається розпаралеленню.

Отже, при розробці паралельних алгоритмів для рішення задачі ЦЛП з БЗ, крім протиріччя між точністю рішення задачі та часом її рішення, виникає ще одне протиріччя – між сильною зв'язністю властивостей даної задачі та необхідністю її розпаралелення.

Таким чином, розробка алгоритмів рішення задачі оптимального розміщення даних в білінгових OLTP-системах на основі рангового підходу (РП) – є актуальною науковою задачею.

Аналіз публікацій за темою дослідження. У роботі [1] сформульований набір математичних постановок задач оптимізації розміщення даних на різних етапах роботи тренажерно-моделюючих комплексів (ТМК) та розроблені моделі розміщення

даних на основі хмарних обчислень і інформаційних кластерів.

Крім того, автором запропоновано алгоритми оптимізації розміщення модельного світу ТМК на основі комбінування підходів початкової стратегії розподілу і еволюційних методів генетичних алгоритмів (ГА). Однак, результати дослідження застосовні тільки до ТМК різного призначення, оскільки запропоновані моделі засновані на специфічних особливостях побудови та функціонування подібних систем.

На даний час, усе більшої значущості набуває процес проектування РБД. У більшості джерел у якості окремого етапу проектування РБД виділяють фрагментацію і розміщення БД (розбиття БД на фрагменти і прийняття рішення про те, де вони будуть зберігатися) [2]. Однак, завдання побудови саме оптимальної структури РБД часто залишається поза увагою, що значно впливає на продуктивність інформаційної системи у цілому.

У роботі [3] коротко викладаються методи оптимального розміщення фрагментів РБД по вузлах обчислювальної мережі. У якості цільових функцій в даній задачі можуть виступати швидкість виконання запитів у системі і загальна вартість трафіку.

У роботі [4] розглядаються підходи до оптимізації структури сховища даних у вузлах інфокомунікаційної мережі хмарного середовища.

Таким чином, проблема оптимального розміщення фрагментів БД описується математично у вигляді цільової функції однієї змінної, що є бінарною матрицею, яка характеризує розташування конкретного фрагмента на конкретному вузлі.

Стратегія виконання запитів представляється у вигляді матриць інтенсивностей запитів з певного вузла до певного фрагменту, матриць довжин запитів, а також матриць довжин відповідей на запити. Пошук глобального оптимуму цільової функції проводиться за допомогою ГА, що використовує автоматичну ідентифікацію параметрів.

Застосування отриманих результатів до будь-якої розподіленої інформаційної системи підвищить продуктивність її роботи за рахунок зменшення часу виконання запитів та обробки оновлень шляхом оптимального розподілу даних по вузлах комп'ютерної мережі.

Однак, у розглянутих працях не розкрито питання щодо розробки ефективних наближених та точних алгоритмів рішення задачі оптимізації структури РБД на основі рангового підходу (РП).

Метою статті є розробка ефективних наближених та точних алгоритмів рішення задачі оптимізації структури РБД на основі РП.

Основна частина

1. Аналіз підходів до оптимального розміщення даних у білінгвових OLTP-системах. Завдання оптимального розміщення даних у білінгвових OLTP-системах, можуть бути вирішені комбінаторними методами, а саме – методами гілок та кордонів, методами відсікання, до яких належать алгоритми Гоморі та Балаша. Проте, такі завдання є NP-

повними, тобто зі зростанням розмірності завдання їхня обчислювальна складність зростає експоненційно. Тому, у багатьох дослідженнях зазначається, що для вирішення подібних завдань кластеризації та компонування найбільш успішно застосовуються ГА, які належать до класу еволюційних алгоритмів (ЕА).

Еволюційні алгоритми оптимізації – це клас методів, які засновані на імітації природного відбору, генетичних мутацій та кросоверів, що використовуються для вирішення оптимізаційних задач.

Сутність ЕА полягає у тому, що вони працюють з популяцією можливих рішень, а не з окремими точками простору параметрів, як це робиться, наприклад, у градієнтних методах. Починаючи з деякої початкової популяції рішень, ЕА виконують ітерації, під час яких здійснюється відбір кращих особин (індивідів), на основі яких створюються нові потомки шляхом зміни генетичного матеріалу (кросоверів та мутацій), тобто проводиться еволюція популяції. Даний процес повторюється доти, доки не буде досягнута задовільна точність або не буде вичерпана видима прогресивність.

Основна перевага ЕА полягає у тому, що вони можуть ефективно вирішувати задачі оптимізації з великою кількістю варіантів вхідних параметрів або складною необхідністю задати точну форму функції витрат. Також, ЕА можуть працювати у обмеженому просторі пошуку, де інші методи можуть втратити свою ефективність.

До недоліків ЕА можна віднести наступні:

- евристична природа алгоритму не гарантує отримання абсолютно раціонального рішення;
- еволюція може зупинитися на непродуктивній гілці (цей недолік можна запобігти за допомогою використання паралельних обчислень);
- невисока ефективність ЕА на останніх стадіях пошуку локального оптимуму.

На основі викладених переваг та недоліків ЕА можна зробити висновок про те, що даний клас обчислювальних методів найбільше підходить для вирішення задач багатовимірної оптимізації, для яких немає адекватних нееволоційних методів рішення та складних завдань комбінаторної оптимізації великої розмірності.

Розглядаються наступні алгоритми вирішення представленої задачі: жадібний алгоритм (ЖА), алгоритм імітації відпалу та кілька модифікацій ГА.

Жадібний алгоритм – є простим алгоритмом оптимізації, який використовується для знаходження локальних оптимальних рішень у задачах оптимізації. Сутність ЖА полягає у тому, що на кожному кроці алгоритму вибирається найкращий з доступних варіантів, що призводить до знаходження локально оптимального рішення.

У більшості випадків ЖА не гарантує знаходження глобально оптимального рішення, а лише дозволяє отримати локально оптимальне рішення. Це пов'язано з тим, що на кожному кроці алгоритму виконується локально оптимальний вибір, не зважаючи на можливі варіанти розвитку подальших кроків алгоритму.

Однак, ЖА має свої переваги. Він є простим у реалізації і може працювати дуже швидко навіть для великих об'ємів даних. Крім того, ЖА може використовуватись як евристика для більш складних алгоритмів оптимізації, які враховують глобальну оптимальність.

Метод імітації відпалу (МІВ, англ. simulated annealing) – це стохастичний алгоритм оптимізації, який базується на аналогії з процесом охолодження металів. Основна ідея полягає у тому, що метал, який нагрівається, потрапляє у стан з високою енергією, що дозволяє металу рухатися й приймати нові форми. Після цього, метал поступово охолоджується, його енергія зменшується, що призводить до зниження рухомості атомів та, у кінці кінців, метал застигає у новій структурі з меншою енергією.

Таким же чином, МІВ починається зі стану системи з високою енергією (наприклад, початкового розв'язку) та, поступово, зменшує енергію системи шляхом зниження температури. У цей час можуть відбуватися випадкові зміни стану системи (наприклад, за допомогою переміщень, додавання або видалення елементів тощо), які можуть збільшувати енергію системи, або зменшувати її, залежно від того, як вони впливають на функцію метрики.

У МІВ, на відміну від градієнтних методів, зміна стану системи може бути не тільки у напрямку зменшення значення функції метрики, а й у зворотному напрямку, залежно від температури та інших параметрів.

Метод імітації відпалу має кілька переваг порівняно з іншими методами оптимізації:

- здатність уникнути застрягання у локальному оптимумі: МІВ здатний перескакувати через локальні мінімуми і шукати глобальний мінімум. Це досягається завдяки використанню стохастичного процесу;

- можливість роботи з функціями складної форми: МІВ здатний працювати з функціями складної форми, які мають багато локальних мінімумів та максимумів. Це досягається завдяки тому, що метод використовує випадкові зміщення;

- швидкість роботи: МІВ зазвичай швидший за генетичні алгоритми та інші методи оптимізації, оскільки він не вимагає створення та збереження повної популяції;

- не вимагає похідних: МІВ не вимагає вирахування похідних функції, що робить його корисним для оптимізації функцій, які важко або неможливо диференціювати;

- здатність до паралельної обробки: МІВ здатний до паралельної обробки, що дозволяє використовувати багатоядерні процесори та комп'ютерні кластери для прискорення процесу оптимізації.

До недоліків методу імітації відпалу можна віднести наступні:

- залежність від параметрів: МІВ має декілька параметрів, таких як температура, розмір шагу тощо, і правильний вибір цих параметрів може сильно впливати на ефективність методу;

- можливість застрягання у локальному мінімумі: МІВ може застрягнути у локальному мінімумі,

як і більшість методів оптимізації. Хоча існують різні стратегії для уникнення цього, такі як додавання випадкових складових. Однак, це все ще може бути проблемою;

- обчислювальні витрати: МІВ може бути досить обчислювально витратним, особливо для складних оптимізаційних задач. Це може стати проблемою при використанні методу для великих даних або у реальному часі;

- не гарантує глобальний оптимум: хоча МІВ зазвичай знаходить гарні рішення, він не гарантує знаходження глобального оптимуму, особливо у випадку складних оптимізаційних задач з багатьма локальними мінімумами.

Таким чином, МІВ – є потужним та ефективним методом оптимізації, який може бути використаний для рішення різноманітних задач оптимізації.

На даний час набули широкого поширення адаптивні ГА. Це пов'язано з тим, що класичний ГА може бути успішно застосований далеко не до усіх завдань, а для отримання оптимального рішення у конкретній області часто потрібна зміна будь-яких параметрів алгоритму, застосування нових підходів до конструювання хромосоми та подання фітнес-функції. У зв'язку з цим, доцільним є використання адаптивних алгоритмів – тобто алгоритмів, які здатні змінювати свої параметри у процесі роботи.

Під час дослідження ГА для вирішення задач оптимізації у розподілених системах використовувалась низка підходів та модифікацій класичного ГА. В класичному ГА, який є одним з основних видів ЕА, кожен індивід має свою фітнес-функцію, яка визначає його пристосованість до середовища. Індивіди з більш високою фітнес-функцією мають більші шанси на відбір для репродукції та наступних поколінь. У ГА з відносною фітнес-функцією, також відомому як пропорційний відбір, пристосованість індивіда визначається не його абсолютним значенням фітнес-функції, а відносною величиною у порівнянні з іншими індивідами у популяції. Індивіди з вищим значенням фітнесу у порівнянні з рештою популяції отримують більшу ймовірність на відбір для наступного покоління.

Отже, в основі обох видів ГА лежить ідея відбору найкращих рішень за допомогою певної фітнес-функції. Відмінність полягає у тому, як саме вимірюється пристосованість індивіда та як це відображається на процесі відбору та репродукції.

У стандартному ГА мутація проводиться шляхом зміни випадково обраного біту у геномі (тобто у рядку бінарних чисел, які представляють рішення). Така мутація може не вплинути на конкретну змінну, яка повинна бути змінена, а також може вплинути на змінну, яка не потребує зміни. Це може призвести до втрати цінної інформації про оптимальність рішення.

У ГА з роздільною мутацією кожна змінна мутується окремо. Тому, можна досягти більш точної оптимізації кожної змінної. Даний метод дозволяє збільшити швидкість збіжності алгоритму, оскільки окрема мутація не впливає на усі змінні одразу, а лише на ті, які потребують зміни.

Однак, необхідно мати на увазі, що роздільна мутація може привести до того, що різні змінні не взаємодіють між собою у оптимальний спосіб. Тому доцільно використовувати її у поєднанні з іншими методами оптимізації, наприклад, з мутацією зміни точки перетину або зі зміною розміру популяції.

ГА Уїтлі (Whitley) – є одним з варіантів стандартного ГА і включає у себе додаткову операцію, що називається "локальною оптимізацією". Така операція дозволяє зменшити кількість ітерацій ГА, необхідних для знаходження оптимального рішення.

Основна суть ГА Уїтлі полягає у тому, що на кожній ітерації алгоритму застосовується додаткова операція, яка полягає у виборі декількох найкращих особин з попередньої популяції і подальшому запуску на них локального пошуку. Локальний пошук може бути будь-яким ефективним алгоритмом пошуку оптимуму, наприклад, методом найшвидшого спуску, методом Ньютона або методом золотого перетину. Після локального пошуку отримані розв'язки вносяться у нову популяцію разом з найкращими розв'язками з попередньої ітерації. Потім, на новій популяції виконуються стандартні операції ГА, такі як селекція, кросингвер та мутація, і так далі до знаходження оптимального розв'язку.

Головною перевагою ГА Уїтлі є те, що він дозволяє збільшити швидкість збіжності алгоритму за рахунок використання локального пошуку. Проте, його недоліком може бути те, що локальний пошук може застрягти у локальному оптимумі, що може привести до погіршення результатів алгоритму.

ГА Ешельмана (англ. Eshelman's genetic algorithm) – це вид ГА, який був запропонований Кеннетом Ешельманом у 1991 році. Його основна ідея полягає у тому, щоб використовувати "турнірну" селекцію, у якій кожен індивідуум конкурує з іншими за право на наступне покоління, а не зважається його абсолютна пристосованість.

Головна особливість ГА Ешельмана – це використання методу "без заміщення" при проведенні турнірів, що означає, що кожен індивідуум може бути відібраний лише один раз для участі у турнірі, що унеможливує відбір дублікатів.

Другою особливістю ГА Ешельмана є використання оператора "головного гена", який гарантує збереження кращих індивідуумів з попереднього покоління у наступному без яких-небудь змін, що дозволяє уникнути можливої втрати корисної інформації та зменшення різноманітності у популяції.

Таким чином, ГА Ешельмана – є ефективним методом оптимізації, який дозволяє досягти більшої різноманітності у популяції та забезпечити надійну збереженість кращих рішень від покоління до покоління.

2. Модель рангового підходу до рішення задачі оптимального розміщення даних у білінгових OLTP-системах. Оптимізація структури РБД заснована на моделі, яка має такий вигляд [5-11]:

$$f(\bar{x}) = \sum_{j=1}^n c_j \cdot x_j \Rightarrow \max, \quad (1)$$

$$L = \sum_{j=1}^n c_j x_j \rightarrow \min, \quad (2)$$

$$\sum_{j=1}^n a_{ij} \cdot x_j \leq b_i, \quad (3)$$

$$\sum_{j=1}^n a_{ij} x_j \geq 1, \quad i = \overline{1, m}; \quad (4)$$

$$x_j \in \{0, 1\}; c_j \geq 0; i = \overline{1, m}; j = \overline{1, n}. \quad (5)$$

Ідея РП основана на представленні n -мірного одиничного куба у виді графа GA , геометричний зміст якого полягає у наступному [5-11]. Геометрично вершина k графа GA рангу r – це множина векторів $(x_1, x_2, \dots, x_k, \dots, x_n)$, у яких $x_k=1$, а на позиціях від 1 до k знаходиться r одиниць. Ребру, що входить у вершину k графу GA , відповідає одиничний вектор $(0, 0, \dots, 0, 1, 0, \dots, 0)$ n -мірного одиничного куба B_n з одиницею у k -й позиції. Тоді, шляху μ_{sj}^r рангу r у графі GA відповідає вектор \bar{x} , який дорівнює сумі одиничних векторів ребер, по яким він досяг вершину j рангу r , починаючи з вершини s . Множину шляхів у графі GA до вершин j , розташованих на ярусах від вершини s , можна представити наступним чином:

$$m_s^r(j) = m_{sj}^{r=1} \cup m_{sj}^{r=2} \cup \dots \cup m_{sj}^{r=n}, \quad j = \overline{1, n}, \quad (6)$$

де $m_s^r(j)$ – множина шляхів у графі GA від вершини s до вершин j , розташованих на r -х ярусах графа GA , (ранг шляху $\mu_{sj}^r \in m_{sj}^r$ визначається числом ребер, що утворюють цей шлях).

Варто мати на увазі те, що множині шляхів $m_{sj}^{r=k}$ у графі GA відповідає множині векторів, що містять k одиниць. Отже, $|m_{sj}^r| = C_n^{r=k}$, тобто кожному шляху у множині $m_s^r(j)$ відповідає деякий вектор (x_1, x_2, \dots, x_n) . Використавши (6) можливо записати наступне:

$$|m_s^r(j)| = C_n^{r=1} + C_n^{r=2} + \dots + C_n^{r=n} = 2^n - 1. \quad (7)$$

Таким чином, граф GA є упорядкований по рангах еквівалент n -мірного одиничного куба B_n , у якому шляхи $\mu_{sj}^r \in m_{sj}^r$ відповідають вершинам B_n . Для рішення задач у [1-4] запропонована система каліброваних векторів, що дозволяє звести рішення цих задач до визначення екстремальних шляхів у графі GA . Сформульовано нтаке:

– принцип оптимізації за напрямком, обумовлений співвідношенням:

$$\forall (\mu_{sj}^r \in m_{sj}^r) \left[\mu_{sp}^{r=r+1} = L_w \{ \mu_{sj}^r \cup (j, p) \} \right]; \quad (8)$$

$$p = \overline{(r+1, n)}; \quad j = \overline{(r, n)},$$

– принцип виділення коридору у підмножинах:

$$m_{sp}^{r=r+1} = \left\{ L_w (\forall (\mu_{sj}^r \cup (j, p))) \right\}, \quad (9)$$

де $\{L_w\}$ – правила відсікання шляхів μ_{sj}^r у множинах m_{sj}^r ; $\mu_{sj}^r \cup (j, p)$ – шлях з вершини s графу GA шлях у вершину p , що проходить через проміжну вершину j та задовольняє правилам $\{L_w\}$, тобто шлях μ_{sp}^r отримано за рахунок приєднання до шляху μ_{sj}^r ребра (j, p) , якщо таке з'єднання не суперечить правилам $\{L_w\}$.

3. Розробка наближених алгоритмів оптимального розміщення даних у білінгових ОЛТР-системах. Для рішення задач типу (1-5), пропонується використовувати узагальнену процедуру A_0 , яка дозволяє визначити локальні екстремуми у W -областях графу GA щораз і, потім, виділяти глобальний екстремум з $n(n+1)/2$ локальних, що отримані на основі принципу оптимізації за напрямком з використанням правил відсікання $\{L_w\}$ шляхів у множинах, що вводяться.

Узагальнена процедура A_0

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}$, $j = \overline{(1, n)}$, що задовольняють властивості v . Виділяються шляхи $\mu_{sj}^{*r=1}$, що визначають локальні екстремуми областей Ω_j .

КРОК 2. Формуються множини шляхів $m_{sp}^{r=r+1}$ $p = \overline{(r+1, n)}$ наступного рангу, що задовольняють властивості v , на базі множини шляхів m_{sj}^r попереднього рангу відповідно до співвідношення (8). В утворених множинах $m_{sp}^{r=r+1}$ здійснюється відсікання шляхів відповідно до обраного правила відсікання $\{L_w\}$ та виділяються шляхи $\mu_{sp}^{*r=r+1}$, що визначають локальні екстремуми областей Ω_p .

КРОК 3. Перевіряємо, чи усі множини $m_{sp}^{r=r+1}$ наступного рангу порожні. За умови, якщо це так, то переходимо до кроку 4, якщо ні, то перевіряємо $r = (n - 1)$. У випадку виконання рівності переходимо до кроку 4, інакше збільшуємо r на 1 та виконуємо крок 2.

КРОК 4. Виділяємо серед множин локальних екстремумів Ω_j $j = \overline{(1, n^2/2)}$ глобальний і процедура A_0 закінчує роботу.

Узагальнена процедура A_0 дозволяє визначити локальні екстремуми у Ω -областях графу GA щораз на кроці 2 і потім на кроці 4 виділити глобальний екстремум з $n^2/2$ локальних, які отримуються на основі принципу оптимізації за напрямком (8) з використанням правил відсікання, що вводять, $\{L_w\}$ шляхів у m_{sj}^r множинах.

При побудові наближених алгоритмів із множини правил $\{L_w\}$, розглянутих у [1-6], будемо розглядати таке:

$$L1: \mu_{sp}^{r=r+1} = \max_{\{c_j\}} \{ \mu_{sj}^r \cup (j, p) \}, \quad (10)$$

$$p = \overline{r+1, n}; \quad j = \overline{r, n}; \quad j \neq p;$$

$$L2: \mu_{sp}^{r=r+1} = \min_{\{\alpha_j\}} \{ \mu_{sj}^r \cup (j, p) \}, \quad (11)$$

$$p = \overline{r+1, n}; \quad j = \overline{r, n}; \quad j \neq p;$$

$$L3: \begin{cases} \mu_{sp}^{r=r+1} = \max_{c_j} \{ \mu_{sj}^r \cup (j, p) \}; \\ \mu_{sp}^{r=r+1} = \min_{\alpha_j} \{ \mu_{sj}^r \cup (j, p) \}. \end{cases} \quad (12)$$

З множин m_{sj}^r будемо виключати як не перспективні шляху μ_{sp}^r , які задовольняють нерівності:

$$d_c(\mu_{sp}^r) + \gamma_p < \max_{\{c_j\}} \{ d_c(\mu_{sp}^{*r}) \}, \quad (13)$$

$$\text{де} \quad \gamma_p = c_{p+1} + c_{p+2} + \dots + c_n \quad (14)$$

(для вершини $j \neq n$ вага γ_j – приймається рівною нулю); $d_s(\mu_{sp}^r)$ – довжина шляху до вершини p рангу r по вагах функціоналу.

При цьому, покроковий опис наближених алгоритмів, позначимо їх MAX та MIN, буде мати наступний вигляд

Алгоритм MAX

Крок 1. Виконання сортування значень функціоналу й обмежень вигляду:

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n; \quad (15)$$

$$c_1 \geq c_2 \geq \dots \geq c_n; \quad (16)$$

$$\frac{c_1}{\alpha_1} \geq \frac{c_2}{\alpha_2} \geq \dots \geq \frac{c_n}{\alpha_n}. \quad (17)$$

Крок 2. З вершини s будуються множини шляхів $\{ \mu_{sj}^{r=1} \}$, $j = \overline{(1, n)}$, першого рангу r , що задовольняють властивості γ , і визначаються в множинах $m_{sj}^{r=1}$ шляхи максимальної довжини $\{ \mu_{sj}^{*r} \}$ по вагах функціоналу $\{c_j\}$. Для кожної вершини j визначається вага γ_j відповідно до (14). Для вершини $j=n$ вагу γ_j приймаємо рівною 0.

Крок 3. Виключаються $\{ \mu_{sp}^r \}$, $p = \overline{(r, n)}$ у множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівності (13).

Крок 4. Формуються множини шляхів $m_{sj}^{r=r+1}$, $j = \overline{(r+1, n)}$ такого рангу, що задовольняють властивості γ , на базі множин шляхів попереднього рангу m_{sp}^r на основі рекурентного співвідношення (10). У множинах $m_{sj}^{r=r+1}$, що утворилися,

виділяємо самі довгі $\{\mu_{sj}^{*r=r+1}\}$ шляхи. За умови, якщо виявиться декілька шляхів максимальної довжини, то серед них вибирається шлях із меншим значенням довжини по вагах обмежень.

Крок 5. Перевіряється, чи усі множини шляхів такого $(r+1)$ -го рангу порожні. За умови, якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r=(n-1)$. У випадку виконання рівняння переходимо до кроку 5, інакше збільшуємо r на 1 та виконуємо крок 2.

Крок 6. Виділяємо у $\{\mu_{sj}^{*r}, \mu_{sp}^{*r=2}, \dots, \mu_{sw}^{*r=n}\}$ шлях максимальної довжини, кінець роботи.

Алгоритм MIN

Крок 1. Виконання сортування значень функціоналу й обмежень вигляду (15-17).

Крок 2. З вершини s будуються множини шляхів $\{\mu_{sj}^{r=1}\}$, $j = (\overline{1, n})$, першого рангу r , що задовольняють властивості γ та визначаються у множинах $m_{sj}^{r=1}$ шляхи мінімальної довжини $\{\mu_{sj}^{*r}\}$ по вагах обмеження $\{a_j\}$. Для кожної вершини j визначається вага γ_j відповідно до (14). Для вершини $j=n$ вагу γ_j приймаємо рівною 0.

Крок 3. Виключаються $\{\mu_{sp}^r\}$, $p = (\overline{r, n})$ в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівності (13).

Крок 4. Формуються множини шляхів $m_{sj}^{r=r+1}$, $j = (\overline{r+1, n})$ такого рангу, що задовольняють властивості γ , на базі множин шляхів попереднього рангу m_{sp}^r на основі рекурентного співвідношення (11). У множинах $m_{sj}^{r=r+1}$, що утворилися, виділяємо шляхи мінімальної $\{\mu_{sj}^{*r=r+1}\}$ довжини. За умови, якщо виявиться декілька шляхів мінімальної довжини по вагах обмежень, то серед них вибирається шлях із більшим значенням довжини по вагах функціоналу.

Крок 5. Перевіряється, чи усі множини шляхів такого $(r+1)$ -го рангу порожні. За умови, якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r=(n-1)$. У випадку виконання рівняння переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

Крок 6. Виділяємо у $\{\mu_{sj}^{*r}, \mu_{sp}^{*r=2}, \dots, \mu_{sw}^{*r=n}\}$ шлях мінімальної довжини по вагах обмежень і алгоритм закінчує роботу.

Крім цього становить інтерес модифікація цих алгоритмів, яку позначимо алгоритмом MAX-MIN, у якому на кроці 3 виділяється щоразу два шляхи відповідно до співвідношення (12).

4. Розробка точних алгоритмів оптимально розміщення даних у білінгових OLTP-системах

Побудуємо точний алгоритм рішення одновимірної задачі ЦЛП з БЗ.

КРОК 1. З вершини s будується множина шляхів $m_{sj}^{r=1}$, $j = (\overline{1, n})$ першого рангу r , що задовольняє властивості та визначаються у множині $m_{sj}^{r=1}$ шляху максимальної довжини μ_{sj}^{*r} за вагою функціонала c_j . Для кожної вершини j формуються калібрувальні вектори \bar{y}_j (18) та \bar{z}_j (19):

$$y_{jk} = a_{jk} + y_{j(k-1)}; \quad k = (\overline{1, n-j}); \quad y_{j0} = 0; \quad y_{n0} = b_1; \quad j = (\overline{1, n-1}). \quad (18)$$

$$z_{jk}^e = c_{j+k} + z_{j(k-1)}^e; \quad k = (\overline{1, n-j}); \quad z_{j0}^e = 0; \quad z_{n0}^e = 0; \quad j = (\overline{1, n-1}). \quad (19)$$

КРОК 2. Виключаються шляхи у множині m_{sj}^r поточного рангу r , довжини якої $d_c(\mu_{sp}^r)$ задовольняють нерівності (14).

КРОК 3. Для кожного шляху $\{\mu_{sj}^r\}$, $j = (\overline{r, n})$ поточного рангу r визначається за правилом K_1 значення \hat{r}_e [11]. Виключаються шляхи, довжини яких задовольняють нерівності:

$$d_c(\mu_{sp}^r) + \hat{z}_{p\hat{r}_e}^e(\mu_{sp}^r) < \max_{c_j} \left\{ d_c(\mu_{sp}^{*r}) \right\}. \quad (20)$$

КРОК 4. Формується $m_{sp}^{r=r+1}$, $p = (\overline{1, n})$ наступного рангу, що задовольняє властивості, на базі множини шляхів m_{sj}^r попереднього рангу на основі принципу оптимізації за напрямком з виділенням коридору по стратегії L_6 та виключенням векторів усередині коридору відповідно до стратегії L_8 . Шлях у множині $m_{sp}^{r=r+1}$ може бути сформований, якщо він задовольняє властивості. За умови, якщо властивість v не виконується, то шлях виключається з подальшого аналізу. У освічених множинах $m_{sp}^{r=r+1}$ виділяємо щонайдовші шляхи $\{\mu_{sp}^{*r=r+1}\}$.

КРОК 5. Перевіряється, чи уся множина шляхів $(r+1)$ -го рангу порожня. За умови, якщо це так, то переходимо до кроку 6, якщо ні, перевіряємо $r = (n - 1)$. У разі виконання рівності переходимо до кроку 6, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 6. Виділяємо у множині шлях максимальної довжини і алгоритм закінчує роботу.

Висновки

Таким чином, запропоновано стратегії відсікання $\{L_w\}$ безперспективних шляхів у множинах, що призводять до наближених і точних рішень задачі ЦЛП з БЗ. Побудовано ефективні точні та наближені алгоритми рішення задач ЦЛП з БЗ. Як показали результати експериментального дослідження, важливою перевагою розроблених алгоритмів на основі РП є той факт, що збільшення числа обмежень прак-

тично не впливає на погіршеність рішень алгоритмів, тоді як для методів рішення задач дискретної оптимізації, що засновані на ідеях методу гілок та кордонів, зростання числа обмежень до декількох сотень приводить фактично до неможливості їхнього практичного застосування [12].

СПИСОК ЛІТЕРАТУРИ

1. Янюшкин, В.В. Математические модели оптимизации распределенных информационных систем тренажерно-моделирующих комплексов : автореф. дис. ... канд. техн. наук : 05.13.18 / Янюшкин Вадим Вадимович. – Новочеркасск, 2010. – 19 с.
2. Многоатрибутивное формирование оптимальных по составу высоконадежных сложных систем / И.В. Ковалев [и др.]. – Красноярск: Краснояр. гос. аграр.ун-т., 2009. – 166 с.
3. Жуков, В.С. Исследование методов оптимального размещения базы данных по узлам вычислительной сети / В.С. Жуков // В мире научных открытий. – 2010. – № 4 (10). – С. 75-76.
4. Третяк В.Ф., Пашнева А.А. Оптимізація структури сховища даних у вузлах інфокомунікаційної мережі хмарного середовища // Системи управління, навігації та зв'язку. – 2017. – №. 4 (44). – С. 122-128.
5. Голубничий Д.Ю. Інформаційна технологія відсікання неперспективних варіантів в алгоритмах рішення задачі цілочисельного лінійного програмування з булевими змінними на основі рангового підходу // Theoretical foundations in research in Engineering: collective monograph / Д.Ю. Голубничий, О.В. Коломійцев, В.Ф. Третяк, А.О. Рибальченко [та ін.]; International Science Group. – Boston, 2022. – С. 96-133.
6. Голубничий Д.Ю. Архітектура системи обміну медичними даними пацієнтів з лікарями на основі ІОТА / Д.Ю. Голубничий, О.В. Коломійцев, В.Ф. Третяк, Я.О. Ключка, А.О. Рибальченко // Системи управління, навігації та зв'язку. – Полтава: Полтавський національний технічний університет ім. Кондратюка, 2022. – Вип. 1(67). – С. 56-61.
7. Коломійцев О.В. Метод рішення задачі оптимізації маршрутів для спеціалізованих машин логістичного забезпечення в автоматизованій інформаційній системі складського обліку на основі рангового підходу / О.В. Коломійцев, В.В. Старцев, В.Ф. Третяк, А.І. Нікорчук, О.І. Шаповалов, З.З. Закіров, Е.М. Полтавський, П.В. Черненко, О.А. Крамар, А.О. Рибальченко // InterConf. – Прага: Author-publishers miscellaneous, 2022. – Вип. 27(133), – С. 417-434.
8. Третяк В.Ф. Математична модель процесу виконання MDX-запитів на основі рангового підходу до рішення задачі цілочисельного лінійного програмування з булевими змінними / В.Ф. Третяк, Д.М. Запара, С.В. Новіченко, О.В. Коломійцев, А.М. Савельєв, В.І. Кривчун, М.М. Охромович, Н.М. Шамрай, А.О. Рибальченко, О.А. Крамар // Modern Directions and Movements in Science: I міжн. НПК., 06-08 жовтня 2022 р. – Люксембург, 2022. – С. 281-292.
9. Коломійцев О.В. Задачі дискретної оптимізації та їх постановка / О.В. Коломійцев, С.В. Осієвський, В.Ф. Третяк, З.З. Закіров, А.О. Романюк С.М. Логвиненко, А.О. Лисиця // Scientific trends and trends in the context of globalization: II МНПК., 19-20 вересня 2021 р. – Рим, 2021. – С. 285-302. – DOI: <https://doi.org/10.51582/interconf.19-20.09.2021.033>.
10. Третяк В.Ф. Аналіз сучасних систем управління базами даних / В.Ф. Третяк, В., О.В. Коломійцев, Д.І. Євстрат, С.В. Ворошилов, В.М., В., Логвиненко, А.О. Лисиця, В.О. Місюра // Scientific goals and purposes in XXI century: II міжн. НПК, 07-08 жовтня 2021 р. – Сіетл, 2021. – С. 453-465. – DOI: <https://doi.org/10.51582/interconf.7-8.10.2021.050>.
11. Голубничий Д.Ю. Інформаційна технологія відсікання неперспективних варіантів в алгоритмах рішення задачі цілочисельного лінійного програмування з булевими змінними на основі рангового підходу // Theoretical foundations in research in Engineering: collective monograph / Д.Ю. Голубничий, О.В. Коломійцев, В.Ф. Третяк, А.О. Рибальченко [та ін.]; International Science Group. – Boston, 2022. – С. 96-133.
12. Technical and agricultural sciences in modern realities: problems, prospects and solutions: collective monograph / Hladyshev D., Brodskyi M., Lisnykh L., Rybalchenko A. – etc. – International Science Group. – Boston : Primedia eLaunch, 2023. 461 p. Available at : DOI – 10.46299/ISG.2023.MONO.TECH.2.

Received (Надійшла) 23.03.2023

Accepted for publication (Прийнята до друку) 31.05.2023

Algorithms for solving the problem of optimal placement of data in billing OLTP systems based on the implementation of the ranked approach

Alina Rybalchenko

Abstract. The article presents the results of the development of approximate and exact algorithms for solving the problem of optimal data placement in billing OLTP systems based on the implementation of the rank approach. This type of problem belongs to the class of integer linear programming (ILP) with Boolean variables. Combinatorial methods and evolutionary algorithms currently occupy a dominant place in the methods of solving such problems. The practical application of these methods is complicated when solving large-scale problems. To eliminate this problem, it is suggested to use the idea of the rank approach. The analysis of approaches to the optimal placement of data in billing OLTP systems, the model of the rank approach, as well as approximate and exact algorithms are given. **The object** of research is the algorithms of functioning and placement of information resources in the cloud environment, the concept of cloud computing and multi-level information systems. **The subject** of the study is the principles of data placement stored in distributed databases and circulating in the cloud network, as well as the specifics of subscriber service processes in modern implementations of OLTP systems. **The purpose** of the research work is to develop approximate and accurate algorithms for optimizing the placement of RDB fragments on network nodes of the cloud structure, which will allow to increase the productivity of the information system due to the rational distribution of data. **Conclusions.** Strategies for cutting off unpromising paths in sets are proposed, leading to approximate and exact solutions of the problem of integer linear programming with Boolean variables, and efficient exact and approximate algorithms are constructed. It is shown that an important advantage of the developed algorithms based on the rank approach is the fact that an increase in the number of constraints practically does not affect the error of algorithm solutions, while for methods of solving discrete optimization problems based on the ideas of the branch-and-bound method, an increase in the number of constraints to several hundreds actually leads to the impossibility of their practical application.

Keywords: rank approach, integer linear programming, boolean variables, optimization by direction, approximate and exact algorithms.

Д. М. Росінський, К. О. Свірщевський

Харківський національний технічний університет радіоелектроніки, Харків, Україна

ОПТИМІЗАЦІЯ ЗБОРУ ОПЕРАТИВНИХ ДАНИХ В ПІРИНГОВІЙ МЕРЕЖІ МЕДИЧНОЇ УСТАНОВИ

Анотація. **Актуальність.** В сучасних медичних інформаційних системах обсяг даних, (наприклад, фізіологічних показників, зібраних від пацієнтів), є величезним, оскільки аномальні та нормальні дані збираються разом. Це призводить до затримки надання допомоги невідкладним пацієнтам. Щоб вирішити цю проблему, потрібні оптимальні схеми збору оперативних даних для коротко- і довгострокового прогнозування. При цьому доцільно використовувати пріоритизацію і фільтрацію даних, які безперервно збираються з переносних датчиків на пацієнтах і надходять до інформаційної системи медичної установи. **Метою даної роботи** є розробка моделі інформаційної системи медичної установи на основі пірингової мережі з використанням ефективних методів збору оперативних даних. **Об'єктом дослідження** є автоматизований процес збору і обробки масивів даних, що відображають стан здоров'я пацієнта медичної установи. **Предметом дослідження** є методи та алгоритми збору і обробки оперативних даних, що характеризують стан пацієнта медичної установи, для формування короткострокового прогнозу. **Результати.** Запропоновані нові алгоритми для оптимізації процесів збору і обробки даних шляхом введення критерія терміновості пацієнтів, що сприятиме зменшенню обсягу даних, які необхідно передати, скороченню часу очікування в постановці діагнозу. Для оцінки терміновості пацієнтів прийняті загальні клінічні критерії. **Висновок.** Запропоновано модель інформаційної системи медичної установи, яка ефективна при зборі даних і оптимізує порядок їхньої обробки в прогнозуванні. Досягнуто зменшення обсягів медичних даних, які надходять від пацієнтів, і налаштування часу очікування для даних, що потрібні в прогнозуванні, на основі різних пріоритетів.

Ключові слова: медична інформаційна система, вектор ознак, класифікація часових рядів, пірингова мережа, пріоритизація даних.

Вступ

Удосконалення інформаційних систем в сфері охорони здоров'я пов'язане з підвищенням ефективності збору даних і оптимізацією потоків їхньої обробки [1, 2]. Основною задачею стає впровадження пріоритизації даних на основі терміновості стану пацієнтів, що надасть можливість регулювати час очікування даних перед прогнозуванням. Отже завданнями є оцінка пріоритету на основі послідовностей даних у часі. Разом з тим, виникає інша задача, яка полягає в пришвидшенні передачі оперативних даних, що характеризують поточний стан пацієнтів. Розв'язання цієї задачі можливе за рахунок зменшення обсягу даних, що збираються. Проте, намагаючись зібрати меншу кількість даних, для точного діагнозу необхідно надати клініцистам достатню інформацію. В різних дослідженнях [3-6] для прогнозування аномалій враховуються численні фізіологічні показники (наприклад, артеріальний тиск, частота серцевих скорочень, частота дихання, рівень сатурації, тощо). Але більшість з цих досліджень і досі знаходяться на теоретичному рівні і непридатні для широкого використання. Для визначення того, які дані потрібно передати, окремі автори [7] пропонують до використання приховану марковську модель (НММ), фокусуючись, в основному, на оцінці особистого стану на основі спеціальних правил. Для того, щоб вирішити проблему зберігання та обробки даних, пов'язану з їхньою величезною кількістю, пропонується [8] використовувати хмарну інфраструктуру. Цікавою є пропозиція [3, 7] використовувати мобільний пристрій для безперервного збирання та передачі необроблених фізіологічних показників на сервер. Ще один практичний приклад – апарат BioSign [9], який може мінімізувати імовірність виникнення критичної клінічної ситуації, але не здатен формувати прогноз.

Недоліком наведених вище систем є те, що всі вони обробляють медичні показники пацієнтів за принципом FIFO. Однак у реальних клінічних випадках, щоб визначити оптимальний порядок надання медичних послуг, терміновість пацієнта зазвичай розглядається на початку. Існує приклад [10] ефективної системи сортування, яка використовується у відділенні невідкладної допомоги під час прийняття рішення про скерування пацієнтів до відділення інтенсивної терапії. Отже, під час обробки медичних даних пацієнтів в інформаційній медичній системі потрібно, щоб черговість надходження даних на обробку відображала терміновість стану пацієнтів.

Іншим аспектом дослідження є реалізація концепції машинного навчання з використанням моделі взаємодії типу "Peer-to-Peer" (P2P) [11]. Відомо, що використання централізованих моделей пов'язане із значними витратами на зв'язок через обмін величезними обсягами особистих життєво важливих медичних показників з метою навчання. Використання децентралізованої моделі P2P створює основу для розподіленого обміну знаннями між медичними установами і дозволяє розробити швидку систему машинного навчання, за допомогою якої кожна медична установа може підвищити точність діагностики, вивчаючи знання, отримані від інших установ, і може оптимізувати потік обробки даних пацієнтів на основі різних значень терміновості.

Зважаючи на стан і результати досліджень різних авторів, можна зробити висновок про актуальність питання підвищення ефективності передачі даних і оптимізації потоку даних на основі пріоритетів.

Метою роботи є розробка моделі інформаційної системи медичної установи на основі пірингової мережі з використанням ефективних методів збору оперативних даних.

Основна частина

Модель інформаційної системи медичної установи на основі пірингової мережі продемонстровано на рис. 1. Можна виділити дві частини системи: клієнт для збору даних і сервер для аналізу даних.

Клієнт використовує мобільні пристрої (смартфон, смарт-годинник, тощо) для постійного збирання даних, що відображають фізіологічні показники пацієнтів. Ці дані надходять від різноманітних датчиків на тілі пацієнта. Клієнт також здійснює оцінку пріоритету даних на основі певного часового вікна (наприклад, 5 хвилин). Пріоритет даних відповідає терміновості стану пацієнта: більш «важкий» пацієнт має вищий пріоритет даних. Крім того, ці медичні дані в межах часового вікна з їхніми пріоритетами становлять фрейми показників (ФП). Задачею клієнта є відбір фреймів, які потрібно передати на сервер, виходячи з їхніх пріоритетів.

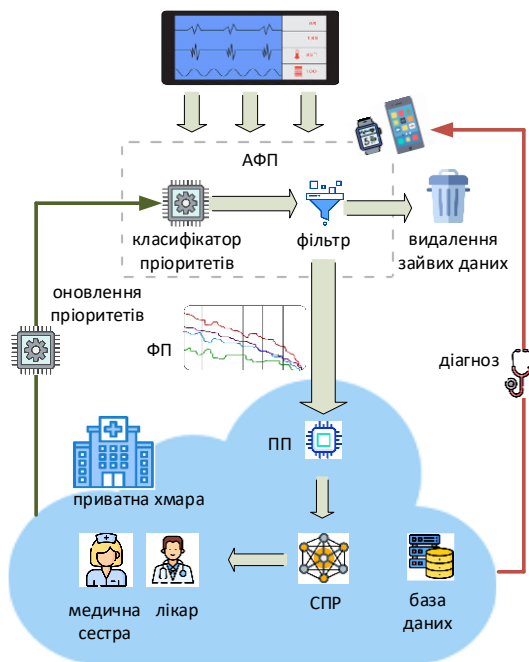


Рис. 1. Модель інформаційної системи медичної установи

Сервер призначений для формування точного короткочасного прогнозу стану здоров'я пацієнтів на основі їхніх фізіологічних показників, отриманих від клієнта. Порядок обробки даних пацієнтів оптимізується за допомогою пріоритетів, що гарантує отримання рішення щодо діагнозу у відповідний час залежно від терміновості пацієнтів. Крім того, завдяки накопиченню даних про пацієнтів, сервер може надати більш точне прогнозування різних медичних станів за допомогою навчання на основі використання пірингової мережі. Основним компонентом клієнта є аналізатор фреймів показників (АФП), який містить два основні компоненти: класифікатор пріоритетів для оцінки даних і фільтр для відбору даних.

Класифікатор пріоритетів використовує метод машинного навчання (МН) для автоматичної класифікації ФП за різними пріоритетами. Пріоритети відіграють важливу роль в оптимізації: класифікатор МН

навчається за допомогою зразків з 4 фізіологічних показників, а мітки навчальних даних ідентифікуються алгоритмом оцінки даних в реальних умовах.

Фільтр допомагає системі ефективніше збирати медичні дані пацієнтів з урахуванням різних пріоритетів. Слід зазначити, що аномальні дані є більш значущими, ніж нормальні, оскільки відповідні діагностичні рішення повинні формуватися під час виявлення аномалій. Таким чином, фільтр націлений на аномальні дані, які мають вищий пріоритет. При цьому потрібна гарантія повноти медичної інформації, яка передається на сервер для діагностики клініцистами.

Сервер містить три компоненти: базу даних для зберігання даних, процесор пріоритетів (ПП) для керування чергою прогнозування та систему прийняття рішень (СПР) для прогнозування. База даних використовується для зберігання записів ФП від різних пацієнтів. Ці записи можна використовувати для встановлення довгострокових діагнозів пацієнтам і видобутку певних корисних медичних знань.

ПП встановлює порядок прогнозування на основі пріоритету даних, що сприяє скороченню часу очікування термінових даних пацієнтів. Довгострокові пріоритети даних визначаються на основі пріоритетів ФП. Наприклад, якщо потрібно обчислити пріоритет даних за годину, потрібні пріоритети 12 часових рядів ФП із використанням 5-хвилинного вікна.

СПР є машинним класифікатором, який може точно прогнозувати стан здоров'я пацієнтів, використовуючи медичні дані, зібрані від пацієнтів у режимі реального часу. З метою забезпечення відповідності змінам у клінічному середовищі (наприклад, появи нових захворювань), для СПР передбачено Р2Р-навчання.

Робота клієнта реалізована у три етапи.

1. Мобільний пристрій постійно збирає дані з датчиків на тілі пацієнта. Вони відображають значення таких фізіологічних показників, як частота дихальних рухів (ЧДР), рівень сатурації киснем (SpO_2), частота серцевих скорочень (ЧСС) та систолічний артеріальний тиск (САТ).

2. АФП у пристрої обробляє дані в межах часових вікон (наприклад, 10 хвилин). Дані у вікні визначаються як ФП. АФП виконує класифікацію пріоритету ФП і фільтрацію ФП.

3. Після фільтрації АФП передає фрейми ФП з індикаторами пріоритетів на сервер медичного закладу.

На сервері обробка даних виконується в 3 етапи.

1. Після отримання достатньої кількості ФП протягом певного вікна спостереження (наприклад, 60 хвилин) сервер групує їх разом і обчислює загальний пріоритет, враховуючи всі ФП в інтервалі спостереження. Далі сервер розміщує ці згруповані дані в списку очікування на основі пріоритету. Вищий пріоритет має вищий індекс, що означає менше часу на обробку.

2. СПР прогнозує медичні стани за допомогою згрупованих ФП і надсилає дані клініцистам після виявлення аномальної клінічної події. Потім пацієнту надаються правильні діагностичні рішення.

3. ФП зберігаються в базі даних для резервного копіювання та подальшого видобутку знань. Дані також використовуються для оновлення класифікатора пріоритетів.

Як показано в табл. 1, бали фізіологічним показником призначаються на основі їхніх порогових значень. Наприклад, якщо значення ЧДР становить від 9 до 11, то йому надається оцінка 1. Важливо зауважити, що всі значення є цілими. Оцінка 0 означає, що значення знаходиться в межах норми.

Таблиця 1 – Оцінка фізіологічних показників

Оцінка	Значення показника			
	ЧДР, 1/хв.	SpO ₂ , %	ЧСС, 1/хв.	САТ, мм рт./ст.
-3	< 8	< 91	< 40	< 90
-2	-	91-92	-	91-100
-1	9- 11	94-95	41-50	101-110
0	12-20	> 96	51-90	110-120
+1	-	-	91-110	121-170
+2	21-24	-	111-130	171-219
+3	> 25	-	> 131	> 220

Загальний бал усіх показників класифікується за 4 пріоритетами (табл. 2). Критерії класифікації базуються на часі очікування. Зокрема, якщо загальна оцінка дорівнює 0, система розглядає можливість відхилення даних, оскільки всі життєво важливі показники знаходяться в межах норми. Діапазон кожного пріоритету використовується для оцінки послідовної групи пріоритетів, що пояснюється далі.

Таблиця 2 – Визначення пріоритету даних за оцінкою фізіологічних показників

Пріоритет	Мітка	Час очікування	Діапазон	Загальний рахунок
1	норма	-	0	0
2	увага	<1 год.	(0,1]	1-3
3	невідкладна	< 0,5 год.	(1,2]	4-6
4	екстрена	0	(2, 3]	≥ 7 або оцінка +3/-3

Основна роль класифікатора пріоритетів у АФП полягає в позначенні ФП відповідним пріоритетом даних. Припустимо, що дискретна послідовність даних (D) тривалості часу (T) розбивається на L вікон однакового розміру. У кожному вікні є $N = D/L$ зразків. Тоді дані часового ряду можна розглядати як послідовність ФП $D_1(t), D_2(t), \dots, D_L(t)$. Наприклад, якщо в системі використовується 10-хвилинне вікно, 60-хвилинну серію даних можна розділити на послідовність із 6 ФП. Щоб визначити пріоритет ФП, розраховують середнє значення кожного показника із зразків у вікні:

$$f = \sum_{l=1}^N D_l(t) / N. \tag{1}$$

Середні значення усіх фізіологічних показників групуються у вектор ознак ($f_1, f_2, f_3, \dots, f_c$), де c – загальна кількість показників. Оскільки передбачено обробку лише цілих чисел, усі дробові значення компонент вектора ознак округлюються до найближчого цілого значення. Після виділення ознак ФП класифікатор пріоритетів може визначити його пріоритет, використовуючи будь-які види алгоритмів машинного навчання.

У запропонованій системі фільтр у АФП виділяє значущі дані на основі пріоритету, що сприяє зменшенню обсягу даних. Пріоритет 0 означає, що всі показники належать до нормального діапазону. Якщо в буфері присутній аномальний ФП, усі наступні ФП мають бути взяті до уваги (рис. 2). Коли новий ФП надходить у АФП кожного періоду часу, вікно буфера передачі зсувається ліворуч на крок, що відповідає часовому інтервалу t . Можливі три ідентифікатори, якими позначається ФП: надсилання, очікування та відхилення. Якщо пріоритет ФП у буфері перевищує 0, всі ці 6 ФП надсилаються на сервер. Якщо пріоритети нового ФП та решти 5 наступних дорівнюють 0, він позначається як очікуваний. Коли буфер зміщується від очікуючого ФП, той відкидається. Таким чином, алгоритм коригує процес збору даних, фокусуючись на аномальних ФП.

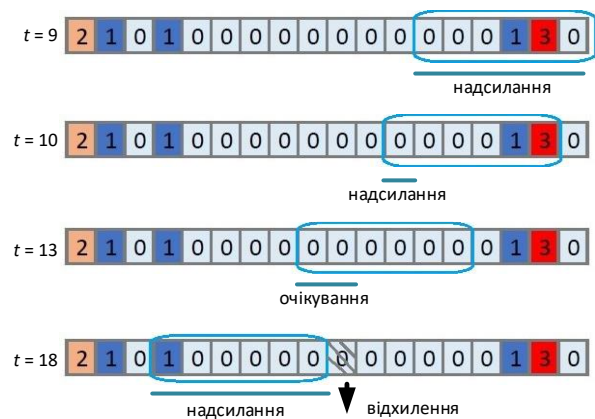


Рис. 2. Приклад передачі даних

Для визначення загального пріоритету ФП використовуються всього L пріоритетів фрагментів послідовностей даних. Замість усереднення всіх пріоритетів, використовуються вагові коефіцієнти часу:

$$w_i = T_i^2 / \sum_{i=1}^L T_i^2; T_i = i \times W. \tag{2}$$

Позначимо множину всіх пріоритетів як P . Тоді $P = (p_1, p_2, \dots, p_L)$. Загальний пріоритет p' безперервних ФП обчислюється наступним чином:

$$p' = P \times W = \sum_{i=1}^L p_i \times w_i. \tag{3}$$

Система поміщає всі видобуті функції послідовних ФП у список очікування на основі їхнього пріоритету. Для обробки вибираються дані з найвищим пріоритетом. У кожному ФП виділяється 5 ознак, якими є середнє значення, стандартне відхилення, медіана, кількість тенденцій зростання та тенденцій до зменшення.

Критеріями ефективності запропонованих рішень можуть бути: точність класифікації пріоритетів; продуктивність збору даних; середній час очікування.

Таблиця 3 – Матриця помилок класифікації

Пріоритет	1	2	3	4
1	445	136	10	1
2	4	248	18	8
3	1	66	380	63
4	0	0	42	378

Використано реальні медичні дані з МІМІС-II [11]. Щоб збалансувати набір даних, випадковим чином вибираються 1500 зразків кожного пріоритету. Набір даних нормалізується за допомогою лінійного перетворення z-показника. Була отримана оцінка точності класифікації пріоритетів. Для навчання використано 70% вибірок, решта – для тестування (табл. 3). Крім загальної матриці помилок, можуть застосовуватися інші показники точності для кожного пріоритету.

Висновки

Досліджено автоматизований процес збору і обробки масивів даних, що відображає стан здоров'я пацієнта медичної установи. Запропонована модель ефективна при зборі даних, оптимізує порядок обробки даних в прогнозуванні. Вона сприяє зменшенню обсягів медичних даних, зібраних від пацієнтів, фільтруючи звичайні дані, і може налаштувати час очікування даних перед прогнозуванням на основі різних пріоритетів. Використання децентралізованої моделі P2P надає можли-

вість розподіленого обміну знаннями між різними медичними установами і дозволяє розробити швидку систему машинного навчання, за допомогою якої кожна медична установа може підвищити точність діагностики, вивчаючи знання, отримані від інших установ, і може оптимізувати потік обробки даних пацієнтів.

Отже, отримано модель системи з оцінкою пріоритету даних, що оптимізує збір даних і прогнозування щодо розміру даних і часу очікування. Розроблено новий алгоритм оцінки пріоритету даних, який може представляти терміновість пацієнтів. Завдяки пріоритетизації даних, у запропонованій системі не тільки збирається менше даних від пацієнтів, але також надається повна медична інформація, необхідна для довгострокового встановлення точного діагнозу.

Розглядаючи різні розподіли пріоритету даних у реальному випадку, експериментальні результати свідчать про те, що запропоновані рішення підвищують ефективність збору даних і оптимізують час очікування відповідно до терміновості пацієнтів.

СПИСОК ЛІТЕРАТУРИ

1. Frick N., Mirbabaie M., Stieglitz S., and Salomon J. (2021). "Maneuvering through the stormy seas of digital transformation: the impact of empowering leadership on the AI readiness of enterprises." *Journal Of Decision Systems*, 1-24.
2. Bohr A., and Memarzadeh K. (2020). "Artificial intelligence in healthcare." San Diego: Elsevier Science & Technology.
3. Forkan A. R. M., and Khalil I. (2017). "Peace-home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring." *Pervasive and Mobile Computing*, 38:296-311.
4. Li H., and Boulanger P. (2020). "A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG)." *Sensors*, 20(5), 1461. doi: 10.3390/s20051461
5. Cao H., Eshelman L., Chbat N., Nielsen L., Gross B., and Saeed M. (2008). "Predicting icu hemodynamic instability using continuous multiparameter trends." In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3803-3806.
6. Salem O., Liu Y., Mehaoua A., and Boutaba R. (2014). "Online anomaly detection in wireless body area networks for reliable healthcare monitoring." *IEEE J. Biomedical and Health Informatics*, 18(5):1541-1551.
7. Jiang P., Winkley J., Zhao C., Munnoch R., Min G., and Yang L. T. (2016). "An intelligent information forwarder for healthcare big data systems with distributed wearable sensors." *IEEE Systems Journal*, 10(3):1147-1159.
8. Holm S., Stanton C., and Bartlett B. (2021). "A New Argument for No-Fault Compensation in Health Care: The Introduction of Artificial Intelligence Systems." *Health Care Analysis*. doi: 10.1007/s10728-021-00430-4
9. Tarassenko L., Hann A., Patterson A., Braithwaite E., Davidson K., Barber V., and Young D. (2005). "Biosign: multi-parameter monitoring for early warning of patient deterioration." pages 71-76.
10. Xie R., Khalil I., Badsha S., and Atiquzzaman M. (2018). "Fast and peer-to-peer vital signal learning system for cloud-based healthcare." *Future Generation Computer Systems*, 88:220-233.
11. Saeed M., Villarroel M., Reisner A. T., Clifford G., Lehman L.-W., Moody G., Heldt T., Kyaw T. H., Moody B., and Mark R. G. (2019). "Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database." *Critical care medicine*, 39(5):952.

Received (Надійшла) 23.03.2023

Accepted for publication (Прийнята до друку) 17.05.2023

Optimizing operational data collection in a medical institution's peer-to-peer network

Dmytro Rosinskiy, Kyryl Svirshchevskyi

Abstract. Topicality. In modern medical information systems, the volume of data (for example, physiological parameters collected from patients) is enormous, as abnormal and normal data are collected together. This leads to delays in providing care to emergency patients. To solve this problem, optimal operational data collection schemes for short- and long-term forecasting are needed. At the same time, it is advisable to use prioritization and filtering of data that are continuously collected from portable sensors on patients and sent to the information system of the medical institution. **The goal of this work** is to develop a model of the information system of a medical institution based on a peering network using effective methods of collecting operational data. **The object of research** is the automated process of collecting and processing data arrays, which will reflect the state of health of a patient of a medical institution. **The subject of research** is methods and algorithms for collecting and processing operational data characterizing the condition of a patient in a medical institution, for the formation of a short-term prognosis. **Results.** New algorithms are proposed to optimize the processes of data collection and processing by introducing a criterion of urgency for patients, which will contribute to reducing the amount of data that must be transferred, reducing the waiting time for making a diagnosis. Common clinical criteria are used to assess the urgency of patients. **Conclusions.** A model of the information system of a medical institution is proposed, which is effective in collecting data and optimizes the order of their processing in forecasting. Achieved a reduction in the amount of medical data received from patients and the setting of the waiting time for the data required for forecasting.

Keywords: medical information system, feature vector, time series classification, peering network, data prioritization.

Shafag Samadova

Azerbaijan State Oil and Industry University, Baku, Azerbaijan

ALGORITHM FOR AUTOMATIC RECOGNITION OF CARDIAC ARRHYTHMIAS

Abstract. Problem statement: Cardiovascular diseases currently remain one of the leading causes of death. It is important to monitor the state of the cardiovascular system in early stages of pathology in order to diagnose these diseases in a timely manner. A special place is occupied by various arrhythmias among diseases of the cardiovascular system. The most common ones of various arrhythmias are extrasystoles. Ventricular extrasystoles are considered the most life-threatening among extrasystoles. In order to diagnose ventricular extrasystoles at an early stage of their development, it is necessary to process and analyze large amounts of electrocardiographic data. In this regard, the development and software implementation of algorithms for automatic recognition of ventricular arrhythmias based on electrocardiographic data through modern computer technologies is an urgent task. Work objective is developing an algorithm for automatic recognition of ventricular arrhythmias and its software implementation. Results: An algorithm for automatic recognition of ventricular extrasystoles, which is characterized by simplicity of implementation and minimal requirements for computing resources, has been developed. At the same time, high values of sensitivity and specificity are maintained for ECG signals with single ventricular extrasystoles. The algorithm is implemented in the LabVIEW software environment and tested using ECG files taken from international databases on cardiac arrhythmias, as well as using stimulated ECG signal models. Practical significance: The developed algorithm can be used in automatic processing and analysis of long-term ECG recordings and recognition of ventricular arrhythmias.

Keywords: Electrocardiographic signal, arrhythmia, ventricular extrasystole, phase space, phase portrait, RR intervals, LabVIEW software, heart rate.

Introduction

Currently, cardiovascular disease ranks first among all diseases in terms of danger to human life. The most common symptoms, precursors of cardiovascular diseases are cardiac arrhythmias - heart rhythm disturbances. Characteristic features of arrhythmias are determined based on the results of the analysis of heart rate variability (HRV). Heart rhythm disturbances that occur abnormally related to the main heart rhythm and manifest as excitation of the heart as a whole or its individual parts are called extrasystoles.

The cause of extrasystole is explained by the presence of active heterotopic foci that generate electrical impulses of sufficient power that disrupt the main sinus rhythm. Depending on the localization in the heart, there are supraventricular extrasystoles: sinus, atrial, atrioventricular and ventricular extrasystoles. The distinction between ventricular extrasystoles and supraventricular extrasystoles should be considered very important, since some medications for the treatment of supraventricular extrasystoles can worsen the clinical condition if the rhythm is ventricular.

Ventricular extrasystole is the most common cardiac arrhythmia.

The prevalence of arrhythmias directly depends on age and the presence of heart pathology. Ventricular extrasystole is an alarm signal in patients with cardiomyopathies, valvular heart disease, severe myocardial ischemia, as well as with cases of sudden death in a family history [1].

Timely diagnosis of cardiac arrhythmias can help reduce the risk of adverse outcomes for patients with cardiovascular disease. One of the main directions in the development of modern electrocardiography is the expansion in the use of automated analysis of ECG signals. Automated ECG analysis enables cardiologists to more accurately determine the parameters of

electrocardiographic signals, objectively and quickly assess the state of the heart, and increases the probability of making a right decision about the patient and recommendations for his treatment in the future. Recently, there has been a significant increase in the number of algorithms assigned for automated ECG analysis, and at the same time, the expansion in the scope of their application. This is due to two main factors. First, unfortunately, the number of cardiovascular diseases is increasing, most of which are caused by arrhythmias. Secondly, a significant increase in the capabilities, power and speed of computer technologies, which makes it possible to implement even mathematically complex signal processing algorithms without fundamental difficulties, including in real time.

Goal setting and research objectives

Cardiac arrhythmias are usually diagnosed by electrocardiogram (more reliably by Holter ECG monitoring) based on P, QRS, and T wave characteristics of ECG. Reliable recognition of these complex ECG characteristics, which are specifically associated with cardiac arrhythmias, requires considerable competence and training of medical specialists. However, processing such a large amount of information and making a correct diagnosis in the limited time allotted to the patient is a difficult task even for experienced doctors. The problem of detection and recognition of arrhythmias can be effectively solved using computer algorithms for automatic diagnosis of arrhythmias. Based on these considerations, the objective of this work is to develop an algorithm for automatic recognition of ventricular arrhythmias and software implementation of this algorithm.

The question of accurately determining the boundaries of QRS complexes and the allocation of R peaks in ECG signals is of fundamental importance in the diag-

nosis of arrhythmias and in the recognition of signals in clinical manifestations in general.

Existing methods for detecting QRS complexes can be divided into two large groups: the first group includes high-precision methods designed for basic clinical examinations using several divisions; the second group includes methods used in mobile electrocardiographic devices allowing real-time analysis and are intended for patients with cardiac failure. Automating the classification of arrhythmias by ECG is very important for making a quick and objective decision about the class of arrhythmia.

The main requirements for an automated system are not the complexity of the algorithm, making a quick decision and less memory.

Research methods and materials

Generally, the algorithm used to automatically classify arrhythmias consists of three steps: preliminary processing, extraction and classification of feature. Preliminary processing of recorded ECG signals is performed to remove noise that degrades classifier performance, such as baseline deviation, motion artifact, power line interference, and high frequency noise.

Preliminary processing was carried out on the basis of wavelet transform. Wavelet analysis remains a popular technique for both filtering and extraction of feature.

Extractable features of ECG include:

a) temporal features of palpitation, such as P-Q interval, QRS interval, S-T interval, Q-R interval, R-S interval, and R-R interval between adjacent cardiac contractions,

b) amplitude characteristics, such as amplitudespeak of P, Q, R, S and T characteristics based on wavelet transform, at different levels of decomposition.

A comparative analysis of the main technological methods for constructing algorithms to detect QRS complexes in ECG signals was considered at different times in numerous studies, for example, in [2-5].

A number of approaches have been proposed for the detection of QRS complexes, such as:

wavelet transforms [5, 7],

algorithms from the field of artificial neural networks [8, 9],

genetic algorithms [10],

filter banks [11],

as well as heuristic methods based on non-linear transformations of ECG elements [12].

Experimental results

Today, there are many methods for detecting normal and pathological QRS complexes [5,13-18], which practically represent improvement of previously known methods.

These improvements are aimed at eliminating various interferences, applying various transformations for reliable detection and recognition of arrhythmias. However, the problem associated with unpredictability of the rhythm behavior even in one patient still remains open and raises the question of improving the algorithms for identifying QRS complexes and creating algorithms that

are weakly dependent on a particular patient. For that purpose, in contrast to the existing algorithms, which are mainly based on a comparative analysis of differences in RR intervals with their averaged values, we have proposed an algorithm based on the analysis of the ratios of these intervals for the recognition of arrhythmias.

When interpreting the results of Holter monitoring, the followings are realized under the general name "high-grade ventricular arrhythmias" [1]:

1) single extrasystoles;

2) polytopic extrasystoles;

3) extrasystoles of type <<R to T>>;

4) ≥ 2 consecutive ventricular extrasystoles.

The article proposes an algorithm for recognizing single ventricular extrasystoles.

As it is known, at the time of rhythm disturbances, the change in RR interval is $\geq 10\%$.

Based on this, it can be said that in the case of a normal rhythm, the ratio of two adjacent intervals ΔR_{i-1} and ΔR_i should satisfy the condition:

$$0,9 \leq \alpha_i \leq 1,0, \quad (1)$$

where

$$\alpha_i = \frac{\Delta R_i}{\Delta R_{i-1}}.$$

In the case of a normal rhythm, regardless of the absolute lengths of adjacent intervals, condition (1) should be preserved (weakly depends on a specific healthy rhythm). So, the essence of the algorithm proposed by us is as follows:

1) the investigated file of ECG data is loaded (or registered online);

2) amplitudes and localizations of R peaks are distinguished;

3) the sequence of lengths of RR intervals is determined

$$\{\Delta R_i = \Delta R_i - \Delta R_{i-1}\};$$

4) according to the obtained series of sequences of RR interval lengths, a new sequence of such a_i elements, which are defined as the ratio of adjacent RR intervals, is determined:

$$\left\{ \alpha_i = \frac{\Delta R_i}{\Delta R_{i-1}} \right\}.$$

5) according to the number of a_i values satisfying conditions $a_i \leq 0,9$ or $a_i \geq 1$, the number of extrasystoles or arrhythmias is calculated.

These conditions correspond to the appearance of pathological intervals, i.e., change in the frequency of contractions is not less than 10%.

If the number of a_k values corresponding to pathological intervals is exactly n_k , then the number of arrhythmias (extrasystoles) n_e , will be determined by formula

$$n_e = \frac{n_k}{3}.$$

As it is seen from Fig. 1, this ratio follows from ECG with a single ventricular extrasystole.

6) according to the number of rhythm disturbances, diagnostic parameters (sensitivity, specificity) and temporal localization of extrasystoles are determined. This algorithm was implemented in LabVIEW 2014 software environment.

Fig. 2 shows the front panel of the program. The program consists of three subroutines: reading a file, generating cardiointervalogram (CIG), signal analysis.

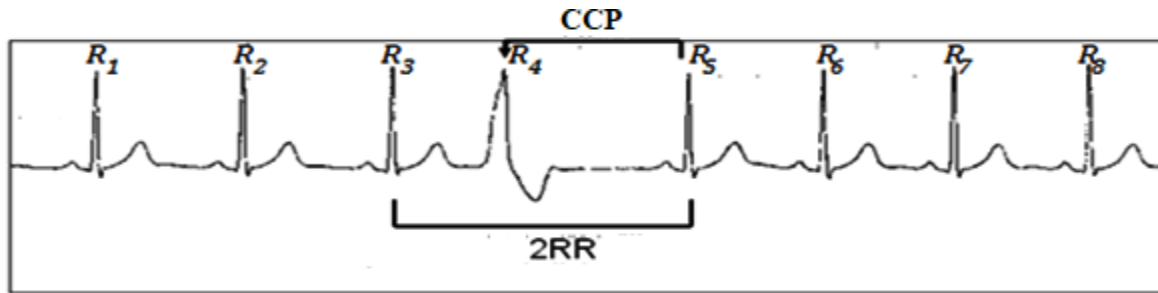


Fig. 1. Single ventricular extrasystole (indicated by an arrow). CCP is a complete compensatory pause. 2RR - two normal cardiac cycles. (Holter monitoring ECG recording) [19]

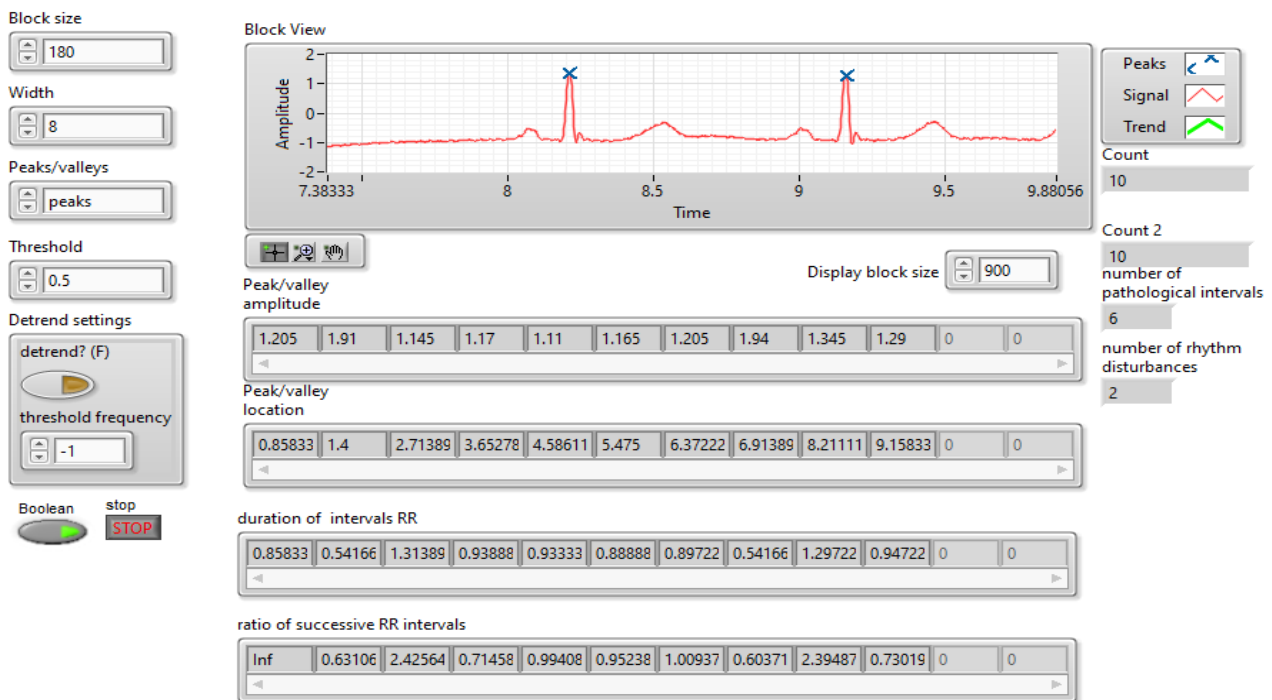


Fig. 2. Program for detecting arrhythmias in the environment of LabVIEW

The file is read using the function of *Read Biosignal Express VI*.

Configuration virtual instrument of *Read Biosignal Express VI* allows to access the folder where the file under study is available, cut the desired part from the file, select the desired lead, present the file as a sequence of reports (Waveform Values).

CIG formation subroutine was implemented using the function of *WA Online Multiscale Peak Detection VI*, which allows to identify the peaks of R and their localization in real time based on wavelet analysis.

Using the obtained CIG, an analysis of HRV and recognition of arrhythmias was carried out.

Discussion of the results

In order to test the performance of the proposed algorithm, real signals were used from the annotated

international databases of Physio net site [20] - from the database of Massachusetts Institute of Technology (MIT-BIH), in accordance with the recommendations of AAMI. EC57:2012 [21].

Electrocardiography signals from this database are considered today as a generally recognized standard for testing software tools designed for cardiology applications.

MIT-BIH Arrhythmia Database contains 48 half-hour extracts from dual-channel ambulatory ECG recordings.

Characteristics of the signals of this base are: the number of leads - 2 (II standard, V1), sampling frequency - 360 Hz; ADC resolution - 11 bits.

In addition, stimulated models of ECG signals were also used to test the program.

Table 1 presents the results of testing the program using fragments of various lengths of 119.hea file, taken from MIT-BIH database.

The results of testing the program, as it is seen from tables 1 and 2, gives satisfactory results.

However, further research showed that the program for signals with sinus rhythm and single ventricular extrasystoles gives high results, the reliability of

detecting pathological QRS complexes in which is 97-98%.

In order to recognize more complex extrasystoles, this algorithm should be improved, taking into account the structural features of extrasystoles.

Table 2 shows the results of testing the program for fragment:

00.00.00-00.00.10 sec of file 119.hea.

Table 1 – Results of testing the program

Length of signal fragment	Number of complexes	Real number of extrasystoles	The number of extrasystoles detected by the program	Detection of error
00.00.00- 00.00.10 sec	10	2	2	0
00.00.00- 00.00.20 sec	21	3	4	+1
00.00.00- 00.00.50 sec	55	14	12	-2
00.00.00- 00.00.100 sec	108	25	21	-4

Table 2 – The results of testing the program for fragment 00.00.00- 00.00.10 sec of file 119.hea

Localization of observed peaks of R, sec	Length of RR intervals $\{\Delta R_i = R_i - R_{i-1}\}$	Ratio of adjacent RR intervals: $\{a_i = \Delta R_i / \Delta R_{i-1}\}$	Belonging to pathological interval	Number of rhythm disturbances (extrasystoles)
0.858333	0.858333			2
1.4	0.541667	0.631068	yes	
2.71389	1.31389	2.42564	no	
3.65278	0.938889	0.714588	yes	
4.58611	0.933333	0.994083	no	
5.475	0.888889	0.952381	no	
6.37222	0.897222	1.00937	no	
6.91389	0.541667	0.603715	yes	
8.21111	1.29722	2.39487	yes	
9.15833	0.947222	0.730193	yes	

Conclusions

An algorithm was developed to perform ECG signal processing procedures and detect QRS complex in accordance with the requirements of AAMI EC57:2012 documents [21] for reliable identification of arrhythmias. The proposed algorithm for recognition of arrhythmias (ventricular extrasystoles) is characterized by simplicity of implementation and minimal requirements for

computing resources, while maintaining high values of sensitivity and specificity for ECG signals with sinus rhythm and single ventricular extrasystoles.

The developed program was tested using ECG signal of 119.hea from MIT BIH database [20] as well as models of stimulated ECG signals in LabVIEW environment.

The results obtained during the processing and analysis of ECG files taken from databases [20, 22] (di-

agnostic features of which were known in advance) with space” and “HRV analysis system”) implemented in both systems (“ECG signal analysis system in phase LabVIEW environment confirm their adequacy.

REFERENCES

1. Guide to cardiology/edited by V.N. Kovalenko. — K.: MORION, 2008. — 1424 p.
2. G. D.Clifford Signal processing methods for heart rate variability analysis / G. D. Clifford. —PhD Thesis michaelmas Term., 2002. — 244 p.
3. Rangayyan, P. M. Analysis of biomedical signals. Practical Approach/ P. M. Rangayyan. — M. : Fizmatlit, 2007. — 440 p.
4. Luz E.J.S. et. al. ECG-based heartbeat classification for arrhythmia detection: A survey //Computer methods and programs in biomedicine. — 2016. — T. 127. — P. 144-164.
5. Khalaydzhii A.K., Muchnik I.B. Methods for classifying cardiac arrhythmias based on coding sequences of RR-intervals of the ECG signal. Proceedings of NSTU im. R.E. Alekseeva 2021. No. 1 (132), pp. 38-53.
6. Kadambe S., Murray R. and Boudreaux-Bartels G. F. Wavelet transform-based QRS complex detector // IEEE Trans. Biomed. Eng. — 1999. — Vol. 46. — P. 838-848.
7. Li C., Zheng C. and Tai C. Detection of ECG characteristic points using wavelet transforms // IEEE Trans. Biomed. Eng. — 1995. — Vol. 42. — P. 21-28.
8. Xiang *et al.* Automatic QRS complex detection using two-level convolutional neural network // *BioMed Eng OnLine* <https://biomedical-engineering-online.biomedcentral.com/track/pdf/10.1186/s12938-018-0441-4.pdf>
9. Hu Y.H., Tompkins W.J., Urrusti J.L., and Afonso V.X. Applications of artificial neuralnetworks for ECG signal detection and classification // *J. Electrocardiology*. — 1993. Vol. 26 (Suppl.). p. 66-73.
10. Poli R., Cagnoni S. and Valli G. Genetic design of optimum linear and nonlinear QRS detectors // IEEE Trans. Biomed. Eng. — 1995. — Vol. 42. — P. 1137-1141.
11. Afonso V.X., Tompkins W.J., Nguyen T.Q., Luo S. ECG beat detection using filter banks // *IEEE Transactions on Biomedical Engineering*. — 1999. — Vol. 46. — No. 2. — P. 192-202.
12. Köhler B.-U., Hennig C. and Orglmeister R. QRS detection using zero crossing counts // *Applied genomics and proteomics*. — 2003. — T. 2. — №. 2. — P. 138-145.
13. O. N. Bodin, M. N. Kramm, L. Yu. Krivonogov, V. G. Polosin, N. S. Shilov A new technology for noise interference in electrocardiosignals//*Bulletin of Cybernetics*. 2017. № 4 (28), p.122-130.
14. A. A. Chupov, A. E. Jdanov, S. T. Knyazev, F. K. Rakhmatullof, R. F. Rakhmatullof, A. Y. Dolganov. Processing of ECG signals using wavelet analysis: diagnostic possibilities//*Ural Radio Engineering Journal*. 2021;5(4): pp 337–352.
15. A.V. Leonova, A.A. Ageichenko Modification of Pan-Tompkins on Real-Time QRS Complex Recognition Algorithm //*Engineering Bulletin of Don*, №2, p.2 (2015) <http://ivdon.ru/magazine/archive/n2p2y2015/3058>
16. I. O. Khokhlov, A. N. Kalinichenko. Algorithm for predicting the parameters of ECG cardiocomplexes for noise-interference detection of R-wave // *News of St. Petersburg Electrotechnical University “LETI”* № 1/2017, pp.72-77.
17. Petrov S.P., Epishina E.V., Voronin V.V. Evaluation of pattern recognition algorithms for tasks of automatic analysis of electrocardiograms // *Eurasian Union of Scientists (ESU) # VIII, 2014 | Technical sciences*, pp.27-29.
18. D. A. Lipchak, A. A. Chupov Review of methods for automatic diagnosis of cardiac arrhythmia for making decisions on the need for defibrillation, *Ural Radio Engineering Journal*. 2021;5(4):380–409.
19. <https://medspecial.ru> (date of the application 03.01.2023).
20. <http://ecg.mit.edu/george/publications/mitdb-embs-2001.pdf>
21. ANSI/AAMI EC57:2012 (ANSI/AAMI EC 57:2012). Testing And Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. URL: <https://webstore.ansi.org/Standards/AAMI/ansiaamiec572012ec57>
22. <https://physionet.org/content/edb/1.0.0/>

Received (Надійшла) 14.02.2023

Accepted for publication (Прийнята до друку) 26.04.2023

Алгоритм автоматичного розпізнавання порушень ритму серця

Самедова Шафаг

Анотація. Постановка проблеми: Серцево-судинні захворювання нині залишаються однією з провідних причин смертності. Важливо стежити за станом серцево-судинної системи на ранніх стадіях патології, щоб своєчасно діагностувати ці захворювання. Особливе місце серед захворювань серцево-судинної системи посідають різні аритмії. Найбільш частими з різних аритмій є екстрасистоли. Шлуночкові екстрасистоли вважаються найбільш небезпечними для життя екстрасистолами. Для діагностики шлуночкових екстрасистол на ранній стадії їх розвитку потрібна обробка та аналіз великих обсягів електрокардіографічних даних. У зв'язку з цим розробка та програмна реалізація алгоритмів автоматичного розпізнавання шлуночкових аритмій на основі даних електрокардіографії за допомогою сучасних комп'ютерних технологій є актуальним завданням. Мета роботи – розробка алгоритму автоматичного розпізнавання шлуночкових аритмій та його програмна реалізація. Результати: Розроблено алгоритм автоматичного розпізнавання шлуночкових екстрасистол, що відрізняється простотою реалізації та мінімальними вимогами до обчислювальних ресурсів. У той же час високі значення чутливості та специфічності зберігаються для сигналів ЕКГ з одиничними шлуночковими екстрасистолами. Алгоритм реалізований у програмній середовищі LabVIEW та протестований з використанням файлів ЕКГ, взятих із міжнародних баз даних за серцевими аритміями, а також із використанням моделей стимульованих сигналів ЕКГ. Практична значущість: Розроблений алгоритм може бути використаний для автоматичної обробки та аналізу тривалих записів ЕКГ та розпізнавання шлуночкових аритмій.

Ключові слова: електрокардіографічний сигнал, аритмія, шлуночкова екстрасистоля, фазовий простір, фазовий портрет, RR-інтервали, програма LabVIEW, частота серцевих скорочень.

С. В. Суліма, О. Д. Єрмолаєв

Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна

МЕТОД ОПТИМІЗАЦІЇ SQL ЗАПИТІВ СИСТЕМИ УПРАВЛІННЯ БАЗАМИ ДАНИХ

Анотація. Актуальність. Розмір баз даних, який відноситься до обсягу збережених даних, може значно варіюватися. Зрозуміло, що чим більший розмір бази даних, тим більше часу потрібно на пошук необхідної інформації, що призводить до збільшення часу обробки запитів сервером. З одного боку, цю проблему вирішують шляхом підвищення продуктивності комп'ютерів, на яких розташовані системи управління базами даних (СУБД). Однак, просто підвищення продуктивності комп'ютерів недостатньо; часто значно кращі результати можна досягти шляхом зміни алгоритмів обробки SQL-запитів. Таким чином, незважаючи на те, що роботи з оптимізації SQL-запитів ведуться десятиліттями, через зростання темпів накопичення інформації та навантаження на сервери баз даних, така робота стала ще більш актуальною. **Мета.** Основною метою цієї роботи є покращення швидкості виконання вхідних запитів у реляційній базі даних, забезпечуючи високу продуктивність та зручну реалізацію для користувача. Це буде досягнуто шляхом розробки вдосконаленого методу оптимізації, який синтезує складні SQL-запити з великої кількості простіших запитів, тим самим підвищуючи загальну ефективність та зручність використання. **Метод.** У статті представлено метод оптимізації синтезу складних SQL-запитів з безлічі простих, що дозволяє підвищити швидкість виконання вхідного запиту реляційною базою даних при одночасному забезпеченні високої продуктивності та легкості застосування. **Результати.** Розроблено метод оптимізації SQL-запитів спеціально для ситуацій, коли швидкість вибірки даних погіршується з часом. Цей метод включає заміну оператора IN на тимчасову таблицю та використання не кластеризованого індексу. Таким чином, він прискорює процес вибірки даних, зменшуючи логічні звернення. **Висновки.** Основні цілі дослідження були визначені та успішно досягнуті: проаналізовано існуючі підходи до оптимізації SQL-запитів та основних засад роботи оптимізатора запитів як компонента СУБД; розроблено метод оптимізації SQL-запитів; оцінено ефективність запропонованого удосконаленого методу.

Ключові слова: бази даних, оптимізація SQL-запитів, індекси, оператори IN.

Аббревіатури

SQL – Structured Query Language;

БД – база даних;

СУБД – система управління базами даних.

Вступ

Обробка запитів SQL включає в себе синтаксичний аналіз, оптимізацію, генерацію коду та виконання SQL запитів. У сфері автоматичного налаштування SQL існує велика кількість інструментів, які допомагають програмістам баз даних виявляти запити, які призводять до проблем, і пропонують рішення для їх виправлення [1]. Однак налаштування SQL все ще потребує значної уваги людини, незважаючи на наявність таких інструментів.

Перелічимо типові проблеми, що впливають на продуктивність SQL-запитів:

1. Недостатня оптимізація запитів: SQL-запити повинні бути оптимізовані для швидкого виконання. Якщо запити не оптимізовані, вони можуть вимагати велику кількість обчислювальних ресурсів і часу для обробки даних. Це може бути викликано неправильними або неефективними виразами, відсутністю необхідних індексів або недостатньою кількістю фільтрів для зменшення обсягу даних, які потрібно обробити.

2. Відсутність необхідних індексів: Індекси використовуються для швидкого пошуку даних у таблицях. Відсутність необхідних індексів або неправильне використання індексів може призвести до повільних запитів. Наприклад, якщо часто виконуються запити з умовою WHERE на певному стовпці, то варто розглянути створення індексу для цього стовпця.

3. Погана структура бази даних: Якщо база даних не має оптимальної структури, це може призводити до повільних запитів. Наприклад, якщо таблиці мають занадто багато зв'язків або неправильні типи даних, то це може впливати на продуктивність запитів.

4. Великий обсяг даних: Якщо база даних містить великий обсяг даних, це може призвести до повільних запитів, оскільки системі потрібно багато часу для обробки такої кількості інформації. У таких випадках можна розглянути використання підходів, таких як розподілена обробка даних або використання партицій для розділення таблиць на менші сегменти.

5. Недостатня оптимізація сервера бази даних: Неправильна конфігурація сервера бази даних, недостатній розмір буферів, погано налаштовані параметри пам'яті або недостатня кількість ресурсів можуть впливати на продуктивність SQL-запитів. Важливо виконати належне налаштування сервера бази даних для забезпечення оптимальної продуктивності.

6. Несправні апаратні засоби: Якщо сервер бази даних або мережа мають проблеми зі швидкістю або стабільністю, це може призвести до повільних відповідей на SQL-запити. Необхідно перевірити апаратне забезпечення, забезпечити достатню пропускну здатність і знизити відсоток непередбачуваних перебоїв у роботі сервера та мережі.

7. Конфлікти паралельного виконання: При одночасному виконанні багатьох запитів можуть виникати конфлікти, які знижують продуктивність. Це можуть бути блокування, колізії або конкуренція за ресурси. Варто розглянути використання транзакцій, блокування на рівні рядків або ізоляційні рівні для уникнення таких проблем.

Найкращий спосіб підготовки набору запитів з оптимальною продуктивністю - написати запити кількома різними способами та порівняти їх планування та виконання, що називається методом проб і помилок. Для підготовки та тестування оптимізації продуктивності запитів, дослідники пропонують безліч методів та порад, які можна використовувати [1-6]. Існує багато різних способів визначити найкращі можливості написання запитів, проте два найбільш часто використовуваних методи - це аналіз кількості логічних даних, створених запитами, та перегляд графічних планів.

Об'єктом дослідження є SQL-запити до бази даних, з метою вдосконалення їх швидкості та ефективності. **Предметом** вивчення є методи вдосконалення швидкості запитів та ефективності.

Метою роботи є збільшити швидкість виконання SQL-запиту базою даних при одночасній підтримці високої продуктивності та легкості застосування рішення шляхом розробки покращеного методу оптимізації синтезу складних SQL-запитів з простих.

1 Постановка задачі

Зараз реляційні бази даних широко використовуються для зберігання та обробки даних, при цьому розмір баз може коливатися від кілобайтів до терабайтів. Збільшення розміру бази даних призводить до збільшення часу, потрібного для пошуку потрібної інформації та обробки запитів сервером. Бази даних також використовуються для публікації даних в Інтернеті, що може призводити до значного зростання кількості користувачів, що звертаються до бази даних в режимі реального часу. Збільшення продуктивності комп'ютерів, на яких працюють системи управління базами даних, не завжди може вирішити ці проблеми. Іноді зміна алгоритмів обробки SQL-запитів, наприклад, використання однорівневих індексів замість сканування таблиці, може дати значно кращий результат. У випадку використання збалансованих дерев індексів, вираш у продуктивності може бути ще більшим. Однак підвищення швидкості роботи зовнішньої пам'яті призводить до лінійного збільшення продуктивності обробки запитів.

Існує низка критеріїв, за якими можна оптимізувати виконання SQL-запитів, такі як швидкість виконання, завантаження процесора та використання пам'яті. Зазвичай, головною метою при оптимізації є мінімізація часу виконання запиту, проте звернення до диска може бути найдорожчою операцією, тому кількість звернень до диска є одним з основних критеріїв оптимізації. Інший важливий критерій - процесорний час.

Для збільшення швидкості виконання запитів можна використовувати різні методи, такі як використання індексів, хеш-функцій, матеріалізованих і секціонованих уявлень, алгоритмів виконання операцій з'єднання та обмеження, ведення статистики даних та інші. Наприклад:

1. Оптимізація запитів. Перевірити SQL-запити та спробувати їх оптимізувати. Слід використовувати ефективні вирази, уникати зайвих обчис-

лень і функцій, які можуть затримувати виконання запитів.

2. Створення індексів. Відповідно до потреб доцільно створити індекси для стовпців, що часто використовуються в операторах WHERE або JOIN. Індекси дозволяють швидше знаходити та фільтрувати дані, покращуючи продуктивність запиту.

3. Перегляд структури бази даних. Доцільно перевірити чи правильно спроектована структура БД. Таблиці повинні мати належні первинні та зовнішні ключі, і оптимальні зв'язки між таблицями. Також слід використовувати нормалізацію, де це необхідно, щоб забезпечити ефективну організацію даних.

4. Кешування запитів. Механізми кешування дозволяють зберігати результати запитів у пам'яті для швидкого доступу. Це особливо корисно для запитів, які виконуються часто та повторюються.

5. Поділ запиту на менші частини. Якщо запит складний або має багато об'єднань, можливо є сенс розбити його на менші запити або підзапити. Це може знизити навантаження і покращити продуктивність.

6. Налаштування сервера бази даних. Перевірити налаштування сервера бази даних, такі як розмір буферів, параметри пам'яті, кількість одночасних з'єднань і т. д. Правильна конфігурація може вплинути на швидкість виконання SQL-запитів.

7. Використання попереднього завантаження даних. Якщо відомо, що певні дані будуть потрібні для виконання запитів, доцільно завантажити їх заздалегідь у пам'ять або тимчасові таблиці. Це допоможе уникнути зайвих запитів до бази даних і прискорити виконання запитів.

8. Паралельне виконання. Використовувати можливість паралельного виконання запитів, де це можливо. Розглянути розділення великих запитів на менші частини, які можуть бути виконані паралельно.

Ці методи можуть бути використані окремо або в поєднанні один з одним для досягнення кращої продуктивності SQL-запитів системи.

Оптимізатор - спеціальний компонент СУБД, що займається побудовою плану запиту, тобто його компіляцією. Оптимізатор розглядає різні критерії та наявність допоміжних структур, встановлює кілька планів виконання, оцінює вартість виконання кожного з них і вибирає найкращий. Однак, скласти всі можливі плани та оцінювати їх було б занадто дорого та часоно затратно, тому складається тільки підмножина можливих планів, з якої вибирається найкращий.

Зараз проводяться роботи по розробці нових структур, що дозволяють зменшити кількість звернень до диска та підвищенню ефективності оптимізатора запитів, тобто алгоритму вибору можливих планів виконання.

Незважаючи на тривалість робіт з оптимізації SQL-запитів протягом багатьох років, вони залишаються важливими й сьогодні. Справді, із зростанням обсягу інформації й навантаження на бази даних, такі завдання набули ще більшої актуальності.

2 Огляд літератури

Обробка SQL складається з кількох етапів, які включають синтаксичний розбір, оптимізацію, генерацію рядків джерела та виконання оператора SQL. Рис. 1 показує загальні етапи обробки SQL запиту. Існує широкий асортимент інструментів для автоматичного налаштування SQL, які допомагають розробникам баз даних виявляти проблеми в SQL-операторах та пропонують їх виправлення [1]. Хоча такі інструменти можуть знизити навантаження на розробників баз даних, налаштування SQL все ж вимагає значної участі людини.

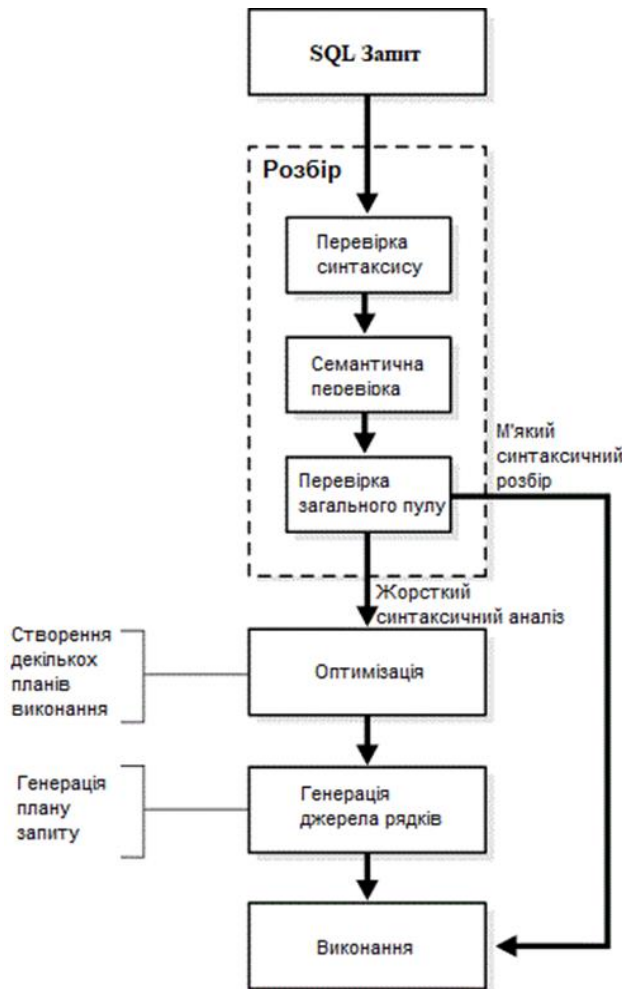


Рис. 1. Загальні етапи обробки SQL запиту

Щоб забезпечити оптимальну продуктивність, найкращий спосіб скласти набір запитів - це написати їх кількома різними способами та порівняти плани планування та виконання, використовуючи метод проб і помилок. Існує безліч методів та порад для підготовки та тестування оптимізації продуктивності запитів [1-6]. Два найбільш часто використовуваних методи - це аналіз кількості логічних даних, створених запитом, та перегляд графічних планів.

В [3] Р. Karthik та співавтори висвітлюють важливість налаштування запитів SQL для досягнення високої продуктивності систем управління базами даних. Перш ніж запропонувати методи налаштування SQL-запитів, вони досліджують найкращі підходи

до одночасного налаштування кількох операторів SQL. Вони пропонують широкий підхід, який дозволяє паралельно налаштовувати SQL-оператори та може значно зекономити час на налаштування на пізніших етапах. Автори не надають жодних конкретних методів налаштування SQL, але підтверджують, що використання ключового слова DISTINCT у запитах SQL не є доцільним, коли пошук здійснюється на основі атрибуту первинного ключа.

У статті [2] R. Bhajipale та інші провели докладне дослідження необхідності налаштування запитів у великих базах даних з метою підвищення продуктивності. Вони запропонували кілька технік налаштування запитів, що є частиною їх рекомендацій. Дослідники зазначили, що оптимізація запитів - це глибока тема, і їх поради є лише загальними настановами, які охоплюють найважливіші аспекти. Рекомендації, запропоновані у статті, не є вичерпним списком технік або типів складних запитів, тому ще є багато роботи для інших дослідників у цій галузі.

У статті S. S. Srinivas [7] було представлено кілька методів налаштування та оптимізації SQL-запитів. Незважаючи на те, що ці методи є важливими, список їх не є повним і не охоплює всі доступні методи. Це дає можливість дослідникам продовжувати розробку своїх пропозицій з налаштування SQL у майбутньому. У статті не було надано оцінок ефективності будь-якого з запропонованих методів, що є недоліком. Тому не можна гарантувати, що запропоновані у дослідженні методи у реальних ситуаціях будуть ефективними.

У своїй статті D. Patel та P. Patel [8] запропонували метод оптимізації запитів, що ґрунтується на перегляді базового об'єкта Schema для великих баз даних. Вони вважають, що їх схема приводить до скорочення часу виконання запиту та зменшення обсягу запиту, порівняно з іншими методами, такими як Query Graph, Tableaus або Optimization of Queries with Aggregates. Проте, у статті не було представлено жодного методу для покращення синтаксису SQL з метою підвищення продуктивності.

У статті S. Patil та інші [9], було розглянуто 9 різних методів, які можна використовувати для налаштування SQL-запитів, проте, стаття не містить конкретних пропозицій або порад щодо того, як писати ефективні оператори SQL.

Дослідження D. Colley та C. Stanier зосереджені на нових методах, які були запропоновані недавно для збільшення продуктивності баз даних [20]. Вони дослідили три ключові фактори, які впливають на продуктивність запитів до БД, включаючи вплив оператора JOIN на продуктивність, але їхній аналіз не є повним, оскільки вони не розглядають різні способи використання оператора JOIN в операторі SQL.

Натомість, стаття [6] пропонує оптимізацію запитів з використанням перетворень SQL, яка є одним з чотирьох кроків оптимізатора запитів Oracle. Перетворення SQL виконується оптимізатором бази даних, щоб замінити даний оператор SQL на інший, який може бути виконаний більш ефективно, але повертає ті ж результати. Стаття розглядає різні методи перетворення SQL, які реалізовані в Oracle, але

не надає порад щодо написання ефективних SQL-запитів для конкретної ситуації або вимоги.

У роботі [10] С. G. Corlatan та інші досліджували методи оптимізації запитів для сервера Microsoft SQL і з'ясували, що декілька факторів можуть впливати на продуктивність СУБД, такі як відсутність індексів, неточна статистика, погано написані запити, взаємоблокування, використання курсорів в операціях TransactSQL, фрагментація індексів і часта перекомпіляція запитів. Хоча дослідники запропонували кілька порад щодо оптимізації запитів SELECT, вони не були перевірені в реальних експериментах, тому їх можна розглядати лише як загальні рекомендації.

У статтях [11, 12] розглянуто використання підказок при оптимізації вкладених SQL-запитів. Оскільки користувачі не мають прямого доступу до оптимізатора запитів СУБД, підказки, вставлені у вигляді коментарів в оператор SQL, можуть дати оптимізатору вказівку виконати певні операції відповідно до потреб користувача, не вдаючись до автоматичної оптимізації. Дослідники вказують на те, що використання підказок може збільшити продуктивність запитів та що метод є простим у використанні та зрозумілим, через те що підказки пишуться простою мовою. Однак головним недоліком цього методу є те, що підказки синтаксису залежать від певної реалізації СУБД і не можуть бути типізовані, тому що підказки не є частиною SQL-стандартів.

У роботі J. Habimana [4] надав поради до написання ефективних та швидких SQL-запитів. Для дослідження ефективності методів, що пропонуються, було проведено порівняльний аналіз продуктивності налаштованих та не налаштованих запитів. Результати свідчать про те, що налаштовані запити мають деякі переваги. Однак, в роботі було виявлено недоліки, зокрема, всі поради стосуються запитів, які отримують дані з однієї таблиці та не включають в себе вкладені запити. Крім того, відсутня оцінка ефективності запропонованих порад, тому їх ефективність не може бути визначена. Автор не надав жодних атрибутів продуктивності, які можна було б використовувати для перевірки ефективності запропонованих методів іншими дослідниками.

3 Матеріали і методи

Оператор у SQL - це слово або символ, яке застосовується переважно у реченні WHERE для виконання операцій, таких як порівняння та арифметичні операції. Ці оператори допомагають встановлювати умови в запиті та поєднувати декілька умов в одному запиті. Хоча оператори є дуже корисними у SQL, деякі з них можуть вимагати значну кількість ресурсів для виконання.

Далі запропоновано покращений метод [13], що намагається уникати оператор IN в реченні WHERE.

Хід розв'язку:

Зробіть тимчасову таблицю, вставте потрібні дані та виконайте JOIN до основного запиту.

У випадку якщо складений запит містить оператор IN, його слід вилучити, а для створення тим-

часової таблиці та додавання її в запит слід здійснити наступне:

1. Визначити атрибути SELECT замість SELECT *.

Якщо таблиця містить багато атрибутів та кортежів, вибірка усіх полів (за допомогою SELECT *) перевикористовує ресурси бази даних для запиту великого об'єму зайвих даних. Одночасно з цим, зазначення стовпців у операторі SELECT буде видавати тільки важливі дані.

2. Обминати SELECT DISTINCT при можливості.

SELECT DISTINCT групує всі поля в запиті, для створення чіткої відповіді. Проте щоб досягти цю ціль треба значний об'єм обчислювальної потужності.

3. Застосовувати WHERE замість HAVING для фільтрації даних.

Згідно з порядком операцій SQL, оператор HAVING виконується після оператора WHERE. Якщо ж треба відфільтрувати запит на базі певних умов, оператор WHERE є ефективнішим.

4. Застосовувати оператор LIMIT для скорочення результатів запиту.

Застосовуючи LIMIT до першого запуску запиту, слід переконатися, що будуть отримані значущі і бажані результати. При цьому якщо у СУБД оператор LIMIT відсутній, то можна використати наступний синтаксис:

```
SELECT TOP 3 * FROM Table;
```

5. Змінити підзапит на оператор з'єднання JOIN.

Підзапити можна замінити з'єднанням, яке буде виконуватися швидше. Наприклад:

```
SELECT authors.id, (SELECT MAX(created)
```

```
FROM posts
```

```
WHERE author_id = authors.id)
```

```
AS latest_post FROM authors
```

Цей запит можливо переписати з використанням JOIN:

```
SELECT authors.id, MAX(posts.created) AS
```

```
latest_post
```

```
FROM authors
```

```
INNER JOIN posts
```

```
ON (authors.id = posts.author_id)
```

```
GROUP BY authors.id
```

6. Прибрати оператор IN, створити таблицю #temp і вставити в неї потрібні дані.

7. Визначити індекс для таблиці #temp.

8. З'єднати #temp з основним відношенням.

4 Експерименти

Перед початком виконання запитів [13] було обрано малу та середню по розміру базу даних з [4], для маніпулювання БД було обрано Microsoft SQL Server Management Studio.

Схема БД представлена на рис. 2, а опис – в табл. 1.

Розмір малої бази даних складає біля 10 ГБ, а середньої 50 ГБ.

Характеристики комп'ютера, на якому проводяться експерименти опичані у табл. 2.

Таблиця 1 – Опис баз даних, що використовуються

Назва таблиці	Кількість рядків		Кількість зв'язків
	Маленька БД	Середня БД	
LinkTypes	2	2	1
PostLinks	161519	1421208	2
Posts	3729195	17142169	4
PostTypes	8	8	1
Votes	10143364	52928720	3
VoteTypes	15	15	1
Comments	3875183	24534730	2
Users	299398	2465713	3
Badges	1102019	8042005	1

Таблиця 2 – Опис комп'ютера, що використовуються

Характеристика	Значення
Процесор	Intel Pentium N3542 2.16 GHz
Опер. система	Windows 10
Пам'ять	HDD читування – 90 MB/s , записування – 85 MB/s
Опер. пам'ять	4 Гб DDR3–1333 MHz

Оцінку методу було проведено шляхом створення кількох SQL-запитів з оператором IN та використання Temp Tables замість оператора IN, як запропоновано у цьому дослідженні. Процес застосування методу показано на лістингу 1.

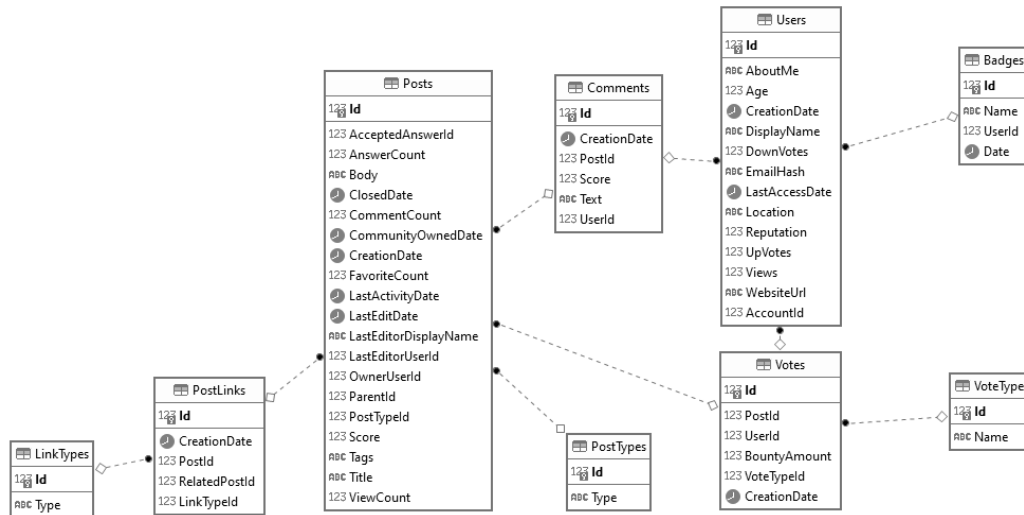


Рис. 2. Схема БД

5 Обговорення

Результати запитів 1 та 2 показані на рис. 3, результати запитів 3 та 4 – на рис. 4, результати запитів 5 та 6 – на рис. 5.

Якщо дивитись на плани виконання запитів, видно, що при використанні запропонованого удосконаленого методу кількість логічних звернень зменшилась, що в результаті дозволило скоротити час виконання запиту.

Тому, на підставі отриманих результатів виконання, можна зробити висновок, що в малій базі даних запит виконався на 15% швидше, а в середній базі даних запит виконався на 17% швидше.

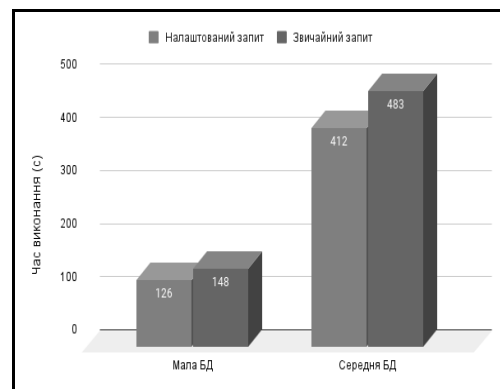


Рис. 4. Час виконання запитів 3 та 4

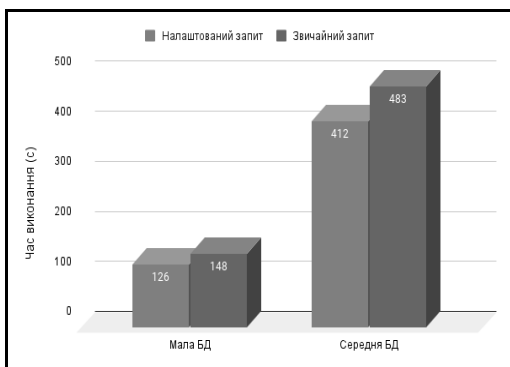


Рис. 3. Час виконання запитів 1 та 2

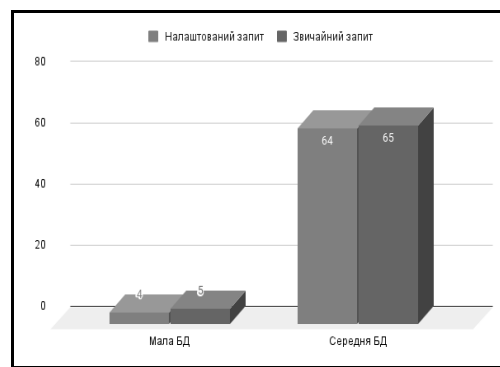


Рис. 5. Час виконання запитів 5 та 6

Запит 1 з використанням IN:

- вибираємо стовпці та об'єднуємо такі таблиці як: Users, Posts, PostTypes, Comments, Votes

```
select U.DisplayName, u.AboutMe, p.Body, pt.type, c.text from Users u
join Posts p on u.Id = p.OwnerUserId
join PostTypes pt on p.PostTypeId = pt.Id
join Comments c on c.PostId = p.Id
join Votes v on u.Id = v.UserId
```
- задаємо умову використовуючи оператори between та in;

```
where u.Views between 5000 and 50000
and Location in ('El Cerrito, CA', 'Corvallis, OR')
and v.UserId in (select UserId from Votes where VoteTypeId in(5,8))
```
- При виконанні запиту індексів додано не було.

Запит 2 з використанням Temp Tables:

- створюємо тимчасову таблицю

```
create table #temptable (UserID int)
```
- вставляємо відповідні дані в тимчасову таблицю

```
insert into #temptable (UserID) select UserId from Votes where VoteTypeId in(5,8)
```
- створення індексу на таблицю #temptable

```
create INDEX INX_temp on #temptable (UserID)
```
- вибираємо стовпці та об'єднуємо такі таблиці як: Users, Posts, PostTypes, Comments та #temptable

```
select U.DisplayName, u.AboutMe, p.Body, pt.type, c.text from Users u join Posts p on u.Id = p.OwnerUserId
join PostTypes pt on p.PostTypeId = pt.Id
join Comments c on c.PostId = p.Id
join #temptable t on t.UserID = u.Id
```
- задаємо умову використовуючи оператори between та in;

```
where u.Views between 5000 and 50000
and Location in ('El Cerrito, CA', 'Corvallis, OR')
```
- видаляємо тимчасову таблицю

```
drop table #temptable
```

Запит 3 з використанням IN:

- вибираємо стовпці та об'єднуємо таблиці Users та Badges

```
select DisplayName, Location, B.name from Users U
join Badges B on U.Id = B.UserId
```
- задаємо умову використовуючи оператори IN та between

```
where Location in ('Israel', 'United States', 'Germany')
and Reputation between 50 and 200
```

Запит 4 з використанням Temp Tables:

- створюємо тимчасову таблицю

```
create table #temptable (UserID int, DisplayName varchar(255), Location varchar(255))
```
- вставляємо відповідні дані в тимчасову таблицю

```
insert into #temptable (UserID, DisplayName, Location) select Id, DisplayName, Location from Users where Location in
('Israel', 'United States', 'Germany') and Reputation between 50 and 200
```
- створення індексу на таблицю #temptable

```
create index INdx on #temptable (Location, UserID)
```
- вибираємо стовпці та об'єднуємо такі таблиці як: #temptable та Badges

```
select t.DisplayName, t.Location, B.name from #temptable t join Badges B on t.UserID = B.UserId
```
- видаляємо тимчасову таблицю

```
drop table #temptable
```

Запит 5 з використанням IN:

- В наступному запиті було вирішено дослідити метод при використанні під запиту
- вибираємо стовпці та задаємо умову за допомогою IN та WHERE

```
select Votes.PostId, Votes.UserId from Votes
where UserId in (
select Users.Id from Users
where Users.DisplayName in ('Jeff Atwood', 'Nick', 'Gaius'))
```

Запит 6 з використанням Temp Tables:

- створюємо тимчасову таблицю вставляємо дані та створюємо індекс

```
create table #temptable (UserID int)
insert into #temptable (UserID) select Users.Id from Users where Users.DisplayName in ('Jeff Atwood', 'Nick', 'Gaius')
create index INdx on #temptable (UserID)
```
- вибираємо стовпці та використовуємо temptable в підзапиті

```
select Votes.PostId, Votes.UserId from Votes
where UserId in (select UserId from #temptable)
drop table #temptable
```

Лістинг 1. Процес застосування запропонованого удосконаленого методу

Висновки

Було проведено огляд та аналіз існуючих методів оптимізації SQL-запитів та основних принципів роботи оптимізатора запитів як компонента СУБД. Було обрано напрямок зменшення логічних операцій та вирішено проблему оптимізації SQL-запитів до реляційної бази даних.

Розроблено метод оптимізації SQL запитів у випадках, коли швидкість вибірки даних починала погіршуватись, який полягає у заміні оператора IN на тимчасову таблицю та використанні некластери-

зованого індексу, що дозволяє швидше вибирати дані за рахунок зменшення логічних операцій порівняно з існуючими методами.

Були виконані запити, що використовували запропонований метод на основі тимчасової таблиці та некластеризованого індексу, що підтверджує життєздатність запропонованого рішення. Була проведена експериментальна оцінка, яка показала, що виведення даних було на 15% швидше при використанні малої бази даних загальним обсягом 10 ГБ і на 17% швидше при використанні середньої бази даних обсягом 50 ГБ.

СПИСОК ЛІТЕРАТУРИ

1. Rupley, M. L. (2008). Introduction to query processing and optimization. Indiana University, South Bend, South Bend, IN, USA, TechReport TR-20080105-1.
2. Bhajipale, R., Bisen, P., Meshram, A., & Thakur, S. S. (2016). SQL tuner. International Journal of Computer Trends and Technology, 33(1), 29-32.
3. Karthik, P., Reddy, G. T., & Vanan, E. K. (2012). Tuning the SQL query in order to reduce time consumption. International Journal of Computer Science Issues, 9(4/3), 418-423.
4. Habimana, J. (2015). Query optimization techniques – tips for writing efficient and faster SQL queries. International Journal of Scientific & Technology Research, 4(10), 22-26.
5. Sahal, R., Nihad, M., Khafagy, M. H., & Omara, F. A. (2018). iHOME: Index based JOIN query optimization for limited big data storage. Journal of Grid Computing, 16(2), 345-380.
6. Sharma, M. (2012). Query optimization using SQL transformations. International Journal of IT, Engineering and Applied Sciences Research, 1(1), 100-104.
7. Srinivas, S. S., Naik, B. V., & Kumar, J. S. A. (2017). Query minimization methods. International Journal of Scientific & Engineering Research, 8(5), 30-33.
8. Patel, D., & Patel, P. (2015). An approach for query optimization by using schema object base view. International Journal of Computer Applications, 119(16), 21-24.
9. Patil, S., Damare, P., Sonawane, J., & Maitre, N. (2015). Study of performance tuning techniques. Journal of Emerging Technologies and Innovative Research (JETIR), 2(3), 499-502.
10. Corlatan, C. G., Lazar, M. M., Luca, V., & Petricica, O. T. (2014). Query optimization techniques in Microsoft SQL server. Database Systems Journal, 5(2), 33-48.
11. Lokhande, A. D., & Shete, R. M. (2012). The use of hints in SQL-Nested query optimization. Journal of Data Mining and Knowledge Discovery, 3(1), 54-57.
12. Brent Ozar Unlimited. (2023). How to Download the Stack Overflow Database via BitTorrent. Retrieved April 3, 2023, from <https://www.brentozar.com/archive/2015/10/how-to-download-the-stack-overflow-database-via-bittorrent/>.
13. Єрмолаєв, О. Д. Метод оптимізації SQL запитів : дипломна робота ... бакалавра : 172 Телекомунікації та радіотехніка / Єрмолаєв Олександр Дмитрович. – Київ, 2021. – 56 с.

Received (Надійшла) 05.04.2023

Accepted for publication (Прийнята до друку) 17.05.2023

The method of optimizing sql queries of the database management system

S. Sulima, O. Iermolaiev

Abstract. Context. The size of a database, which refers to the volume of stored data, can vary significantly. It is clear that the larger the size of the database, the more time is required to search for the necessary information, resulting in increased query processing time for the server. On one hand, this problem can be addressed by improving the performance of the computers hosting the database management systems (DBMS). However, simply enhancing computer performance is insufficient; often, much better results can be achieved by changing the algorithms for processing SQL queries. Therefore, despite the fact that optimization of SQL queries has been conducted for decades, due to the rapid accumulation of information and the workload on database servers, such work has become even more relevant. **Objective.** The main objective of this work is to improve the speed of executing incoming queries in a relational database, ensuring high performance and user convenience. This will be accomplished by developing an enhanced optimization method that synthesizes complex SQL queries from a large number of simpler queries, thereby increasing overall efficiency and usability. **Method.** This article presents a method for optimizing the synthesis of complex SQL queries from numerous simple queries, which enhances the speed of executing the incoming query by a relational database while simultaneously ensuring high performance and ease of use. **Results.** A method for optimizing SQL queries specifically for situations where data retrieval speed deteriorates over time has been developed. This method involves replacing the IN operator with a temporary table and utilizing a non-clustered index. Consequently, it accelerates the data retrieval process by reducing logical references. **Conclusions.** The main research objectives were identified and successfully achieved: the analysis of existing approaches to SQL query optimization and the fundamental principles of query optimizers as components of DBMS were conducted, a method for optimizing SQL queries was developed, and the effectiveness of the proposed enhanced method was evaluated.

Keywords: databases, SQL query optimization, indexes, IN operators.

О. О. Фомічов, В. С. Бурцев

Харківський національний технічний університет радіоелектроніки, Харків, Україна

ДОСЛІДЖЕННЯ ІМУННИХ ОПЕРАТОРІВ В МОДЕЛІ ШТУЧНОЇ ІМУННОЇ МЕРЕЖІ

Анотація. Актуальність. Значною актуальністю в області інформаційних технологій користуються моделі та методи, які дозволяють проводити інтелектуальну обробку та систематизацію даних. Серед таких моделей можна виділити моделі, які працюють на основі біологічних принципів організації обчислень, такі як штучні нейронні мережі, генетичні алгоритми та штучні імунні системи. На сьогоднішній день існує декілька видів моделей штучних імунних систем, які використовують специфічні імунні оператори, що впливають на швидкість роботи моделі та якість вирішення практичних задач. **Метою даної роботи** є дослідження впливу організації роботи найбільш поширених імунних операторів на швидкість роботи імунних алгоритмів та точність прийняття рішення щодо класифікації об'єктів з контрольованим навчанням. **Об'єктом дослідження** є імунні оператори клонування, мутації та супресії, а також оцінка швидкості процесу формування імунної відповіді мережі антитіл на популяцію антигенів. **Предметом дослідження** є модель штучно імунної мережі та алгоритм aiNET, імунні оператори та їх вплив на процес формування імунної відповіді для класифікації об'єктів з контрольованим навчанням. **Результати.** У даній роботі запропоновано оптимальне налаштування імунних операторів для забезпечення високої швидкості формування імунної відповіді алгоритмом aiNET під час вирішення задачі класифікації. Окрім того запропоновано використання цільових імунних об'єктів для пришвидшення класифікації антитіл, що не набули стану специфічності до антигенів навчальної вибірки. **Висновок.** Модифікований метод aiNET класифікації з контрольованим навчанням планується використовувати у подальшому для керування поведінкою персонажів в ігрових додатках типу action.

Ключові слова: штучні імунні системи, антитіло, антиген, афінність, клонування, супресія, повна специфічність, умовна специфічність, супресія імунної мережі

Вступ

Однією з актуальних задач в області інтелектуальної обробки даних та прийняття рішень є задача класифікації об'єктів. Задля вирішення цієї задачі можуть використовуватися різні види навчання: контрольоване навчання, неконтрольоване навчання, та навчання з підкріпленням. Під час класифікації з контрольованим навчанням передбачається використання навчальної вибірки класифікованих об'єктів, які представляють сукупність класів, до яких необхідно визначити належність об'єктів, що класифікуються [1-3]. У випадку класифікації з неконтрольованим навчанням інформація про класи відсутня або повністю, або представлена у вигляді кількості класів, які необхідно сформувати з набору об'єктів.

Класифікація з контрольованим та неконтрольованим навчанням може відбуватися за допомогою основних принципів функціонування штучних імунних систем, головними робочими елементами яких є антитіла, які формують популяції на різних етапах роботи методу класифікації. У випадку класифікації з контрольованим навчанням навчальна вибірка представляється популяцією антигенів, до яких штучна імунна система формує адаптивну імунну відповідь [2-4]. У випадку класифікації з неконтрольованим навчанням штучна імунна система організовує взаємодію між своїми антитілами за відсутності антигенів навчальної вибірки. Слід зазначити, що серед існуючих моделей штучних імунних систем найбільш поширеними є модель клонального відбору та модель штучної імунної мережі [2-13]. Головною відмінністю між цими моделями є принципова неможливість взаємодії між антитілами в моделі клонального відбору [2-4]. Альтернативою такому підходу є модель штучної імунної мережі, де антитіла можуть взаємодіяти не тільки з анти-

генами, а й між собою [3-5]. Таким чином, модель штучної імунної мережі має більший науково-дослідницький потенціал, ніж модель клонального відбору при вирішенні задачі класифікації.

Модель штучної імунної мережі реалізована низкою методів: aiNET, opt-aiNET та RLAI [3-12]. Незважаючи на те, що ці методи значно відрізняються між собою, всі вони використовують перелік універсальних імунних операторів, кожен з яких відповідає за деякий етап роботи штучної імунної мережі. Серед таких операторів можна виділити оператор клонування, оператор мутації клонів, оператор селекції клонів, оператор супресії мережі. Слід зазначити, що всі перелічені методи, що функціонують за принципами роботи штучної імунної мережі, мають велику кількість зайвих обчислень, особливо в операторах селекції клонів та супресії мережі [12]. На кількість таких обчислень також значно впливають особливості роботи операторів клонування, мутації, та презентації імунних об'єктів. Окрім того, важливою частиною будь-якого імунного методу є спосіб прийняття рішення про зупинку процесу формування імунної відповіді або процесу імунного навчання.

Метою цієї роботи є дослідження впливу імунних операторів на швидкість та якість роботи найбільш поширеного методу aiNET, який буде застосовуватися для вирішення задачі класифікації з контрольованим навчанням. Зміни принципу роботи операторів стосуються принципу селекції антитіл, особливостей клонування, способу мутації, принципу селекції клонів після мутації, а також принципу супресії антитіл при формуванні нових популяцій імунних об'єктів. Окрім того, пропонується використання направленої мутації клонів, а також використання цільових імунних об'єктів, які обираються з популяції антитіл та антигенів.

Основна частина

Класифікація об'єктів з контрольованим навчанням на основі штучних імунних систем проводиться для множини антитіл, що формують початкову популяцію об'єктів імунної мережі, та за допомогою множини класифікованих об'єктів навчальної вибірки, розподілених серед деякою кількістю класів. З об'єктів навчальної вибірки формується популяція антигенів. Слід зазначити, що задля забезпечення функціонування імунного методу класифікації, антитіла та антигени мають представлятися однаковим чином, тобто мати співставний набір ознак. У якості міри близькості між об'єктами використовується критерій афінності, який базується на евклідовому просторі між ними [2-12]:

$$aff_{ij} = (1 + d_{ij})^{-1}, \quad (1)$$

$$d_{ij} = \sqrt{\sum_{m=1}^n (p_{im} - p_{jm})^2}, \quad (2)$$

де d_{ij} – евклідова дистанція між i та j об'єктами, $s_{im} - m$ ознака відповідного об'єкту.

Слід зазначити, що метод aiNET [4-7] є універсальним методом, що моделює мережеву взаємодію антитіл та не може без змін застосовуватися для вирішення задач класифікації. Тому цей алгоритм необхідно модифікувати шляхом додавання низки специфічних операторів вибору класів, що представляються популяцією антигенів.

Роботу модифікованого імунного алгоритму aiNET для класифікації можна умовно розділити на два основних етапи [12]: 1) етап формування специфічної імунної відповіді шляхом відтворення антигенів антитілами в процесі імунного навчання; 2) визначення класів для антитіл, які не досягли високого рівня специфічності до антигенів навчальної вибірки. Відповідно до цього, алгоритм класифікації об'єктів з контрольованим навчанням на основі моделі штучної імунної мережі на рівні імунних операторів визначається наступним чином:

$$\begin{aligned} & class\text{-}aiNET = \\ & = \left[\begin{array}{l} Presentation(AB, AG) \rightarrow \\ Cloning(AB, CL) \rightarrow \\ Mutation(CL) \rightarrow \\ Supression(AB, CL) \rightarrow \\ Termination(AB, AG) \end{array} \right]^{aiNET} \\ & \rightarrow \left[\begin{array}{l} Division(AB, AB', AB'') \rightarrow \\ Specification(AB', AG) \rightarrow \\ Selection(AB'', AB', AG) \end{array} \right]^{Class} \end{aligned} \quad (3)$$

де $Presentation(AB, AG)$ – оператор представлення популяції антитіл мережі до антигенів, з яких сформована навчальна вибірка, а також антитіл між собою; $Cloning(AB, CL)$ – оператор клонування популяції антитіл; $Mutation(CL)$ – оператор мутації сформованих клонів; $Supression(AB, CL)$ – оператор супресії антитіл та клонів мережі; $Termination(AB, AG)$ – оператор перевірки зупинки процесу формування імунної відповіді мережі; $Division(AB, AB', AB'')$ – оператор розділення по-

пуляції антитіл на специфічні та неспецифічні антигенам; $Specification(AB', AG)$ – оператор визначення класів для антитіл, що набули стану специфічності антигенам; $Selection(AB'', AB', AG)$ – оператор визначення класу для антитіл, які не набули стану специфічності антигенам навчальної вибірки.

На швидкість роботи імунного алгоритму класифікації впливає спосіб організації кожного з наведених імунних операторів. Слід зазначити, що етап формування специфічної імунної відповіді за рахунок своєї циклічності займає більшу частину часу роботи алгоритму [2-13]. Час формування імунної відповіді має лінійну залежність від розміру популяції антитіл та популяції антигенів, які представляють класи та формують навчальну вибірку.

Значний вплив на швидкість роботи імунного алгоритму має визначення поняття специфічності антитіла до антигену. Зазвичай під специфічністю розуміють такий стан ознак антитіла, при якому воно повністю відтворює ознаки деякого антигену. Це призводить до того, що афінність між антигеном та специфічним до нього антитілом приймає максимальне значення з діапазону $(0,0; 1,0]$. Слід зазначити, що набуття стану специфічності антитіла до антигену відбувається за рахунок його послідовного клонування, мутації та супресії, що характеризується великою кількістю обчислювальних операцій та уповільнює процес класифікації об'єктів [2-12].

Задля підвищення швидкості процесу формування імунної відповіді пропонується використання двох видів специфічності антитіл, а саме повну та умовну специфічності. У випадку повної специфічності антитіла в процесі формування імунної відповіді повністю відтворює ознаки деякого антигену, отже повна специфічність антитіла можлива тільки до одного антигену і не можлива до групи антигенів. Стан умовної специфічності передбачає те, що антитіло частково адаптується під ознаки одного або декількох антигенів. Відповідно до цього, ознакою повної специфічності антитіла до антигену буде значення афінності у діапазоні $(0,9; 1,0]$, а для визначення ознаки умовної специфічності може використовуватися або критерій авідності антитіла до групи антигенів, або значення достатньої афінності антитіла до деякого окремого антигену, що визначається у діапазоні $(0,75; 0,9)$. Слід зазначити, що набуття стану умовної специфічності антитіла до деякого антигену у вказаних межах афінності може супроводжуватися зменшенням афінності цього антитіла до інших антигенів, що значно спрощує його подальшу класифікацію. Таким чином, у випадку набуття антитілом стану повної або умовної специфічності воно припиняє клонуватися під час формування імунної відповіді, що зменшує кількість обчислень.

Прискорення імунного алгоритму класифікації також може відбуватися за рахунок зменшення кількості антигенів, з якими взаємодіє антитіло та його клони в процесі набуття стану повної або умовної специфічності до одного з них. Під взаємодією між антитілом та антигеном розуміють обчислення афінностей між ними. Отже, зменшення кількості антигенів, до яких антитіло визначає афінності при-

зводить до зменшення обчислень, що позитивно впливає на швидкість алгоритму класифікації. Відповідно до цього, під час першого виконання оператора $Presentation(AB, AG)$ для кожного антитіла відбувається визначення переліку антигенів, з якими воно в подальшому буде взаємодіяти в процесі набуття стану специфічності. Для визначення кількості переліку таких цільових антигенів можна використати значення Natural Affinity Threshold (NAT) [2-7], яке визначається між всіма антитілами мережі на початку роботи алгоритму. Значення NAT визначають як середню афінності між всіма антитілами мережі:

$$NAT_{AB} = \frac{1}{n} \sum_{i=1}^n aff_{ij}, \quad (4)$$

де aff_{ij} – значення афінності між i та j антитілами, n – загальна кількість антитіл у мережі.

Відповідно до цього, робота оператора $Presentation(AB, AG)$ [2-13] відбувається наступним чином: спочатку відбувається визначення критерію NAT, потім для кожного антитіла відбувається визначення афінностей з кожним антигеном навчальної вибірки, і якщо ця афінність перевищує значення NAT – антиген буде обрано у перелік цільових антигенів для такого антитіла [12]. Слід зазначити, що якщо антитіла представлені в багатовимірному просторі ознак за середнім законом розподілення NAT важко ефективно використовувати для визначення цільових об'єктів. Тому задля підвищення значення NAT пропонується використовувати коефіцієнт звуження простору пошуку цільових об'єктів L , який визначається в діапазоні $(0,1; 1,0]$, та використовується у якості вхідного параметру імунного алгоритму класифікації. Таким чином умовою того, щоб антиген був обраний у перелік цільових об'єктів деякого антитіла, є виконання наступного правила:

$$aff_{ij} > \frac{1}{L} NAT_{AB}, \quad (5)$$

де aff_{ij} – значення афінності між i антитілом та j антигеном, NAT_{AB} – значення середньої афінності в популяції антитіл, а L – коефіцієнт звуження простору пошуку цільових об'єктів.

Під час презентації популяції антигенів антитілам існує імовірність визначення стану повної або умовної специфічності для антитіла до одного з його цільових антигенів. У такому випадку це антитіло буде класифіковано одразу і не буде приймати участі у подальшому формуванні імунної відповіді, оскільки воно вже є специфічним до антигенів. Також існує імовірність того, що всі цільові антигени для деякого антитіла будуть належати до одного класу. У такому випадку антитіло також буде класифіковано одразу і не буде приймати участь у формуванні імунної відповіді через набуття своєї специфічності до того чи іншого класу антигенів. Через велику кількість обчислювальних операцій оператор презентації всієї популяції антигенів антитілам використовується тільки один раз за весь час роботи імунного алгоритму класифікації. Під час визначення NAT для кожного антитіла формується перелік умовно специфічних антитіл, що відбувається відповідно

до (5). Цей перелік антитіл буде використовуватися після закінчення формування імунної відповіді на етапі визначення класів задля прискорення процесу класифікації. Таким чином використання NAT допомагає визначити не тільки цільові антигени задля пришвидшення процесу формування імунної відповіді, але й класифікувати антитіла, які не набули стану повної або умовної специфічності.

Оператор клонування $Cloning(AB, CL)$ [2-13] використовується задля формування набору клонів для кожного неспецифічного або некласифікованого антитіла. Для імунних алгоритмів найбільш поширеними є статичне та пропорційне клонування. У випадку статичного клонування – кількість клонів, яка формується для кожного антитіла, не залежить від його афінностей до антигенів і є вхідним параметром імунного алгоритму. У випадку використання пропорційного клонування кількість клонів антитіла пропорційно залежить від його афінностей з одним або декількома цільовими антигенами. Слід зазначити, використання надмірно великої або малої кількості клонів може мати однаково негативний вплив на швидкість процесу формування специфічної імунної відповіді.

Оператор мутації $Mutation(CL)$ [2-13] має великий вплив на швидкість роботи імунного алгоритму через те, що саме він забезпечує набуття стану повної або умовної специфічності між клонами антитіла та антигенами. Це відбувається завдяки тому, що робота даного оператора призводить до зміни ознак клонів, що може наблизити їх до ознак того чи іншого антигену. За принципом роботи відрізняють декілька видів оператора мутації: статична, пропорційна та зворотньо-пропорційна мутація. При використанні оператора статичної мутації зміни ознак клону антитіла відбуваються у межах одного чітко вказаного діапазону значень, який використовується у якості вхідного аргументу. У випадку використання пропорційної мутації зміни ознак клону відбуваються пропорційно його афінності до одного або групи цільових антигенів:

$$\mu = rand(0,0; aff_{ij}], \quad (6)$$

де μ – коефіцієнт мутації, а aff_{ij} – значення афінності між i антитілом, від якого було сформовано клон для подальшої мутації, та j антигеном.

Особливістю зворотньо-пропорційної мутації [12] є те, що зміни ознак клонів збільшуються у випадку низького рівня його специфічності до одного або групи цільових антигенів, та зменшуються у випадку високого рівня специфічності. Відповідно до цього визначення коефіцієнту мутації відбувається наступним чином:

$$\mu = rand(0,0; 1 - aff_{ij}]. \quad (7)$$

Задля підвищення швидкості формування імунної відповіді пропонується використання направленої мутації. У більшості існуючих імунних алгоритмах використовується випадкова мутація, тобто після визначення коефіцієнту мутації параметри клону або збільшуються, або зменшуються випадковим чином. Особливістю направленої мутації є те, що рішення про збільшення або зменшення ознаки

клону приймається після порівняння ознак цього клону та ознак одного з цільових антигенів [12].

Оператор супресії мережі *Supression(AB, CL)* дозволяє скоротити кількість антитіл та змінених під час мутації клонів до початкового рівня [2-7]. Головною умовою високої швидкості класифікації є зменшення кількості імунних об'єктів для яких необхідно проводити клонування, мутацію та презентацію, але за рахунок формування нових антитіл під час клонування відбувається значне збільшення кількості імунних об'єктів у мережі. Отже робота цього оператора дозволяє скоротити кількість антитіл до тієї кількості, яка була до початку клонування. Відповідно до цього, для кожної групи, яку утворює клоноване антитіло та його клони, відбувається вибір одного імунного об'єкту, який в подальшому буде використовуватися алгоритмом. Таким чином, з кожної такої групи буде обрано по одному антитілу, що має найбільший рівень афінності до антигенів. За принципом роботи відрізняють декілька видів оператора супресії: популяційний та цільовий.

У випадку популяційної супресії кожному антитілу або клону, ознаки якого було змінено під час мутації, представляється вся популяція антигенів, тобто до кожного з існуючих антигенів цей клон визначає афінність [3-5]. З цих афінностей потім визначається авідність як середня афінність антитіла чи клону до всієї популяції антигенів. Після співставлення авідностей в кожній групі антитіла та його клонів обирається один імунний об'єкт, що має найбільший рівень авідності, який і замінює антитіло, від якого він був клонований.

У випадку цільової супресії антитілам та клонам представляється не вся популяція, а тільки їх цільові антигени [12], тобто вони визначають афінності не до всіх антигенів, а до їх невеликої кількості, що визначається на початку роботи імунного алгоритму. Головною умовою обрання імунного об'єкту, який замінить клоноване антитіло є максимальна афінність до одного з цільових антигенів, або набуття стану специфічності до одного з них.

Оператор *Termination(AB, AG)* [2-12] використовується для перевірки можливості зупинки процесу формування імунної відповіді мережі на антигени (імунного навчання). За принципом роботи відрізняють декілька видів оператора зупинки формування імунної відповіді: статичний, критеріальний, повний та комбінований.

Використання статичного оператора зупинки формування імунної відповіді використовується значення максимальної кількості поколінь антитіл, які формуються за рахунок клонування, мутації та супресії. Ця кількість поколінь визначається перед початком роботи імунного алгоритму та використовується у якості його вхідного аргументу.

Використання критеріального оператора зупинки передбачає досягнення кожним антитілом мережі стану умовної специфічності до групи цільових антигенів або до окремого антигену. Це унеможливує вплив на зупинку процесу імунного навчання, що може призвести до збільшення часу на класифікацію об'єктів.

Особливістю повного оператора зупинки імунного навчання є досягнення стану повної специфічності між кожним антитілом імунної мережі та одним з антигенів. Це також унеможливує вплив на зупинку процесу формування імунної відповіді, що призводить до збільшення часу класифікації.

Комбінований оператор зупинки імунного навчання передбачає досягнення повної або умовної специфічності між кожним антитілом мережі та одним або групою антигенів, та використання максимальної кількості поколінь імунних об'єктів у якості вхідного аргументу для алгоритму класифікації. Цей аргумент використовується у якості запобіжника від нескінченного формування популяції антитіл та збільшення часу класифікації об'єктів.

Процес визначення класів для антитіл після зупинки процесу імунного навчання передбачає розділення всієї популяції антитіл на специфічні або умовно специфічні та неспецифічні об'єкти завдяки використанню оператора *Division(AB, AB', AB'')*. Визначення класів для специфічних та умовно специфічних антитіл відбувається за рахунок роботи оператора *Specification(AB', AG)*. При цьому у випадку роботи зі специфічними антитілами – їх належності до класів визначається класами антигенів, до яких вони набули специфічності. У випадку роботи з умовно специфічними антитілами, клас для антитіла визначається класом антигену, до якого це антитіло має найбільшу афінність. Неспецифічні антитіла класифікуються під час виконання оператора *Selection(AB'', AB', AG)*. При цьому для кожного антитіла, яке не набуло повної або умовної специфічності до одного з антигенів визначається перелік цільових об'єктів серед яких використовуються не тільки антигени, але й класифіковані антитіла, що набули стану повної або умовної специфічності до антигенів. Слід зазначити, що на даному етапі немає необхідності у додатковому обчисленні афінностей між імунними об'єктами через те, що афінності між антитілами були визначені на етапі визначення значення NAT ще до запуску процесу формування імунної відповіді мережею антитіл, а визначення афінностей між антитілами та антигенами відбулося під час презентації антигенів антитілам на початку роботи імунного алгоритму.

На етапі *Selection(AB'', AB', AG)* для кожного некласифікованого антитіла відбувається пошук класу по найбільшій середній афінності до цільових антигенів та класифікованих антитіл, що набули стани умовної або повної специфічності. Відповідно до цього для кожного неспецифічного антитіла відбувається пошук значення середньої афінності до кожного з класів, які представлені цільовими об'єктами. Таким чином некласифіковане антитіло буде належати до того класу, до якого воно має максимальну середню афінність, при цьому цей клас представляється переліком цільових об'єктів антитіла, що не набуло специфічності під час процесу формування імунної відповіді.

Випробування методів класифікації на основі імунного алгоритму aiNET та дослідження імунних операторів проводилося для мережі з популяцією в

1000 антигін рівномірно розподілених у тривимірному просторі ознак $1000 \times 1000 \times 1000$ та навчальною вибіркою в 250 антигенів, що були розподілені між 4 класами. Задля перевірки якості класифікації використовується інформація щодо відношення цих антигін до початкових класів, яка використовується виключно на етапі формування висновків про помилки класифікації імунного алгоритму. Слід зазначити, що об'єктом дослідження були імунні оператори, які використовуються алгоритмом класифікації під час формування імунної відповіді на антигени. Тому частина адаптованого для класифікації методу aiNET з операторами визначення класів для антигін була незмінною. Для оцінювання впливу імунних операторів на роботу алгоритму класифікації використовувалися наступні показники: кількість популяцій, що необхідна для досягнення умови припинення формування імунної відповіді – *Populations (P)*, час витрачений на опрацювання популяцій антигін, які створюються при формуванні імунної відповіді – *Time (T)*, відсоток антигін, які набули стану повної або умовної специфічності до антигенів – *Specificity (S)*, та помилки класифікації *Errors (E)* – відсоток антигін, що було помилково класифіковано.

Особливості імунних операторів класичного алгоритму aiNET наведені у табл. 1.

Таблиця 1 – Імунні оператори алгоритму aiNET

Оператор	Особливість
<i>Presentation(AB, AG)</i>	Параметр L відсутній
<i>Cloning(AB, CL)</i>	Статичне клонування, 100 клонів
<i>Mutation(CL)</i>	Статична мутація, коефіцієнт $\mu = 0,1$
<i>Supression(AB, CL)</i>	Популяційна супресія
<i>Termination(AB, AG)</i>	Статичний, формування 100 популяцій антигін

Відповідно до особливостей імунних операторів, наведених у табл. 1 та основних показників, за якими оцінюються модифікації aiNET для класифікації, було проаналізовано результати формування імунної відповіді та класифікації популяції антигін, які наведені у табл. 2.

Таблиця 2 – Оцінка ефективності алгоритму aiNET

Алгоритм	P	T	S	E
aiNET	100	100%	28%	32%

Слід зазначити, що оскільки не модифікований алгоритм aiNET витрачає, порівняно з модифікаціями, які наведені нижче, найбільше часу на формування імунної відповіді його значення Time використовується як 100%. Таким чином, час формування імунної відповіді всіма іншими модифікаціями aiNET буде порівнюватися із цим значенням.

Задля пришвидшення формування імунної відповіді, слід змінити роботу оператора супресії з популяційної на використання цільових антигенів. Разом з цим можна дослідити вплив параметру L на якість роботи імунного алгоритму. Модифікації ал-

горитму aiNET, які використовують під час супресії цільові антигени та різні значення параметру L наведені у табл. 3.

Таблиця 3 – Модифікації алгоритму aiNET із супресією до цільових антигенів

Алгоритм	Значення L
aiNET-ts1	L = 1,0
aiNET-ts2	L = 0,75
aiNET-ts3	L = 0,5
aiNET-ts4	L = 0,35
aiNET-ts5	L = 0,2
aiNET-ts6	L = 0,1
aiNET-ts7	L = 0,05

За результатами формування імунної відповіді та класифікації популяції антигін модифікаціями aiNET із супресією до цільових антигенів та різним значенням параметру L, наведених у табл. 4, можна зробити висновок, що використання оператора супресії на основі цільових антигенів, замість оператора популяційної супресії, призводить до значного пришвидшення роботи алгоритму.

Таблиця 4 – Оцінка ефективності алгоритмів aiNET із супресією до цільових антигенів

Алгоритм	P	T	S	E
aiNET-ts1	100	100%	28%	32%
aiNET-ts2	100	70%	29%	32%
aiNET-ts3	100	40%	28%	31%
aiNET-ts4	100	23%	29%	32%
aiNET-ts5	100	11%	22%	36%
aiNET-ts6	100	4%	18%	39%
aiNET-ts7	100	0,1%	16%	39%

Крім того, з табл. 4 видно, що зменшення параметру L призводить до суттєвого зниження відсотка формування специфічних антигін та зростання помилки класифікації. Тому в інших модифікаціях aiNET буде використовуватися супресія до цільових антигенів із значенням L в діапазоні (0,35; 0,5).

Задля дослідження впливу оператора клонування на якість класифікації було створено декілька модифікацій aiNET, наведених у табл. 5.

Таблиця 5 – Модифікації алгоритму aiNET із різними операторами клонування

Алгоритм	L	Клонування
aiNET-ts3-sc100	L = 0,5	Статичне, 100 клонів
aiNET-ts3-sc200	L = 0,5	Статичне, 200 клонів
aiNET-ts3-sc300	L = 0,5	Статичне, 300 клонів
aiNET-ts3-sc400	L = 0,5	Статичне, 400 клонів
aiNET-ts3-scp	L = 0,5	Пропорційне
aiNET-ts4-sc100	L = 0,35	Статичне, 100 клонів
aiNET-ts4-sc200	L = 0,35	Статичне, 200 клонів
aiNET-ts4-sc300	L = 0,35	Статичне, 300 клонів
aiNET-ts4-sc400	L = 0,35	Статичне, 400 клонів
aiNET-ts4-scp	L = 0,35	Пропорційне

Відповідно до результатів класифікації визначеними алгоритмами, наведеними у табл. 6 можна зробити висновок, що використання пропорційного клонування призводить до зниження точності класифікації та поступається операторам статичного клонування, які формують від 100 до 200 клонів для кожного антитіла, яке не набуло стану умовної або повної специфічності до одного з антигенів.

Таблиця 6 – Оцінка ефективності алгоритмів aiNET із різними видами клонування

Алгоритм	P	T	S	E
aiNET-ts3-sc100	100	40%	28%	31%
aiNET-ts3-sc200	100	45%	29%	31%
aiNET-ts3-sc300	100	52%	31%	30%
aiNET-ts3-sc400	100	60%	33%	30%
aiNET-ts3-scp	100	26%	17%	46%
aiNET-ts4-sc100	100	23%	29%	32%
aiNET-ts4-sc200	100	29%	30%	30%
aiNET-ts4-sc300	100	34%	32%	30%
aiNET-ts4-sc400	100	46%	34%	29%
aiNET-ts4-scp	100	14%	19%	41%

Задля дослідження впливу оператора мутації на швидкість процесу формування імунної відповіді під час класифікації було створено декілька модифікацій aiNET, наведених у табл. 7.

Слід зазначити, що всі модифіковані алгоритми використовують супресію до цільових антигенів, мають параметр звуження пошуку цільових об'єктів $L = 0,35$ та використовують статичне клонування, під час якого для кожного антитіла формується по 100 клонів.

Також всі модифіковані алгоритми формують по 100 популяцій імунних об'єктів перед початком процесу визначення класів для всіх антитіл, що не набули стану повної або умовної специфічності.

Таблиця 7 – Модифікації алгоритму aiNET із різними операторами мутації

Алгоритм	Мутація
aiNET-ts4-sc100-sm1	Статична, $\mu = 0,1$
aiNET-ts4-sc100-sm2	Статична, $\mu = 0,2$
aiNET-ts4-sc100-sm3	Пропорційна
aiNET-ts4-sc100-sm4	Зворотно-пропорційна

Відповідно до результатів класифікації алгоритмами, наведеними у табл. 7, можна зробити висновок, що використання пропорційної мутації призводить до зниження якості класифікації, порівняно з результатами, отриманими при використанні операторів мутації інших видів.

Також слід зазначити, що використання зворотно-пропорційної мутації призводить до підвищення якості результатів роботи імунного алгоритму майже за всіма показниками.

Задля дослідження впливу налаштувань оператора зупинки процесу формування імунної відповіді

було створено декілька модифікацій aiNET, наведених у табл. 9, які використовують супресію до цільових антигенів, мають параметр звуження пошуку цільових об'єктів $L = 0,35$ та використовують статичне клонування (100 клонів) та зворотно-пропорційну мутацію.

Таблиця 8 – Оцінка ефективності алгоритмів aiNET із різними видами мутації

Алгоритм	P	T	S	E
aiNET-ts4-sc100-sm1	100	23%	29%	32%
aiNET-ts4-sc100-sm2	100	23%	31%	31%
aiNET-ts4-sc100-sm3	100	23%	28%	32%
aiNET-ts4-sc100-sm4	100	24%	39%	26%

Таблиця 9 – Модифікації алгоритму aiNET із різними операторами зупинки процесу формування імунної відповіді

Алгоритм	Умови зупинки
aiNET-ts4-sc100-sm4-st100	Обробка 100 популяцій
aiNET-ts4-sc100-sm4-tt	Повна специфічність
aiNET-ts4-sc100-sm4-ct	Специфічність 75% антитіл
aiNET-ts4-sc100-sm4-kt250	Специфічність 75% антитіл або обробка 250 популяцій

Згідно з результатами вказаних алгоритмів класифікації, що наведені у табл. 10, можна зробити висновок, що використання умови повної специфічності для зупинки процесу формування імунної відповіді призводить до необхідності формування надвеликої кількості популяцій імунних об'єктів, що негативно впливає на швидкість класифікації. Використання умови досягнення умовної специфічності всіма антитілами мережі також негативно впливає на швидкість роботи імунного алгоритму, але забезпечує високу точність класифікації. Найбільш збалансованим варіантом оператора зупинки формування імунної відповіді є комбінований варіант, коли умова досягнення умовної специфічності поєднується із використанням максимальної кількості популяцій антитіл.

Таблиця 10 – Оцінка ефективності алгоритмів із різними операторами зупинки процесу формування імунної відповіді

Алгоритм	P	T	S	E
aiNET-ts4-sc100-sm4-st100	100	24%	39%	26%
aiNET-ts4-sc100-sm4-tt	1875	443%	100%	1%
aiNET-ts4-sc100-sm4-ct	1053	251%	100%	4%
aiNET-ts4-sc100-sm4-kt250	250	59%	73%	11%

Висновки

В результаті проведених досліджень виявлено оптимальну комбінацію імунних операторів, яка забезпечує високу швидкість формування імунної відповіді для алгоритмів, що функціонують на основі моделі штучної імунної мережі та реалізуються на

основі поширеного алгоритму aiNET. Згідно з результатами випробувань, найбільшу ефективність показав алгоритм aiNET-ts4-sc100-sm4-kt250, який використовує супрецію до цільових антигенів, має параметр звуження багатовимірного простору L для пошуку цільових об'єктів в діапазоні $(0,35; 0,5)$, використовує статичне клонування та зворотно-пропорційну мутацію, а також має комбінований оператор зупинки процесу формування імунної відповіді на основі умовної специфічності та максимальної кількості популяцій антитіл.

У цій роботі запропоновано використання цільових імунних об'єктів для визначення класів ан-

титіл, використання поняття умовної специфічності між антитілами та антигенами, а також використання критерію звуження простору пошуку цільових об'єктів для класифікації.

Даний метод може бути використаний для вирішення задач класифікації з неконтрольованим навчанням, а також задач прогнозування, розпізнавання образів та керування персонажами в комп'ютерних та мобільних іграх.

Розроблений модифікований метод aiNET-ts4-sc100-sm4-kt250 планується використовувати у подальшому для керування персонажами в комп'ютерних іграх жанру action.

СПИСОК ЛІТЕРАТУРИ

1. R.O. Duda, P.E. Hart, D.G. (2010), Stork, Pattern classification. Wiley & Sons.
2. D. Dasgupta, L.F. Nino. (2009), Immunological computation, Theory, and applications. Taylor & Francis Group.
3. D. Dasgupta, S. Yu, L.F. Nino. (2011), Recent Advanced in Artificial Immune Systems: Models and Applications. Applied Soft Computing, Elsevier P. 1574-1587.
4. M. Read, P.S. Andrews, J.Timms. (2012), An Introduction to Artificial Immune Systems. In Handbook of Natural Computing, Shringar, Berlin, Germany P. 1575-1597.
5. K.B. Bahekar. (2020), Classification techniques based on Artificial immune system algorithms for Heart disease using Principal Component Analysis. International Journal of Scientific Research in Science, Engineering, and Technology, IJSRSET, Vol. 7, Iss. 5, P. 150-160.
6. S. Shekhar, D.K. Sharma, D.K. Agarwal, Y. Pathak. (2022), Artificial Immune Systems-Based Classification Model for Code-Mixed Social Media Data. IRBM, Vol. 43, Iss. 2, P. 120-129.
7. R.M. Mikherskii, M.R. Mikherskii. (2021), Analysis of the Use of Artificial Immune Systems. In: IOP Conference Series: Materials Science and Engineering, P 1-6.
8. H. Park, J.E. Choi, D. Kim, S.J. Hong. (2021), Artificial immune system for fault detection and classification of semiconductor equipment. Electronics, Vol. 10, No. 8, 944, P.1-14.
9. S.S.F. Souza, F.P.A. Lima, F.R. Chavarette. (2020), A New Artificial Immune System Based on Continuous Learning for Pattern Recognition. Revista de Informatica Teorica e Aplicada, RITA. Vol. 27, No. 04, P. 34-44.
10. A.T. Haouari, L. Souici-Meslati, F. Atil, D. Meslati. (2020), Empirical comparison and evaluation of artificial immune systems in inter-release software fault prediction. Applied Soft Computing Journal, 96, P. 1-18
11. V. Cutello, G. Nicosia. (2022), Multiple learning using immune algorithms. In: Proceedings of 4th International Conference on Recent Advances in Soft Computing, RASC, P. 102-107.
12. M. Korablyov, O. Fomichov, N. Axak. (2021), Classification of objects based on a tree-shaped artificial immune network model. Advances in Intelligent Systems and Computing V, Springer, P. 160-172.
13. C. Lan, H. Zhang, X. Sun, Z. Ren. (2020), An intelligent diagnostic method based on optimizing B-cell pool clonal selection classification algorithm. Turkish Journal of Electrical Engineering and Computer Sciences, 28 P. 3270-3284.

Received (Надійшла) 27.03.2023

Accepted for publication (Прийнята до друку) 24.05.2023

The research of immune operators in the artificial immune network model

Oleksandr Fomichov, Vladyslav Burtsev

Abstract. Topicality. Models and methods that allow intelligent processing and systematization of data are of considerable relevance in the field of information technologies. Among such models, one can single out models that work on the basis of biological principles of computing organization, such as artificial neural networks, genetic algorithms, and artificial immune systems. Today, there are several types of models of artificial immune systems that use specific immune operators that affect the speed of the model and the quality of solving practical problems. **The goal of this work** is a study of the influence of the organization of the work of the most common immune operators on the speed of work of immune algorithms and the accuracy of decision-making regarding the classification of objects with supervised learning. **The object of research** are epy immune operators of cloning, mutation and suppression, as well as an assessment of the speed of the process of forming an immune response of a network of antibodies to a population of antigens. **The subject of research** is the model of an artificial immune network and the aiNET algorithm, immune operators and their influence on the process of forming an immune response for the classification of objects with supervised learning. **Results.** In this paper, the optimal setting of immune operators is proposed to ensure a high speed of immune response formation by the aiNET algorithm when solving the classification problem. In addition, it is proposed to use target immune objects to speed up the classification of antibodies that have not acquired the state of specificity to the antigens of the training sample. **Conclusions.** The modified aiNET method of classification with supervised learning is planned to be used in the future to control the behavior of characters in action game applications.

Keywords: artificial immune systems, antibody, antigen, affinity, clonal selection, suppression, complete specificity, conditional specificity, immune network suppression.

С. Ф. Чалий, В. О. Лещинський

Харківський національний університет радіоелектроніки, Харків, Україна

ОЦІНКА ЧУТЛИВОСТІ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНІЙ ІНФОРМАЦІЙНІЙ СИСТЕМІ

Анотація. Предметом вивчення в статті є процеси побудови пояснень щодо отриманих рішень в інтелектуальній інформаційній системі. **Метою** є оцінка чутливості пояснень на основі аналізу властивостей вхідних даних та відповідних рішень в інтелектуальній інформаційній системі для підтримки вибору найкращого пояснення з позицій задоволення інтересів користувача. **Завдання:** структуризація критеріїв кількісної оцінки пояснень при представленні інтелектуальної системи у вигляді чорного ящика; розробка методу оцінки чутливості пояснень в інтелектуальній інформаційній системі. Використовуваними **підходами** є: підходи до побудови пояснень, підходи оцінки пояснень в інтелектуальних інформаційних системах. Отримані наступні **результати**. Структуровано критерії оцінки пояснень для інтелектуальних систем, представлених згідно принципу чорного ящика. Вказані критерії враховують вплив на пояснення вхідних та вихідних даних інтелектуальної системи, відповідність пояснення процесу прийняття рішення в інтелектуальній системі, а також відповідність пояснення і розуміння результатів інтелектуальної системи користувачем. На базі виконаної структуризації запропоновано метод оцінки чутливості пояснень для інтелектуальної системи, представленої згідно принципу чорного ящика. **Висновки.** Наукова новизна отриманих результатів полягає в наступному. Запропоновано метод оцінки чутливості пояснень для інтелектуальної системи, представленої згідно принципу чорного ящика. Метод містить етапи, пов'язані із перевіркою та визначенням схожості вхідних даних і результатів для альтернативних моделей інтелектуальних систем за кількісними та якісними показниками, а також кількісної оцінки вхідних даних та визначення чутливості пояснення. Запропонований метод дає можливість порівняти та вибрати пояснення з урахуванням властивостей та важливості вхідних даних з тим, щоб визначити можливість застосування альтернативних підходів до побудови пояснень щодо результатів інтелектуальної інформаційної системи. Подальший розвиток запропонованого підходу орієнтований на визначення і реалізацію метрик оцінки точності і прозорості пояснень.

Ключові слова: інтелектуальна система, пояснення, процес прийняття рішення, причинно-наслідковий зв'язок, оцінка пояснень.

Вступ

Важливість використання пояснень щодо отриманих в інтелектуальних системах рішень пов'язана із суттєвим підвищенням складності таких систем і використанням складних, непрозорих для користувача алгоритмів при отриманні результатів в цих системах.

Незрозумілість алгоритмів прийняття рішень для користувача може привести до недовіри щодо отриманих результатів і, як наслідок, до невикористання або затримок використання запропонованої інтелектуальною системою рішення. Така невідповідність на практиці може приводити до суттєвих збитків. Наприклад, недовіра користувача до результатів розпізнавання об'єктів в охоронній системі знижує безпекові можливості відповідного підприємства, недовіра до рекомендацій в системі електронної комерції знижує кількість та суму покупок товарів або послуг, тощо [1-3]. Для вирішення цієї проблеми використовується механізм пояснення отриманих в інтелектуальній системі результатів [4]. Пояснення дає можливість користувачеві визначити причинно-наслідкові залежності між вхідними даними і результатом роботи інтелектуальної системи, або ж між окремими діями і отриманими проміжними результатами [5].

В першому випадку при побудові пояснень робота інтелектуальної системи представляється як задача, вирішення якої має вхідні дані і результат, й не представлено проміжними станами. Тобто інтелектуальна система розглядається згідно принципу чорного ящика.

В другому випадку функціонування системи розглядається як процес, який містить проміжні стани. Кожен стан є результатом виконання однієї з дій процесу прийняття рішення. Тобто в другому випадку функціонування інтелектуальної інформаційної системи розглядається на різних рівнях деталізації, у відповідності до принципу білого ящика.

Згідно принципів білого та чорного ящиків можна виділити два напрямки побудови пояснень.

За принципом білого ящика, пояснення є безпосередньо однією із функцій інтелектуальної системи. В даному випадку для функції пояснення є доступними всі внутрішні дані системи, що забезпечують формування рішення. Перевага підходу полягає у можливості формування точних пояснень на основі повного набору внутрішніх даних. Суттєвим обмеженням даного підходу є необхідність включення функціональності пояснення безпосередньо на етапі проектування інтелектуальної системи. Доповнення існуючої системи вказаною функціональністю потребує значних витрат на її перепроєктування.

Побудова пояснень за принципом чорного ящика передбачає розробку окремої підсистеми, яка використовує вхідні дані, результат і, можливо, доступні проміжні дані для формування пояснень. На базі цих даних для користувача формуються відповідні каузальні залежності.

Основні методи побудови пояснень розроблялись згідно програми XAI (Explainable Artificial intelligence), запропонованої DARPA [6]. Ключова увага за цією програмою приділялась визначенню зрозумілих для користувача причинно-наслідкових зв'язків, що

відображають залежності між входами та виходом інтелектуальної системи. Однак проблемі порівняння та обґрунтованого вибору методу побудови пояснень при представленні інтелектуальної системи у вигляді чорного ящика не приділялось достатньо уваги. Для практичного застосування методів побудови пояснень важливо мати критерій їх порівняння. Такий критерій має давати можливість порівняти альтернативних підходів та визначити, який метод точніше визначає залежності для пояснень між характеристиками вхідних даних та отриманим в інтелектуальній системі кінцевим або проміжними результатами.

Таким чином, актуальною задачею при побудові пояснень є формування та використання критеріїв для порівняння їх ефективності при однакових вхідних даних, а також отриманих в інтелектуальній інформаційній системі рішеннях.

Однак існуючі підходи до оцінки ефективності пояснень в інтелектуальних системах орієнтовані в першу чергу на якісне їх порівняння з точки зору людей-користувачів [7]. Також, з використанням принципу білого ящика, використовуються окремі, локальні підходи до перевірки залежностей між властивостями вхідних даних та отриманим в інтелектуальній системі рішенням [8], враховуються особливості моделі прийняття рішень та відповідні властивості вхідних даних інтелектуальної [9].

Таким чином, обґрунтування вибору методу побудови пояснень на основі кількісного критерію оцінки їх ефективності є актуальною задачею.

Такий критерій має враховувати узагальнені властивості пояснень, які забезпечують прозорість останніх для користувача. Ці властивості в першу чергу мають враховувати відповідність пояснення набору вхідних даних, тобто чутливість пояснення до помилок і змін в таких даних. Також необхідно врахувати зрозумілість отриманого пояснення для користувача, та відповідність пояснення отриманим результатам.

Таблиця 1 – Відмінності базових властивостей пояснень

Властивість	Особливості	Якісна/кількісна оцінка
Чутливість до змін	Уточнення пояснення відповідно до розвитку інтелектуальної системи, нечутливість до випадкових змін у вхідних даних внаслідок зовнішніх впливів	Якісна та кількісна (з урахуванням властивостей даних та результатів)
Точність	Відповідність пояснення реальному процесу функціонування інтелектуальної системи: каузальні залежності пояснення мають відповідати каузальним залежностям процесу прийняття рішення	Кількісна
Прозорість представлення користувачеві	Представлення з мінімальною складністю, з урахуванням обмежень сприйняття людини-користувача	Кількісна, за представленими елементами

Згідно підходу «чорного ящика», пояснення *Expl* в інтелектуальній системі представляється як функція, яка задає каузальні залежності між заданим набором вхідних даних $V = \{v_i\}$ та відповіддю системи R за умови, що ці дані характеризуються їх якісними властивостями a_i та кількісними властивостями w_i :

$$Expl = f(V, R). \quad (1)$$

$$v_i = (a_i, w_i). \quad (2)$$

Метою статті є оцінка чутливості пояснень на основі аналізу властивостей вхідних даних та відповідних рішень в інтелектуальній інформаційній системі для підтримки вибору найкращого пояснення з позиції задоволення інтересів користувача.

Для досягнення поставленої мети вирішуються такі задачі:

- структуризація критеріїв кількісної оцінки пояснень при представленні інтелектуальної системи у вигляді чорного ящика;
- розробка методу оцінки чутливості пояснень в інтелектуальній інформаційній системі, яка представлена як чорний ящик.

Критерії оцінки пояснень

Для реалізації кількісної оцінки пояснень доцільно врахувати такі їх загальні властивості:

- чутливість до змін;
- точність;
- прозорість представлення користувачеві.

Обґрунтування вибору вказаних властивостей полягає в наступному. Відповідно до першої властивості, пояснення має, з одного боку, змінюватись при зміні процесу функціонування інтелектуальної системи, а з іншого -бути нечутливим до випадкових або раптових змін у вхідних даних, що не впливають на кінцевий результат роботи системи [10].

Згідно другої властивості, пояснення має максимально точно відобразити процес роботи інтелектуальної системи, у загальному вигляді представленій як чорний ящик.

Третя властивість враховує відомі обмеження оперативної пам'яті людини, яка у більшості випадків забезпечує ефективне оперування сімома предметами одночасно. Тому пояснення має бути мінімальним з позицій кількості об'єктів, з якими воно оперує.

Перелік особливостей наведених властивостей пояснень представлено в табл. 1.

Таке визначення пояснення як функції показує, що пояснення може бути представлено у вигляді залежності між властивостями елементів набору вхідних даних v_i та отриманим в інтелектуальній системі результатом. Вхідні дані можуть мати як якісні, так і кількісні властивості. Наприклад, при якісній оцінці ми можемо виділити підмножину вхідних даних, яка визначає результат роботи інтелектуальної системи. Кількісна оцінка задає вплив цих даних на результат через їх вагу. Іншими словами, кількісна оцінка може бути пов'язана з якісною через апріорно задане

порогове значення ваги. Наприклад, вплив на результат елемента вхідних даних вважається суттєвим у випадку, якщо його вага перевищує задане порогове значення w^* . Таким чином, якісна оцінка може бути отримана на основі кількісної оцінки властивостей вхідних даних. Для вибору пояснень нам необхідно оцінити їх кількісно й якісно. При такій оцінці необхідно

враховувати суб'єктивний та об'єктивний аспекти пояснень. Суб'єктивний аспект пояснення пов'язаний із якісною оцінкою. По-перше, необхідно отримати оцінку задоволеності користувачів цим поясненням. А по друге - ступінь розуміння пояснення користувачами. Структуризацію якісного аспекту пояснень представлено в табл. 2.

Таблиця 2 – Якісна оцінка пояснень

Характеристика	Особливості	Можливість переходу до кількісної оцінки
Задоволеність користувачів	Фактично відображає ступінь довіри користувачів до результатів інтелектуальної інформаційної системи	Через підтвердження вибору запропонованого системою рішення, наприклад рекомендованого товару або послуги
Розуміння	Відображення процесу прийняття рішення в інтелектуальній системі у вигляді множини причинно-наслідкових зв'язків	Через аналіз текстів відгуків на результати використання системи (за наявності), через анкетування

Однак якісна характеристика не дає можливості порівнювати пояснення між собою, оскільки при виборі краще пояснення повинно мати кращу кількісну оцінку. Можливість вибору забезпечує удосконалення підходів до побудови пояснень. Таким чином, об'єктивна характеристика пояснення базується на його кількісній оцінці. Наявність такої оцінки дає можливість перевірити, чи відповідає процес пояснення заданим значенням узагальнених властивостей чутливості до змін, точності та прозорості представлення користувачеві. Властивість точності дає можливість при побудові пояснення визначити підмножину вхідних даних, які значною мірою впливають на отриманий в інтелектуальній системі результат. Тоді встановлення залежності між цією підмножиною властивостей та рішенням системи забезпечує умови для встановлення причинно-наслідкового зв'язку як основи пояснення. Альтернативний підхід полягає у визначенні зв'язку між зваженою сумою властивостей та отриманим результатом. В подальшому, на основі уточнення ваг, можна виділити підмножину даних, що мають властивості з найбільшими значеннями ваг. Такі дані є визначальними для отримання рішення в інтелектуальній системі. Таким чином, з використанням результатів даного підходу також можна отримати підмножину вхідних даних, що суттєво впливають на результат системи.

Метод оцінки чутливості пояснень

Визначимо кількісну оцінку чутливості пояснення на основі такого підходу.

Нехай маємо дві схожі моделі роботи інтелектуальної системи M_1 та M_2 , які вирішують одну й ту ж задачу. При використанні цими моделями схожих вхідних даних V_1 та V_2 й отриманні схожих або однакових результатів R_1 та R_2 пояснення для користувача $Expl_1$ та $Expl_2$ мають бути схожими:

$$Expl_1 \approx Expl_2 | V_1 \approx V_2 \Rightarrow R_1 = R_2 = R. \quad (3)$$

Тоді кількісна чутливість пояснення S визначається через порівняння двох схожих (або однакових) пояснень $S(Expl_1, Expl_2)$, для яких виконується умова (3), з використанням метрики D :

$$S(Expl_1, Expl_2) = D(M_1(V_1, R), M_2(V_2, R)). \quad (4)$$

Схожість між вхідними даними визначимо через врахування відмінності між нормованими значеннями ваг $w_{1,i}$ та $w_{2,i}$ відповідних властивостей вхідних даних $v_{1,i}$ та $v_{2,i}$:

$$\Delta w_i = |w_{1,i} - w_{2,i}|, 0 \leq w_{k,i} \leq 1. \quad (5)$$

Відповідно до (5), показник ΔV схожості/відмінності вхідних даних для двох схожих моделей M_1 та M_2 прийняття рішень в інтелектуальній системі можна визначити як усереднену різницю між вагами властивостей вхідних даних за умови, що ці дані є схожими в якісному і кількісному вимірах:

$$\Delta V_{1,2} = \frac{1}{I} \sum_{i=1}^I \Delta w_i | (\forall i) \exists a_{1,i}, \exists a_{2,i}. \quad (6)$$

Схожість у якісному вимірі полягає в тому, що в обох випадках використовуються однакові набори вхідних даних:

$$(\forall i) a_{1,i} = a_{2,i}. \quad (7)$$

Згідно виразу (7), властивості v_i можуть, наприклад, мати однакові назви a_i . Схожість у кількісному вимірі полягає у незначній, в межах заданого порогу ε , відмінності між вагами вхідних даних:

$$(\forall i) \Delta w_i \leq \varepsilon. \quad (8)$$

Наприклад, в рекомендаційній системі в якості вхідних даних може бути використана інформація про покупки товарів, про їх рейтинги, про дату покупок та виставлення рейтингу товарів, про групи товарів, про їх популярність, демографічна інформація щодо користувача, тощо. Очевидно, що ми не можемо порівнювати пояснення для систем, перша з яких використовує для побудови рекомендацій інформацію про покупки, а інша – про рейтинги. Така ситуація свідчить про невиконання умови (7) щодо схожості вхідних даних у якісному вимірі.

Також важко порівнювати пояснення у системах, перша з яких враховує в першу чергу дані про товари, а друга – дані про користувачів (з відповід-

ною вагою), оскільки може не виконуватись умова (8). Тобто в першій системі будуть мінімальні ваги даних щодо користувачів і максимальні – щодо товарів, а в другій – навпаки. Схожість результатів $R = \{r_j\}$, отриманих в інтелектуальній системі, обчислюється лише у якісному вимірі аналогічно (7):

$$(R_1 = R_2) \equiv (\forall i) r_{1,i} = r_{2,i}. \quad (9)$$

Кількісний вимір елементів рішення є необов'язковим. Наприклад, він може бути відсутнім при вирішенні задачі бінарної класифікації зображень. У випадку, якщо рішення містить кількісну складову ω_i , то визначення еквівалентності рішень є аналогічним (8):

$$R_1 = R_2 \left| (\forall j) r_{1,j} = r_{2,j}, \Delta\omega_j \leq \varepsilon. \quad (10)$$

Якщо ж відхилення кількісної складової перевищує поріг ε , то відмінність рішень ΔR є такою:

$$\Delta R = \sum_{j=1}^J \Delta\omega_j. \quad (11)$$

У відповідності до виразів (4) - (11), відхилення вхідних даних та результату є нормованими і тому чутливість пояснення може бути розрахована через співвідношення відхилень результату роботи інтелектуальної системи та її вхідних даних за умови однакових або дуже близьких пояснень $Expl_1 = Expl_2$:

$$S_2^{(1)} = \left| \frac{\Delta V_1}{\Delta R_1} - \frac{\Delta V_2}{\Delta R_2} \right|. \quad (12)$$

При відсутності кількісної складової рішення згідно (10) вираз (12) спрощується і приймає вигляд:

$$S_2^{(1)} = |\Delta V_1 - \Delta V_2|. \quad (13)$$

Таким чином, для розрахунку чутливості пояснення необхідно визначити відмінності між складовими рішення та між вхідними даними за умови еквівалентності або схожості пояснень. Перші можуть бути отримані з урахуванням доступної із рішення кількісної складової. Для того, щоб отримати кількісну складову для вхідних даних, необхідно знайти латентні фактори, що пов'язують вхідні дані та отримані результати. Постановка цієї задачі має такий вигляд.

Дано розріджену матрицю розміру $I \times J \times N$, де N - кількість значень вхідних даних. В інших вимірах матриця задає елементи вхідних даних, та елементи результату, отриманого в інтелектуальній системі. Оскільки система представлена чорним ящиком, то відомо лише, який результат був для заданих значень вхідних даних.

Ілюстрація даної задачі для одного елементу вхідних даних і одного елементу результату представлена на рис. 1. Стівпчики матриці задають можливі значення результату, а рядки – можливі значення вхідних даних. Хрестики задають відповідність – для яких вхідних даних був отриманий результат.

Потрібно знайти ваги впливу вхідних даних на рішення системи. Наприклад, для рекомендаційної системи на вході маємо перелік користувачів, а на

виході – рекомендацію з одного товару. Маємо знайти і представити у кількісному вимірі зв'язок між користувачем та рекомендованим товаром.

Задача у даній інтерпретації є аналогічною задачі побудови рекомендацій і може бути вирішена відповідними методами. Так, виявлення латентних факторів забезпечується на основі використання методу колаборативної фільтрації. Тому для побудови пояснень на одному із етапів методу оцінки чутливості пояснень доцільно використати колаборативну фільтрацію.

		Результат роботи системи		
		(один елемент)		
		Значення	...	Значення
		1		J
Значення вхідних даних	1			
				X
			X	
	...			
	X			
			X	
I				

Рис. 1. Вхідні залежності для виявлення ваг вхідних даних для оцінки пояснення

Запропонований метод оцінки чутливості містить у собі такі етапи.

Етап 1. Перевірка схожості вхідних даних за якісним критерієм (7). У випадку несхожості виконання методу завершується, оскільки невідповідність вхідних даних не дає можливість оцінити пояснення.

Етап 2. Перевірка схожості результатів за якісним критерієм (9). У випадку несхожості виконання методу завершується.

Етап 3. Визначення схожості результатів за кількісним критерієм (11). Етап виконується за наявності кількісних даних щодо результату роботи інтелектуальної системи.

Етап 4. Формування матриці вхідних даних для методу колаборативної фільтрації. Матриця формується аналогічно прикладу на рис. 1, але має додаткові виміри, які відображають множину вхідних даних та складових результату.

Етап 5. Виявлення ваг вхідних даних за допомогою колаборативної фільтрації.

Етап 6. Визначення чутливості пояснень згідно (12) у випадку кількісної складової результату або згідно (13) у випадку відсутності такої складової.

Експериментальну перевірку методу виконано на наборі даних, що містить інформацію про продажі подарункових товарів у мережі супермаркетів Лондона. Пояснення були сформовані на основі темпоральних залежностей, що відображають динаміку користувацького попиту на вказані товари.

Були відібрані товари з однаковими поясненнями (ріст попиту та продажів у визначений період). Використовувалась інформація про продажі товарів, групована по тижням за період. Чутливість становила від 72 до 79 відсотків в залежності від періоду часу

покупок товарів. Аналіз ваг показав, що ключовим вхідним параметром є період часу.

Таким чином, експериментальна перевірка показала можливість обґрунтованого переходу від кількісної до якісної оцінки чутливості пояснення в інтелектуальній системі.

Висновки

Структуровано критерії оцінки пояснень для інтелектуальних систем, представлених згідно принципу чорного ящика. Ці критерії враховують можливість побудови релевантного пояснення на основі вхідних та вихідних даних інтелектуальної системи (чутливість), відповідність пояснення і процесу роботи інтелектуальної системи (точність), а також ступінь сприйняття пояснення користувачем (прозорість). Сукупність цих критеріїв дає можливість комплексно оцінити весь процес побудови пояснення, що створює умови для раціонального вибору пояс-

нення з урахуванням вхідних даних, процесу прийняття рішення в інтелектуальній системі, а також вимог користувача.

Запропоновано метод оцінки чутливості пояснень для інтелектуальної системи, представленої згідно принципу чорного ящика. Метод містить етапи перевірки та визначення схожості вхідних даних і результатів для альтернативних моделей інтелектуальних систем, а також кількісної оцінки вхідних даних і, на цій основі, визначення чутливості пояснення.

В практичному плані запропонований метод дає можливість порівняти пояснення з урахуванням властивостей вхідних даних і, на цій основі, визначити сферу застосування альтернативних підходів до побудови пояснень щодо роботи інтелектуальної інформаційної системи.

Подальший розвиток запропонованого підходу пов'язаний із визначенням метрик для оцінки точності і прозорості пояснень для користувача.

REFERENCES

1. Castelvechi D. (2016), "Can we open the black box of AI?" *Nature*, Vol. 538 (7623), pp. 20-23.
2. Tintarev N., Masthoff J. (2012), "Evaluating the effectiveness of explanations for recommender systems", *User Model User-Adapt Inter.*, Vol. 22, pp. 399–439, <https://doi.org/10.1007/s11257-011-9117-5>.
3. Miller T. (2019), "Explanation in artificial intelligence: Insights from the social sciences", *Artificial Intelligence*, vol. 267, pp.1-38, DOI: <https://doi.org/10.1016/j.artint.2018.07.007>
4. Adadi, A., Berrada, M. (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6, 52138–52160.
5. Chalyi, S., Leshchynskiy, V. (2020). Method of constructing explanations for recommender systems based on the temporal dynamics of user preferences. *EUREKA: Physics and Engineering*, 3, 43-50. doi: 10.21303/2461-4262.2020.001228. Available at: <http://journal.eu-jr.eu/engineering/article/view/14>.
6. Gunning i D. Aha, (2019) "DARPA's Explainable Artificial Intelligence (XAI) Program", *AI Magazine*, Vol. 40(2), pp.44-58, doi: 10.1609/aimag.v40i2.2850.
7. Isaac Lage, Emily Chen, Jeffrey He, Menaka Narayanan, Been Kim, Samuel J Gershman, Finale Doshi-Velez. (2019) Human evaluation of models built for interpretability. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol.7, pp 59–67.
8. Oana-Maria Camburu, Eleonora Giunchiglia, Jakob Foerster, Thomas Lukasiewicz, Phil Blunsom. (2019) Can I trust the explainer? Verifying post-hoc explanatory methods. arXiv:1910.02065.
9. Mengjiao Yang and Been Kim (2019). BIM: Towards quantitative evaluation of interpretability methods with ground truth. arXiv:1907.09701.
10. Peter Lipton. *Inference to the best explanation*. Routledge, 2003.

Received (Надійшла) 11.04.2023

Accepted for publication (Прийнята до друку) 07.06.2023

Evaluation of the sensitivity of explanations in the intelligent information system

S. Chalyi, V. Leshchynskiy

Abstract. The article's subject matter is the process of constructing explanations for the received decisions in the intellectual information system. The goal is to evaluate the sensitivity of the explanations based on the analysis of the properties of the input data and the corresponding decisions in the intelligent information system to support the selection of the best explanation from the standpoint of satisfying the user's interests. **Task:** structuring of criteria for quantitative assessment of explanations when presenting an intelligent system in the form of a black box; development of a method for assessing the sensitivity of explanations in an intelligent information system. The used approaches are approaches to constructing explanations and approaches to evaluating explanations in intelligent information systems. The following results were obtained. Criteria for evaluating explanations for intelligent systems presented according to the black box principle are structured. The specified criteria consider the impact on the explanation of the input and output data of the intelligent system, the appropriateness of the explanation of the decision-making process in the intelligent system, as well as the appropriateness of the explanation and understanding of the results of the intelligent system by the user. On the basis of the performed structuring, a method for assessing the sensitivity of explanations for an intelligent system presented according to the black box principle is proposed. **Conclusions.** The scientific novelty of the obtained results is as follows. A method for assessing the sensitivity of explanations for an intelligent system presented according to the black box principle is proposed. The method includes steps related to testing and determining the similarity of input data and results for alternative models of intelligent systems by quantitative and qualitative indicators, as well as quantifying the input data and determining the sensitivity of the explanation. The proposed method makes it possible to compare and select explanations considering the properties and importance of input data in order to determine the possibilities of applying alternative approaches to constructing explanations for the results of an intelligent information system. Further development of the proposed approach is focused on the definition and implementation of metrics for assessing the accuracy and transparency of explanations.

Keywords: intellectual system, explanation, decision-making process, causal relationship, evaluation of explanations.

Б. Ю. Вінтенко¹, О. А. Смірнов², О. В. Коваленко², С. А. Смірнов², А. С. Коваленко²

¹ ПАТ “Науково-виробниче підприємство “Радій”, Кропивницький, Україна

²Центральноукраїнський національний технічний університет, Кропивницький, Україна

ДОСЛІДЖЕННЯ НОРМАТИВНИХ ДОКУМЕНТІВ ТА ГАЛУЗЕВИХ СТАНДАРТІВ РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ КОМП’ЮТЕРНИХ СИСТЕМ УПРАВЛІННЯ АЕС, ВАЖЛИВИХ ДЛЯ БЕЗПЕКИ

Анотація. На сьогоднішній день одним з основних джерел електричної енергії є атомні електростанції (АЕС). В сучасному світі для управління технологічними процесами та забезпечення безпеки на АЕС використовуються цифрові системи управління з широким застосуванням комп’ютерів та контролерів. Від якості та надійності програмного забезпечення таких систем залежить безпечна експлуатація АЕС. В даній роботі проведено дослідження нормативних документів та стандартів, які мають відношення до розробки програмного забезпечення комп’ютерних систем управління АЕС, важливих для безпеки. **Метою** роботи є визначення стандартів, що можуть використовуватися розробниками програмного забезпечення для комп’ютерних систем управління АЕС, важливих для безпеки. **Об’єктом дослідження** є процес розробки програмного забезпечення для комп’ютерних систем управління АЕС. **Предметом** є дослідження нормативних документів та галузевих стандартів розробки програмного забезпечення для комп’ютерних систем управління АЕС. **У результаті дослідження** визначені критерії безпеки комп’ютерних систем управління, досліджено стандарти IEC 61508 та IEC 61513, досліджена настанова з безпеки МАГАТЕ SSG-39, визначені категорії функцій управління та класи безпеки систем АЕС, досліджені стандарти, що містять вимоги до програмного забезпечення, інтерфейсів користувача та кодування MISRA C/C++. **Висновки.** З метою створення високонадійного та якісного програмного забезпечення комп’ютерних систем керування АЕС, важливих для безпеки, в світі існують стандарти проектування та галузеві нормативні документи. Результатом дослідження стандартів та галузевих нормативних документів, які мають відношення до розробки програмного забезпечення комп’ютерних систем управління АЕС, важливих для безпеки є наступне: міжнародні стандарти та нормативні документи мають різні рівні спеціалізації. В документах вищого рівня наводяться загальні вимоги до всіх етапів життєвого циклу комп’ютерних систем управління та містяться посилання на документи наступного рівня. Документи наступного рівня є більш спеціалізованими і мають відношення до більш вузьких сфер діяльності; кожна країна світу може визначити свій набір стандартів та нормативних документів, який вимагається використовувати під час проектування та реалізації комп’ютерних систем управління для АЕС; в нормативних документах містяться вимоги та рекомендації як щодо організації процесу розробки, так і загальні рекомендації та приклади написання коду програм різними мовами програмування; через те, що вимоги розглянутих стандартів мають бути актуальними протягом тривалого часу, ці стандарти не містять формалізованих методик написання коду програм; не існує формалізованих методик сертифікації коду на відповідність вимог розглянутих стандартів; кожна організація, яка займається проектуванням, розробкою та верифікацією програмного забезпечення комп’ютерних систем управління для АЕС, повинна розробити власні методики з застосування вимог та оцінки критеріїв виконання вимог стандартів та нормативів. Таким чином, зроблено висновок, що перспективами подальших досліджень є детальний аналіз вимог міжнародних стандартів та галузевих нормативних документів до програмного забезпечення комп’ютерних систем управління АЕС та розробка методики його написання й оцінки відповідності вимогам стандартів.

Ключові слова: програмне забезпечення, комп’ютерні системи управління, АЕС.

Вступ

Постановка проблеми. На сьогоднішній день одним з основних джерел електричної енергії є атомні електростанції (АЕС). В багатьох країнах світу, зокрема в Словаччині, Франції, Україні [1] обсяги атомної генерації становлять понад 50%. Використання атомної енергії є екологічно чистим та економічно вигідним, проте має високі вимоги до безпеки та надійності для забезпечення захисту людини та навколишнього середовища від радіаційного випромінювання ядерного палива. В сучасному світі для управління технологічними процесами та забезпечення безпеки на АЕС використовуються цифрові системи управління з широким застосуванням комп’ютерів та контролерів. Від якості та надійності програмного забезпечення (ПЗ) таких систем залежить безпечна експлуатація АЕС.

Аналіз останніх досліджень і публікацій. Цифрові системи управління потребують особливих підходів до розробки та оцінювання надійності в порівнянні з аналоговими системами, оскільки в силу свого характеру та призначення допускають більшу свободу під час проектування [11]. Ці підходи мають відповідати визначеним вимогам. Вимоги до підходів проектування, реалізації, верифікації, валідації та інших етапів життєвого циклу таких систем описуються в міжнародних стандартах [2, 3], публікаціях МАГАТЕ [11] та галузевих нормативних документах різних країн [16–19].

Ці документи є обов’язковими для використання всіма учасниками процесів проектування, розробки, впровадження та експлуатації комп’ютерних систем управління (КСУ).

Дослідження вимог стандартів проектування та розробки КСУ проведено в роботах [20–23]. Детальні

вимоги до ПЗ КСУ АЕС описуються в міжнародних стандартах [5, 6], а також, наприклад, в галузевому стандарті США [7]. Вимоги до інтерфейсів користувача КСУ АЕС описуються в міжнародних стандартах [7, 8], а також, наприклад, в галузевому стандарті США [14]. Схема взаємозв'язків міжнародних та галузевих стандартів, на які посиляється дана робота, показана на рис. 1.

Постановка задач дослідження

В даній роботі необхідно провести дослідження нормативних документів та стандартів, які мають відношення до розробки програмного забезпечення КСУ АЕС, важливих для безпеки. **Метою** роботи є визначення стандартів, що можуть використовуватися розробниками програмного забезпечення для КСУ АЕС, важливих для безпеки. **Об'єктом дослідження** є процес розробки програмного забезпечення для КСУ АЕС. **Предметом** є дослідження нормативних документів та галузевих стандартів розробки програмного забезпечення для КСУ АЕС.

Безпека комп'ютерних систем управління

Комп'ютерні системи управління (КСУ) згідно з стандартом IEC 61508 є окремим випадком систем контролю та управління (СКУ). СКУ – це система, що приймає вхідні сигнали від процесу (датчиків) та/або оператора, обробляє вхідні параметри згідно з необхідними технологічними алгоритмами та генерує вихідні сигнали, зумовлюючи роботу керованого об'єкту необхідним чином. СКУ, що базуються на КСУ все в більших об'ємах використовуються для виконання функцій безпеки на об'єктах, які можуть створювати ризики для людей та навколишнього середовища (атомні електростанції, газова та нафтова промисловість, транспорт, хімічне виробництво). СКУ на таких об'єктах мають відповідати певним характеристикам безпеки (резервування, самодіагностика, стійкість до відмов). КСУ мають високий рівень програмної складової, для якої існують, зокрема, вимоги інформаційної безпеки (Security Requirements) [20]. Згідно з ними, дані та програмне забезпечення мають бути захищеними від несанкціонованого доступу та несанкціонованих змін. Інформаційна безпека описана в групі стандарті ISO/IEC 27000 [9], нормативному документі NIST SP 800-82 «Guide to Industrial Control Systems (ICS) Security» [10]. Разом з тим, для КСУ важливими є вимоги функціональної безпеки (Safety Requirements). Функціональна безпека системи забезпечує коректність виконання функцій управління об'єктом, у випадку відмови системи – переведення об'єкту керування в безпечний стан. Її вимоги стосуються не тільки програмної складової, а і апаратного забезпечення, електричного та електронного обладнання та механічних пристроїв, а також власне інформаційної безпеки [20].

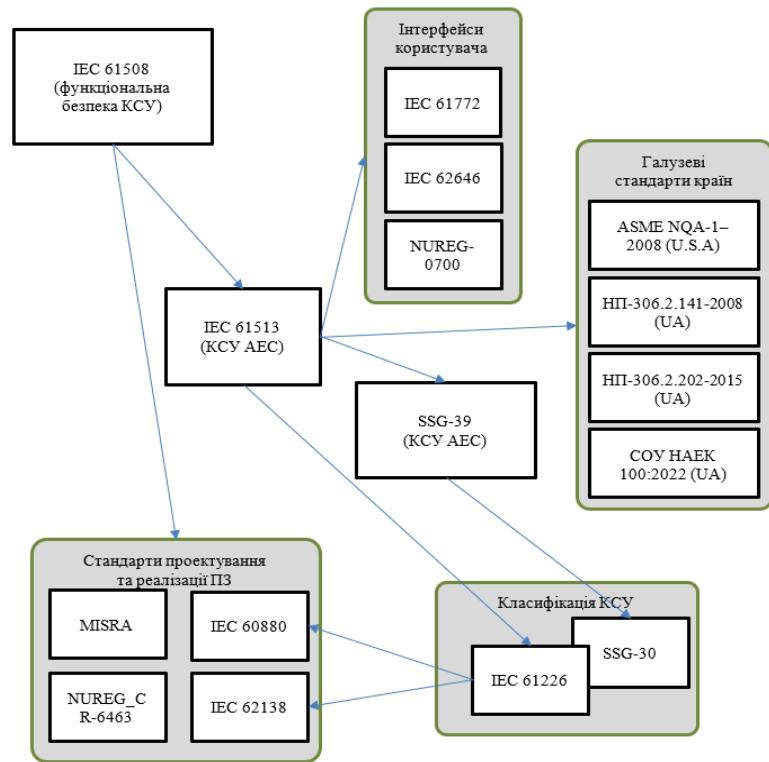


Рис. 1. Схема взаємозв'язків міжнародних та галузевих стандартів

Загальні стандарти функціональної безпеки для КСУ наведено в міжнародних стандартах, таких як IEC 61508 [2].

Стандарт з функціональної безпеки КСУ IEC 61508

Стандарт IEC 61508 «Functional safety of electrical/electronic/programmable electronic safety-related systems» [2] є стандартом верхнього рівня для функціональної безпеки (ФБ) КСУ. Виконання вимог цього стандарту підвищує надійність та життєздатність продукту, а також формалізує та систематизує процеси проектування, розробки, тестування та впровадження. Стандарт IEC 61508 складається з семи частин. Частина IEC 61508-1 описує загальні вимоги до функціональної безпеки, частина IEC 61508-2 – додаткові вимоги до апаратних засобів, частина IEC 61508-3 – додаткові вимоги до програмного забезпечення та документації, частина IEC 61508-4 містить терміни та визначення. Інші частини мають довідковий характер. Частина стандарту IEC 61508-1 визначає вимоги до керування функціональною безпекою та вимоги до життєвого циклу (ЖЦ) системи в контексті промислового об'єкту автоматизації. Вона визначає такі етапи ЖЦ як концепція, розробка, верифікація, валідація, впровадження, експлуатація, супроводження, зняття з експлуатації.

В частині стандарту IEC 61508-2 додаються вимоги до апаратної складової КСУ. Проте ця частина містить додатки, які мають враховувати і розробники програмного забезпечення. Додатки А та В описують підхід до реалізації самодіагностики та захисту від одиничних відмов. Додаток С містить рекомендації з розрахунку діагностичного покриття з метою забезпечення потрібного рівня повноти безпеки (Safety

Integrity Level, SIL). Додаток D містить вимоги до керівництва з експлуатації з урахуванням вимог безпеки (Safety Manual).

Частина стандарту IEC 61508-3 конкретизує вимоги до програмного забезпечення КСУ. Зокрема, розділ 7 наводить деталізовані вимоги до життєвого циклу ПЗ.

У підрозділі 7.1 наводиться підхід до розділення процесу розробки програмного забезпечення на етапи, описані в підрозділах 7.2-7.9 стандарту.

Підрозділ 7.2 містить вимоги щодо етапу специфікації вимог до функцій безпеки програмного забезпечення та стійкості до систематичних відмов. Специфікація вимог має враховувати самоконтроль ПЗ, моніторинг апаратного забезпечення, періодичне тестування функцій безпеки тощо.

Підрозділи 7.3 та 7.7 наводять вимоги до етапу планування та виконання процедур підтвердження відповідності аспектів програмного забезпечення до вимог безпеки системи.

Підрозділ 7.4 містить загальні вимоги до етапу проектування та розробки програмного забезпечення. Це вимоги:

- до проектування архітектури програмного забезпечення;
- до вибору інструментальних засобів розробки;
- до реалізації ПЗ з урахуванням вимог специфікації;
- до підходів тестування програмних модулів.

Підрозділ 7.5 описує підходи до етапу інтеграції програмного забезпечення та апаратних засобів. Метою вимог цього підрозділу є об'єднання ПЗ та апаратних засобів та перевірка їх відповідності вимогам повноти безпеки.

Підрозділ 7.6 описує процедури етапів експлуатації та модифікації програмного забезпечення. Мета цих процедур – впевнитися в тому, що функціональна безпека системи зберігається під час експлуатації та модифікації.

Підрозділ 7.8 детально описує процедури модифікації програмного забезпечення.

Підрозділ 7.9 детально описує процедури верифікації програмного забезпечення.

В додатку А стандарту IEC 61508-3 приведена настанова з вибору методів та засобів проектування програмного забезпечення. В додатку В стандарту IEC 61508-3 приведені детальні вимоги до методів та засобів проектування.

Дослідження стандарту IEC 61508 дозволяє зробити висновок, що він містить важливі вимоги до всіх етапів життєвого циклу програмного забезпечення КСУ. Згідно стандарту IEC 61508 застосування багатьох підходів, залежить від конкретних практик, засобів та мов програмування, що використовуються, але на більшості рівнів повноти безпеки (SIL) можна виділити загальні підходи до розробки ПЗ:

- використання сертифікованих інструментальних засобів;
- використання перевірених елементів ПЗ (за їх наявності);
- застосування модульності та структурного програмування;

- використання стандартів кодування (наприклад, MISRA C/C++);
- застосування формальних та напівформальних методів розробки та тестування;
- застосування засобів діагностування;
- статичний аналіз коду;
- контроль часових характеристик виконання програми;
- верифікація ПЗ.

Особливість верифікації програмного коду модулів полягає в аналізі критичності ПЗ, при якому встановлюється ступінь критичності модулів для виконання функцій безпеки. Це дозволяє знизити об'єм заходів для другорядного ПЗ та сконцентруватися на найбільш відповідальному ПЗ.

Наступний етап включає в себе дослідження стандартів розробки ПЗ, що мають безпосереднє відношення до атомних станцій.

Стандарт IEC 61513. Проектування КСУ АЕС, важливих для безпеки

Вимоги стандарту IEC 61508 [2] не є спеціалізованими виключно для атомної галузі. Для різних областей діяльності вимоги стандарту деталізовані в інших стандартах, наприклад IEC 61511 – для небезпечних виробництв, IEC 62061 – для машинобудування. Стандарт **IEC 61513** «Nuclear power plants – Instrumentation and control important to safety – General requirements for systems» [3] є головним в ієрархії стандартів для КСУ, важливих для безпеки атомних станцій. Він встановлює загальні вимоги для архітектури та обладнання КСУ АЕС, важливих для безпеки, що створюються на основі як аналогових, так і цифрових технологій. В багатьох країнах вимагається сертифікація КСУ атомних станцій згідно з стандартами IEC 61508/IEC 61513.

Загальна модель життєвого циклу КСУ АЕС, що приводиться в стандарті IEC 61513, запозичена з стандарту IEC 61508. Стандарт IEC 61513 забезпечує інтерпретацію частин IEC 61508-1, IEC 61508-2, IEC 61508-4 до ядерної галузі. Вимоги частини IEC 61508-3, що стосуються програмного забезпечення КСУ АЕС ВБ, інтерпретуються в стандартах IEC 60880 [5] та IEC 61238 [6]. Розділ 5 стандарту IEC 61513 присвячено загальній структурі життєвого циклу комп'ютерних систем управління АЕС, важливих для безпеки. Він містить вимоги до наступних етапів та процесів проектування КСУ АЕС:

- визначення функцій контролю та управління АЕС (підрозділ 5.2.2);
- категоризацію функцій контролю та управління АЕС в залежності від впливу на безпеку (підрозділ 5.2.3);
- аналізу обмежень проекту конкретної АЕС (підрозділ 5.2.4);
- декомпозицію загальної архітектури КСУ з розподіленням її на окремі підсистеми (підрозділ 5.4.2);
- визначення функцій та інтерфейсів підсистем КСУ (підрозділ 5.4.3);
- планування забезпечення якості КСУ (підрозділ 5.5.2);

- планування захищеності КСУ (підрозділ 5.5.3);
- планування загальної інтеграції КСУ та вводу її в експлуатацію (підрозділ 5.5.4).

В результаті виконання має бути розроблена вихідна документація з проекту архітектури КСУ та функцій, які будуть виконуватися (підрозділ 5.6).

В розділі 6 стандарту IEC 61513 наводяться вимоги до етапів життєвого циклу окремих підсистем КСУ. Перелік етапів включає в себе такі пункти:

- специфікація вимог до функцій підсистеми, обмежень, інтерфейсів користувача, умов навколишнього середовища;
- специфікація системи, аналіз існуючого обладнання, розробка проекту архітектури підсистеми, призначення прикладних функцій;
- детальне проектування та реалізація апаратних засобів та програмного забезпечення, валідація вимог до прикладних функцій;
- інтеграція компонентів програмного та апаратного забезпечення в єдину систему;
- валідація системи;
- монтаж та випробування системи;
- планування подальшої валідації, експлуатації та модифікації.

Розділи 7 та 8 містять вимоги до процесів загальної інтеграції, експлуатації та технічного обслуговування системи.

Аналіз вимог стандарту IEC 61513 дозволяє зробити висновок, що він не містить прямих вимог до особливостей розробки програмного забезпечення. Проте він визначає процеси, в яких беруть участь розробники програмного забезпечення. Також він вказує на стандарт, що визначає класифікацію КСУ АЕС з функціональної безпеки [4], а також на стандарти, що містять конкретні вимоги до програмного забезпечення в залежності від класу функціональної безпеки [5, 6]. Посилання на стандарт IEC 61513 містяться в різних нормативних документах з безпеки КСУ АЕС. Це публікації міжнародних організацій [11], а також галузеві стандарти різних країн [16–19].

Настанова з безпеки МАГАТЕ SSG-39

Настанова з безпеки МАГАТЕ SSG-39 «Design of Instrumentation and Control Systems for Nuclear Power Plants, Specific Safety Guide No. 39» [11] є публікацією, що спирається на стандарт IEC 61513 та відповідає його вимогам. Цей документ також містить рекомендації щодо всіх етапів розробки КСУ АЕС, важливих для безпеки.

Розділ 2 присвячений менеджменту проектування КСУ. Він містить опис всіх етапів життєвого циклу КСУ. В розділі 3 описуються методи визначення функцій контролю та керування АЕС. Розділ 4 містить рекомендації щодо проектування загальної архітектури КСУ та розділення системи керування та контролю на окремі функціональні підсистеми. Розділ 5 описує класифікацію функцій та систем безпеки.

Розділ 6 містить загальні вимоги до систем, важливих для безпеки. Сюди входять пункти щодо надійного проектування, захищеності, випробувань, здатності до ремонту, технічного обслуговування, маркування тощо.

Розділ 7 описує принципи проектування конкретних систем.

Розділ 8 містить вимоги до інтерфейсу «людина»-«машина».

Розділ 9 містить вимоги щодо програмного забезпечення КСУ. Це загальні вимоги до програмного забезпечення (пункти 9.6-9.15), вимоги до його проектування (пункти 9.16-9.43), реалізації (пункти 9.44-9.66), верифікації (пункти 9.64-9.95). В пунктах 9.96-9.98 містяться вимоги до застосування раніше розробленого ПЗ.

В результаті дослідження публікації МАГАТЕ SSG-39 можна зробити висновок, що цей документ містить важливі вимоги до менеджменту розробки, проектування, апаратного та програмного забезпечення КСУ. Він містить посилання на інші відповідні міжнародні стандарти, що дає можливість проектантам КСУ підібрати необхідні для вивчення документи.

Категорії функцій управління та класи безпеки систем АЕС

Функції систем контролю та управління АЕС класифікуються в категорії, що означають важливість функції для безпеки. Важливість функції для безпеки оцінюється наслідками її відмови тоді, коли необхідне її виконання, та наслідками хибного спрацювання (пункт 5.1 [4], пункт 3.15 [12]).

Методи класифікації систем на категорії безпеки встановлені в стандарті IEC 61226 (Nuclear power plants – Instrumentation and control important to safety – Classification of instrumentation and control functions) [4], на який прямо посилається стандарт IEC 61513. Стандарт IEC 61226 застосовується для всіх інформаційних та управляючих функцій, а також систем, що реалізують ці функції. Згідно з пунктом 5.3 стандарту IEC 61226, існують три категорії функцій КСУ за важливістю для безпеки – «А», «В» та «С».

Методика розподілення систем на класи безпеки також міститься в документах МАГАТЕ SSG-30 «Safety Classification of Structures, Systems and Components in Nuclear Power Plants» [12] та TECDOC-1787 «Application of the Safety Classification of Structures, Systems and Components in Nuclear Power Plants» [13]. Розділ 3.16 [12] описує категорії безпеки 1, 2 та 3 в залежності їх впливу на безпеку. Згідно розділу 3.2.7 [13], для КСУ АЕС ці категорії відповідають категоріям «А», «В», «С» стандарту IEC 61226.

Категорія «А» (клас 1) використовується для систем, що грають основну роль в забезпеченні безпеки АЕС і відмова яких може безпосередньо привести до аварійних умов.

Категорію «В» (клас 2) мають системи, що грають додаткову роль в забезпеченні безпечних умов, зокрема у випадку, коли робота функцій цих систем виключить необхідність спрацювання функцій категорії «А». До цієї категорії також відносяться функції «А». Відмова або хибне спрацювання яких може викликати або підсилити аварійний стан. Через наявність ешелону функцій безпеки категорії «А» вимоги до безпеки систем категорії «В» можуть бути не такими строгими, і функціональність таких систем може бути розширеною.

До категорії «С» (клас 3) відносять функції, що грають додаткову чи допоміжну роль. Вони мають відношення до забезпечення безпеки АЕС, але можуть безпосередньо не використовуватися для реагування на аварійні ситуації.

В свою чергу, нормативний документ **НП 306.2.141-2008 «Загальні положення безпеки атомних станцій»** [17], що діє в Україні, поділяє елементи АЕС на чотири класи безпеки. Клас 1 – це тепловиділяючі елементи (твели) та інші елементи, відмова яких є вихідними подіями, що призводять до пошкодження тепловиділяючих елементів з перевищенням встановлених для проектних аварій меж. Клас 2 – елементи, відмова яких є вихідними подіями, що призводять до пошкодження тепловиділяючих елементів в межах, встановлених для проектних аварій, а також елементи систем безпеки, відмови яких призводять до невиконання цими системами своїх функцій. Клас 3 – це елементи систем, важливих для безпеки, що не увійшли до класів 1 та 2. Клас 4 – це елементи систем нормальної експлуатації АЕС. Також для класів використовуються кваліфікаційні позначення літерою, що означають призначення елемента: Н – нормальна експлуатація, У – управляючі системи, Л – локалізуючі системи безпеки, З – захисні системи.

Нормативний документ України **НП 306.2.202-2015 «Вимоги з ядерної та радіаційної безпеки до інформаційних та керуючих систем, важливих для безпеки атомних станцій»** [18] та стандарт державного підприємства «НАЕК «Енергоатом» **СОУ НАЕК 100:2022 «Інформаційні та керуючі системи, важливі для безпеки атомних електричних станцій. Загальні технічні вимоги»** [19] аналогічно до стандарту ІЕС 61226 поділяють функції КСУ класів безпеки 2 та 3 за **НП 306.2.141-2008** на категорії «А», «В» та «С» залежно від їх ролі в забезпеченні та підтримці безпеки, а також можливим невиконанням або помилковим виконанням своїх функцій.

Класи безпеки систем АЕС за стандартом ІЕС 61513. Крім визначення категорій за функціями безпеки, стандарт ІЕС 61513 відносить КСУ, важливі для безпеки, до одного з трьох класів безпеки (додаток В. 4). Вимоги до функції безпеки найвищої категорії визначають клас безпеки системи. Так, до класу 1 відносяться системи аварійного електропостачання та активації захисту, а до класів 2 та 3 – системи автоматизації та контролю АЕС.

В результаті дослідження стандартів з класифікації систем та стандартів з функціональної безпеки **можна зробити висновок**, існує декілька градацій розділення КСУ за класами безпеки та категоріями впливу на безпеку. Вимоги до проектування та розробки програмного забезпечення можуть відрізнятися залежно від класу безпеки КСУ. Подальше дослідження буде стосуватися стандартів, що містять вимоги до програмного забезпечення з урахуванням категорії функцій безпеки.

Стандарти, що містять вимоги до програмного забезпечення

Стандарти ІЕС 61508, ІЕС 61513 не містять визначених вимог до програмного забезпечення КСУ

АЕС, важливих для безпеки. Проте у вступі до стандарту ІЕС 61513, а також таблиці 1 пункту 7.1 стандарту ІЕС 61226 наведені посилання на інші стандарти, в яких ці вимоги вводяться. Для ПЗ систем категорії функцій безпеки «А» використовується стандарт ІЕС 60880, для систем категорій функцій безпеки «В», «С» – ІЕС 62138. Ці стандарти відповідають загальним вимогам стандарту ІЕС 61508-3 до програмного забезпечення КСУ в ядерній сфері.

Вимоги до програмного забезпечення КСУ, що виконують функції безпеки категорії «А», містяться в стандарті ІЕС 60880 «Nuclear power plants – Instrumentation and control systems important to safety – Software aspects for computer-based systems performing category A functions» [5]. Вимоги стандарту застосовуються до програмного забезпечення, що використовується в КСУ АЕС для виконання функцій безпеки категорії «А». Це ПЗ, яке використовується для функцій захисту, а також запобігання розвитку проектних аварій і приведення станції в безпечний контрольований стан під час аварійних ситуацій, має бути створене з високим ступенем якості та надійності, пройти повний цикл верифікації та атестації. Мета вимог стандарту ІЕС 60880 – розробка ПЗ, що відповідає критеріям якості та надійності. Вимоги стандарту відносяться до всіх етапів життєвого циклу програмного забезпечення, включаючи специфікацію вимог, проектування, вибір інструментальних засобів та мов, розробку, верифікацію, валідацію, модифікацію тощо.

Розділ 3 стандарту ІЕС 60880 описує терміни та визначення. Зокрема, вводяться поняття універсальної мови програмування, проблемно-орієнтованої мови програмування, системного та прикладного програмного забезпечення, прикладної функції, стиснення коду, різноманітності, резервування, стійкості до дефектів та помилок тощо.

Розділ 4 містить список скорочень та позначень.

Розділ 5 містить загальні вимоги до проектам програмного забезпечення. В розділі розглядаються наступні концепції:

- життєвого циклу безпеки ПЗ в додаток до загального життєвого циклу системи керування;
- розподілу ПЗ та конфігураційних даних за типами (системне або прикладне);
- наявності підходів до розробки ПЗ за допомогою універсальних або проблемно-орієнтованих мов програмування;
- керування проектуванням ПЗ;
- керування забезпеченням якості ПЗ;
- захищеності ПЗ.

Розділ 6 присвячений етапу створення специфікації вимог до програмного забезпечення. Вимоги до ПЗ мають визначатися вимогами до системи безпеки і являтися частиною специфікації КСУ. Вони мають описувати те, що повинна робити програма, а не те, як вона повинна це робити. До основних вимог до ПЗ, що входять до специфікації, входять:

- чіткий опис прикладних функцій ПЗ, його поведінки, інтерфейсів користувача, часу відгуку, зв'язків з іншими системами, функціональних обмежень;

- наявність самоконтролю та збору діагностичної інформації;
- передбачення періодичних тестувань функцій безпеки;
- наявність чіткої документації.

Розділ 7 містить вимоги щодо проектування та реалізації програмного забезпечення.

В підрозділі 7.1 описані загальні принципи проектування та реалізації ПЗ. Наведені окремі вимоги до реалізації ПЗ з використанням універсальних та проблемно-орієнтованих мов програмування. Також наводяться вимоги до конфігурування раніше розробленого програмного забезпечення.

В підрозділі 7.2 наведені детальні вимоги щодо вибору мови програмування та інструментальних засобів.

Підрозділ 7.3 містить посилання на додаток В, що містить детальні вимоги і рекомендації з розробки ПЗ. Зазначається, що в процесі розробки ці вимоги та рекомендації мають виконуватися, а їх невиконання має бути документально обґрунтованим.

Розділ 8 визначає вимоги до верифікації ПЗ. В ньому наводяться окремі вимоги до планування процесу верифікації, верифікації проекту та його реалізації.

У розділі 9 знаходиться опис вимог до процесів інтеграції ПЗ до технічного забезпечення та його верифікації під час інтеграції.

Розділ 10 містить опис аспектів валідації ПЗ в складі інтегрованої системи. Має бути створений план валідації, який буде виконуватися особами, які не брали участь в проектуванні та реалізації системи. Для кожної функції має бути проведено комплексне всебічне тестування. За результатами тестувань мають бути складені звіти та визначені процедури усунення виявлених дефектів.

Розділ 11 описує вимоги до процесу модифікації ПЗ. Розглядаються вимоги до процедури запиту на модифікацію та процедури виконання модифікації.

Розділ 12 формує вимоги до установки та експлуатації ПЗ, його захищеності на місці експлуатації та навчання операторів.

Розділ 13 містить вимоги до захисту від дефектів ПЗ, що можуть призвести до відмови з загальної причини. Розглядаються підходи до проектування ПЗ з врахуванням критерію одиничної відмови, реалізація різноманітності.

Розділ 14 містить детальні вимоги до вибору, атестації та конфігурації інструментальних програм.

Стандарт ІЕС 60880 також містить додатки, які деталізують інформацію, наведену в розділах. За об'ємом вони займають половину розміру всього документу стандарту і є обов'язковими доповненнями до відповідних розділів. Зокрема:

- розділ А (обов'язковий) містить детальну інформацію щодо життєвого циклу ПЗ та формування вимог до ПЗ;
- розділ В (обов'язковий) містить деталізовані вимоги та рекомендації щодо процесу проектування ПЗ, структури ПЗ, самоконтролю, детального проектування та кодування;
- розділ С (довідковий) містить приклади технологій прикладного програмування;

– розділ D (довідковий) містить детальні вимоги до мов програмування та інструментальних засобів;

- інші розділи є також довідковими.

В результаті дослідження стандарту ІЕС 60880 **можна зробити висновок**, що він є одним з основних стандартів з розробки та проектування ПЗ КСУ АЕС, важливих для безпеки. Незалежно від технологій та мов програмування, що використовуються, розробники програмного забезпечення мають детально знати та чітко дотримуватись його рекомендацій.

Незважаючи на те, що вимоги цього стандарту відносяться до ПЗ категорії функціональної безпеки «А», вони можуть застосовуватися і при розробці ПЗ нижчих функціональних категорій. Це дозволить покращити якість та надійність ПЗ загалом, а також розширити межі використання розробленого програмного забезпечення до вищої категорії «А».

Стандарт не містить детальних методик досягнення та оцінки відповідності програмного коду своїм вимогам. **Тому в перспективі подальшого дослідження** цього стандарту є детальний аналіз вимог та створення методів написання та оцінки програмного коду.

Вимоги до програмного забезпечення КСУ, що виконують функції безпеки категорій «В» та «С», містяться в стандарті ІЕС 62138 «Nuclear power plants – Instrumentation and control important for safety – Software aspects for computer-based systems performing category B or C functions» [6]. Цей стандарт доповнює стандарт ІЕС 60880, який містить вимоги до програмного забезпечення КСУ, що виконує функції безпеки категорії «А». Також він доповнює загальний стандарт з функціональної безпеки КСУ АЕС ІЕС 61513 в об'ємі вимог до програмного забезпечення.

Мета даного стандарту – максимально знизити імовірність прихованих програмних дефектів, що можуть привести до системних відмов через одиничні або множинні програмні відмови (тобто відмови з загальної причини).

Вимоги стандарту ІЕС 62138 прив'язані не до категорій функцій безпеки, а до класу безпеки КСУ (клас 2 та клас 3 за стандартом ІЕС 61513), оскільки КСУ можуть одночасно реалізовувати як функції безпеки категорій «В», «С», так і не класифіковані за безпекою,

Розділ 3 стандарту ІЕС 62138 описує терміни та визначення.

Розділ 4 містить список скорочень та позначень.

Розділ 5 містить загальні концепції та положення. В ньому виділяються типи програмного забезпечення в залежності від типу робіт, що виконуються ним (системне та прикладне). Розглядаються типи конфігураційних даних. Наводиться архітектура життєвого циклу програмного забезпечення. Вводяться принципи градації вимог до програмного забезпечення в залежності від класу безпеки КСУ (класи безпеки 2 та 3 за ІЕС 61513 відповідно до категорії функцій безпеки «В» та «С»).

Розділ 6 містить вимоги до програмного забезпечення КСУ класів безпеки 2 та 3. В підрозділі 6.1 зазначається, що частина вимог цього розділу

застосовується до ПЗ обох класів безпеки, а окремі вимоги стосуються тільки ПЗ класу 2 або класу 3.

Підрозділ 6.2 містить загальні вимоги до життєвого циклу ПЗ та змісту програми контролю його якості. Зазначається, що кожний етап життєвого циклу ПЗ має включати дії, які необхідні для досягнення необхідної якості, а також дії з верифікації показників досягнення цієї якості. Мають бути заплановані заходи щодо керування конфігурацією та контролю версій ПЗ. Також наводяться загальні рекомендації щодо вибору інструментальних засобів та мов програмування.

Підрозділ 6.3 містить вимоги до вибору та аналізу раніше розробленого програмного забезпечення. Вимагається наявність документації з безпеки, докази коректності та функціональної відповідності цього ПЗ вимогам КСУ, що розробляється.

Підрозділ 6.4 доповнює та уточнює вимоги стандарту IEC 61513 про те, що вимоги до ПЗ повинні бути оформлені документально і сформована специфікація вимог до ПЗ. Вводяться вимоги до початкової інформації, згідно з якою створюється КСУ, та змісту специфікації ПЗ (наявність опису прикладних функцій, режимів роботи, інтерфейсів, обмежень тощо).

Підрозділи 6.5 та 6.6 стандарту IEC 61238 містять детальні вимоги до етапу проектування, початкових даних та етапу реалізації нового програмного забезпечення. Ці підрозділи не містять конкретних вимог щодо написання специфікації та коду програм, але містять інформацію, яким чином має виконуватися проектування та реалізація в залежності від класу безпеки системи. Розглянуті вимоги як для проблемно-орієнтованих, так і універсальних мов програмування.

Подальші підрозділи (6.7-6.11) містять вимоги до процесів інтеграції програмного та апаратного забезпечення, проведення процедур валідації системи, встановлення та експлуатацію програмного забезпечення на місці експлуатації та подальші модифікації. В підрозділі 6.12 вказується на необхідність з захисту КСУ від відмов з загальної причини внаслідок роботи програмного забезпечення та міститься посилання на стандарт IEC 61513, де це питання розглянуто більш детально. Додатки до стандарту є довідковою інформацією та не містять вимог до програмного забезпечення.

З дослідження стандарту IEC 61238 можна зробити висновок, що вимоги до реалізації ПЗ функцій безпеки категорій «В» та «С» є менш формальними та суворими, ніж ПЗ функцій безпеки до категорії «А», а є більш узагальнюючими. Проте в перспективі подальших досліджень для цих вимог є більш детальний їх аналіз з метою розробки методики написання ПЗ та оцінки його відповідності вимогам стандарту IEC 61238.

Стандарти інтерфейсів користувача

Інтерфейс користувача визначає зручність взаємодії людини з програмним забезпеченням. При створенні програмного забезпечення КСУ АЕС його інтерфейс користувача має забезпечувати безпомилковість, ефективність та зручність роботи операторів.

Зокрема, в **стандарті IEC 61772** «Nuclear power plants – Control rooms – Application of visual display

units (VDUs)» [7] наводяться вимоги до пристроїв візуального відображення, які використовуються операторами АЕС. Ці вимоги мають бути враховані під час реалізації інтерфейсів користувача програмного забезпечення. Вони стосуються наступних аспектів:

- відбір інформації, яка має відображатися;
- компонування та розміщення інформації;
- формат, розмір та колір елементів відображення;
- форма відображення інформації: алфавітно-цифрова, образна та символічна;
- забезпечення зручного вибору оператором інформації, необхідної на поточний момент.

Інший **стандарт IEC 62646** «Nuclear power plants – Control rooms – Computer-based procedures» [8] містить узагальнені вимоги до процедур керування АЕС, в яких застосовуються комп'ютери. Зокрема, згідно з цими вимогами, при проектуванні програмного забезпечення КСУ АЕС мають бути реалізовані:

- зручний інтерфейс «людина-машина»;
- чітка навігація між видами інформації;
- автоматизація процедур, що виконуються оператором;
- допомога оператору під час прийняття рішень.

Галузевий нормативний документ **NUREG-0700** «Human-System Interface Design Review Guidelines» розглядає інтерфейси користувачів КСУ з декількох напрямків.

В першій частині (розділи 1-3) розглядаються загальні вимоги до відображення інформації:

- до форматів відображення даних: (написи, списки, таблиці, гістограми, графіки, діаграми тощо);
- до елементів інтерфейсу користувача: символи, мітки, іконки, шкали, стрілки, кольорові схеми тощо;
- до швидкості оновлення інформації та часу відгуку ПЗ;
- до засобів інтерактивної взаємодії ПЗ з користувачем (команди, пункти меню, клавіші, курсори);
- до засобів вибору необхідної інформації;
- до перевірки дій та захисту від помилок оператора, а також до довідкової інформації.

В цій частині також наводяться вимоги до аналогових засобів відображення та контролю (прилади вимірювання, індикатори, цифрові табло), що не містять програмного забезпечення.

В другій частині (розділи 4-10) наводяться вимоги до реалізації інтерфейсів користувача для систем з різною функціональністю: системи захисту, системи відображення критичних параметрів, системи відображення колективного користування, комплекси автоматизації, системи передачі інформації.

Третя частина (розділи 11, 12) містить вимоги до розміщення та обладнання робочих місць операторів.

Четверта частина (розділи 13, 14) містить вимоги до безпеки, надійності та здатності до ремонту обладнання робочих місць операторів.

В результаті досліджень вимог стандартів IEC 61772 та IEC 62646, а також нормативного документу NUREG-0700 можна зробити висновок, що

вони визначають загальні підходи до проектування інтерфейсів користувача програмного забезпечення КСУ як частини системи взаємодії «людина-машина». Ці вимоги відносяться до всіх етапів життєвого циклу ПЗ КСУ: проектування, реалізації, верифікації, валідації, експлуатації, навчання персоналу. Дотримання вимог цих документів дозволяє створити програмне забезпечення, яке буде відповідати критеріям надійності, захищеності, зручності та ергономічності.

Стандарти кодування MISRA C/C++

Мови програмування C та C++ широко використовуються при створенні програмного забезпечення КСУ. Стандарт IEC 61508 містить рекомендації щодо відповідності коду програм до правил безпечного та ефективного програмування. Прикладом такого стандарту для цих мов є стандарти MISRA C та MISRA C++.

Стандарти кодування MISRA C/C++ розроблені Motor Industry Software Reliability Association – організацією, що створює стандарти розробки програмного забезпечення для автомобільної промисловості [24]. Ці стандарти спрямовані на підвищення функціональної безпеки, інформаційної безпеки та надійності вбудованого ПЗ. Вони використовуються в різноманітних сферах розробки КСУ – космічна діяльність, оборонні технології, автоматизація виробництва та медицина.

Стандарти MISRA C/C++ містить набір правил. Кожне правило має певну категорію: обов'язкове (mandatory), необхідне (requirement) та рекомендоване (advisory). Також вони розподіляються за функціональним призначенням (врахування особливостей компіляторів, уникнення використання деяких функцій, рекомендації щодо найменувань змінних та коментування тощо). Для відповідності ПЗ стандарту обов'язкові та необхідні правила мають бути впроваджені, а рекомендовані правила можуть не виконуватися з обґрунтуванням таких рішень в документації.

Існують програмні інструменти, що використовуються для статичного аналізу коду програми стандарту MISRA: Helix QAC, PVS-Studio, Polyspace [25].

З дослідження змісту стандартів MISRA можна отримати висновок, що їх рекомендації підвищують якість та надійність коду ПЗ КСУ АЕС. Високі

значення показника якості та надійності необхідні для забезпечення високого рівня функціональної безпеки КСУ АЕС.

Висновки

З метою створення високонадійного та якісного програмного забезпечення комп'ютерних систем керування АЕС, важливих для безпеки, в світі існують стандарти проектування та галузеві нормативні документи. Результатом дослідження стандартів та галузевих нормативних документів, які мають відношення до розробки ПЗ КСУ АЕС, важливих для безпеки є наступне:

- міжнародні стандарти та нормативні документи мають різні рівні спеціалізації. В документах вищого рівня наводяться загальні вимоги до всіх етапів життєвого циклу КСУ та містяться посилання на документи наступного рівня. Документи наступного рівня є більш спеціалізованими і мають відношення до більш вузьких сфер діяльності;

- кожна країна світу може визначити свій набір стандартів та нормативних документів, який вимагається використовувати під час проектування та реалізації КСУ для АЕС;

- в нормативних документах містяться вимоги та рекомендації як щодо організації процесу розробки, так і загальні рекомендації та приклади написання коду програм різними мовами програмування;

- через те, що вимоги розглянутих стандартів мають бути актуальними протягом тривалого часу, ці стандарти не містять формалізованих методик написання коду програм;

- не існує формалізованих методик сертифікації коду на відповідність вимог розглянутих стандартів;

- кожна організація, яка займається проектуванням, розробкою та верифікацією програмного забезпечення КСУ для АЕС, повинна розробити власні методики з застосування вимог та оцінки критеріїв виконання вимог стандартів та нормативів.

Таким чином, виходячи з вищеперерахованого, перспективами подальших досліджень є детальний аналіз вимог міжнародних стандартів та галузевих нормативних документів до програмного забезпечення КСУ АЕС та розробка методики його написання та оцінки відповідності вимогам стандартів.

СПИСОК ЛІТЕРАТУРИ

1. Nuclear power by country. [Електронний ресурс] – Режим доступу: https://en.wikipedia.org/wiki/Nuclear_power_by_country.
2. IEC 61508-2010: Functional safety of electrical/electronic/programmable electronic safety-related systems – Part 1: General requirements. Geneva, International Electrotechnical Commission (IEC), 2010.
3. IEC 61513-2011: Nuclear power plants – Instrumentation and control important to safety – General requirements for systems requirements. Geneva, International Electrotechnical Commission (IEC), 2011.
4. IEC 61226-2009: Nuclear power plants – Instrumentation and control important to safety – Classification of instrumentation and control functions. Geneva, International Electrotechnical Commission (IEC), 2009.
5. IEC 60880-2006: Nuclear power plants – Instrumentation and control systems important to safety – Software aspects for computer-based systems performing category A functions. Geneva, International Electrotechnical Commission (IEC), 2006.
6. IEC 62138-2004: Nuclear power plants – Instrumentation and control important for safety – Software aspects for computer-based systems performing category B or C functions. Geneva, International Electrotechnical Commission (IEC), 2004.
7. IEC 61772:2009: Nuclear power plants – Control rooms – Application of visual display units (VDUs). Geneva, International Electrotechnical Commission (IEC), 2009.
8. IEC 62646-2019: Nuclear power plants – Control rooms – Computer based procedures. Geneva, International Electrotechnical Commission (IEC), 2012.

9. ISO/IEC 27000-2018: Information technology – Security techniques – Information security management systems – Overview and vocabulary. 2018.
10. Keith Stouffer, Victoria Pillitteri, Suzanne Lightman, Marshall Abrams, Adam Hahn. NIST SP 800-82 Revision 2: Guide to Industrial Control Systems (ICS) Security. National Institute of Standards and Technology, 2015.
11. Design of Instrumentation and Control Systems for Nuclear Power Plants: Specific Safety Guide No. SSG-39. Vienna, IAEA, 2016.
12. Safety Classification of Structures, Systems and Components in Nuclear Power Plants: Specific Safety Guide No. SSG-30. Vienna, IAEA, 2014.
13. Application of the Safety Classification of Structures, Systems, and components in Nuclear Power Plants: IAEA-TECDOC-1787. Vienna, IAEA, 2016.
14. ASME NQA-1–2008. Quality Assurance Requirements for Nuclear Facility Applications. An American National Standard. American Society of Mechanical Engineers, 2008.
15. NUREG/CR-6463. Review Guidelines on Software Languages for Use in Nuclear Power Plant Safety Systems. U.S. Nuclear Regulatory Commission, 1996.
16. NUREG-0700 Revision 3. Human-System Interface Design Review Guidelines. U.S. Nuclear Regulatory Commission, 2020.
17. НП 306.2.141–2008. Загальні положення безпеки атомних станцій. Київ, Держатомрегулювання України, 2007.
18. НП 306.2.202-2015. Вимоги з ядерної та радіаційної безпеки до інформаційних та керуючих систем, важливих для безпеки атомних станцій. Київ, Держатомрегулювання України, 2015.
19. СОУ НАЕК 100:2022. Інформаційні та керуючі системи, важливі для безпеки атомних електричних станцій: загальні технічні вимоги. Стандарт державного підприємства «Національна атомна енергогенеруюча компанія «Енергоатом». Київ, ДП НАЕК «Енергоатом», 2022.
20. В. В. Складар. Обеспечение безопасности АСУТП в соответствии с современными стандартами. Москва – Вологда, Инфра-инженерия, 2018.
21. David J. Smith, Kenneth G. L. Simpson. The Safety Critical Systems Handbook. A Straightforward Guide to Functional Safety IEC 61508 (2010 Edition), IEC 61511 (2015 Edition) and Related Guidance: Forth Edition. Elsevier Ltd, 2016.
22. Chris Hobbs. Embedded Software Development for Safety-Critical Systems: Second Edition. CRC Press, 2020.
23. А. Л. Клевцов, М. А. Ястребенецкий, С. А. Трубочанинов. Компьютерная безопасность информационных и управляющих систем АЭС: нормативная база. Ядерная та радіаційна безпека 4(68), 2015.
24. MISRA website, the official resource for information on MISRA's publications and activities. [Електронний ресурс] – Режим доступу: <https://www.misra.org.uk/>
25. MISRA C. [Електронний ресурс] – Режим доступу: https://en.wikipedia.org/wiki/MISRA_C

Received (Надійшла) 28.01.2023

Accepted for publication (Прийнята до друку) 12.04.2023

Research of regulatory documents and industry standards for the development of safety-important computer systems for NPP control systems

Boris Vintenko, Oleksii Smirnov, Oleksandr Kovalenko, Serhii Smirnov, Anna Kovalenko

Abstract. Today, one of the main sources of electrical energy are nuclear power plants (NPP). In today's world, digital control systems with widespread use of computers and controllers are used to manage technological processes and ensure safety at nuclear power plants. The safe operation of the NPP depends on the quality and reliability of the software of such systems. In this work, a study of normative documents and standards related to the development of software for computer control systems of nuclear power plants, important for safety, was conducted. **The purpose of the work** is to define standards that can be used by software developers for computerized NPP control systems important for safety. **The object of research** is the process of developing software for computer control systems of nuclear power plants. **The subject is the study** of regulatory documents and industry standards for the development of software for computer control systems of nuclear power plants. **As a result of the study**, the safety criteria of computer control systems were determined, the IEC 61508 and IEC 61513 standards were studied, the IAEA SSG-39 safety instruction was studied, the categories of control functions and safety classes of NPP systems were determined, the standards containing software requirements were studied, user interfaces and MISRA C/C++ coding. **Conclusions.** In order to create highly reliable and high-quality software for computer control systems of nuclear power plants, which are important for safety, there are design standards and industry regulatory documents in the world. The result of the study of standards and industry regulatory documents relevant to the development of software for NPP computer control systems important for safety is the following: international standards and regulatory documents have different levels of specialization. The higher-level documents specify the general requirements for all stages of the life cycle of computer control systems and contain references to the documents of the next level. Documents of the next level are more specialized and relate to narrower fields of activity; each country of the world can determine its own set of standards and regulatory documents, which is required to be used during the design and implementation of computer control systems for nuclear power plants; regulatory documents contain requirements and recommendations both regarding the organization of the development process, as well as general recommendations and examples of writing program code in various programming languages; due to the fact that the requirements of the considered standards must be relevant for a long time, these standards do not contain formalized methods of writing program code; there are no formalized methods of code certification for compliance with the requirements of the considered standards; each organization that is engaged in the design, development and verification of software for computer control systems for nuclear power plants must develop its own methods for applying requirements and evaluating criteria for meeting the requirements of standards and regulations. Thus, it was concluded that the prospects for further research are a detailed analysis of the requirements of international standards and industry normative documents for the software of computer systems of NPP management and the development of a methodology for its writing and assessment of compliance with the requirements of the standards.

Keywords: software, computer control systems, nuclear power plant.

В. А. Глива¹, О. М. Тихенко², С. В. Зозуля², О. О. Козлітін²

¹ Київський національний університет будівництва і архітектури, Київ, Україна

² Національний авіаційний університет, Київ, Україна

ДОСЛІДЖЕННЯ ВПЛИВУ ЕЛЕКТРОСТАТИЧНИХ ПОЛІВ НА КОНЦЕНТРАЦІЇ АЕРОІОНІВ НА КОМП'ЮТЕРИЗОВАНИХ РОБОЧИХ МІСЦЯХ

Анотація. В статті наведено результати вимірювань концентрацій іонів та напруженості електростатичних полів на комп'ютеризованих робочих місцях. Виявлено пряму залежність концентрацій іонів від рівнів електростатичних полів. Визначено залежності концентрацій іонів від відстані до комп'ютера. Результати натурних вимірювань свідчать, що, не зважаючи на допустимі рівні напруженості електростатичних полів, електростатичне поле комп'ютера та іншого обладнання істотно впливає на концентрації іонів. Суттєвим чинником електризації поверхонь є також крісло оператора та випадкові тертя полімерних поверхонь, які періодично здійснюються користувачем і є невідворотними для будь-якої сидячої роботи. Зв'язок концентрацій аероіонів з електростатичними полями експериментально підтверджується зниженням концентрацій аероіонів знаку протилежного знаку електростатичного заряду. В результаті досліджень встановлено, що знак електростатичного заряду є непередбачуваним. Він визначається трибоелектричним ефектом внаслідок контакту різнорідних полімерних матеріалів, які генерують знаки зарядів, що обумовлюються контактною парою. Тобто, у процесі проектування заходів безпеки необхідно попередньо визначити можливі ступені електризації полімерних покриттів.

Ключові слова: іони, електростатичне поле, деіонізація, оператор.

Вступ

Фізико-хімічні властивості повітря, зокрема, концентрація іонів у повітрі, є одними із основних показників якості виробничого середовища. Якщо концентрації не відповідають нормам, то умови праці відносять до класу 3.1 за показниками шкідливості та небезпечності факторів виробничого середовища [1]. У сучасних виробничих процесах досить широко використовуються комп'ютери, повітря навколо яких, згідно даних [2-10] характеризується наднизькими концентраціями іонів, які нижчі за мінімально необхідний рівень. Відповідно, нормалізація іонного складу повітря робочих приміщень є актуальним науковим завданням.

Для вирішення завдань з нормалізації концентрацій іонів можна застосувати два підходи. Перший – класичний, полягає у застосуванні іонізаторів повітря або ж встановленні припливно-витяжної системи вентиляції. Але застосування таких заходів не завжди є можливим у виробничих умовах. А подекуди, навіть, не дає позитивних результатів для вирішення проблеми. Другий підхід полягає в мінімізації впливу факторів деіонізації повітря. Інформація про застосування цього підходу практично відсутня в науковій літературі, однак пропонується авторами, оскільки дає можливість нормалізувати концентрації іонів без витрат на встановлення спеціального обладнання.

Для мінімізації впливу деіонізуючих факторів, головною задачею є виділення основних чинників деіонізації та визначення їх питомої ваги у процесах деіонізації. До відомих чинників деіонізації повітря у приміщенні належать високі концентрації дрібнодисперсного пилу, аерозолів у повітрі, фактори життєдіяльності людини та електростатичні заряди. Сутність завдання полягає у визначенні внеску кожного з них у динаміку концентрацій іонів.

Аналіз літературних джерел. У низці робіт [2-10] наведено численні дані щодо результатів вимі-

рювань концентрацій іонів у комп'ютеризованих виробничих приміщеннях, які свідчать про невідповідність концентрацій іонів мінімально необхідному рівню. У деяких роботах причиною цих невідповідностей визначене електростатичне поле моніторів комп'ютерів [2, 6, 7]. Але при цьому не наводиться вагомої аргументації, що саме ці чинники стали причиною деіонізації. У деяких інших роботах [3-5, 8, 9] наводяться виключно дані моніторингу без аналізу причин, які спричинили деіонізацію повітря.

Дослідження [10] свідчать, що в деяких випадках є вплив зовнішнього повітря (особливо це стосується урбанізованих територій), що внаслідок високої концентрації аерозолів є повністю деіонізованим і при повітрообміні у приміщенні (внаслідок провітрювання) знижує концентрації іонів до нульових. Встановлено, що вплив комп'ютерної техніки на деіонізацію повітря в приміщенні значно менший за вплив зовнішнього повітря, яке є деіонізованим [10]. Тривале перебування оператора за комп'ютером призводить до зниження концентрацій іонів. Згідно класичної роботи О. Л. Чижевського причиною цього є утворення аерозолів у повітрі, яке видихається оператором. Проте, як показали дослідження [11], причиною є не дихання операторів, а осідання іонів на поверхні одягу, меблів або обладнання. тіла та одягом людини. Аргументом щодо цього висновку є результати моніторингу концентрацій іонів у зоні дихання оператора та у просторі поблизу оператора, у місцях, куди надходить повітря, що видихав оператор. За результатами досліджень встановлено відсутність різниці між цими концентраціями. Системні дослідження впливу користувача комп'ютера на концентрації іонів, свідчать, що різниця між концентраціями при наявності та за відсутності оператора за комп'ютером відрізнялася в межах похибки вимірювань. Зниження концентрацій негативних іонів за присутності оператора за комп'ютером, порівняно з вихідними кон-

центраціями, що були до початку його роботи на комп'ютері, було незначним (до 60 іонів/куб. см). Більш значним було зниження концентрацій позитивних аероіонів (до 100 іонів/см³).

Встановлено динаміку зміни концентрацій аероіонів при працюючому комп'ютері за відсутності оператора та до увімкнення. Результати досліджень свідчать, що робота комп'ютера призвела до зниження концентрацій позитивних легких іонів (до 200 іонів/см³). Наведені факти свідчать про можливість припущення, що електростатичні заряди є причиною зниження концентрації позитивних іонів. Яким належить провідна роль у деіонізації повітря при дослідженні впливу оператора на іони. Наведене вище дозволяє дійти висновку, що електростатичні заряди як чинник деіонізації повітря потребують додаткового дослідження з метою визначення його питомої ваги у загальній деіонізації повітря при роботі оператора.

Мета та завдання дослідження. Мета – дослідження динаміки концентрації аероіонів під впливом електростатичних зарядів, які виникають у процесі роботи оператора. Для досягнення цієї мети були поставлені наступні завдання:

- виконати вимірювання концентрацій іонів на робочому місці оператора за різних виробничих умов (вимкненого електрообладнання в приміщенні, увімкнених стаціонарних комп'ютерів);
- виміряти напруженості електростатичних полів поверхневих зарядів потенційних джерел деіонізації;
- виконати вимірювання концентрацій іонів на різних відстанях від комп'ютера.

Викладення основного матеріалу

Дослідження проводилися у виробничому приміщенні, розташованому на 6-ому поверсі промислової будівлі (8 × 9 × 3,5 м). У приміщенні відсутні системами припливної вентиляції та кондиціонування. В період вимірювань природна вентиляція була відсутня, повітря надходило у приміщення внаслідок

інфільтрації. Дослідження проводилося за відсутності людей та у неробочий час. В приміщенні були присутній лише дослідник, який фіксував результати вимірювань. Вимірювання концентрацій іонів здійснювалось з використанням «Сапфір-3к» та проводилося відповідно до інструкції з експлуатації приладу. Вимірювання здійснювались в зоні дихання оператора, на висоті 20 см над робочою поверхнею. Вимірювання напруженості електростатичного поля виконували з використанням вимірювача електростатичних зарядів «ІЕЗ-П» відповідно до паспорту приладу. Одночасно із вимірюванням концентрацій іонів проводились вимірювання потужності еквівалентної дози гамма-випромінювання (0,13 мкЗв/год) за допомогою дозиметра. Дослідження проводились упродовж двох днів.

Результати досліджень. Одночасно вимірювалися концентрації іонів та напруженості електростатичного поля (E) на робочому місці оператора та обладнання, яке знаходилося поряд. Результати вимірювань наведені в таблиці 1. Встановлено, що напруженість електростатичного поля на кріслі була різною в різних точках крісла: на сидінні – від 4 до 16 кВ/м, на спинці – від 2 до 4 кВ/м. Розташовані поряд монітор, принтер та БФП, генерували значне електростатичне поле. Найбільший внесок у формування якого вносив багатофункціональний пристрій.

Напруженість загального електростатичного поля на відстані 2 м від робочого місця оператора з боку комп'ютера та багатофункціонального пристрою складала 4 кВ/м (знак заряду - позитивний).

Результати досліджень динаміки концентрацій іонів на кожній стадії експерименту разом із характеристиками електростатичних полів наведено у табл. 2.

В результаті незалежно проведених вимірювань на відстані 30 см від крісла оператора були отримані такі концентрації іонів (негативні – 440 іонів/см³, позитивні – 290 іонів/см³). Встановлено пряму залежність між концентраціями іонів та відстанню до комп'ютера (рис. 1).

Таблиця 1 – Результати досліджень напруженостей електростатичних полів

Об'єкт дослідження	Напруженість електростатичного поля, кВ/м (відстань від об'єкту дослідження, см)	Знак заряду
Робоча поверхня (без комп'ютера)	0 (1)	
Дисплей	4 (1)	+
Клавіатура	9 (1); 4,7 (2); 4 (3)	-
Спинка крісла (після сидіння на кріслі користувача)	2-4 (1)	-
Сидіння крісла (після сидіння на кріслі користувача)	≈ 6,5 (1)	-
Принтер (був під'єднаний до мережі електроживлення, але не використовувався; розташований поряд з монітором)	4 (1)	-
Монітор комп'ютера		
Задня стінка	4 (1)	-
Верхня стінка	60 (1); 50 (3); 28 (5); 18 (7); 12 (10); 10 (15); 8 (20); 7,2 (25); 7,2 (30)	+
Багатофункціональний пристрій, (був під'єднаний до мережі електроживлення, але не використовувався; розташований поряд з монітором, але ближче до робочого столу, ніж сам монітор)		
Верхня кришка	9,2 (1); 9,2 (3); 12(5); 16,8 (7); 14 (10); 13,6 (15); 11,2 (20); 11,2 (25); 10 (30)	+
Бічна стінка (зі сторони оператора)	8,8 (1); 10,4 (3); 10,8 (5); 9,2 (7); 8,4 (10); 6,4 (15); 4,4 (20); 4 (25); 4 (30)	+

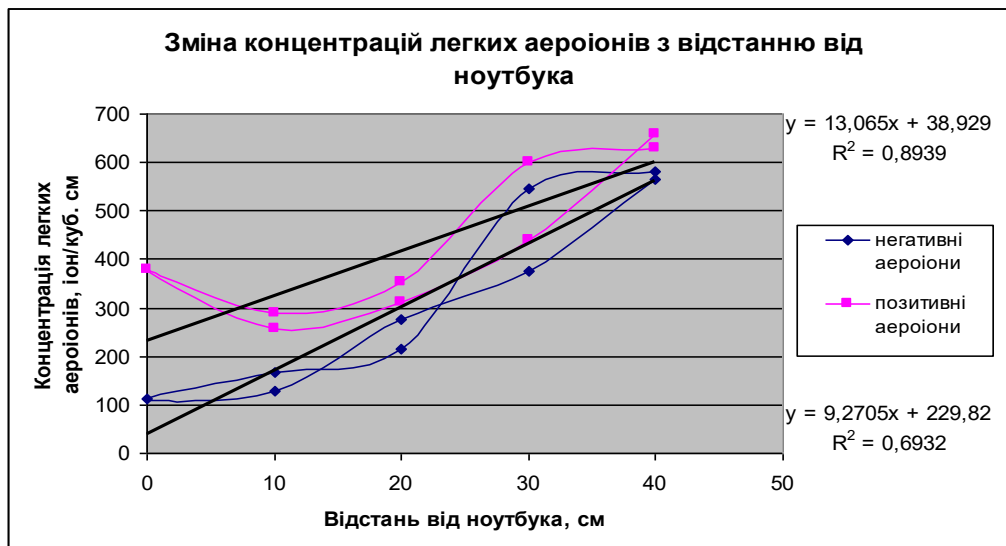


Рис. 1. Динаміка концентрацій іонів поблизу комп'ютера

Вихідні концентрації іонів у виробничому приміщенні, виміряні за вимкненого електричного обладнання, були більшими за мінімально необхідні значення. Показник полярності був притаманний природним концентраціям іонів, але ці концентрації були значно нижчими за оптимальні значення, встановлені нормативними документами. Напруженості полів електростатичних зарядів не перевищу-

вали гранично допустимого рівня. Також було встановлено, що при увімкненні обладнання (монітора комп'ютера, принтера та багатофункціонального пристрою) в електромережу, яке було розташоване на відстанях 2,5–3,0 м від контрольного робочого місця, спостерігалось зниження концентрацій іонів. При цьому переважно знижувалися концентрації негативних іонів.

Таблиця 2 – Результати досліджень концентрацій іонів на робочому місці оператора

Умови проведення досліджень	Концентрація іонів (різниця із попередньою концентрацією), іон/куб. см		Показник полярності, П	Коротка характеристика електростатичних полів
	негативні	позитивні		
Усі прилади в приміщенні вимкнені, результати вимірювання реєструвалися в паперовому журналі, на робочому столі встановлений лише лічильник аероіонів	910	1000	0,05	Е полів у приміщенні = 0
На відстані 3 м від робочого місця ввімкнено в електромережу комп'ютер (разом з принтером та багатофункціональним пристроєм) для реєстрації показів на окремому столі розташований лічильник аероіонів	770	890	0,07	На відстані 2,5-3 м від робочого столу – сильне електростатичне поле монітору та БФП (знак заряду +)
На відстані 3 м від робочого місця ввімкнені в електромережу комп'ютер (разом з принтером та багатофункціональним пристроєм) для реєстрації даних, на робочому столі розташовані ноутбук та лічильник аероіонів	520	800	0,2	До електростатичного поля монітору та багатофункціонального пристрою додається вплив поля дисплея та клавіатури ноутбуку
На відстані 3 м від робочого місця працює комп'ютер (разом з принтером та багатофункціональним пристроєм), оператор працює за ноутбуком	420	390	-0,04	До вказаних вище електростатичних полів додається поля, які генеруються на кріслі оператора (знак заряду негативний)

Це було обумовлено впливом електростатичного поля увімкненого обладнання, поверхневий заряд якого був позитивний.

Підключений до електромережі комп'ютер був причиною поступового зниження концентрацій, причому у більшості негативних іонів через вплив елек-

тростатичного поля (з позитивними зарядами) дисплею комп'ютера, який перебував ближче до зони дихання оператора, ніж клавіатура.

Електростатичне поле клавіатури певним чином також вплинуло на зниження концентрацій позитивних аероіонів.

Висновки

1. У контрольному приміщенні, за умови відсутності деіонізації повітря технічними засобами концентрації іонів обох полярностей у відповідають мінімально допустимим значенням. Для підтримання цього показника на нормативному рівні необхідно тільки мінімізувати вплив чинників деіонізації повітря. При цьому можна не застосовувати аероіонізуюче обладнання. Для досягнення оптимальних концентрацій аероіонів таке обладнання застосовувати необхідно.

2. За результатами виконаних досліджень встановлено, що не дивлячись на відповідність напруженості електростатичних полів чинним нормативним актам електростатичні поля поверхневих зарядів, які накопичуються на полімерних поверхнях комп'ютерів та іншого устаткування змінюють кон-

центрації іонів. Критичним джерелом електростатичного поля є поверхня крісла оператора, виготовлена із синтетичного матеріалу. Ці заряди виникають внаслідок трибоелектричного ефекту через випадкові рухи оператора (тобто, ковзання матеріалу одягу по поверхні). Прямий зв'язок концентрацій аероіонів з напруженостями електростатичних полів підтверджується переважним зниженням концентрацій аероіонів полярності, яка протилежна знаку поверхневого заряду.

3. Перспективою подальших досліджень є розроблення заходів зі зниження генерації електростатичних полів, які деіонізують повітря. Доцільно розглянути методологію нейтралізації поверхневих електростатичних зарядів за допомогою підвищення концентрацій аероіонів потрібної полярності. Це можливо реалізувати за рахунок застосування біполярного іонізатора повітря з керованою переважною полярністю генерації.

СПИСОК ЛІТЕРАТУРИ

1. ДСНтаП «Гігієнічна класифікація праці за показниками шкідливості та небезпечності факторів виробничого середовища, важкості та напруженості трудового процесу», затв. наказом МОЗ України від 8.04.2014 р № 248.
2. Ahmad I., Sattar Ab., Nawaz Al. Occupational health and safety in industries in developing world. *Gomal Journal of Medical Sciences*. 2016. Vol. 14. 223–228.
3. Глива В. А., Ніколаєв К. Д., Тихенко О. М., Тимошенко О. П. Дослідження рівнів фізичних факторів у приміщеннях диспетчерських служб аеропортів цивільної авіації. *Системи управління, навігації та зв'язку*. Полтава, 2019. Вип. 1(53). С. 32–35.
4. Wallner P., Kundi M., Panny M., Tappler P., Hutter H.-P. Exposure to Air Ions in Indoor Environments: Experimental Study with Healthy Adults *International Journal of Environmental Research and Public Health*. 2015. Vol. 12, Issue 11. P. 14301–14311. doi: <https://doi.org/10.3390/ijerph121114301>
5. Терещенко П. С. Гігієнічна оцінка фізичних факторів в сучасних офісних приміщеннях. *Гігієна населених місць*. 2011. № 57. С. 216–219.
6. Глива В.А., Левченко Л.О., Тихенко О.М. Методи визначення концентрацій аероіонів у приміщеннях та моделювання їх змін. *Системи управління, навігації та зв'язку*. 2018. Вип. 4 (50). С. 157–160.
7. Magnier-Bergeron L., Derome D., Zmeureanu R. Three-dimensional model of air speed in the secondary zone of displacement ventilation jet. *Building and Environment*. 2017. Vol. 114. P. 483–494. doi: <https://doi.org/10.1016/j.buildenv.2020.101603>
8. Панова О. В., Левченко Л. О., Теслицький І. А. Дослідження аероіонізації повітря у приміщеннях з експлуатації комп'ютерної техніки. *Науково-технічний збірник «Комунальне господарство міст»*. 2021. Т. 4 № 164. С.215–219.
9. Ченчевой В., Данова К., Ченчева О., Перекрест А., Григор'єва Д. Наукове обґрунтування оптимізації складу гідроаероіонізатору в громадських приміщеннях для осіб з особливими потребами. *Проблеми охорони праці в Україні*. 2020. 36 (4), 8–15. <https://doi.org/10.36804/ndipbop.36-4.2020.8-15>
10. Глива В. А., Левченко Л. О., Панова О. В., Тихенко О. М. Вплив мікрокліматичних параметрів на аероіонізацію повітря виробничого середовища. *«Містобудування та територіальне планування»*. 2018. Вип. № 68, С. 108–116.
11. Ченчевой В. В., Сукач С. В., Ченчева О. О., Федорова Н. С., Григор'єва Д. С. Дослідження параметрів гідроаероіонізатору робочого приміщення з ультразвуковою іонізацією. *Вісті Донецького гірничого інституту*. 2020. Вип. № 2(47). С. 168–174. <https://doi.org/10.31474/1999-981x-2020-2-168-175>

Received (Надійшла) 10.04.2023

Accepted for publication (Прийнята до друку) 07.06.2023

Study of the influence of electrostatic fields on the concentration of aeroions at computerized workplaces

Valentyn Glyva, Oksana Tykhenko, Sergey Zozulya, Oleksii Kozlitiin

Abstract. The article presents the results of measurements of ion concentrations and electrostatic field strengths at computerized workplaces. A direct dependence of ion concentrations on the levels of electrostatic fields was revealed. The dependence of ion concentrations on the distance to the computer was determined. The results of field measurements show that, regardless of the permissible levels of electrostatic field intensity, the electrostatic field of the computer and other equipment significantly affects the concentration of ions. A significant factor in the electrification of surfaces is also the operator's chair and accidental friction of polymer surfaces, which are periodically carried out by the user and are inevitable for any sedentary work. The connection of aeroion concentrations with electrostatic fields is experimentally confirmed by a decrease in aeroion concentrations of the opposite sign of the electrostatic charge. As a result of research, it was established that the sign of the electrostatic charge is unpredictable. It is determined by the triboelectric effect as a result of the contact of dissimilar polymer materials that generate charge signs determined by the contact pair. That is, in the process of designing safety measures, it is necessary to determine in advance the possible degrees of electrification of polymer coatings.

Keywords: ions, electrostatic field, deionization, operator.

Г. М. Кочетов, Д. М. Самченко, О. В. Ластівка

Київський національний університет будівництва і архітектури, Київ, Україна

МЕТОДОЛОГІЧНІ ЗАСАДИ РОЗРОБЛЕННЯ РІДКИХ СУМІШЕЙ ДЛЯ ЕКРАНУВАННЯ ЕЛЕКТРОМАГНІТНИХ ПОЛІВ

Анотація. Показано, що в умовах підвищення амплітуд та розширення частотного спектра техногенних електромагнітних полів найбільш прийнятними екрануючими матеріалами є композити з потрібними властивостями. Особливістю композиційних матеріалів усіх класів є відсутність стандартних довідкових даних щодо їх ефективності та електрофізичних і магнітних властивостей, які визначають коефіцієнти екранування. Для спрощення процесів проектування екрануючих сумішей доцільне попереднє розрахункове оцінювання значень діелектричної і магнітної проникностей та питомої провідності кінцевого матеріалу. Проаналізовано можливість використання для цих розрахунків співвідношень Лорентца, Максвелла-Гарнета, Оделевського та Ліхтенеккера. Показано пов'язані з цим труднощі і межі застосування та необхідність врахування морфології екрануючих частинок у матриці. Наголошено, що у процесі проектування слід враховувати залежність електрофізичних властивостей компонентів від частоти поля, яке потребує екранування. Такий підхід дозволяє отримати рідкі матеріали з різними властивостями для формування багатошарової структури: верхній шар може мати мінімальні коефіцієнти відбиття, а нижній – максимальні поглинальні властивості. Надано послідовність дій при проектуванні захисних композицій для раціоналізації проектних робіт.

Ключові слова: електромагнітне поле, екранування, композиційний матеріал, електрофізичні властивості.

Вступ

Пріоритетним напрямом досліджень та прикладних розробок у галузі екранування електромагнітних полів є створення захисних композицій на основі лакофарбових та полімерних матеріалів. Найбільш прийнятними основами (матрицями) є фарби, що обумовлюються їх високими функціональними властивостями – зчеплення з поверхнею нанесення, незначна деградація внаслідок фізико-хімічних впливів. Але застосування фарб у процесі проектування захисних композицій має низку проблем. Додавання до фарби екрануючих домішок може критично змінювати адгезійні та когезійні властивості носія, а також робить непередбачуваними електрофізичні та магнітні властивості суміші (питому електропровідність, діелектричну та магнітну проникність). Визначити функціональні параметри кінцевого продукту можливо експериментально, але це вимагає виготовлення та випробування великої кількості дослідних зразків, значної витрати коштів і часу, і не є раціональним. Тому доцільно розробити цілісну систему попереднього оцінювання необхідних властивостей композиції, які забезпечують потрібні захисні властивості матеріалу у залежності від амплітудно-частотних характеристик електромагнітних полів, які потребують екранування.

Огляд останніх публікацій за темою дослідження. Розробленню рідких композицій для екранування електромагнітних полів широкого частотного діапазону та електричних і магнітних полів наднизьких частот приділяється багато уваги. Більшість робіт є експериментальними, що пояснюється зазначеним вище. Так, у дослідженнях [1, 2] на стандартному синтетичному носії отримані матеріали малої товщини (до 100 мкм) з високими коефіцієнтами екранування (20 дБ і вище). Їх недоліком є висока вартість наповнювачів (нанотрубки, графені, графітізована сажа) та наявність кількох компонентів наповнювача, що робить їх метастабільними. У роботі [3] пропонується носій на водній основі. Слід зазначити, що водно-емульсійна фарба, у якості матриці показала найгірші результати [4]. Крім того,

у матеріалі на основі залізородного концентрату низькі коефіцієнти екранування і не відповідають сучасним вимогам. Значна частина досліджень присвячена застосуванню наноматеріалів у якості екрануючого наповнювача [5]. Такі матеріали мають велику вартість через складність технологій отримання магнітної рідини з вмістом наночастинок. Коефіцієнти відбиття електромагнітних хвиль композиційних матеріалів принаймні у декілька разів нижчі, ніж у металів та сплавів. Узгодженість імпедансу поверхні з показником повітря (377 Ом) знижує загальний коефіцієнт екранування. У розробці [6] пропонується створення градієнтних магнітних властивостей матеріалу. Але оброблення рідкої композиції неоднорідним полем дозволяє отримати градієнт тільки під час виготовлення матеріалу, а не у процесі його нанесення на поверхню. Тому доцільно розглядати багатошарові структури з малими товщинами шарів, але різними фізичними властивостями [7]. Для забезпечення високої ефективності багатошарових структур потрібне визначення електрофізичних і магнітних властивостей кожного шару, що потребує коректних розрахункових методів. Тому доцільним є обґрунтування та розроблення певного алгоритму щодо проектування захисних матеріалів (у тому числі і багатошарових) для безпосереднього нанесення на елементи будівель і конструкцій.

Мета роботи – розроблення концептуальних засад проектування рідких композицій для екранування електромагнітних полів широких частотних діапазонів.

Викладення основного матеріалу

Для проектування захисних властивостей композиції необхідно попередньо оцінити її електрофізичні та магнітні властивості з урахуванням об'ємної частки екрануючого наповнювача у діелектричній матриці. Для визначення магнітної проникності суміші існує декілька можливостей розрахунку. Теоретичною основою розрахунків для матеріалів з впливом геометричної структури, наприклад магнітодіелектриків, є поняття структурного фактору розмагнічування:

$$\mu_k = B/(\mu_0 H) = \mu_n / (1 + N(\mu_n - 1)),$$

де μ_k – магнітна проникність композиту, μ_0 – магнітна стала, H – напруженість магнітного поля за відсутності магнітного матеріалу, μ_n – магнітна проникність частинок наповнювача, N – коефіцієнт розмагнічування.

$$N = (1 + 1/(\mu_n - 1)) / (\mu_k - 1/(\mu_n - 1)).$$

У випадку, коли μ_k набагато нижча за проникність феромагнітної фази, це співвідношення може бути використане для оцінювання структурного коефіцієнта розмагнічування. Наприклад, за $\mu_k=5$, а $\mu_n=200$ (конструкційна сталь) похибка розрахунку складає 2,5 %.

Практичне значення мають співвідношення, визначені для малого (до 0,3) заповнення матриці феромагнітним матеріалом. Розрахунок для системи регулярно розташованих частинок (формула Лорентца):

$$\mu_k = 1 + V_n(\mu_n - 1) / (1 + (1 - \mathcal{G}_n(\mu_n - 1))/3),$$

де V_n – відносна об'ємна частина наповнювача у матеріалі. Не дивлячись на певну громіздкість співвідношення, за умови немагнітності матеріалу матриці та $\mu_n \gg 0$ магнітні властивості матеріалу визначаються тільки його структурою:

$$\mu_k = (1 + 2V_n) / (1 - V_n),$$

Але для малих значень магнітної проникності (10–30), характерної для феритів, доцільно користуватися формулою Максвелла:

$$\mu_k = 1 + 3V_n(\mu_n - 1) / (\mu_n + 2),$$

Її перевагою є достатня точність за умови, що магнітна проникність матеріалу матриці відмінна від одиниці. Прийнятний збіг з експериментом дають розрахунки за формулою Ліхтенеккера:

$$\mu_k = \mu_n^{V_n} \mu_m^{(1-V_n)},$$

де μ_m – магнітна проникність матеріалу матриці.

$$\text{За } \mu_m=1, \mu_k = \mu_n^{V_n}.$$

У загальному випадку це співвідношення має обмеження – воно передбачає постійне зростання магнітної проникності композиту із збільшенням магнітної проникності наповнювача, що не зовсім коректно. Крім того передбачається, що частинки наповнювача мають форму, близьку до сферичної, що на практиці не завжди відповідає реальним формам наповнювача. Ці особливості розглянемо нижче.

Формула Ліхтенеккера дає прийнятний результат у разі заміни магнітної проникності на провідність σ . А цей параметр σ критичний за досягнення перколяційного ефекту – зростання провідності композиту внаслідок формування провідних витягнутих структур:

$$\sigma_k = \sigma_n^{V_n} \sigma_m^{(1-V_n)},$$

де σ_k , σ_n , σ_m – питомі провідності композиту, наповнювача та матриці.

Як зазначалося, для більш точного визначення магнітних властивостей магнітодіелектрика необхідно

враховувати морфологію частинок наповнювача – співвідношення довжини та діаметру частинок. Як показано у роботі [8], найприйнятнішими співвідношеннями є значення 2/1. Цей параметр враховує коефіцієнт деполяризації F :

$$F = \ln(2l/d) / (1/d)^2,$$

де l – довжина частинки, d – діаметр.

Зазвичай частинки різняться за параметрами, тому у практичних розрахунках доцільно користуватися середніми значеннями для подовжених структур. Ще одним важливим параметром композиту, принаймні для визначення коефіцієнтів відбиття електромагнітних хвиль, є діелектрична проникність матеріалу. Ефективну діелектричну проникність можна оцінити, виходячи зі співвідношення Максвелла-Гарнета:

$$\frac{\varepsilon_k - \varepsilon_m}{\varepsilon_k + 2\varepsilon_m} = V_n \frac{\varepsilon_n - \varepsilon_m}{\varepsilon_n + 2\varepsilon_m},$$

де ε_k , ε_m , ε_n – відповідно діелектричні проникності композиту, матриці та наповнювача. Співвідношення дає прийнятні результати за $V_n < 1/3$. Такі пропорції прийнятні з огляду на те, що велика кількість наповнювача може критично впливати на адгезійні і когезійні властивості кінцевого матеріалу.

Для визначення критичного вмісту наповнювача, за якого різко зростає провідність матеріалу, доцільно скористатися формулою Оделевського:

$$\varepsilon_k = \varepsilon_m \left[1 + \frac{V_n(\varepsilon_n - \varepsilon_m)}{(1 - V_n/V_k)F(\varepsilon_n - \varepsilon_m) + \varepsilon_m} \right],$$

де V_k – критична об'ємна концентрація наповнювача, за якої екрануючі частинки контактують між собою.

V_k визначається як:

$$V_k = [3F(1-F)]^{0,6} / [4F\sqrt{l/d} + (1-F)/\sqrt{l/d}]^{0,6}.$$

Застосування наведених вище співвідношень допомагає певним чином раціоналізувати як обирання компонентів композиції, так і їх співвідношення у залежності від мети розробки. Слід враховувати, що усі розрахунки містять об'ємні частки матеріалів, що не завжди зручно. За додавання у рідкий матеріал насипного наповнювача необхідно перераховувати густини у твердому стані у об'ємні, виходячи з густин наповнювача і матриці. При цьому важливою може бути зміна густини матеріалу матриці у процесі висихання або полімеризації.

У процесі проектування захисних сумішей існує ще одна проблема, пов'язана з вихідними даними – деяка відмінність стандартних проникностей матеріалів від проникностей у дрібнодисперсній субстанції (вона зменшується, наприклад, для традиційних феромагнітних матеріалів – заліза, нікелю, пермалоїв). Тому у процесі проектування слід робити деякий запас по потрібним властивостям.

Застосування співвідношення Ліхтенеккера може давати некоректний результат через окислення поверхонь екрануючих частинок, що заважає прояву перколяційного ефекту. Це треба враховувати обираючи компоненти композиції. Ще однією проблемою є

частотна залежність магнітної проникності. Наприклад, для заліза, проникність по поглинанню за підвищення частоти поля від 1 ГГц до 10 ГГц знижується з 20–22 до 1–2. Тому у практичній діяльності необхідні дані потрібно визначати експериментально, або з'ясувати із спеціальних наукових джерел.

Проектування рідких екрануючих сумішей доцільно здійснювати за певним алгоритмом. На першому етапі необхідно з'ясувати амплітудно-частотні характеристики електромагнітного поля, яке потребує екранування та можливе співвідношення захисту за рахунок відбиття та поглинання електромагнітних хвиль. На другому етапі визначається кількість шарів матеріалу та електрофізичні і магнітні властивості кожного шару. На третьому етапі розраховуються співвідношення параметрів компонентів та обираються прийнятні для цього матеріали матриці та наповнювача.

Враховуючи практично повну відсутність довідкових даних щодо сучасних композиційних матеріалів та багатьох прийнятних для їх виготовлення компонентів такий підхід певним чином спростить проектування композицій для екранування електромагнітних полів визначених частот і смуг частот потрібної ефективності.

Висновки

1. Відсутність довідкових даних щодо екрануючих властивостей композиційних матеріалів та їх електрофізичних і магнітних властивостей ускладнює процес розроблення захисних матеріалів і конструкцій та потребує попереднього розрахункового оцінювання прийнятних параметрів.

2. Визначення діелектричної та магнітної проникностей і питомої провідності кінцевого матеріалу доцільно здійснювати з використанням співвідношень Лорентца, Максвелла-Гарнета, Оделевського та Ліхтенеккера. При цьому слід враховувати форму і розміри екрануючих частинок у матриці та визначати межу перколяційного ефекту – різкого зростання провідності внаслідок контакту частинок наповнювача. Слід враховувати частотну залежність властивостей наповнювача.

3. Проектування композиційного матеріалу слід здійснювати у певній послідовності, виходячи з необхідного співвідношення захисту за рахунок відбиття та поглинання електромагнітних хвиль і пов'язаної з цим кількості та характеристик окремих шарів матеріалу.

СПИСОК ЛІТЕРАТУРИ

1. Butenko, O., Boychuk, V., Savchenko, B., V Kotsyubynsky, Barsukov, V. Pure ultrafine magnetite from carbon steel wastes. *Materials Today: Proc.*, 2019, v. 6, pp. 270–278 <https://www.sciencedirect.com/science/article/pii/S221478531832399X>
2. Senyk, I., Kuryptia, Y., Barsukov, V., Butenko, O., Khomenko, V. Development and application of thin wide-band screening composite materials. *Physics and Chemistry of Solid State*, 2020, 21(4), pp. 771–778.
3. Tudose Ioan Valentin, Mouratis Kyriakos, Ionescu Octavian Narcis, Romanitan Cosmin, Pachiu, Cristina, Popescu Marian, Khomenko Volodymyr, Butenko Oksana, Chernysh Oksana, Kenanakis George, Barsukov Viacheslav Z., Sucheia Mirela Petruta, Koudoumas Emmanouel, *Novel Water-Based Paints for Composite Materials Used in Electromagnetic Shielding Applications*, *Nanomaterials*, 2022, 12(3), 487, <https://www.mdpi.com/2079-4991/12/3/487>
4. V. Glyva, V. Bakharev, N. Kasatkina, O. Levchenko, L. Levchenko, N. Burdeina, S. Guzii, O. Panova, O. Tykhenko, Y. Biruk. 2021. Design and study of protective properties of liquid composite materials for screening electromagnetic fields *European Journal of Enterprise Technologies. Materials Science*. Vol 3/6, No (111), pp.25-31. DOI: 10.15587/1729-4061.2021.231479
5. Glyva V.A., Podoltsev A.D., Bolibrukh B.V. A Thin Electromagnetic Shield of a Composite Structure Made On the Basis of a Magnetic Fluid. *Tekhnichna elektrodynamika*. 2018. № 4. P.14–18. DOI: <https://doi.org/10.15407/technd2018.04.014>
6. Градієнтний електромагнітний екран Пат. 144619, Україна МПК G12B 17/00 (2020.01). Глива В.А., Кажан К.І., Левченко Л.О., Панова О.В., Тихенко О.М., Халмуродов Б.Д., Володілець: Глива В.А., Кажан К.І., Левченко Л.О., Панова О.В., Тихенко О.М., Халмуродов Б.Д. № u 2020 03224; заявл. 28.05.2020; опубл. 12.11.2020, Бюл. № 19.
7. Бурдейна Н.Б., Бірук, Я.І., Ніколаєв К.Д. (2023). Розроблення матеріалів багат шарової структури градієнтного типу на основі рідких композицій для екранування електромагнітних полів. Екологічна безпека та природокористування, 45 (1), 68–75. <https://doi.org/10.32347/2411-4049.2023.1.68-75>
8. Касаткіна Н. В., Тихенко О. М., Панова О. В., Бірук Я. І. Підвищення ефективності композиційних електромагнітних екранів регулюванням морфології феромагнітного наповнювача. «Системи управління навігації та зв'язку», - 2020. Вип. № 3(61), С. 115-119. doi: 10.26906/SUNZ.2020.3.115

Received (Надійшла) 28.02.2023

Accepted for publication (Прийнята до друку) 10.05.2023

Methodological principles of development of liquid mixtures for the shielding of electromagnetic fields

G. Kochetov, D. Samchenko, O. Lastivka

Abstract. It is shown that in conditions of increasing amplitudes and expanding the frequency spectrum of man-made electromagnetic fields, the most suitable shielding materials are composites with the required properties. A feature of composite materials of all classes is the lack of standard reference data on their efficiency and electrophysical and magnetic properties, which determine the shielding coefficients. To simplify the design processes of shielding mixtures, it is advisable to make a preliminary calculation of the values of the dielectric and magnetic permeability and the specific conductivity of the final material. The possibility of using the Lorentz, Maxwell-Garnett, Odelevsky and Lichtenecker ratios for these calculations was analyzed. The related difficulties and limits of application and the need to take into account the morphology of shielding particles in the matrix are shown. It is emphasized that the dependence of electrophysical properties of components on the frequency of the field that requires shielding should be taken into account in the design process. This approach makes it possible to obtain liquid materials with different properties for the formation of a multilayer structure: the upper layer can have the minimum reflection coefficient, and the lower one - the maximum absorption properties. The sequence of actions in the design of protective compositions for the rationalization of project works is given.

Keywords: electromagnetic field, shielding, composite material, electrophysical properties.

Т. М. Ткаченко¹, Н. Б. Бурдейна¹, О. О. Ченчева²

¹Київський національний університет будівництва і архітектури, Київ, Україна

²Кременчуцький національний університет імені М. Остроградського, Кременчук, Україна

ЕКРАНУВАННЯ ЕЛЕКТРОМАГНІТНИХ ПОЛІВ ТА ШУМУ У БУДІВЛЯХ І СПОРУДАХ

Анотація. Аналіз сучасного стану електромагнітного та акустичного забруднення виробничого та навколишнього середовища свідчить про необхідність обґрунтування загальних засад захисту будівель універсальними захисними конструкціями. Показана можливість застосування витких рослин для покриття поверхонь будівель. Каркасом для поширення витких рослин доцільно обрати регулярну структуру – сітчасту або шпарувату. Ефективність таких структур розраховано, виходячи з фундаментальних співвідношень електродинаміки. Отримані співвідношення для розрахунку коефіцієнтів екранування для нормально падаючої електромагнітної хвилі та для хвилі, яка падає на екран під довільним кутом. Показано співвідношення довжин хвиль електромагнітного та акустичного полів, які екрануються регулярною структурою і визначається через швидкості розповсюдження електромагнітних та акустичних хвиль. Впровадження витких рослин для захисту від фізичних впливів необхідними є експериментальні дослідження, які дозволять розробити емпіричні або напівемпіричні співвідношення щодо попереднього оцінювання ефективності захисту. Надано розрахунковий апарат для обчислення параметрів перфорованого або шпаруватого екрана для екранування шуму. Його структура ґрунтується на визначенні параметрів конструкції, виходячи з частоти звуку з найбільшою амплітудою, яка розглядається як резонансна. Така конструкція також може бути носієм витких рослин. Перевагою таких конструкцій є пристосованість для екранування низькочастотного звуку, а виткі рослини розширюють діапазон захисту. Крім того, виткі рослини сприяють енергозощадженню.

Ключові слова: екранування, електромагнітне поле, шум, виткі рослини.

Вступ

Електромагнітні та акустичні поля техногенного походження є найбільш вагомими фізичними факторами несприятливого впливу на працюючих та населення. Найбільш ефективним засобом захисту від змінних електромагнітних полів та шуму є застосування екранування відповідними матеріалами та конструкціями з них. Такий захист гарантовано ефективний для ультрависоких та вищих частот електромагнітних полів високих частот акустичних полів, принаймні для невеликої ділянки або приміщення, що обумовлене малими довжинами електромагнітних і акустичних хвиль. Для них дифракційні явища на кромках екранів не є суттєвими, а проникність у захищену зону – мінімальна. Це ж стосується засобів індивідуального захисту працюючих. Для електромагнітних полів дуже високих частот і частот нижчих діапазонів та звукових хвиль середніх і низьких частот ситуація ускладнюється. Особливо це стосується електромагнітних полів наднизької частоти (в основному – 50 Гц) та акустичних низької – до 250–300 Гц. Навантаження на виробниче та навколишнє середовище з боку полів таких частот зростає. Цей процес супроводжується підвищенням внеску у електромагнітне навантаження частот бездротового зв'язку різного призначення. У таких умовах найбільш раціональним є застосування засобів колективного захисту. Таким захистом є екранування усєї будівлі або її частини з урахуванням амплітудно-частотних характеристик фізичних полів, які розповсюджуються від найбільш критичних джерел. Доцільно розглянути комбіновані засоби захисту, ефективні щодо екранування як електромагнітних, так і акустичних полів.

Стан питання. В умовах поступового переходу України на загальноєвропейські стандарти у галузі

безпеки працюючих та населення, при виконанні досліджень і прикладних розробок доцільно керуватися вимогами відповідних директив ЄС, які крім загальних вимог безпеки містять кількісні дані щодо гранично допустимих значень шкідливих та небезпечних чинників виробничого середовища. Це загальна вимога [1] та директиви щодо захисту від впливу електромагнітних полів та шуму [2–4]. Методам та засобам захисту від цих фізичних чинників присвячено багато досліджень. Тенденцією останніх років є розроблення та дослідження захисних властивостей композиційних матеріалів для екранування фізичних полів. Більшість з них стосуються електромагнітних полів [5–7]. У цих роботах надано коефіцієнт екранування електромагнітних полів широкого частотного діапазону. При цьому підкреслюється, що дослідження з екранування магнітних полів промислової частоти виконувалися для геометрично замкнених конструкцій, що обумовлене квазістаціонарністю цих полів. У реальних виробничих умовах, принаймні для окремого приміщення, суцільне екранування не завжди прийнятне. Дослідження щодо композиційних матеріалів для захисту від шуму практично відсутні. Це пояснюється тим, що поглинання механічних хвиль залежить від значень модулів – пружних характеристик матеріалів, а ці дані для композитів не є довідковими величинами через їх значну залежність від складу оброблення сумішей. У дослідженні [8] наведено результати розроблення і випробування захисних властивостей універсального електромагнітного та акустичного екрана. Але екранування акустичних хвиль досягається виключно за рахунок неоднорідності матеріалу. Ця неоднорідність досягається за рахунок застосування латексу у вигляді пінолатексу та додаванням у латекс гранульованого пінополістиролу. Неоднорідні структури також використовуються для захисту від електромагнітних полів

як наднизьких частот [9], так і ультрависоких і вищих частот [10]. При цьому приблизні розрахунки щодо перфорованих неоднорідних звукопоглинальних панелей відомі і наведені у довідковій літературі. Загальновідомо, що зелені насадження знижують рівні шуму від автомобільних шляхів, аеродромів цивільної авіації. Але їх застосування для захисту окремих будівель проблематичні, наприклад, через геометричні характеристики захисних смуг з дерев. У роботах [11, 12] наведено результати досліджень щодо застосування витких рослин для терморегуляції приміщень і будівель. Застосування таких конструкцій може мати комплексний характер [13]. Виходячи з цього, доцільно розглянути можливість одночасного використання для екранування електромагнітних та акустичних полів металевих регулярних структур у комплексі з виткими рослинами.

Метою роботи є розроблення концептуальних засад забезпечення електромагнітної та акустичної безпеки працюючих та населення за рахунок екранування електромагнітних полів та шуму у захисних конструкціях.

Викладення основного матеріалу

Ідея роботи полягає у застосуванні регулярної металевої структури для екранування електромагнітних полів, яка слугує каркасом для поширення витких рослин, що екранують акустичний шум.

Перевагою регулярних металевих структур є те, що їх ефективність можна розрахувати, виходячи із фундаментальних співвідношень електродинаміки. Розрахунки ґрунтуються на дифракції електромагнітних хвиль на решітках або ґратках, вироблених з провідних матеріалів. Для електромагнітних полів, принаймні дуже високих і вищих частот, питома провідність дротів, з яких виготовлено сітку не є критичною. Тому для виготовлення такої сітки можна використовувати дріт-катанку з номенклатури, яка виробляється металургійною промисловістю. Це виявилось можливим навіть для екранування магнітного поля промислової частоти [9]. При нормальному падінні на металеву сітку електромагнітної хвилі коефіцієнт екранування визначається як:

$$K_s = 10 \lg \left(4 \left(\frac{d}{\lambda} \ln \frac{d}{2\pi r} \right)^2 \right) / \left(1 + 4 \left(\frac{d}{\lambda} \ln \frac{d}{2\pi r} \right)^2 \right),$$

де λ – довжина падаючої хвилі, d – відстань між дротами (крок сітки), r – радіус дротів.

При падінні хвилі під кутом φ :

$$K_r = 10 \lg \left(4 \left(\frac{d \cos \varphi}{\lambda} \ln \frac{d}{2\pi r} \right)^2 \right) / \left(1 + 4 \left(\frac{d \cos \varphi}{\lambda} \ln \frac{d}{2\pi r} \right)^2 \right),$$

$$K_s = 10 \lg \frac{(1 - \cos \varphi)^2 + 4 \left(\frac{d}{\lambda} \ln \left(\frac{d}{2\pi r} \right) \right)}{1 + 4 \left(\frac{d}{\lambda} \ln \left(\frac{d}{2\pi r} \right) \right)}.$$

Слід врахувати, що ці співвідношення коректні за умов:

$$d/\lambda < 10, \quad r/\lambda < 0,04, \quad r/d < 0,1.$$

При цьому сітка повинна перебувати у хвильовій зоні електромагнітного поля. Існує експериментально визначена вимога, що лінійні розміри екрана (довжина і ширина) повинні перевищувати довжину падаючої хвилі принаймні у 5 разів, що виконується у разі облицювання поверхонь будівель для електромагнітних полів усіх типових джерел, крім полів промислової частоти. Для поширених довжин хвиль (радіотехнічних об'єктів цивільної авіації, мобільного та радіорелейного зв'язку) та типових сіток залежності екрануючих властивостей від цих параметрів наведені на рис. 1–3.

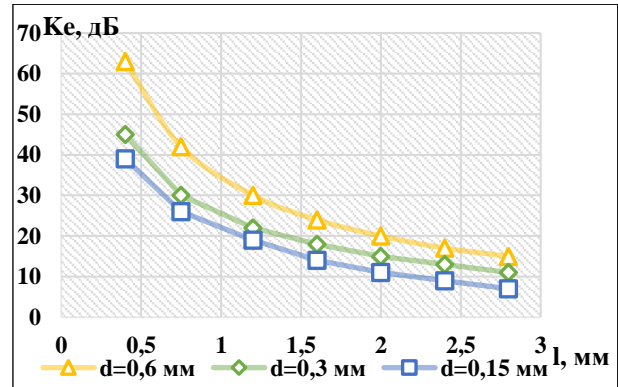


Рис. 1. Залежність коефіцієнта екранування від параметрів металевої сітки $\lambda=0,03$ м (довжина хвилі); l – розмір комірки сітки; d – діаметр дротів, з яких виготовлено сітки

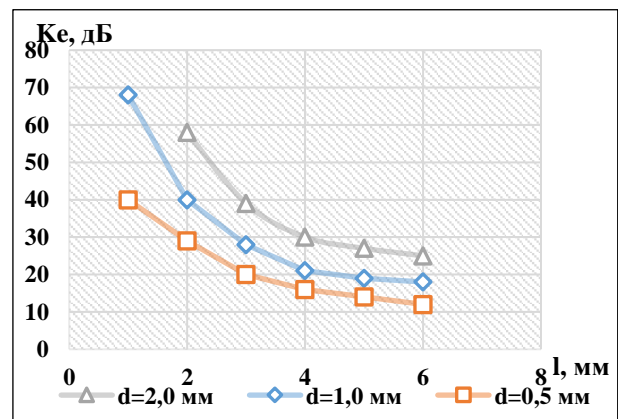


Рис. 2. Залежність коефіцієнта екранування від параметрів металевої сітки $\lambda=0,1$ м (довжина хвилі); l – розмір комірки сітки; d – діаметр дротів, з яких виготовлено сітки

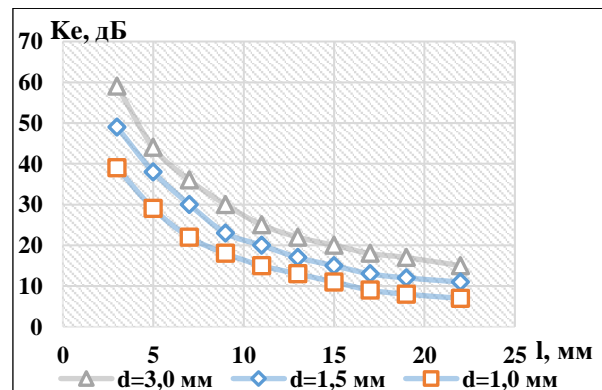


Рис. 3. Залежність коефіцієнтів екранування електромагнітного випромінювання частотою 0,9 ГГц від розміру комірки сітчастої структури, виготовленої з дротів діаметрами 1,0 мм, 1,5 мм, 3,0 мм

Досвід натурних вимірювань свідчить, що за кроку сітки у $0,1\lambda$ структура стає повністю непроникною для електромагнітного поля. Це дозволяє обрати параметри захисту, які забезпечують мінімально необхідні сигнали базових станцій мобільного зв'язку. Такі структури можуть бути використані як каркас для поширення витких рослин. Ці рослини через великий вміст води частково екранують електромагнітні поля (принаймні ультрависоких і вищих частот), але розрахувати цей параметр практично неможливо через непередбачуваність товщини і щільності покриву.

Зелені насадження на прийнятному рівні здійснюють звукопоглинання. Перевагою витких рослин є покриття безпосередньо об'єкту захисту, що за великих довжин звукових хвиль дозволяє мінімізувати дифракційні явища. На сьогоднішній день навіть приблизний розрахунок ефективності звукопоглинання рослинами не уявляється можливим, тому доцільно спиратися на вимірювання індексів зниження шуму зеленими панелями. За даними професора Вінчука М.М. шумозахисні властивості витких рослин значні (табл. 1).

Таблиця 1 – Зниження рівнів шуму виткими рослинами

Дівочий виноград'ятилисточковий	Відстань від джерела шуму, м	Показання шумометра, dB	
		Без рослин	З рослиною
	1	79,9	72,7
	3	73,8	69,1
	5	69,7	65,6
Плющ звичайний		Без рослин	З рослиною
	1	81,3	76,8
	3	75,2	68,3
Жимолость в'юнка		Без рослин	З рослиною
	1	71,7	64,6
	3	66,8	62,2
	5	63,2	57,6

Знаючи параметри несучого сітчастого каркасу можливо оцінити його внесок у захист від шуму. Це визначається фундаментальним співвідношенням:

$$f = v/\lambda,$$

де λ – довжина падаючої хвилі, f – частота, v – швидкість поширення хвилі.

Різниця коефіцієнтів екранування на однаковій структурі електромагнітних та звукових хвиль у загальному випадку визначається різними швидкостями їх поширення. Наприклад, структура ефективна для екранування електромагнітного поля частотою 1,8 ГГц (найбільш поширена в Україні частота мобільного зв'язку), буде ефективною для звукових хвиль частотою приблизно 2 кГц. Ця частота зростає зі зменшення кроку сітки. Ще однією можливістю комбінованого захисту від впливу двох фізичних чинників є застосування перфорованих та шпаринчастих регулярних структур. Основою їх застосування є теоретичні дослідження [14], за якими можна з мінімальною похибкою розрахувати ефективність екранування електромагнітних полів визначених частот.

Перфорований або шпаринчастий екран можна застосувати для захисту від впливу шуму [15]. Його ефективність, принаймні на переважних частотах, можна розрахувати. Параметри екрана обираються виходячи з частоти найбільшої амплітуди. Екрануюча панель працює за резонансним принципом, а резонансною частотою є частота з найбільшою амплітудою коливань. Така панель може бути каркасом для витких рослин, які забезпечують широкосмуговість звукопоглинання. Перфорований металевий лист розташовується на деякій відстані від стіни будівлі і є одним з параметрів, який визначає ефективність захисту. Зазор між панеллю та стінкою сприяє провітрюванню і відводу вологи.

Перфорована конструкція може розглядатися як ряд резонаторів. При рівномірній перфорації листа поглинач має типову резонансу криву поглинання. При нерівномірній – рівномірну криву поглинання. Для рівномірного розташування отворів резонансна частота визначається зі співвідношення:

$$f_p = \frac{v}{2\pi} \sqrt{\frac{S}{l_e d^2 h}},$$

де v – швидкість звуку, S – переріз отвору, $l_e = \Delta + 0,5\sqrt{\pi S}$ – ефективна товщина панелі, Δ – товщина листа, h – відстань панелі від поверхні, на якій вона змонтована, d – відстань між центрами отворів.

Якщо отвори у вигляді шпарин, то резонансна частота визначається як:

$$f_p = \frac{v}{2\pi} \sqrt{\frac{b}{ldh}}.$$

де b – ширина шпарини, d – відстань між шпаринами.

Найбільшою перевагою таких панелей є придатність для захисту від низькочастотного звуку, ефективність яких можна розрахувати. Це важливо з точки зору розповсюдження низькочастотного звуку на великі відстані через його повільне згасання з відстанню. Викладене є загальними засадами проектування і впровадження комплексного захисту будівель від впливу електромагнітних полів та шуму широких частотних діапазонів. Крім звукопоглинання виткі рослини певним чином сприяють енергозощадженню через позитивний вплив на температурний режим будівель.

Висновки

1. Підвищення електромагнітного та акустичного навантаження на виробниче та навколишнє середовище робить актуальними розроблення та впровадження універсальних засобів комплексного захисту від впливу техногенних фізичних полів. Це знімає проблему застосування двох окремих захисних конструкцій.

2. Ефективним засобом звукопоглинання є покриття поверхонь будівель виткими рослинами, які безпосередньо прилягають до стін. Це мінімізує прояви дифракційних явищ на кромках екрануючих конструкцій.

3. Каркасом для поширення витких рослин доцільно обирати сітчасті або шпаруваті металеві

конструкції для екранування електромагнітних полів. Коефіцієнти екранування розраховано виходячи з фундаментальних співвідношень електродинаміки. Такі конструкції частково екранують акустичні хвилі. Переважні частоти екранування визначаються співвідношеннями швидкостей поширення електромагнітних та акустичних хвиль.

4. Каркасом для поширення витких рослин можуть бути і перфоровані звукопоглинальні конструкції резонансного типу. Перевагою таких конструкцій є можливість екранування низькочастотного шуму, захист від якого найбільш складний. Показана можливість аналітичного попереднього визначення параметрів захисних панелей.

СПИСОК ЛІТЕРАТУРИ

1. Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions on an EU Strategic Framework on Health and Safety at Work 2014-2020. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2014:332:FIN>
2. Dir. 2013/35/EU - electromagnetic fields. <https://osha.europa.eu/en/legislation/directives/directive-2013-35-eu-electromagnetic-fields>
3. Directive 2003/10/EC - noise. <https://osha.europa.eu/en/legislation/directives/82>
4. Directive 2000/14/EC - noise - equipment for use outdoors. <https://osha.europa.eu/en/legislation/directives/directive-2000-14-ec>
5. Seon-Chil Kim and Sung-Hyoun Cho., 2019. Analysis of the Correlation between Shielding Material Blending Characteristics and Porosity for Radiation Shielding Films. *Journals Applied Sciences* 9(9), 1765; <https://doi.org/10.3390/app9091765>
6. Development and application of thin wide-band screening composite materials. Senyk, I., Kuryptia, Y., Barsukov, V., Butenko, O., Khomenko, V. *Physics and Chemistry of Solid State*, 2020, 21(4), pp. 771–778. <https://journals.pnu.edu.ua/index.php/pcss/article/view/4451/5175>
7. Glyva V., Kasatkina N., Nazarenko V., Burdeina N., Karaieva N., Levchenko L., Panova O., Tykhenko O., Khalmuradov B., Khodakovskyy O. Development and research of protective properties of composite materials for screening electromagnetic fields of a wide frequency range. *Eastern-European Journal of Enterprise Technologies*. 2020. Iss. 2/12 (104). PP. 40–47.
8. Glyva V., Lyashok J., Matvieieva I., Frolov V., Levchenko L., Tykhenko O., Panova O., Khodakovskyy O., Khalmuradov B., Nikolaiev K. Development and investigation of protective properties of the electromagnetic and soundproofing screen. *Eastern-European Journal of Enterprise Technologies*. 2018. Iss. 6/5 (96). P. 54–61.
9. Grinchenko V. S. Mitigation of three-phase power line magnetic field by grid electromagnetic shield // *Tekhnichna Elektrodynamika*. 2018. Vol. 2018, Issue 4. P. 29–32. doi: <https://doi.org/10.15407/techned2018.04.029>
10. Glyva V., Levchenko L., Panova O., Tykhenko O., Radomska M. The composite facing material for electromagnetic fields shielding. *Innovative Technology in Architecture and Design (ITAD 2020): IOP Conference Series: Materials Science and Engineering*. 2020. Vol. 907. URL: <https://iopscience.iop.org/article/10.1088/1757-899X/907/1/012043>.
11. Tkachenko Tetiana Assessment of Light Transmission for Comfort and Energy Efficient Insolation by “Green Structures”// *Advances in Intelligent Systems and Computing*, 2021, No 1296, p. 139–151. https://link.springer.com/chapter/10.1007/978-3-030-63403-2_13
12. Tkachenko Tetiana The Role of "Green Structures" in Reducing the Environmental Footprint of Urbocenes. Tetiana Tkachenko, Olena Voloshkina. *International Journal of Engineering & Technology*. – Vol.7, No 4,8 (2018). – P. 214-220. <https://www.sciencepubco.com/index.php/ijet/issue/view/452>
13. Abu Deeb S. Environmental Assessment of Relationships and Mutual Influences in the System "Protective Forest Plantations – Anthropogenic Landscapes". S.Abu Deeb, Tetiana Tkachenko, Viktor Mileikovskiy. 2nd International Symposium of Earth, Energy, Environmental Science and Sustainable Development (JEESD 2021) 25th-26th September 2021, Jakarta, Indonesia. - IOP Conference Series: Earth and Environmental Science. - 2021. - Vol. 940. - <https://doi.org/10.1088/1755-1315/940/1/012083>
14. Глива В. А., Тихенко О. М., Ходаковський О. В. Методологія проектування неоднорідних електромагнітних екранів. *Системи управління, навігації та зв'язку*. Полтава, 2019. Вип. 4(56). С. 122–125.
15. Glyva, V., Kasatkina, N., Levchenko, L., Tykhenko, O., Nazarenko, V., Burdeina, N., Panova, O., Bahrii, M., Nikolaiev, K., & Biruk, Y. (2022). Determining the dynamics of electromagnetic fields, air ionization, low-frequency sound and their normalization in premises for computer equipment. *Eastern-European Journal of Enterprise Technologies*, 3(10 (117)), 47–55. <https://doi.org/10.15587/1729-4061.2022.258939>

Received (Надійшла) 21.02.2023

Accepted for publication (Прийнята до друку) 10.05.2023

Shielding of electromagnetic fields and noise in buildings and structures

T. Tkachenko, N. Burdeina, O. Chenchava

Abstract. Analysis of the current state of electromagnetic and acoustic pollution of the industrial and environmental environment indicates the need to substantiate the general principles of building protection with universal protective structures. The possibility of using climbing plants to cover the surfaces of buildings is shown. It is advisable to choose a regular structure (mesh or slatted) as a frame for the propagation of creeping plants. The efficiency of such structures is calculated based on the fundamental relations of electrodynamics. Obtained relations for calculation of shielding coefficients for a normally incident electromagnetic wave and for a wave incident on the shield at an arbitrary angle. The ratio of the wavelengths of electromagnetic and acoustic fields, which are shielded by a regular structure, is shown. It is determined by the ratio of propagation speeds of electromagnetic and acoustic waves. For the introduction of creeping plants for protection against physical influences, experimental studies are necessary, which will allow to develop empirical or semi-empirical relationships for the preliminary assessment of the effectiveness of protection. A calculator is provided for calculating the parameters of a perforated or slotted screen for noise shielding. It is based on the determination of design parameters based on the sound frequency with the largest amplitude, which is considered resonant. Such a structure can also be a carrier of creeping plants. The advantage of such structures is their suitability for shielding low-frequency sound, and winding plants extend the range of protection. In addition, winding plants contribute to energy saving.

Keywords: shielding, electromagnetic field, noise, winding plants.

В. А. Цопа¹, С. І. Чеберячко¹, О. О. Яворська¹, О. В. Дерюгін¹, О. О. Ченчева², С. В. Сукач², Д. В. Резнік²

¹ Національний технічний університет «Дніпровська політехніка», Дніпро, Україна

² Кременчуцький національний університет імені Михайла Остроградського, Кременчук, Україна

АНАЛІЗ ВИМОГ РАМКОВОЇ СТРАТЕГІЇ ЄС ДО СИСТЕМ УПРАВЛІННЯ ОХОРОНОЮ ЗДОРОВ'Я ТА БЕЗПЕКОЮ ПРАЦІ

Анотація. Мета. Дослідження вимог Рамкової стратегії ЄС із безпеки та здоров'я на роботі на 2021-2027 рр. для визначення основних цілей системи управління охорони здоров'я та безпеки праці в організації щодо покращення здоров'я та безпеки працівників. **Методика.** Для дослідження був використаний формалізований контент-аналіз, який представляє собою якісно-кількісний метод вивчення документів, який полягає в обробці тексту з подальшою інтерпретацією результатів. **Результати.** Встановлено, що відповідно до Рамкової стратегії ЄС із безпеки та здоров'я на роботі на період 2021-2027 рр. для підвищення результативності системи управління охороною здоров'я та безпекою праці в організаціях необхідно передбачити управління змінами, заходи для профілактики нещасних випадків та захворювань на виробництві та готовності до майбутніх потенційних викликів. Визначено, що системи управління охороною здоров'я та безпекою праці повинні відобразити оцінку професійних ризиків, заміну (усунення) небезпечних речовини з виробничих технологій, захист від гендерного насильства та створення інспекцій праці. Запропоновано цілі Рамкової стратегії ЄС із безпеки та здоров'я на роботі на 2021-2027 р. включити в стратегічний план розвитку організації, що дозволить передбачити відповідні заходи на тактичному і операційному рівні систем управління організацією з наголосом на забезпечення й підтримку добробуту працівників, що спонукає формуванню усвідомленому відношенню працівників до своїх зобов'язань. **Наукова новизна** полягає у науковому обґрунтуванні вимог до розробки стратегія розвитку системи управління охорони здоров'я та безпеки праці в організації на основі управління змінами та готовістю до майбутніх викликів та потенційних загроз. **Практична цінність** полягає у розробці задач, які відповідають стратегії розвитку організації відповідно до вимог Рамкової стратегії ЄС із безпеки та здоров'я на роботі на 2021-2027 р.

Ключові слова: рамкова стратегія, система охорони здоров'я і безпеки праці, професійний ризик, небезпека.

Вступ

У Брюсселі (Бельгія) 28 червня 2021 року Комісія ЄС прийняла нову Рамкову стратегію з охорони здоров'я і безпеки праці (далі - ОЗіБП) на період 2021-2027 рр. У ній визначено ключові дії для урядів країн, які потрібні для поліпшення охорони здоров'я і безпеки праці працівників протягом наступних років. У Рамковій стратегії ЄС на 2021-2027 рр. наголошено, що за останні десятиліття було досягнуто прогресу. Наприклад, кількість нещасних випадків на виробництві зі смертельними наслідками в ЄС за період 1994-2018 рр. зменшилася приблизно на 70%, але цього замало, заявляють європейські фахівці. Незважаючи на цей прогрес, у ЄС у 2022 р. сталося понад 3,3 тис. нещасних випадків зі смертельними наслідками. Щороку понад 200 тис. працівників помирають від хвороб, пов'язаних з роботою [1].

Отже, підтримка та вдосконалення стандартів захисту працівників на виробництві є постійною проблемою і водночас нагальною потребою. Крім поліпшення здоров'я та добробуту є вагомими економічними аргументами щодо підвищення рівня захисту працівників. Нещасні випадки та захворювання, які пов'язані з роботою, коштують економіці ЄС понад 3,3% ВВП щороку. Це потребує відповідного реагування, принаймні, розгляду питання на національному рівні щодо реалізації вимог Рамкової стратегії ЄС при переході на зелені та цифрові технології, а також покращення профілактики та готовності до майбутніх викликів під час настання надзвичайних ситуацій, щоб гарантувати, що оновлені заходи та підходи досягнуть кожного робочого місяця.

Аналіз літературних джерел. Аналіз досліджень з проблем адаптації національного законо-

давства до вимог ЄС показав значний інтерес до цього питання. Так, автором статті [2] наголошується, що адаптація законодавства України до законодавства ЄС повинна базуватись, виключно на розвитку підприємницької, соціальної, культурної активності громадян України, що сприятиме поступовому зростанню рівня життя населення, при чому пропонується відкинути всі положення, які навпаки, призведуть до зменшення економічного зростання країни. В іншому дослідженні [3] говориться, що необхідно проводити поетапне наближення національного законодавства до вимог ЄС, забезпечуючи відповідну модернізацію вітчизняного правового поля і забезпечуючи прозорість змін, що гарантуватиме зменшення кількості як зовнішніх, так і внутрішніх конфліктних ситуацій. Автор наголошує, що потрібно привести власні амбіції до спільної мети шляхом посилення міжнародної співпраці з приватно-державного партнерства, збалансованого фінансування та реформаційних рішень, направлених на запровадження механізмів адаптації всієї законодавчої системи України у відповідність до вимог ЄС. Схожу думку щодо поетапної адаптації законодавства у сфері поводження з відходами до вимог ЄС висловила авторка ще однієї публікації [4], в якій запропоновано відповідний алгоритм дій. В той же час, реалізація подібного алгоритму стикається з купою різних суперечок між гілками влади в Україні, що пов'язана з гострою політичною боротьбою [5]. Для зменшення негативних наслідків автор дослідження говорить про розробку дієвого механізму імплементації чинних правових норм ЄС на основі "acquis communautaire".

Відмітимо, що група міжнародних стандартів із забезпечення безпечних умов праці, які необхідно

адаптувати до національного законодавства, нажаль, знаходяться «останніми» в черзі [6], на що вказує повільна реалізація вимог Концепції реформування системи управління охороною праці в Україні, яка затверджена Кабінетом Міністрів України ще у 2018 році. Можна сказати, що сповільнення відбулось через появу світової пандемії, хоча в країнах ЄС вона, навпаки, стала основним рушієм змін законодавчої бази з безпеки праці [7]. Зокрема, серед основних спільних рухів держав-членів ЄС до розвитку системи управління (далі - СУ) ОЗіБП в умовах пандемії COVID-19 виділено вакцинацію, посилення вимог психосоціальних ризиків [8], розробка нормативів для збільшення тривалості життя [9].

Проведений аналіз літературних джерел оголив проблему з адаптації міжнародного законодавства у сфері СУ ОЗіБП до реалій вітчизняних компаній. Це вимагає прийняття збалансованих рішень, за принципом «не нашкодити» економічному зростанню. Однак забезпечення відповідної якості життя, його тривалості, які є пріоритетом розвитку України [10], вимагає імплементації вимог Рамкової стратегії з охорони здоров'я і безпеки праці на період 2021-2027 рр. Тому виникає задача з проведення аналізу вимог Рамкової Стратегії ЄС та визначення пріоритетних заходів з її реалізації в українських організаціях для формування Національної Стратегії охорони праці в Україні.

Метою статті є дослідження вимог рамкової стратегії ЄС з СУ ОЗіБП на період 2021-2027 рр. та визначення основних цілей для систем управління охорони здоров'я та безпеки праці в організації для зменшення травматизму та професійних захворювань.

Матеріали і методи. Для дослідження був використаний формалізований контент-аналіз, який представляє собою якісно-кількісний метод вивчення документів, який полягає в обробці тексту з подальшою інтерпретацією результатів. Послідовність дій складається з огляду документу (зовнішнє дослідження), вивчення його змісту, визначення ключових формулювань й внутрішнє дослідження, а саме його контексту, смислового навантаження. Внутрішній аналіз складається зі змістовного, юридичного та психологічного оцінювання документу з'ясування змістових одиниць, які можна однозначно фіксувати та перекладати у кількісні показники. Змістові одиниці відображають провідну ідею тексту через окремі поняття, явища, події тощо. Особливістю цього наукового методу є той факт, що інформація, яка міститься в документах зібрана для дослідження певної проблеми, яку дослідник не може на власний розсуд змінювати, доповнювати інформацію, яка міститься у них.

Результати дослідження

Україна отримали статус кандидата на членство у ЄС Рамкова Стратегія ЄС щодо здоров'я та безпеки праці 2021-2027 рр. має назву "Безпека праці та охорона здоров'я на роботі в мінливому світі праці", що підкреслює необхідність управляти змінами в СУ ОЗіБП. Рамкова Стратегія ЄС базується на наступних п'яти стовпах СУ ОЗіБП ЄС (рис. 1):

1. Зміцнена база доказів;
2. Потужний соціальний діалог;
3. Мобілізоване фінансування;
4. Покращення правозастосування;
5. Підвищення обізнаності.

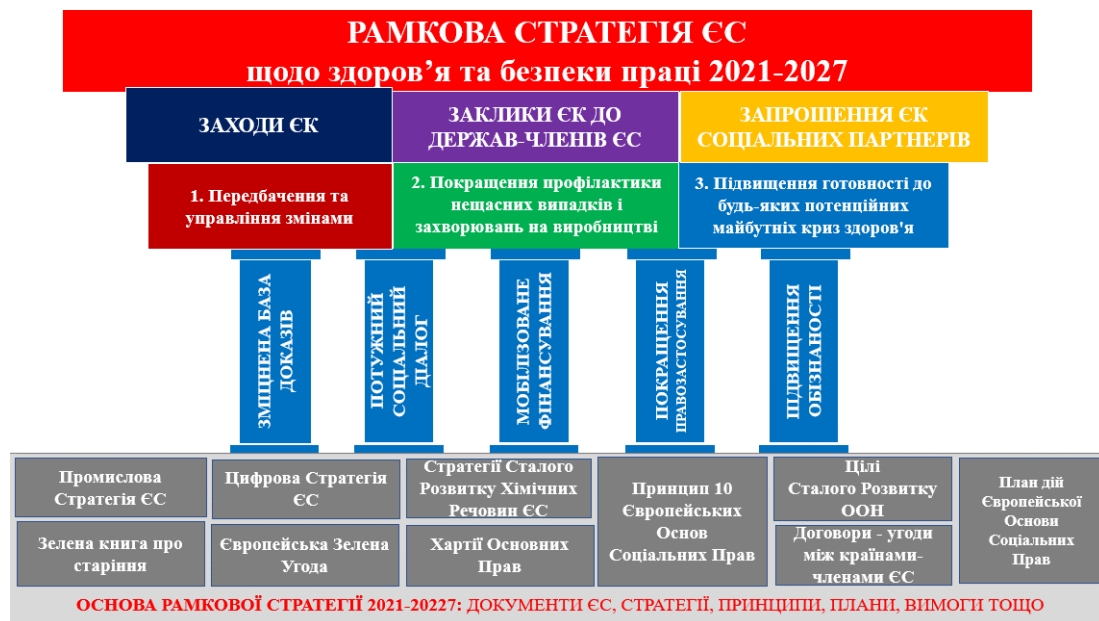


Рис. 1. Основа Рамкової Стратегії ЄС

Розробка Рамкової стратегії ЄС розроблялась на основі загальних етапів (рис. 2), з урахуванням Європейських і Міжнародних документів при участі "American Council on Science and Health" (ACSH), "European Agency for Safety and Health at Work" (EU-OSHA), "Committee of Senior Labor Inspectors"

(CSLI). Це дозволило сформулювати модель оновленого бачення системи управління з охорони здоров'я і безпеки праці на трьох інтегрованих ключових цілях-напрямах, що дають синергетичний ефект в частині зменшення втрат життя та здоров'я працівників в ЄС (рис. 3):

1. Передбачення та управління змінами в новому світі праці, викликаними переходом на зелені та цифрові технології та демографічною ситуацією - старіння населення ЄС;

2. Покращення профілактики нещасних випадків і захворювань на виробництві;

3. Підвищення готовності до будь-яких потенційних майбутніх криз здоров'я.

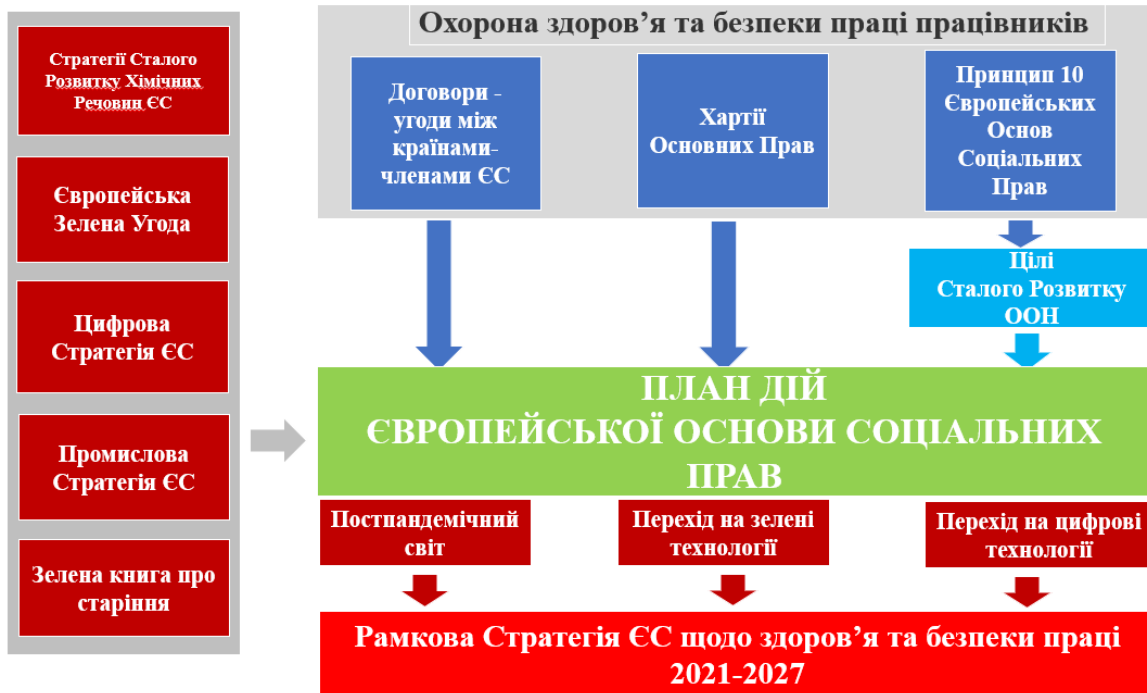


Рис. 2. Основні документи на основі яких була розроблена Рамкової стратегії СУ ОЗіБП ЄС 2021-2027

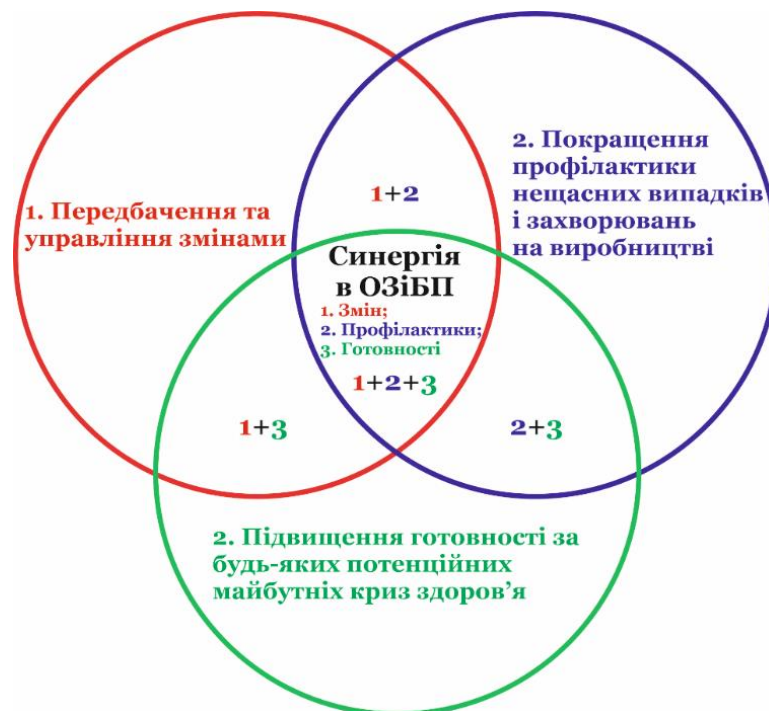


Рис. 3. Модель синергії ключових цілей-напрямків Рамкової Стратегії СУ ОЗіБП ЄС 2021-2027

Виходячи з моделі синергії ключові цілі-напрямки було пов'язано із заходами ЄК та закликами до держав-членів ЄС, які разом із запрошеннями до соціальних партнерів направлені на покращення СУ ОЗіБП. Це дозволило сформуванню відповідних заходів для підвищення безпеки праці (табл. 1 – 3) на основі

управління змінами. Зокрема, необхідно оновити законодавчу базу з безпеки праці, розпочати "Кампанію EU-OSHA за здорові робочі місця" та розробити аналітичну основу, електронні інструменти та вказівки для оцінки ризиків, пов'язаних психосоціальними та ергономічними небезпечними чинниками.

Аналізуючи дані в табл. 1-3 за областями розвитку можна зробити висновок, що основною ціллю-напрямком рамкової стратегії під назвою «Передбачення та управління змінами», являються заходи ЄК,

які в першу чергу направлені на розвиток та покращення системи управління ОЗіБП, по друге це захист від небезпечних речовин, по третє зменшення ризиків (табл. 4).

Таблиця 1 – Заходи Комісії ЄС для цілі-напрямку "Передбачення та управління змінами"

№ з/п	Що буде робитися	З якою сферою пов'язано	Примітка
1.	Оновити законодавчу базу щодо СУ ОЗіБП, пов'язану з цифровізацією, шляхом перегляду Директиви про робочі місця та Директиви про екранне обладнання (термін виконання - до 2023 року)	СУ ОЗіБП (перехід на цифрові технології)	Директива Ради 89/654/ЄЕС від 30 листопада 1989 р. щодо мінімальних вимог безпеки та гігієни праці на робочому місці.
2.	Запропонувати захисні граничні значення для:		
2.1	Азбест у Директиві про азбест на виробництві (термін виконання - 2022 рік)	Небезпечні речовини	Директива 2009/148/ЄС Європейського парламенту та Ради від 30 листопада 2009 року про захист працівників від ризиків, пов'язаних із впливом азбесту на роботі.
2.2	Свинець і дізоціанати в Директиві про хімічні агенти (термін виконання - 2022 рік)	Небезпечні речовини	Директива Ради 98/24/ЄС від 7 квітня 1998 року про захист здоров'я та безпеки працівників від ризиків, пов'язаних з хімічними речовинами на роботі.
2.3	Кобальту в Директиві щодо канцерогенів і мутагенів (термін виконання - I квартал 2024 року)	Небезпечні речовини	Директива 2004/37/ЄС - канцерогени, мутагени або токсичні для репродукції речовини на роботі
3.	Розпочати «Кампанію EU-OSHA за здорові робочі місця», щодо створення безпечного та здорового цифрового майбутнього, зокрема охоплюючи психосоціальні та ергономічні ризики (термін виконання - 2023-2025 рр.)	СУ ОЗіБП (перехід на цифрові технології)	«Кампанія EU-OSHA за здорові робочі місця»
4.	У співпраці з державами-членами та соціальними партнерами підготувати незаконодавчу ініціативу на рівні ЄС, пов'язану з психічним здоров'ям на роботі, яка оцінює нові проблеми, пов'язані з психічним здоров'ям (термін виконання – 2022 рік)	Захворювання (психічне здоров'я на роботі)	Розробка вказівок та рекомендацій, що до охорони психічного здоров'я працівників.
5.	Розробити аналітичну основу, електронні інструменти та вказівки для оцінки ризиків, пов'язаних із зеленими та цифровими роботами включаючи психосоціальні та ергономічні ризики (термін виконання – 2021 рік)	СУ ОЗіБП (перехід на цифрові та зелені технології)	Розробка вказівок та рекомендацій, що до оцінки ризиків небезпек: при переході на зелені та цифрові технології
6.	Попросити групу експертів, щодо ефективних способів інвестування в охорону здоров'я, висловити думку, щодо підтримки психічного здоров'я працівників охорони здоров'я та інших основних працівників (термін виконання - 2021 рік)	Захворювання (психічне здоров'я працівників)	Розробка запобіжних заходів щодо інфекційних хвороб

Таблиця 2 – Запрошення Комісії ЄС до соціальних партнерів до заходів для цілі-напрямку "передбачення та управління змінами"

№№	Запрошення що робитися	З якою сферою пов'язано
1.	Вжити заходів і оновити існуючі угоди на міжгалузевому та галузевому рівнях для вирішення нових питань СУ ОЗіБП, пов'язаних із цифровим ринком праці, зокрема психосоціальних та ергономічних ризиків (термін виконання 2023 рік)	СУ ОЗіБП (перехід на цифрові технології)
2.	Знайти узгоджені рішення для вирішення проблем, пов'язаних із дистанційною роботою, цифровізацією та правом на відключення, спираючись на Рамкову Угоду Європейських Соціальних Партнерів щодо цифровізації	СУ ОЗіБП (перехід на цифрові технології)

В таблицях 5-7 представлені заходи, які рекомендується запровадити у державах-членів ЄС з досягнення цілі «покращення профілактики нещасних випадків і захворювань на виробництві», що пов'язане в першу чергу з сприйняттям підходу «Нульового Бачення» до смертей, пов'язаних з роботою, шляхом

покращення збору даних про нещасні випадки, створення спеціалізованої тристоронньої робочої групи «Нульового Бачення» з громадськими захисними організаціями (наприклад Комітет інспекторів праці (SLIC)) та Американською радою з питань науки та охорони здоров'я (ACSH).

Важливою умовою для зменшення травматизму являється оновлення правил ЄС щодо небезпечних речовин для боротьби з раком, репродуктивними та респіраторними захворюваннями шляхом консультацій із соціальними партнерами щодо знижених граничних значень шкідливих речовин у повітрі робочих зон, Визначення пріоритетного переліку репротоксикантів та інше. Крім того, для виконання підходу «Нульового Бачення» щодо смертності, пов'язаної з роботою, в ЄС передбачається забезпечити покращене керівництво та навчання щодо оцінки ризиків та заходів із запобігання нещасних випадків, зокрема для мікропідприємств. Аналізуючи дані з табл. 5-7 встановили, що основною задачею для реалізації цілі

Рамкової Стратегії - "Покращення профілактики нещасних випадків і захворювань на виробництві" являється, знову ж таки заклик до покращення СУ ОЗіБП, а на другому місці – це зниження "ризиків", а на третьому – зниження кількості використання небезпечних речовин (табл. 8). В табл. 9 і 10 представлені заходи щодо досягнення цілі з "підвищення готовності до що будь-яких потенційних майбутніх криз здоров'я". Так, передбачається, що в організаціях необхідно провести поглиблену оцінку наслідків пандемії та ефективності роботи СУ ОЗіБП для розробки надзвичайних процедур і вказівок для швидкого розгортання, впровадження та моніторингу заходів у потенційних майбутніх кризах.

Таблиця 3 – Заклики Комісії ЄС до держав-членів ЄС для цілі-напряму "Передбачення та управління змінами"

№ з/п	Заклик що повинно робитися	З якою сферою пов'язано
1.	Оновити свою національну законодавчу базу, консультуючись із соціальними партнерами, щоб усунути ризики та використати можливості СУ ОЗіБП, пов'язані з переходом на зелені та цифрові технології. Держави-члени повинні зосередитися на використанні цифрових інструментів, щоб зробити інспекцію праці більш ефективною шляхом як запобігання, так і виявлення порушень законодавства.	СУ ОЗіБП (перехід на цифрові технології) СУ ОЗіБП (перехід на цифрові технології)
2.	Проводити «експертні оцінки», що стосуються професійних психосоціальних та ергономічних проблем.	Ризики (психосоціальні та ергономічні проблеми)
3.	Посилити моніторинг і збір даних щодо психічних і психосоціальних ризиків у різних секторах (галузях).	Ризики (психічні і психосоціальні ризики)

Таблиця 4 – Кількість заходів, закликів та запрошень ЄК по цілі-напряму Рамкової Стратегії "Передбачення та управління змінами"

№ з/п	Область розвитку та покращення ОЗіБП	Заходи ЄК	Заклики до держав членів ЄС	Запрошення до соціальних партнерів	Всі (заходи, заклики та запрошення)	
					кількість	%
1.	СУ ОЗіБП	4	1	2	7	50,0
2.	Захворювання	2	0	0	2	14,3
3.	Небезпечні речовини	3	0	0	3	21,4
4.	Ризики	0	2	0	2	14,3
5.	Система охорона здоров'я	0	0	0	0	0
6.	Гендерне насильство	0	0	0	0	0
7.	Інспекція праці	0	0	0	0	0
Всього, в кількості		9	3	2	14	100,0
Всього, в відсотках (%)		64,3	21,4	14,3		
В підсумку, в кількості					14	

Таблиця 5 – Заклики Комісії ЄС до держав-членів ЄС для досягнення цілі-напряму "Покращення профілактики нещасних випадків і захворювань на виробництві"

№ з/п	Заклик що повинно робитися	З якою сферою пов'язано
1.	Працювати над виконанням підходу «Нульового Бачення» щодо смертності, пов'язаної з роботою, в ЄС.	СУ ОЗіБП (Підхід «Нульового Бачення» до смертей)
2.	Звертати увагу на професійні ризики, пов'язані із захворюваннями системи кровообігу.	Ризики системи кровообігу
3.	Просувати Європейський Протираковий Кодекс серед працівників, щоб підвищити їхню медичну грамотність і зменшити ризик раку.	Ризики раку
4.	Оцінка та вирішення ризиків, приділяючи особливу увагу групам, які найбільше постраждали від пандемії, таким як люди з обмеженими можливостями.	Ризик COVID-19
5.	Активно підтримувати реінтеграцію, недискримінацію та адаптацію умов праці працівників, хворих на рак або тих, хто пережив рак.	СУ ОЗіБП
6.	Сприяти гендерним міркуванням у розробці, реалізації та звітності.	Гендерне насильство
7.	Активно боротися з небезпеками в системі охорони здоров'я, запроваджуючи та впроваджуючи безпечні робочі процедури та забезпечуючи відповідне навчання.	Ризики безпеки в системі охорони здоров'я
8.	Забезпечити покращене керівництво та навчання щодо оцінки ризиків та заходів із запобігання, зокрема для мікропідприємств.	Ризики для мікропідприємств
9.	Проводити навчання фермерів через консультаційні служби для фермерів, щоб підвищити їхні навички та обізнаність щодо правил охорони здоров'я та безпеки на фермах, включаючи безпечне використання хімічних речовин, зокрема засобів захисту рослин.	СУ ОЗіБП (правила охорони праці, включаючи хімічні речовини)

Таблиця 6 – Заходи Комісія ЄС для досягнення цілі "Покращення профілактики нещасних випадків і захворювань на виробництві"

№ з/п	Що буде робитися	З якою сферою пов'язано
1.	Сприяти підходу «Нульового Бачення» до смертей, пов'язаних з роботою, шляхом:	
1.1.	Покращення збору даних про нещасні випадки на виробництві та професійні захворювання, а також аналіз основних причин кожної смерті чи травми, пов'язаної з виробництвом;	СУ ОЗіБП (Підхід "Нульового Бачення")
1.2.	Створення спеціалізованої тристоронньої робочої групи «Нульового Бачення» ACSH та розробка цільових інформаційних дій та інструментів для підвищення обізнаності;	СУ ОЗіБП (Підхід "Нульового Бачення")
1.3.	Посилення правозастосування шляхом підтримки SLIC у підвищенні обізнаності щодо зменшення смертності, пов'язаної з роботою, на рівні компанії, обміну передовими практиками та підтримкою підвищення кваліфікації для інспекцій праці	СУ ОЗіБП (Підхід "Нульового Бачення")
2.	Оновити правила ЄС щодо небезпечних речовин для боротьби з раком, репродуктивними та респіраторними захворюваннями шляхом:	
2.1.	Започаткування консультацій із соціальними партнерами щодо знижених граничних значень для зварювальних димів, поліциклічних ароматичних вуглеводнів, ізопрену та 1,4-діоксану (термін виконання – 2023 рік)	Небезпечні речовини
2.2.	Визначення пріоритетного переліку репротоксикантів, які мають бути розглянуті через відповідні директиви (термін виконання – 2021 рік)	Небезпечні речовини
3.	Надати оновлені рекомендації, зокрема щодо навчання, протоколів, нагляду та моніторингу, для захисту працівників від впливу небезпечних речовин (термін виконання – 2022 рік)	Небезпечні речовини
4.	Підготовка нагляду безпеки праці та охорони здоров'я в системі охорони здоров'я та догляду у співпраці з EU-OSHA (термін виконання – I квартал 2024 рік)	Система охорони здоров'я
5.	Підтримувати підвищення обізнаності щодо захворювань опорно-рухового апарату, раку та психічного здоров'я, а також домагань на робочому місці та гендерних упереджень.	Захворювання, гендерне насильство
6.	Запропонувати законодавчу ініціативу щодо запобігання та протидії гендерному насильству щодо жінок та домашньому насильству (термін виконання – 2021 рік)	Гендерне насильство

Таблиця 7 – Запрошення Комісії ЄС до соціальних партнерів для досягнення «Покращення профілактики нещасних випадків і захворювань на виробництві»

№ з/п	Що буде робитися	Сфера
1.	Розробити керівництво для системи охорони здоров'я, зосереджуючись, серед іншого, на захисті працівників від впливу небезпечних лікарських засобів, відкриваючи шлях для розробки "online tool for interactive risk assessment" (OiRA) спеціально для системи охорони здоров'я у співпраці з EU-OSHA.	Ризики небезпечних лікарських засобів

Крім того, скласти плани готовності до майбутніх криз у національних стратегіях ОЗіБП, включаючи розробку механізму координації між органами системи охорони здоров'я (МОЗ) та охорони здоров'я та безпеки праці. Також виникає потреба у розробці Керівництво для інспекторів праці щодо оцінки якості оцінки ризиків та заходів з керування ризиками відповідно до Директиви про біологічні агенти.

Аналіз за областями розвитку та покращення в сфері СУ ОЗіБП (табл. 9 і 10) дозволяє зробити висновок, що основним дієвим механізмом для підвищення готовності до будь-яких потенційних майбутніх криз здоров'я являється підвищення результативності СУ ОЗіБП та розробка заходів щодо зменшення інфекційних захворювань. Крім того, існує рекомендація із започаткування Інспекції праці (табл. 11).

Таблиця 8 – Кількість заходів, закликів та запрошень ЄК для цілі "Покращення профілактики нещасних випадків і захворювань на виробництві"

№ з/п	Область розвитку та покращення системи управління ОЗіБП	Заходи ЄК	Заклики до держав членів ЄС	Запрошення до соціальних партнерів	Всі заходи, заклики та запрошення	
					кількість	%
1.	СУ ОЗіБП	4	3	0	7	35
2.	Захворювання	1	0	0	1	5
3.	Небезпечні речовини	3	0	0	3	15
4.	Ризики	0	5	1	6	30
5.	Система охорони здоров'я	1	0	0	1	5
6.	Гендерне насильство	1	1	0	2	10
7.	Інспекція праці	0	0	0	0	0
Всього, в кількості		10	9	1	20	100
Всього, в відсотках (%)		50	45	5		
В підсумку, в кількості					20	

Таблиця 9 – Заходи Комісії ЄС в цілі-напрямку "Підвищення готовності до будь-яких потенційних майбутніх криз здоров'я"

№ з/п	Що буде робитися	Сфера
1.	Негайно розпочати поглиблену оцінку наслідків пандемії та ефективності ЄС СУ ОЗіБП для розробки надзвичайних процедур і вказівок для швидкого розгортання, впровадження та моніторингу заходів у сфері охорони здоров'я (термін виконання – 2021 рік)	Захворювання COVID-19
2.	Оновлення Рекомендації Комісії щодо професійних захворювань щодо включення COVID-19 (термін виконання – 2022 рік)	Захворювання COVID-19
3.	Розробити Керівництво для інспекторів праці щодо оцінки якості оцінки ризиків та заходів з керування ризиками відповідно до Директиви про біологічні агенти (термін виконання – 2022 рік)	Інспекція праці

Таблиця 10 – Заклики Комісії ЄС до заходів для цілі-напрямку "Підвищення готовності до будь-яких потенційних майбутніх криз здоров'я"

№ з/п	Заклик що повинно робитися	Сфера
1.	Скласти плани готовності до майбутніх криз у національних стратегіях ОЗіБП, включаючи імплементацію керівництва та інструментів ЄС	СУ ОЗіБП (Швидке реагування)
2.	Розробити механізми координації між органами системи охорони здоров'я та охорони здоров'я та безпеки праці (термін виконання - 2023 рік)	СУ ОЗіБП
3.	Посилити моніторинг та ефективні перевірки виконання зобов'язань щодо СУ БПіОЗ щодо сезонних працівників професій з підвищеним ризиком.	Інспекція праці
4.	Посилити співпрацю та обмін інформацією між інспекціями праці та іншими відповідними національними органами для покращення стандартів з безпеки праці	Інспекція праці

Таблиця 11 – Кількість заходів, закликів та запрошень ЄС для цілі-напрямку "Підвищення готовності до будь-яких потенційних майбутніх криз здоров'я"

№ з/п	Область розвитку та покращення системи управління СУ ОЗіБП	Заходи ЄС	Заклики до держав членів ЄС	Запрошення до соціальних партнерів	Всі заходи, заклики та запрошення	
					кількість	%
1.	СУ ОЗіБП	0	1	0	1	14,3
2.	Захворювання	2	0	0	2	28,1
3.	Небезпечні речовини	0	0	0	0	0
4.	Ризики	0	0	0	0	0
5.	Система охорони здоров'я	0	1	0	1	14,3
6.	Гендерне насильство	0	0	0	0	0
7.	Інспекція праці	1	2	0	3	42,9
Всього, в кількості		3	4	0	7	100
Всього, в відсотках (%)		42,9	57,1	0		
В підсумку, в кількості					7	

В результаті проведеного аналізу визначили основні пріоритети у стратегічному плануванні розвитку будь-якої організації, що передбачає формування відповідних цілей у системах управління охороною здоров'я і безпекою праці. При чому прагнення до зменшення травматизму і кількості професійних захворювань потребує спільне проектування технічних і соціальних систем, на основі відповідності поставленої основній стратегічній меті, яка стоять перед організацією на основі вимог Рамкової стратегії. В подальшому стратегічна мета дозволяє сформулювати завдання на тактичному і операційному рівнях підприємства, рис. 4.

Для прикладу: основна стратегічна ціль СУ ОЗіБП являється створення здорового і безпечного робочого місця, яка потребує відповідального управління, залученості керівного складу організації, формування культури безпеки на основі прийнятого підходу «Нуль травм».

Реалізація стратегічної мети передбачає формулювання необхідних задач на тактичному рівні, наприклад, складання плану реагування на надзвичайні

події чи розробки концепції нульового травматизму чи формування групи з інспекції праці, що дозволить угодити відповідні прийняті рішення на всіх рівнях і системах для зменшення травматизму, а також передбачить необхідну кількість ресурсів для реалізації стратегічних задумів для урахування поставлених стратегічних вимог. Наступний рівень – оперативний, де передбачено конкретні кроки з реалізації визначених задач з урахуванням наявних ресурсів. Обов'язково на цьому рівні виникає необхідність у оцінці професійних ризиків, проведення навчання працівників, моніторингу умов праці.

Як запобігти травматизму, виникненню надзвичайних ситуацій, зменшити кількість професійних захворювань. Питання досить складне, тому відповідь на нього знаходиться в різних глобальних міжнародних документах: концепції сталого розвитку чи рамковій стратегії з охорони здоров'я та безпеки праці. Проведений аналіз дозволяє визначити ключові елементи розбудови як національних стратегій, так і формування конкретних цілей для будь-якого підприємства.

Нажаль діюча система управління охороною здоров'я і безпекою праці на всіх рівнях сформована за принципом «коригувальних дій» (реактивним принципом), тобто реагування на небезпечні випадки

та ситуації, а не за принципом «запобіжних дій» (проактивним підходом), тобто профілактики небезпечних випадків та ситуацій, про що саме йдеться у Рамковій стратегії ЄС.



Рис. 4. Основні задачі на стратегічному, тактичному та оперативному рівнях управління охороною праці

Реактивне реагування на виклики і загрози у сучасних умовах є непридатним для запровадження сучасного інструментарію для зменшення травматизму [11, 12], оскільки виникає значне непорозуміння, несприйняття працівниками, а особливо їх керівниками необхідності змін, щодо до створення належних, безпечних і здорових умов праці.

Рамкова Стратегія СУ ОЗіБП ЄС 2021-2027 зосереджена на моделі оновленого бачення СУ ОЗіБП (рис. 3) і на трьох інтегрованих ключових цілях-напрямах, що дають синергетичний ефект в частині зменшення втрат життя та здоров'я працівників в ЄС: передбачення та управління змінами в новому світі праці, викликаними переходом на зелені та цифрові технології та демографічною ситуацією – старіння населення ЄС; покращення профілактики нещасних випадків і захворювань на виробництві; підвищення готовності до будь-яких потенційних майбутніх криз здоров'я. Це дозволяє сформувати необхідні стратегічні цілі для розвитку організацій: побудови відповідних систем управління, які забезпечать вирішення головної задачі будь якої організації є економічна складова - заробити кошти для прибутку та зменшити фінансові витрати через керування ризиками безпеки праці, здоров'я працівників і екологічної безпеки [13, 14, 15]. Зроблений висновок підтверджується сформованими цілями напрямками, щодо «передбачення та управління змінами» у Рамковій Стратегії СУ ОЗіБП.

Висновки

1. Рамкова Стратегія ЄС вимагає, щоб Державичлени ЄС розглянути питання ОЗіБП в цільових

напрямах: передбачення та управління змінами, покращення профілактики та готовності до майбутніх викликів в ОЗіБП ЄС в таких областях, як: система управління ОЗіБП, професійні ризики, небезпечні речовини, захворювання, система охорони здоров'я, гендерне насильство та інспекція праці.

2. На основі концепції з синергії ключових цілей-направків Рамкової Стратегії ОЗіБП пропонується нова модель для сталого розвитку організації де присутній є взаємозв'язок трьох складових: економічної, безпеки праці та екології більш характерних для організації, де край кожного кола зон безпеки праці та екологічної безпеки це кордон прийняттого ризику (втрат), зони економічного розвитку це нульовий прибуток.

3. Досягнення сталого розвитку організації вимагає використання відповідних заходів не лише на стратегічному рівні, який задається відповідними вимогами міжнародних нормативних актів, але й проектування їх на тактичний, а особливо на оперативний рівень діяльності організації з особливим наголосом на забезпечення й підтримку добробуту працівників, що спонукає формуванню усвідомленому відношенню до своїх зобов'язань.

4. Стратегія розвитку організації, вирішення тактичних задач для економічного росту та забезпечення якісного функціонування людини на операційному рівні дозволила сформувати відповідну систему управління безпекою праці та здоров'ям працівників, особливістю якої є забезпечення сталого розвитку, який базується на рівні розвитку організаційної культури.

СПИСОК ЛІТЕРАТУРИ

1. Офіційний сайт International Labour Organization (ILO). World Statistic. The enormous burden of poor working conditions. Режим доступу: https://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang-en/index.htm.
2. Чорномаз, О.Б. (2016). Проблеми адаптації законодавства України до законодавства Європейського Союзу. *Вісник Національного університету "Львівська політехніка". Серія: Юридичні науки*, 845, 188-192. Режим доступу: http://nbuv.gov.ua/UJRN/vnulpurn_2016_845_31.
3. Француз, А.Й., Лазебна, К.Ю. (2020). Механізми адоптації законодавства України до правової системи Європейського Союзу. *Правничий вісник Університету «КРОК»*, 38, 66-76. Режим доступу: <https://lbku.krok.edu.ua/index.php/krok-university-law-journal/article/view/248>.
4. Федунь, М. (2014). Адаптація законодавства України до стандартів Європейського союзу у сфері поводження з відходами. *Вісник Львівського університету. Серія міжнародні відносини*, 36(2), 191-199. Режим доступу: https://intrel.lnu.edu.ua/wp-content/uploads/2015/09/VLNU_Mv_2014_362_23.pdf.
5. Мороз О.Б. (2020). Проблеми адаптації законодавства України до законодавства європейського союзу. *Альманах міжнародного права*, 23, 180-186. <https://doi.org/10.32841/IL.A.2020.23.03>.
6. Никифорова, О.А., Сидоренко, Г.Г. (2015). Проблеми трансформації національного законодавства в сфері охорони праці до стандартів ЄС. *Збірник наукових праць Дніпропетровського національного університету залізничного транспорту імені академіка В. Лазаряна. Транспортні системи та технології перевезень*, 10, 93-97. Режим доступу: http://nbuv.gov.ua/UJRN/znpdnu_tstp_2015_10_17.
7. Гавриченко, Д.Г. (2022). Охорона здоров'я в умовах загроз від Covid-19: законодавчі механізми Європейського союзу. *Вчені записки ТНУ імені В.І. Вернадського. Серія: Публічне управління та адміністрування*, 33(72-1), 35-39. <https://doi.org/10.32838/TNU-2663-6468/2022.1/07>.
8. Dahler-Larsen, P., Sundby, A., Boodhoo, A. (2020). Can occupational health and safety management systems address psychosocial risk factors? An empirical study. *Safety Science*, 130, 104878. <https://doi.org/10.1016/j.ssci.2020.104878>.
9. Warman, D.J., Jia, H., Kato, H. (2022). The Potential Roles of Probiotics, Resistant Starch, and Resistant Proteins in Ameliorating Inflammation during Aging (Inflammaging). *Nutrients*, 14(4), 747. <https://doi.org/10.3390/nu14040747>.
10. Ставицький, А., Молоканова, К. (2020). Моделювання якості життя населення в Україні та країнах Європи. *Вісник Київського національного університету імені Тараса Шевченка*, 6(213), 20-31. <https://doi.org/10.17721/1728-2667.2020/213-6/3>.
11. Gorgenyi-Hegyes, E., Nathan, R.J., Fekete-Farkas, M. (2021). Workplace Health Promotion, Employee Wellbeing and Loyalty during Covid-19 Pandemic-Large Scale Empirical Evidence from Hungary. *Economies*, 9(2), 55. <https://doi.org/10.3390/economies9020055>.
12. Schoukens, P., De Becker, E., Beke, S.J. (2015). Fighting social exclusion under the Europe 2020 strategy: Which legal nature for social inclusion recommendations? *International Comparative Jurisprudence*, 1(1), 11-23. <https://doi.org/10.1016/j.icj.2015.10.003>.
13. Becker, W., Norlén, H., Dijkstra, L., Athanasoglou, S. (2020). Wrapping up the Europe 2020 strategy: A multidimensional indicator analysis. *Environmental and Sustainability Indicators*, 8, 100075. <https://doi.org/10.1016/j.indic.2020.100075>.
14. Pasimeni, F., Pasimeni, P. (2016). An Institutional Analysis of the Europe 2020 Strategy. *Social Indicators Research*, 127(3), 1021-1038. <https://doi.org/10.1007/s11205-015-1013-7>.
15. Anastasiou, A., Marietta, P. (2020). Sustainable Development at the Frames of the Strategy "Europe 2020". *Theoretical Economics Letters*, 10, 443-457. <https://doi.org/10.4236/tel.2020.103028>.

Received (Надійшла) 24.03.2023

Accepted for publication (Прийнята до друку) 31.05.2023

Analysis of the requirements of the eu framework strategy for occupational health and safety management systems

Vitaliy Tsopa, Serhii Cheberichko, Olena Yavorska, Oleg Deryugin, Olga Chencheva, Serhii Sukach, Dmytro Riezniak

Abstract. Purpose. The purpose is to study the requirements of the EU Framework Strategy for Safety and Health at Work for 2021-2027 to determine the main objectives of the occupational health and safety management system in the organization to improve the health and safety of workers. **Methodology.** For the research, formalized content analysis was used, which is a qualitative-quantitative method of studying documents, which consists in processing the text with further interpretation of the results. **Findings.** It has been established that in accordance with the EU Framework Strategy for Safety and Health at Work for the period 2021-2027, in order to increase the effectiveness of the health and safety management system in organizations, it is necessary to provide for change management, measures for the prevention of accidents and diseases at production and readiness for future potential challenges. It was determined that occupational health and safety management systems should reflect the assessment of occupational risks, the replacement (elimination) of hazardous substances from production technologies, protection against gender-based violence and the creation of labor inspections. It is proposed to include the goals of the EU Framework Strategy for Safety and Health at Work for 2021-2027 in the strategic plan of the organization's development, which will allow for the provision of appropriate measures at the tactical and operational level of the organization's management systems with an emphasis on ensuring and supporting the well-being of employees, which encourages formation of the conscious attitude of employees to their obligations. **Originality.** The scientific novelty consists in the scientific substantiation of the requirements for the development of a strategy for the development of the health and safety management system in the organization based on change management and readiness for future challenges and potential threats. **Practical value.** The practical value lies in the development of tasks that correspond to the organization's development strategy in accordance with the requirements of the EU Framework Strategy for Safety and Health at Work for 2021-2027.

Keywords: framework strategy, occupational health and safety system, occupational risk, danger.

В. І. Мордюк, Ю. В. Синицький

Науково-дослідний інститут воєнної розвідки, Київ, Україна

МЕТОДИЧНІ АСПЕКТИ ВИЗНАЧЕННЯ ОСНОВНИХ ПАРАМЕТРІВ СИГНАЛЬНО-КODOVIХ КОНСТРУКЦІЙ СИСТЕМИ АВТОМАТИЗОВАНОГО ВСТАНОВЛЕННЯ ЗВ'ЯЗКУ У СЕАНСАХ З КОРЕСПОНДЕНТАМИ

Анотація. Радіозв'язок вважається найважливішим засобом управління військовими підрозділами, а також є необхідною умовою для досягнення інформаційної переваги над супротивником. З метою підвищення ефективності використання сил та засобів, а також для успішного виконання поставлених завдань підрозділами доцільно використовувати системи автоматизованого встановлення зв'язку. Однак недосконалість відповідних систем встановлення декаметрового радіозв'язку з урахуванням сучасних вимог до військового радіозв'язку виключає можливість використання відповідних систем. Метою статті є розроблення методичних аспектів визначення параметрів сигнально-кодкових конструкцій системи автоматизованого встановлення зв'язку, які доцільно створювати і впроваджувати для потреб воєнного часу. У статті виконаний огляд основних елементів систем автоматичного встановлення зв'язку та здійснено аналіз сигнально-кодкових конструкцій як однієї з головних частин системи. Розроблено методичні аспекти визначення основних параметрів сигнально-кодкових конструкцій системи автоматизованого встановлення зв'язку, що дає змогу побудови відповідних систем з урахуванням сучасних вимог до військового радіозв'язку. Доцільність та актуальність використання систем автоматизованого встановлення зв'язку підтверджують технічні стандарти MIL-STD-188-110D та MIL-STD-188-141D.

Ключові слова: системи встановлення радіозв'язку; декаметровий радіозв'язок; системи передачі інформації; військовий радіозв'язок; вимоги до військового радіозв'язку; елементи системи; сигнально-кодкові конструкції; стандарти MIL-STD.

Вступ

Постановка проблеми в загальному вигляді.

Сьогодні радіозв'язок є найважливішим засобом управління підрозділами, які беруть у спеціальних операціях та в безпосередньому зіткненні з ворогом. В умовах маневреного швидкоплинного бою надійний радіозв'язок є гарантією стійкого і гнучкого управління, а з урахуванням швидкої модернізації автоматизованих систем управління підрозділами – необхідною умовою досягнення інформаційної переваги над супротивником.

Суттєве ускладнення і розширення цільових завдань у сфері управління призвело до зростання кількості завдань щодо забезпечення управління підрозділами, вирішення яких досягається розгортанням систем передачі інформації (СПІ). Як наслідок - підвищуються вимоги до систем радіозв'язку, зокрема щодо своєчасності, достовірності та прихованості. Для виконання зазначених вимог необхідно вжити низку заходів, спрямованих на підвищення ефективності функціонування СПІ. Одним із них є імплементація системи автоматизованого встановлення зв'язку (АВЗ), доцільність та актуальність якої обґрунтовують технічні стандарти MIL-STD-188-110D та MIL-STD-188-141D.

Таким чином, відсутність систем АВЗ з урахуванням наведених вище вимог виключає можливість використання відповідних систем для забезпечення нормального функціонування СПІ.

Аналіз останніх досліджень і публікацій.

У [1] викладено основні принципи функціонування завадозахищеної системи АВЗ короткохвильового

(КХ) діапазону, а також наведено відповідні складові цієї системи без розкриття параметрів сигнально-кодкових конструкцій.

У [3, 4] проаналізовано концепції систем Automatic Link Establishment (ALE). У зазначених джерелах міститься детальна інформація про встановлені форми сигналів, їх структуру, протоколи та вимоги до продуктивності систем АВЗ.

Результати аналізу сучасних публікацій [3-5] показали, що тенденцією розвитку систем АВЗ є забезпечення повної автономності радіостанцій (радіоліній) зв'язку, і як наслідок – оптимізація заходів та виключення (мінімізація) помилок оператора під час сеансу зв'язку.

Джерело [6] описує створення перспективних систем АВЗ із визначеними варіантами завадостійких кодів, способів модуляцій та перемешувань. Також наведено аналіз продуктивності запропонованої моделі.

Робота [7] описує результати дослідження системи АВЗ, у якому для обміну інформацією використовувалися широкосмугові сигнали у поєднанні з системою автоматичного встановлення зв'язку четвертого покоління відповідно MIL-STD-188-110D.

Із проаналізованого випливає, що наявні системи АВЗ загалом та їх складові (сигнально-кодкові конструкції) не враховують особливостей військового радіозв'язку та вимог до нього, у зв'язку з чим немає умов для якісного їх впровадження у системи радіозв'язку, що використовується для воєнних потреб.

Для забезпечення ефективного та якісного функціонування системи АВЗ та подальшого обміну

інформацією у ході ведення сеансу зв'язку неабияк важливо обґрунтовано обирати параметри сигнально-кодових конструкцій.

Формулювання мети статті. Мета статті - розробити методичні аспекти визначення параметрів сигнально-кодових конструкцій системи автоматизованого встановлення зв'язку для потреб воєнного часу. Для досягнення цієї мети насамперед слід проаналізувати сигнально-кодові конструкції наявних систем АВЗ та їх параметри.

Виклад основного матеріалу дослідження

Умови розповсюдження радіохвиль декаметрового діапазону визначаються станом іоносфери, якій характерні регулярні (добові, сезонні, циклічні) і нерегулярні (короткотривалі) зміни, що обумовлюють необхідність планування робочих частот протягом доби, року, циклу сонячної активності, а також характеристиками антенно-фідерних систем (способом розгортання, напрямком діаграми спрямованості, коефіцієнтом підсилення тощо). У зв'язку з цим декаметровий (ДКМ) зв'язок іоносферною хвилею можна встановити лише при неперервному стеженні за станом іоносфери, її спроможністю відбивати радіохвилі і забезпечувати їх розповсюдження на необхідну відстань та при врахуванні особливостей антенно-фідерних систем, розгорнутих на вузлах зв'язку чи окремих радіостанціях.

Своєчасність та прихованість радіозв'язку при плануванні й у ході проведення сеансу зв'язку насамперед залежать від правильного вибору робочих частот, видів сигналів та антен, що відповідають умовам роботи кореспондентів та їх місцезнаходженню. При нівелюванні критеріїв правильного вибору сигналів, частот і антен збільшення потужності не допоможе значно поліпшити характеристики каналу зв'язку або відновити його у разі втрати.

Досвід застосування засобів зв'язку під впливом систем та комплексів радіоелектронної боротьби противника багаторазово підтверджував необхідність автоматизації процесів встановлення та ведення зв'язку між радіостанціями, тобто використання систем АВЗ. Результати аналізу існуючих систем АВЗ показують, що з урахуванням вимог до військового радіозв'язку та його особливостей процес встановлення зв'язку слід розділити на дві складові:

автоматичне встановлення зв'язку;

вибір оптимальних параметрів сигнально-кодових конструкцій, необхідних для ведення зв'язку на визначених на частотах.

Нині автоматичне встановлення зв'язку зумовлює такі дії [1]:

автоматичний підбір радіочастоти (декількох радіочастот) для виклику кореспондента;

автоматичний підбір радіочастоти (декількох радіочастот) для відповіді кореспондента;

вибірковий виклик кореспондента;

автоматичне оцінювання якості проходження інформації на обраних радіочастотах та вибір оптимальної радіочастоти (субчастоти).

Якість здійснення автоматичного встановлення зв'язку визначається алгоритмом та типом сигналь-

но-кодової конструкції, обраної для забезпечення зв'язку.

Застосування однакових сигнально-кодових конструкцій на всіх етапах ведення зв'язку, як це реалізовано в стандарті MIL-STD-188-141D [3], не є оптимальним, зважаючи на необхідність забезпечення гарантованого встановлення зв'язку в тих каналах, де обрані сигнально-кодові конструкції здатні забезпечувати протидію значному погіршенню якості каналу, зберігаючи при цьому можливість детектування. Такий підхід доцільно застосовувати для зв'язку кореспондентів, які перебувають відносно недалеко один від одного, мають схожу завадозахищеність обстановку, та коли пріоритет швидкості встановлення зв'язку переважає над його завадозахищеністю.

Вибір сигнально-кодової конструкції.

Особливістю сигнально-кодової конструкції, необхідної для автоматичного встановлення зв'язку, є її гарантоване детектування на визначених радіочастотах в умовах відсутності апріорної інформації про стан проходження іоносферної хвилі між радіостанціями.

Специфіка розповсюдження ДКМ-хвиль впливає на характеристики тих сигналів, які можуть ефективно застосовуватися для ведення зв'язку іоносферною хвилею, адже до сигналу не тільки додається шум ефіру, а й відбуваються амплітудні та фазові спотворення – доплерівський зсув, інтерференційні завмирання, радіоехо, багатопроменевість сигналів [1]. Сигнально-кодова конструкція має забезпечувати проходження сигналу через радіоканал із мінімальними спотвореннями (високондійне низькошвидкісне передавання сигналів виклику), що гарантуватиме встановлення зв'язку в складних умовах.

Сигнально-кодова конструкція визначає детальні вимоги до класу випромінювання, часових і частотних параметрів сигналу, виду завадозахищеного кодування та структури повідомлень.

Клас випромінювання.

Радіосигнал при розповсюдженні іоносферною хвилею зазнає фазових спотворень, тому для нього немає можливості наперед спрогнозувати фазу, але можна оцінити її. Завдяки оцінюванню параметрів сигналу на відносно коротких відрізках часу радіоканал із довільно змінною фазою слід вважати локально стаціонарним. У цьому разі для демодуляції радіосигналу можна застосувати когерентний (квазікогерентний) метод і використовувати фазову маніпуляцію як ефективний клас випромінювання. Такий принцип закладено в системи АВЗ типу HFDL та ALE третього покоління, що описаний в стандарті MIL-STD-188-141B: застосовується метод високошвидкісної передачі однотоновим фазоманіпульованим сигналом, який має мінімальний пік-фактор (це дає змогу знизити вимоги до підсилювача потужності радіостанцій), синхронізуючі послідовності, а також метод прогнозування характеристик каналу. Водночас при зменшенні рівня сигналу до рівня, коли однозначно фазу сигналу визначити неможливо - через відносно велику швидкість зміни фази сигналу, або через неідеальну реалізацію коге-

рентного демодулятора, відбувається різке зростання імовірності помилки – надійність системи може бути під загрозою, прийом сигналів виклику стає неможливим. У такому разі єдиною можливістю стає некогерентний спосіб прийому сигналів з активною паузою та ортогональними складовими, тобто, частотна маніпуляція [9].

Швидкість маніпуляції.

Вважаючи, що основний спосіб застосування радіозасобів декаметрового діапазону – стаціонарне розміщення, яке визначається робочими довжинами хвиль і ефективними розмірами антенно-фідерних систем більш ніж 10 м на місцевості, а доплерівський зсув частот визначається в основному рухом іонізованих частинок, на яких здійснюється відбиття радіохвиль, і тоді інтервал когерентності складатиме 40 – 100 мс [1].

Комбінації класу випромінювання та швидкості маніпуляції.

Груповий сигнал багаточастотної маніпуляції з ортогональними складовими являє собою однотоновий сигнал постійної амплітуди зі змінною частотою, що має мінімальний пік-фактор (це дає змогу знизити вимоги до підсилювача потужності радіостанції), а в частотній області – на нескінченному інтервалі часу відстеження являє собою суму N гармонічних коливань з інтервалом Δf . Для ортогональних складових інтервал Δf дорівнює швидкості маніпуляції, а загальна ширина смуги частот багаточастотної маніпуляції складає $\Delta F = (N+1) \times \Delta f$ [8].

Для визначення оптимальної комбінації класу випромінювання та швидкості маніпуляції слід розглянути всі можливі варіанти (швидкості 25, 50, 100 Бод оптимальні для частоти дискретизації 8000 Гц, швидкості 46.875, 75, 93.75, 125 Бод оптимальні для частоти дискретизації 6000 Гц та 12000 Гц.).

Основними аналогами є такі існуючі системи:

ALE 2G відповідно MIL-STD-188-141B (MFSK-8, $V=125$ Бод, $\Delta F= 1125$ Гц);

CROWD-36 (MFSK-32, $V=10/40$ Бод, $\Delta F= 1400$ Гц);

CIS FTM-4 (MFSK-4, $V=150$ Бод, $\Delta F= 12300$ Гц);

Olivia (MFSK-32, $V=31,25$ Бод, $\Delta F= 1030$ Гц);

Piccolo-12 (MFSK-12, $V=20,15$ Бод, $\Delta F= 260$ Гц);

CIS MFSK-68 (MFSK-32, $V=46,875/93,75$ Бод, $\Delta F= 6400$ Гц);

CHINESE-32 (MFSK-32, $V=75$ Бод, $\Delta F= 2475$ Гц).

З метою виконання вимог Сектору стандартизації електрозв'язку Міжнародного союзу електрозв'язку (ITU-T) необхідно, щоб сигнал радіостанції не виходив за межі каналу тональної частоти (300 – 3400 Гц) чи подвоєного каналу (режим двосмугової роботи -3400...+3400 Гц) – загальна ширина смуги частот багаточастотної маніпуляції має бути не більше ніж 3100 чи 6800 Гц.

Завадостійке кодування.

Завадостійке кодування застосовується для підвищення якості зв'язку та зменшення впливу шуму, завад, завмирань радіосигналів завдяки дода-

ванню контрольованої надлишковості до інформації, що підлягає передаванню [8].

Результати наявних досягнень у галузі кодування інформації дають змогу сучасним системам зв'язку наблизитися впритул до межі Шеннона [10] – верхньої межі максимального обсягу інформації, що передається без помилок.

Застосування ефективного завадостійкого кодування дає змогу досягти визначеної ймовірності помилки на біт за меншого значення відношення “сигнал/шум”, що відповідає зменшенню необхідної потужності. Тобто “виграш” у кодуванні означає можливість використання меншої потужності та менш ефективних антен, або збільшення дальності зв'язку при збереженні рівня потужності.

Склад та обсяг команд, необхідних для встановлення зв'язку.

Результати аналізу сучасних систем АВЗ показують, що для передавання команд виклику та відповіді застосовуються пакетні поділи фіксованої довжини, які після формування підлягають завадостійкому кодуванню, та перестановці біт (interleaving). Це забезпечує єдиний алгоритм оброблення повідомлень у сигнальних процесорах, а також оптимізацію процесів кодування й декодування інформації. При цьому ці пакети постійно передаються (циклічно) до отримання підтвердження чи виконання вимог пакета приймальною станцією. У системах типу HF DL та ALE третього покоління використовується 26 біт, в перспективній системі 4G відповідно стандарту MIL-STD-188-141B – 96 біт у пакеті [3].

Отож під час аналізу основних складових наявних систем АВЗ та їх параметрів досліджуються сигнально-кодові конструкції та розробляються методичні аспекти визначення основних параметрів відповідних конструкцій (з урахуванням вимог до військового радіозв'язку) які містять наступні етапи:

1. Визначення класу випромінювання.
2. Визначення швидкості маніпуляції.
3. Визначення комбінації класу випромінювання та швидкості маніпуляції.
4. Визначення методу кодування.
5. Визначення складу та обсягу команд встановлення зв'язку.
6. Аналіз параметрів каналу зв'язку.
7. Адаптація сигнально-кодової конструкції.

Етап 1. Визначення класу випромінювання. Однією з характеристик радіоканалу декаметрового діапазону хвиль є багатопрореневість, середній час затримки радіосигналів зазвичай складає 1 – 3 мс, у найгіршому разі – 4 мс. Основним способом боротьби з запізненням радіосигналів є збільшення тривалості елементарної одиниці інформації, що легко реалізується в системах із частотною маніпуляцією. У зв'язку з цим оптимальним класом випромінювання для системи АВЗ, є частотна маніпуляція (FSK). Згідно з [9], ефективність систем із FSK зростає зі збільшенням кількості частотних складових, тому доцільно обирати багаточастотну маніпуляцію (multiple frequency shift keying – MFSK). Схожий принцип закладено в систему АВЗ типу ALE 2G відповідно стандарту MIL-STD-188-141B, у якій застосовується метод

відносно повільної передачі інформації за допомогою багаточастотної маніпуляції (MFSK-8, швидкість 125 Бод, рознесення частот 250 Гц) та в системі CROWD-36 (MFSK-32, швидкість 10/40 Бод, рознесення частот 40 Гц), що дає змогу ефективно боротись з міжсимвольною інтерференцією.

Етап 2. Визначення швидкості маніпуляції.

Зважаючи на особливості декаметрового зв'язку та способи використання його для вирішення поставлених завдань, оптимальний діапазон тривалості елементарної одиниці інформації для системи автоматичного зв'язку, складає 8 – 40 мс, що визначає межі швидкості маніпуляції 25 – 125 Бод.

Етап 3. Визначення комбінації класу випромінювання та швидкості маніпуляції. При виборі конкретної комбінації класу випромінювання та швидкості маніпуляції слід зважати на: оптимальний діапазон тривалості елементарної одиниці інформації, баланс між інформаційною швидкістю та завадозахищеністю, дозволена чи бажана ширина смуги частот, кратність обраній частоті дискретизації, стабільність частот опорних генераторів радіопередавального та радіоприймального пристроїв, наявність доплерівського зсуву частот.

Зважаючи на проаналізовані обмеження та вимоги, слід зазначити, що оптимальними є такі комбінації класу випромінювання та швидкості маніпуляції, а саме:

MFSK-64, V=25 Бод;

MFSK-64, V=50 Бод;

MFSK-64, V=75 Бод.

Відповідно, на початку процесу встановлення зв'язку обирається варіант із швидкістю 50 Бод. На етапі адаптації комбінація може змінюватися на одну з представлених.

Етап 4. Визначення методу кодування для забезпечення стійкості до різних видів завад. Для виконання визначених вимог до забезпечення встановлення зв'язку необхідно застосовувати сучасні види завадостійкого кодування, такі як:

згорткове кодування (в системах типу HF DL та ALE третього покоління – код зі швидкістю $\frac{1}{2}$ та кодовим обмеженням 7);

каскадний код (згідно з [10], для ймовірності помилки на біт, що дорівнює 10^{-3} - каскадний код ефективніше за згортковий на 5 дБ);

турбокод (згідно з [10], для ймовірності помилки на біт, що дорівнює 10^{-3} - турбокод ефективніше за каскадний на 2,5 дБ);

код із низькою щільністю перевірки на парність (LDPC – код, згідно з [11], для ймовірності помилки на біт рівної 10^{-3} LDPC – код ефективніше турбокоду на 1 – 2 дБ).

Водночас при виборі конкретного виду завадостійкого кодування слід зважати на складність реалізації кодера та декодера [11].

Етап 5. Визначення складу та обсягу команд встановлення зв'язку. З метою адресного виклику кореспондента та забезпечення встановлення зв'язку на оптимальних частотах в складі команд виклику та відповіді повинні бути наступні елементи:

тип команди;

адреса кореспондента, якого викликають;
адреса кореспондента, який викликає;
тип даних до передачі;
тип апаратури (режим роботи);
частота відповіді кореспондента, якого викликають (команда на здійснення трасового випробування);

стисле повідомлення (за потреби);
контрольна сума повідомлення.
Тип команди має містити такі складові:
виклик;
відповідь на виклик;
відмова від зв'язку (розірвання);
передача тексту;
передача бітового потоку (дає змогу передавати додаткову інформацію, що не зазначена в певному протоколі – точний час, режими роботи, квитанції на пакети тощо)

Тип даних до передачі може включати такі складові:

стисле повідомлення;
повідомлення середнього обсягу;
велике за обсягом повідомлення;
передача файлів;
передача голосу в цифровій формі;
передача голосу в аналоговій формі;
потоківне відео;
відповідь на виклик;
завершення (розірвання) виклику;
елемент трасового випробування;
резерв.

Тип апаратури (режим роботи).

Частота відповіді кореспондента, якого викликають (команда на здійснення трасового випробування) має забезпечувати призначення частот ДКМ хвиль із кроком 1 кГц.

Контрольна сума повідомлення на основі циклічних надлишкових кодів CRC (Cyclic redundancy code) дає мінімальну надлишковість у дані, що передаються, та ідеально підходить для виявлення помилок, спричинених шумом та завадами в радіоканалі.

Етап 6. Аналіз параметрів каналу зв'язку. Після першого обміну інформацією між кореспондентами є змога провести аналіз наступних характеристик каналу зв'язку: частотні характеристики, рівень шумів, наявність завад тощо. Відповідний аналіз здійснюється у ході встановлення зв'язку та під час сеансу зв'язку для визначення оптимальних параметрів та адаптації сигнально-кової конструкції.

Етап 7. Адаптація сигнально-кової конструкції. З урахуванням проведеного аналізу характеристик каналу зв'язку проводиться адаптація конструкції по наступним параметрам:

комбінація класу випромінювання та швидкості маніпуляції;
метод кодування;
Відповідно проводиться зміна наступних параметрів випромінювання (за можливістю):
потужність передачі;
зміна передавальних та приймальних антен, їх спрямованість.

Висновки

У процесі огляду основних складових існуючих систем встановлення зв'язку в декаметровому діапазоні хвиль здійснено аналіз сигнально-кодових конструкцій та їх параметрів, на основі чого представлено оптимальні характеристики відповідних параметрів. Отож досліджено сигнально-кодові конструкції наявних систем АВЗ як однієї з головних частин, що впливає на якість системи.

Результатом роботи є розроблені методичні аспекти визначення параметрів сигнально-кодових конструкцій системи автоматизованого встановлення зв'язку для потреб воєнного часу, що містять в собі 7 етапів: визначення класу випромінювання, визначення швидкості маніпуляції, визначення комбінації класу випромінювання та швидкості маніпуляції, визначення методу кодування, визначення складу та обсягу команд встановлення зв'язку, аналіз

параметрів каналу зв'язку, адаптація сигнально-кодової конструкції. Новизна майбутніх розробок полягає у врахуванні вимог до систем автоматизованого встановлення зв'язку при використанні методичних аспектів визначення основних параметрів сигнально-кодових конструкцій для побудови систем АВЗ. Це дасть змогу удосконалити методичний апарат розробки систем військового радіозв'язку та уніфікувати процедури їх проектування.

Доцільність та актуальність використання систем АВЗ підтверджують стандарти MIL-STD-188-110D та MIL-STD-188-141D, однак виникає необхідність виконання сучасних вимог до військового радіозв'язку, а також врахування його особливостей.

Подальший напрямок дослідження полягає у розробці методики визначення основних параметрів сигнально-кодових конструкцій для побудови перспективних систем АВЗ з урахуванням вимог та особливостей військового радіозв'язку.

СПИСОК ЛІТЕРАТУРИ

1. Маковий В.А. Принципы построения помехоустойчивой системы АУС диапазона ДКМВ / В.А. Маковий, С.А. Шкуров // Теория и техника радиосвязи: наук.-техн. журнал / ВАТ "Концерн "Созвездие". – Воронеж, 2011. - № 4. – С. 33-43.
2. П. Эльдаров. Перспективная автоматизированная система связи WIN-T сухопутных войск США. Зарубежное военное обозрение. 2016. №1. С. 57 – 61.
3. Department of Defense Interface Standard, 2017. MIL-STD-188-141D INTEROPERABILITY AND PERFORMANCE STANDARDS FOR MEDIUM AND HIGH FREQUENCY RADIO SYSTEMS. USA.
4. Department of Defense Interface Standard, 2017. MIL-STD-188-110D INTEROPERABILITY AND PERFORMANCE STANDARDS FOR DATA MODEMS. USA.
5. Four generations of ALE – an HF COMM tutorial: Collins Aerospace, United Technologies Company, 2018. 23 с.
6. Bilal, A., Sun, G.: Automatic link establishment for HF radios. In: Proceedings 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, pp. 640–643 (2017)
7. Jorgenson, M.B., Cook, N.T.: Results from a wideband HF usability study. In: Proceedings IEEE Military Communications Conference, Tampa, FL, pp. 1454–1459 (2015)
8. Al-Dulaimi S.S., Kandeeran S.: Automatic Link Establishment for High Frequency Networks with Eavesdropping and Jamming. In: IEEE Transactions on Vehicular Technology, Vol. 65, No. 7, pp. 5701-5716 (2016)
9. Скляр Б. Цифровая связь. Теоретические основы и практическое применение. Изд. 2. испр. Пер с англ. – М.: Изд. Дом. «Вильямс», 2004. – 1104 с.
10. Пятін І. С. Дослідження касадних кодів та турбокодів / І.С. Пятін, В.В. Мішан, Д.В. Вершков // Вісник Хмельницького національного університету. Технічні науки. -2018. – №6. – С. 85-91.
11. Пятін І.С. Порівняння продуктивності завадостійких кодів на основі програмного HDL моделювання для захищених інформаційних технологій / І.С. Пятін, Ю.М. Бойко // Інформаційні та комп'ютерні технології / Університет «УКРАЇНА». – Київ, 2022. - № 1(03). – С. 43-68.

Received (Надійшла) 22.03.2023

Accepted for publication (Прийнята до друку) 24.05.2023

Methodological aspects of determining the main parameters of the signal-code constructions of the automated contact establishment system in sessions with correspondents

V. Mordiyuk, Y. Synytskiy

Abstract. Radio communication is considered the most important means of commanding military units, and is also a necessary condition for achieving information superiority over the enemy. In order to increase the efficiency of the use of forces and means, as well as for the successful performance of assigned tasks by units, it is advisable to use automated communication systems. However, the imperfection of the appropriate systems for establishing decameter radio communication, taking into account modern requirements for military radio communication, excludes the possibility of using the appropriate systems. The purpose of the article is to develop methodological aspects of determining the parameters of the signal-code structures of the system of automated communication establishment, which are expedient to create and implement for the needs of wartime. The article reviews the main elements of automatic communication establishment systems and analyzes signal-code structures as one of the main parts of the system. The methodological aspects of determining the main parameters of the signal-code structures of the automated communication establishment system have been developed, which makes it possible to build appropriate systems taking into account modern requirements for military radio communication. The expediency and relevance of the use of automated communication establishment systems is confirmed by technical standards MIL-STD-188 -110D and MIL-STD-188-141D.

Keywords: radio communication installation systems; decameter radio communication; information transmission systems; military radio communication; requirements for military radio communication; elements of the system; signal-code structures; MIL-STD standards.

Ye. Peleshok¹, M. Diedov¹, B. Nikolaenko²¹ Research Institute of Military Intelligence, Kyiv² Institute of Special Communications and Information Protection National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

INCOHERENT DEMODULATION OF TWO SYNCHRONOUS MUTUALLY NON-ORTHOGONAL DIGITAL SIGNALS WITH FREQUENCY SHIFT KEYING

Abstract. The synthesis method of the incoherent demodulation procedure of two synchronous mutually non-orthogonal digital signals with frequency modulation is considered. In the absence of interference this procedure degenerates into the procedure of classical incoherent demodulation of a digital signal with frequency modulation. When the instantaneous power of one of the signals significantly exceeds the instantaneous power of the other, the interference immunity of the latter approaches the immunity of reception in a channel with additive white Gaussian noise without interference. This procedure can be used in the development of modem compensators that ensure repeated use of the frequency resource as well as in the development of promising interference-protected radio communication devices.

Keywords: radio communication, digital signal, incoherent demodulation, frequency modulation.

Introduction

In modern conditions, reception of radio signals is carried out in an a priori uncertain signal-interference situation, which is due to the limitation of the radio frequency resource and the increase in the number and power of structural radiations of various origins. Therefore, the problem of demodulation of signals under the influence of interference was and remains relevant and a large number of works are aimed at its solution [1, 2].

In this article it is proposed to use demodulators of receiving devices synthesized on the basis of mathematical models of compensation procedures [3–6] to improve the immunity of the reception of a useful signal which is observed against the background of such strong interference.

The purpose and main content of the article is to solve the problem of synthesis of a mathematical model of the procedure of incoherent demodulation of mutually non-orthogonal digital signals with frequency shift keying (FSK). To achieve the goal, we will finalize and use the methodology given in [2, 7].

The procedure of incoherent demodulation of two synchronous mutually non-orthogonal digital signal with FSK

Let's write down the values of the useful FSK signal corresponding ($m = 2$) to two possible values of its discrete parameter $r_1 = 0, 1$. Suppose that the discrete parameter $r_1 = 1$ is modulated and transmitted at the frequency ω_1 , and $r_1 = 0$ at ω_2 then the general form of the useful FSK signal will be written as follows:

$$s_1(r_1, \varphi_{1c}, \varphi_{2c}, t) = r_1 [A_0 \cos(\omega_1 t + \varphi_{1c})] + (1 - r_1) [A_0 \cos(\omega_2 t + \varphi_{2c})],$$

where φ_{1c} , φ_{2c} – are the initial phases of the useful signal at frequencies ω_1 and ω_2 respectively which are random parameters due to the fluctuation of the spread time in the communication channel; A_0 – is the

amplitude of the useful signal, which is constant in frequency.

In turn, the powerful and similar interference FSK also takes two values of a discrete parameter $r_2 = 0, 1$. For our observation model let the discrete parameter $r_2 = 1$ be transmitted at frequency ω_1 and $r_2 = 0$ at ω_2 . Also, will be writing down the expression for a similar FSK interference as follows:

$$s_2(r_2, \varphi_{13}, \varphi_{23}, t) = r_2 A_{21} \cos(\omega_1 t + \varphi_{13}) + (1 - r_2) A_{22} \cos(\omega_2 t + \varphi_{23}),$$

where φ_{13} , φ_{23} – initial phases of interference at frequencies ω_1 and ω_2 ; A_{21} , A_{22} – amplitude of interference at frequencies ω_1 and ω_2 .

We will assume that the frequency positions and clock points of the signal and interference coincide and the modulation of the interference at each of the two frequency positions is carried out without a phase break. The last condition makes it possible to use coherent (quasi-coherent) interference processing, and we will process the useful signal incoherently (quadrature). We will also assume that additive interference in the form of additive white Gaussian noise acts in the communication channel.

On the Fig. 1 shows the useful signal in vector form $s_1(r_1, \varphi_{1c}, t)$ and interference signal $s_2(r_2, \varphi_{13}, t)$ which rotate in the positive direction with the same angular velocities ω_1 , but with different total phase values relative to the real axis.

First of all, in order to present a useful signal $s_1(r_1, \varphi_{1c}, t)$ against the background of powerful and structurally similar interference $s_2(r_2, \varphi_{13}, t)$ it is necessary to obtain the value of the in-phase and quadrature components of the amplitude of the useful signal relative to the interference.

To obtain the in-phase component of the amplitude, we project the vector of the useful signal onto the length A_0 of the signal-interference vector,

and to obtain the quadrature component, we draw the normal to the beginning of the vector of the obstacle and get the projection of the vector of the useful signal on it.

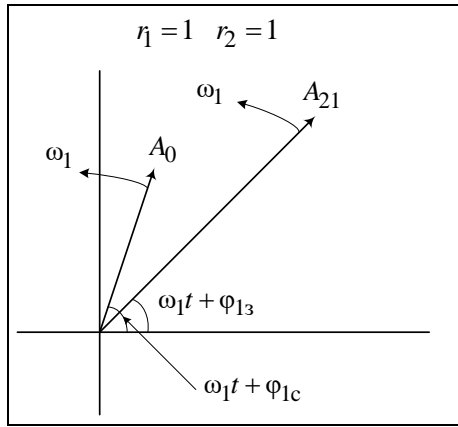


Fig. 1. Vector representation of useful signal and interference

Let us denote the angle between the vector of the useful signal and the interference $\Delta\varphi$ which in turn, will be equal to the difference of the complete phases of these vectors, namely $\Delta\varphi = \varphi_{1c} - \varphi_{13}$ (see Fig. 2).

The value of the in-phase A_1^s and quadrature components A_1^k of the amplitude of the useful signal $s_1(r_1, \varphi_{1c}, t)$ against the background of such interference $s_2(r_2, \varphi_{13}, t)$ is written as follows:

$$A_1^s = A_0 \cos(\varphi_{1c} - \varphi_{13});$$

$$A_1^k = A_0 \sin(\varphi_{1c} - \varphi_{13}).$$

Values A_2^s and A_2^k components of the useful signal $s_1(r_1, \varphi_{2c}, t)$ corresponding to the transmission of a discrete parameter $r_1 = 0$ against the background of such interference $s_2(r_2, \varphi_{23}, t)$, are obtained similarly

$$A_2^s = A_0 \cos(\varphi_{2c} - \varphi_{23}),$$

$$A_2^k = A_0 \sin(\varphi_{2c} - \varphi_{23}).$$

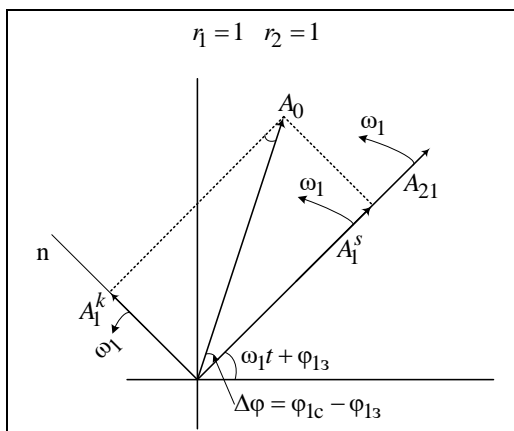


Fig. 2. Angular relations between signal and interference

In Fig. 3 signal interference is represented in vector form $s_2(r_2, \varphi_{13}, t)$, in-phase A_1^s and quadrature

A_1^k components of the amplitude of the useful signal $s_1(r_1, \varphi_{1c}, t)$ rotate with the same angular velocity ω_1 .

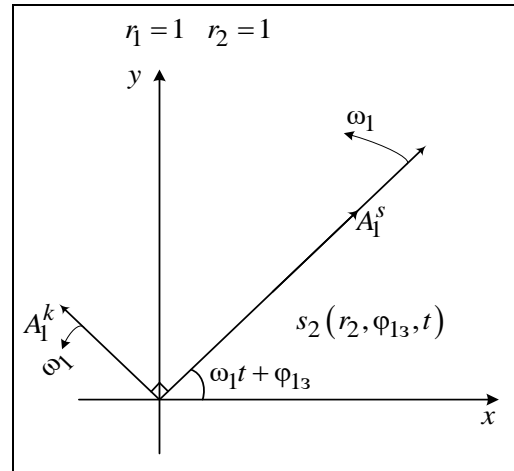


Fig. 3. Display of in-phase and quadrature components of the amplitude of the useful signal and interference

The value of the full phase relative to the axis x of the in-phase component A_1^s coincides with the value of the full phase $\omega_1 t + \varphi_{13}$ of the interference vector, the value of the full phase relative to the axis x of the quadrature component A_1^k is greater $\pi/2$ than the full phase of the interference vector.

The use of the formula for summing trigonometric functions $\cos(\pi/2 + \alpha) = -\sin \alpha$, write down the full expression of the useful signal $s_1(r_1, \varphi_{1c}, t)$ that is observed against the background of a powerful and structurally similar interference $s_2(r_2, \varphi_{13}, t)$:

$$s_1(r_1, \varphi_{1c}, t) = r_1 [A_1^s \cos(\omega_1 t + \varphi_{13}) - A_1^k \sin(\omega_1 t + \varphi_{13})] =$$

$$= r_1 [A_0 \cos(\varphi_{1c} - \varphi_{13}) \cos(\omega_1 t + \varphi_{13}) -$$

$$- A_0 \sin(\varphi_{1c} - \varphi_{13}) \sin(\omega_1 t + \varphi_{13})]. \quad (1)$$

Expression for the useful signal $s_1(r_1, \varphi_{2c}, t)$ that is observed against the background of a powerful interference $s_2(r_2, \varphi_{23}, t)$ is obtained in a similar way and has the following form:

$$s_1(r_1, \varphi_{2c}, t) = (1 - r_1) [A_2^s \cos(\omega_2 t + \varphi_{23}) -$$

$$- A_2^k \sin(\omega_2 t + \varphi_{23})] = (1 - r_1) [A_0 \cos(\varphi_{2c} - \varphi_{23}) \times$$

$$\times \cos(\omega_2 t + \varphi_{23}) - A_0 \sin(\varphi_{2c} - \varphi_{23}) \sin(\omega_2 t + \varphi_{23})]. \quad (2)$$

The general model of monitoring the duration of the clock interval $T = t_k - t_{k-1}$ will be presented as:

$$y(t) = s_1(r_1, \varphi_{1c}, \varphi_{2c}, t) + s_2(r_2, \varphi_{13}, \varphi_{23}, t) + n(t) =$$

$$= r_1 [A_1^s \cos(\omega_1 t + \varphi_{13}) - A_1^k \sin(\omega_1 t + \varphi_{13})] + \quad (3)$$

$$+ (1 - r_1) [A_2^s \cos(\omega_2 t + \varphi_{23}) - A_2^k \sin(\omega_2 t + \varphi_{23})] +$$

$$+ r_2 A_{21} \cos(\omega_1 t + \varphi_{13}) + (1 - r_2) A_{22} \cos(\omega_2 t + \varphi_{23}) + n(t),$$

where $n(t)$ – additive white Gaussian noise.

We will also assume that the states of the discrete parameters r_1 and r_2 are equally likely and mutually independent and the initial phases $\varphi_{1,2,c,3}$ are uniformly distributed over the interval $[0, 2\pi]$.

In addition, with the already proposed rejection of the estimation A_0 of the amplitude of the useful signal, the equation is obvious $h_1^2 = h_2^2$. However, the assumption of abandoning the estimation of the amplitude of the useful signal and replacing it with its value $A_0 \ll A_{21}, A_0 \ll A_{22}$ does not allow to neglect the fact that in the general case $A_{21} \neq A_{22}$, because the case is possible when the difference in amplitudes $|A_{21} - A_{22}|$ commensurate with A_0 .

In the future, for the sake of minimizing records and ease of understanding, we will assume that the useful signal is transmitted at a frequency ω_1 corresponding to the value of the discrete parameter $r_1 = 1$.

Received signal is considered as the sum of the useful signal and interference. For our case, the average power will be write as follows [8]:

$$P_{r_1, r_2, \varphi_{1c, 3}} = \frac{1}{T} \int_{t_{k-1}}^{t_k} [s_1(r_1, \varphi_{1c}, t) + s_2(r_2, \varphi_{13}, t)]^2 dt. \quad (4)$$

Using the formula of the likelihood function for a signal with a random initial phase [9, 10]

$$\Lambda_r[y(t); \varphi] = \exp\left\{-\frac{P_{r, \varphi} \cdot T}{N_0}\right\} \cdot \exp\left\{\frac{2T}{N_0} \cdot b_r[y(t); \varphi]\right\},$$

where $P_{r, \varphi}$ – the average power of the received signal

$$s(r, \varphi, t); \quad b_r[y(t), \varphi] = b_r = \frac{1}{T} \int_{t_{k-1}}^{t_k} y(t) \cdot s(r, \varphi, t) dt -$$

scalar product of the input observation $y(t)$ and $s(r, \varphi, t)$, has the following write of the conditional probability functional for the observation (3)

$$\Lambda_{r_1=1, r_2=1}[y(t); \varphi_{1c}, \varphi_{13}] = \exp\left\{-\frac{P_{r_1, r_2, \varphi_{1c, 3}}}{N_0}\right\} \times \exp\left\{\frac{2}{N_0} \cdot b_{r_1}[y(t), \varphi_{1c}]\right\} \cdot \exp\left\{\frac{2T}{N_0} \cdot b_{r_2}[y(t), \varphi_{13}]\right\}, \quad (5)$$

$$\text{where } b_{r_1}[y(t), \varphi_{1c}] = b_{r_1} = \int_{t_{k-1}}^{t_k} y(t) \cdot s_1(r_1, \varphi_{1c}, t) dt,$$

$$b_{r_2}[y(t), \varphi_{13}] = b_{r_2} = \int_{t_{k-1}}^{t_k} y(t) \cdot s_2(r_2, \varphi_{13}, t) dt. \quad (6)$$

Substitute (4) and (6) into (5)

$$\Lambda_{r_1=1, r_2=1}[y(t); \varphi_{1c, 3}] = \exp\left[\frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) \cdot s_1(r_1, \varphi_{1c}, t) dt + \frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) \cdot s_2(r_2, \varphi_{13}, t) dt - \frac{1}{N_0} \int_{t_{k-1}}^{t_k} s_1^2(r_1, \varphi_{1c}, t) dt -$$

$$-\frac{1}{N_0} \int_{t_{k-1}}^{t_k} s_2^2(r_2, \varphi_{13}, t) dt - \frac{2}{N_0} \int_{t_{k-1}}^{t_k} s_1(r_1, \varphi_{1c}, t) \cdot s_2(r_2, \varphi_{13}, t) dt\right]. \quad (7)$$

Substitute the expression for the useful signal (1) and the interference value $s_2(r_2, \varphi_{13}, t)$ which corresponds to the transmission of the discrete parameter $r_2 = 1$ in (7)

$$\Lambda_{r_1=1, r_2=1}[y(t); \varphi_{1c}; \varphi_{13}] = \exp\left\{\left[\frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) A_0 \cos(\varphi_{1c} - \varphi_{13}) \cos(\omega_1 t + \varphi_{13}) dt - \frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) A_0 \sin(\varphi_{1c} - \varphi_{13}) \sin(\omega_1 t + \varphi_{13}) dt\right] + \left[\frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) A_{21} \cos(\omega_2 t + \varphi_{13}) dt - \left[\frac{1}{N_0} \int_{t_{k-1}}^{t_k} A_0^2 \cos^2(\varphi_{1c} - \varphi_{13}) \cos^2(\omega_1 t + \varphi_{13}) dt - \frac{1}{N_0} \int_{t_{k-1}}^{t_k} A_0^2 \sin^2(\varphi_{1c} - \varphi_{13}) \sin^2(\omega_1 t + \varphi_{13}) dt - \frac{1}{N_0} \int_{t_{k-1}}^{t_k} A_{21}^2 \cos^2(\omega_2 t + \varphi_{13}) dt - \frac{2}{N_0} \left[\int_{t_{k-1}}^{t_k} A_{21} A_0 \cos(\varphi_{1c} - \varphi_{13}) \cos^2(\omega_1 t + \varphi_{13}) dt - \int_{t_{k-1}}^{t_k} A_{21} A_0 \sin(\varphi_{1c} - \varphi_{13}) \sin(\omega_1 t + \varphi_{13}) \cos(\omega_1 t + \varphi_{13}) dt\right]\right]\right\}. \quad (8)$$

Write the right-hand part of (8) as follows:

$$-\frac{2}{N_0} \int_{t_{k-1}}^{t_k} A_{21} A_0 \cos(\varphi_{1c} - \varphi_{13}) \cos^2(\omega_1 t + \varphi_{13}) dt, \int_{t_{k-1}}^{t_k} \sin(\omega_1 t + \varphi_{13}) \cos(\omega_1 t + \varphi_{13}) dt = 0.$$

Introduce the notation in (8) taking into account (5) and (6):

$$b_{r_1=1}^s = \frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) A_0 \cos(\varphi_{1c} - \varphi_{13}) \cos(\omega_1 t + \varphi_{13}) dt = b_{r_1=1}^{s0} \cos(\varphi_{1c} - \varphi_{13});$$

$$b_{r_1=1}^k = \frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) A_0 \sin(\varphi_{1c} - \varphi_{13}) \sin(\omega_1 t + \varphi_{13}) dt = b_{r_1=1}^{k0} \sin(\varphi_{1c} - \varphi_{13});$$

$$b_{r_2=1} = \frac{2}{N_0} \int_{t_{k-1}}^{t_k} y(t) A_{21} \cos(\omega_2 t + \varphi_{13}) dt;$$

$$\begin{aligned}
 h_{\eta_1=1}^2 &= \frac{1}{N_0} \int_{t_{k-1}}^{t_k} A_0^2 \cos^2(\varphi_{1c} - \varphi_{13}) \cos^2(\omega_1 t + \varphi_{13}) dt - \\
 &- \frac{1}{N_0} \int_{t_{k-1}}^{t_k} A_0^2 \sin^2(\varphi_{1c} - \varphi_{13}) \sin^2(\omega_1 t + \varphi_{13}) dt = \\
 &= h_{\eta_1=1,s}^2 + h_{\eta_1=1,k}^2; \\
 h_{\eta_2=1}^2 &= \frac{1}{N_0} \int_{t_{k-1}}^{t_k} A_{21}^2 \cos^2(\omega_1 t + \varphi_{13}) dt; \\
 R_{\eta_1=1, \eta_2=1} &= \frac{1}{N_0} \int_{t_{k-1}}^{t_k} A_{21} A_0 \cos(\varphi_{1c} - \varphi_{13}) \cos^2(\omega_1 t + \varphi_{13}) dt = \\
 &= R_{\eta_1=1, \eta_2=1}^0 \cos(\varphi_{1c} - \varphi_{13}). \quad (9)
 \end{aligned}$$

Taking into account notation (9), the conditional probability functional (8) has presented as follows:

$$\begin{aligned}
 \Lambda_{\eta_1=1, \eta_2=1} [y(t); \varphi_{1c}, \varphi_{13}] &= \\
 = \exp \left[\left[b_{\eta_1=1}^{s0} \cos(\varphi_{1c} - \varphi_{13}) - b_{\eta_1=1}^{k0} \sin(\varphi_{1c} - \varphi_{13}) \right] + b_{\eta_2=1} - \right. \\
 \left. - h_{\eta_1=1}^2 - h_{\eta_2=1}^2 - 2R_{\eta_1=1, \eta_2=1}^0 \cos(\varphi_{1c} - \varphi_{13}) \right]. \quad (10)
 \end{aligned}$$

In case of quasi-coherent processing of interference and under the condition $h_{\eta_2=1}^2 \gg 1$ the initial phase estimation error can be neglected $\varphi_{13} = 0$. Average (10) over φ_{1c} the interval $[0, 2\pi]$ and get the unconditional probability functional:

$$\begin{aligned}
 \Lambda_{\eta_1=1, \eta_2=1} [y(t); \varphi_{1c}] &= \frac{\exp(-h_{\eta_1=1}^2)}{2\pi} \exp(b_{\eta_2=1} - h_{\eta_2=1}^2) \times \\
 \times \int_0^{2\pi} \exp \left[\left(b_{\eta_1=1}^{s0} \cos \varphi_{1c} - b_{\eta_1=1}^{k0} \sin \varphi_{1c} \right) - \right. \\
 \left. - 2R_{\eta_1=1, \eta_2=1}^0 \cos \varphi_{1c} \right] d\varphi_{1c}. \quad (11)
 \end{aligned}$$

In the following, we will ignore the factor $\exp(-h_{\eta_1=1}^2) / 2\pi$ which does not depend on the values of the discrete parameter r_1 and r_2 . Taking into account the previously introduced notations (9) we write:

$$R_{\eta_1=1, \eta_2=1}^0 = \frac{h_{\eta_2=1}^2 A_0}{A_{21}} = \alpha_{21} h_{\eta_2=1}^2. \quad (12)$$

Introduce the notation that facilitates the integration procedure (11):

$$B_1 = \sqrt{\left(b_{\eta_1=1}^{s0} \right)^2 + \left(b_{\eta_1=1}^{k0} \right)^2}; \quad \psi_1 = \arctg \frac{b_{\eta_1=1}^{k0}}{b_{\eta_1=1}^{s0}}, \quad (13)$$

$$b_{\eta_1=1}^{s0} = B_1 \cos \psi_1; \quad b_{\eta_1=1}^{k0} = B_1 \sin \psi_1. \quad (14)$$

Taking into account (12)–(14), (11) will be rewritten in the following form:

$$\begin{aligned}
 \Lambda_{\eta_1=1, \eta_2=1} [y(t); \varphi_{1c}] &= \exp(b_{\eta_2=1} - h_{\eta_2=1}^2) \times \\
 \times \int_0^{2\pi} \exp \left\{ B_1 (\cos \psi_1 \cos \varphi_{1c} - \sin \psi_1 \sin \varphi_{1c}) - \right.
 \end{aligned}$$

$$\left. - 2\alpha_{21} h_{\eta_2=1}^2 \cos \varphi_{1c} \right\} d\varphi_{1c}. \quad (15)$$

Values for unconditional probability functionals for the following options for transmitting discrete parameters: $r_1 = 1 \quad r_2 = 0$; $r_1 = 0 \quad r_2 = 1$; $r_1 = 0 \quad r_2 = 0$, useful signal and interference obtained in a similar way:

$$\begin{aligned}
 \Lambda_{\eta_1=1, \eta_2=0} [y(t); \varphi_{1c}] &= \exp(b_{\eta_2=0} - h_{\eta_2=0}^2) \times \\
 \times \int_0^{2\pi} \exp \left\{ B_1 (\cos \psi_1 \cos \varphi_{1c} - \sin \psi_1 \sin \varphi_{1c}) \right\} d\varphi_{1c}; \\
 \Lambda_{\eta_1=0, \eta_2=1} [y(t); \varphi_{2c}] &= \exp(b_{\eta_2=1} - h_{\eta_2=1}^2) \times \\
 \times \int_0^{2\pi} \exp \left\{ B_2 (\cos \psi_2 \cos \varphi_{2c} - \sin \psi_2 \sin \varphi_{2c}) \right\} d\varphi_{2c}; \\
 \Lambda_{\eta_1=0, \eta_2=0} [y(t); \varphi_{2c}] &= \exp(b_{\eta_2=0} - h_{\eta_2=0}^2) \times \\
 \times \int_0^{2\pi} \exp \left\{ B_2 (\cos \psi_2 \cos \varphi_{2c} - \sin \psi_2 \sin \varphi_{2c}) - \right. \\
 \left. - 2\alpha_{22} h_{\eta_2=0}^2 \cos \varphi_{2c} \right\} d\varphi_{2c}. \quad (16)
 \end{aligned}$$

Let us write the unconditional likelihood functional in the general sense, taking into account (15) and (16)

$$\begin{aligned}
 \Lambda_{\eta_1, \eta_2} [y(t); \varphi_{1c}, \varphi_{2c}] &= \\
 = \exp \left\{ r_2 (b_{\eta_2=1} - h_{\eta_2=1}^2) + (1 - r_2) (b_{\eta_2=0} - h_{\eta_2=0}^2) \right\} \times \\
 \times \int_0^{2\pi} \int_0^{2\pi} \exp \left\{ r_1 B_1 (\cos \psi_1 \cos \varphi_{1c} - \sin \psi_1 \sin \varphi_{1c}) + \right. \\
 \left. + (1 - r_1) B_2 (\cos \psi_2 \cos \varphi_{2c} - \sin \psi_2 \sin \varphi_{2c}) - \right. \\
 \left. - 2r_1 r_2 \alpha_{21} h_{\eta_2=1}^2 \cos \varphi_{1c} - 2(1 - r_1)(1 - r_2) \times \right. \\
 \left. \times \alpha_{22} h_{\eta_2=0}^2 \cos \varphi_{2c} \right\} d\varphi_{1c} d\varphi_{2c}.
 \end{aligned}$$

For an equally probable discrete parameter of the useful signal the decision rule has the form:

$$r_1^* = \text{rect} \left[\Lambda_{\eta_1=1, \eta_2=0} [y(t); \varphi_{1c}] + \Lambda_{\eta_1=1, \eta_2=1} [y(t); \varphi_{1c}] - \right. \\
 \left. - \Lambda_{\eta_1=0, \eta_2=1} [y(t); \varphi_{2c}] - \Lambda_{\eta_1=0, \eta_2=0} [y(t); \varphi_{2c}] \right], \quad (17)$$

where $\text{rect}(x \geq 0) = 1$; $\text{rect}(x < 0) = 0$ – decisive function.

Taking into account (15) and (16), the decision rule (17) for the equally probable discrete parameter of the useful signal:

$$\begin{aligned}
 r_1^* &= \text{rect} \left[\exp(b_{\eta_2=0} - h_{\eta_2=0}^2) \times \right. \\
 \times \int_0^{2\pi} \exp \left\{ B_1 (\cos \psi_1 \cos \varphi_{1c} - \sin \psi_1 \sin \varphi_{1c}) \right\} d\varphi_{1c} + \\
 + \exp(b_{\eta_2=1} - h_{\eta_2=1}^2) \int_0^{2\pi} \exp \left\{ B_1 (\cos \psi_1 \cos \varphi_{1c} - \sin \psi_1 \sin \varphi_{1c}) - \right. \\
 \left. - 2\alpha_{21} h_{\eta_2=1}^2 \cos \varphi_{1c} \right\} d\varphi_{1c} - \exp(b_{\eta_2=1} - h_{\eta_2=1}^2) \times \\
 \times \int_0^{2\pi} \exp \left\{ B_2 (\cos \psi_2 \cos \varphi_{2c} - \sin \psi_2 \sin \varphi_{2c}) \right\} d\varphi_{2c} -
 \end{aligned}$$

$$-\exp\left(b_{r_2=0} - h_{r_2=0}^2\right) \int_0^{2\pi} \exp\left\{B_2(\cos\psi_2 \cos\varphi_{2c} - \sin\psi_2 \sin\varphi_{2c}) - 2\alpha_{22}h_{r_2=0}^2 \cos\varphi_{2c}\right\} d\varphi_{2c} \Big]. \quad (18)$$

Equivalent and simplifying transformations of the procedure for incoherent demodulation of a useful signal with a FSK observed against the background of strong similar interference

Change the variable of integration in (18) to $\xi = \psi_{1,2} + \varphi_{1c,2c}$ after which we get from (18):

$$r_1^* = \text{rect} \left\{ \exp\left(b_{r_2=0} - h_{r_2=0}^2\right) \left[\int_0^{2\pi} \exp(B_1 \cos\xi) d\xi - \int_0^{2\pi} \exp\left[\left(b_{\eta=0}^{s0} - 2\alpha_{22}h_{r_2=0}^2\right) \cos\varphi_{2c} - b_{\eta=0}^{k0} \sin\varphi_{2c}\right] d\varphi_{2c} \right] - \exp\left(b_{r_2=1} - h_{r_2=1}^2\right) \left[\int_0^{2\pi} \exp(B_2 \cos\xi) d\xi - \int_0^{2\pi} \exp\left[\left(b_{\eta=1}^{s0} - 2\alpha_{21}h_{r_2=1}^2\right) \cos\varphi_{1c} - b_{\eta=1}^{k0} \sin\varphi_{1c}\right] d\varphi_{1c} \right] \right\}. \quad (19)$$

Introduce the notation similarly to (13), (14):

$$\begin{aligned} b_{\eta=1,e}^{s0} &= b_{\eta=1}^{s0} - 2\alpha_{21}h_{r_2=1}^2; \\ B_{1e} &= \sqrt{\left(b_{\eta=1,e}^{s0}\right)^2 + \left(b_{\eta=1}^{k0}\right)^2}; \\ b_{\eta=0,e}^{s0} &= b_{\eta=0}^{s0} - 2\alpha_{22}h_{r_2=0}^2; \\ B_{2e} &= \sqrt{\left(b_{\eta=0,e}^{s0}\right)^2 + \left(b_{\eta=0}^{k0}\right)^2}; \\ \eta_1 &= \arctg \frac{b_{\eta=1}^{k0}}{b_{\eta=1,e}^{s0}}; \quad \eta_2 = \arctg \frac{b_{\eta=0}^{k0}}{b_{\eta=0,e}^{s0}}, \end{aligned}$$

whence follows

$$\begin{aligned} b_{\eta=1,e}^{s0} &= B_{1e} \cos\eta_1; & b_{\eta=1}^{k0} &= B_{1e} \sin\eta_1; \\ b_{\eta=0,e}^{s0} &= B_{2e} \cos\eta_2; & b_{\eta=0}^{k0} &= B_{2e} \sin\eta_2. \end{aligned}$$

The decision rule (19) will have the following form:

$$r_1^* = \text{rect} \left[\exp\left(b_{r_2=0} - h_{r_2=0}^2\right) \left(\int_0^{2\pi} \exp(B_1 \cos\xi) d\xi - \int_0^{2\pi} \exp\left\{B_{2e}(\cos\eta_2 \cos\varphi_{2c} - \sin\eta_2 \sin\varphi_{2c})\right\} d\varphi_{2c} \right) - \exp\left(b_{r_2=1} - h_{r_2=1}^2\right) \left(\int_0^{2\pi} \exp(B_2 \cos\xi) d\xi - \int_0^{2\pi} \exp\left\{B_{1e}(\cos\eta_1 \cos\varphi_{1c} - \sin\eta_1 \sin\varphi_{1c})\right\} d\varphi_{1c} \right) \right]. \quad (20)$$

After replacing the integration variables in the second and fourth integrals the decision rule (20) into $\eta_{1,2} + \varphi_{1c,2c}$ is obtained [11]:

$$r_1^* = \text{rect} \left[\exp\left(b_{r_2=0} - h_{r_2=0}^2\right) \left[I_0(B_1) - I_0(B_{2e}) \right] + \exp\left(b_{r_2=1} - h_{r_2=1}^2\right) \left[I_0(B_{1e}) - I_0(B_2) \right] \right], \quad (21)$$

where $I_0(\dots)$ – zero-order Bessel function.

We can see what

$$h_{r_2=1,r_2=0}^2 \gg 1, \quad h_{r_2=1,r_2=0}^2 \gg h_{r_1=1,r_1=0}^2$$

$$\exp\left(b_{r_2=1} - h_{r_2=1}^2\right) \Big|_{r_2=1} \gg 1;$$

$$\exp\left(b_{r_2=0} - h_{r_2=0}^2\right) \Big|_{r_2=1} \cong 0;$$

$$\exp\left(b_{r_2=1} - h_{r_2=1}^2\right) \Big|_{r_2=0} \cong 0;$$

$$\exp\left(b_{r_2=0} - h_{r_2=0}^2\right) \Big|_{r_2=0} \gg 1.$$

Then the decision rule (21) can be replaced by an asymptotically equivalent one:

$$r_1^* = \text{rect} \left[\left(b_{r_2=0} - h_{r_2=0}^2 \right) \left[I_0(B_1) - I_0(B_{2e}) \right] + \text{rect} \left[\left(b_{r_2=1} - h_{r_2=1}^2 \right) \left[I_0(B_{1e}) - I_0(B_2) \right] \right] \right]. \quad (22)$$

The approximate decision-making procedure (22) r_1^* is two-stage where at the first stage a decision is made on which of the frequencies the interference is emitted $s_2(r_2, \varphi_{13}, \varphi_{23}, t)$.

If the interference energy significantly exceeds the useful signal energy $s_1(r_1, \varphi_{1c}, \varphi_{2c}, t)$ then the value

$\text{rect}\left(b_{r_2=1,r_2=0} - h_{r_2=1,r_2=0}^2\right)$ in (22) due to the small influence of errors on the overall decision r_1^* should be replaced by one decision-making rule for coherent (quasi-coherent) reception of the FSK signal [9,10]:

$$r_2^* = \text{rect}\left(b_{r_2=1} - b_{r_2=0}\right).$$

As a result (22) will become:

$$r_1^* = \text{rect} \left[\text{rect}\left(b_{r_2=0} - b_{r_2=1}\right) (B_1 - B_{2e}) + \text{rect}\left(b_{r_2=1} - b_{r_2=0}\right) (B_{1e} - B_2) \right], \quad (23)$$

where it is taken into account that the function $I_0(x)$ is monotonic at $x > 0$.

In the absence of interference $s_2(r_2, \varphi_{13}, \varphi_{23}, t)$ that is when $h_{r_2=1,r_2=0}^2 = 0$ the decision rule (21)–(23) degenerates into the classical rules of incoherent reception of the FSK signal.

We will evaluate the interference resistance of the received decision rule for the asymptotic case of an unlimited increase in the average interference power $s_2(r_2, \varphi_{13}, \varphi_{23}, t)$.

Assuming that the estimation errors of continuous $(A_{21}, A_{22}, \varphi_{1,23})$ parameters and the discrete disturbance parameter will approach zero, we obtain the following expressions for $b_{\eta=1}^s, b_{\eta=0}^s$:

$$b_{\eta=1}^s \Big|_{r_2=1} = \frac{2}{N_0} \int_{t_{k-1}}^{t_k} \left[(A_1^s + A_{21}) \cos(\omega_1 t + \varphi_{13}) + n(t) \right] \times \\ \times A_1^s \cos(\omega_1 t + \varphi_{13}) dt = 2h_{\eta=1,s}^2 + 2a_{21}h_{r_2=1}^2 \cos(\varphi_{1c} - \varphi_{13}) + n_{\text{ш1}} ; \\ b_{\eta=0}^s \Big|_{r_2=0} = \frac{2}{N_0} \int_{t_{k-1}}^{t_k} \left[(A_2^s + A_{22}) \cos(\omega_2 t + \varphi_{23}) + n(t) \right] \times \\ \times A_2^s \cos(\omega_2 t + \varphi_{23}) dt = 2h_{\eta=0,s}^2 + \\ + 2a_{22}h_{r_2=0}^2 \cos(\varphi_{2c} - \varphi_{23}) + n_{\text{ш2}} . \quad (24)$$

It can be seen from comparison (23) and (24) that under the above-mentioned assumptions (about the absence of errors in the estimation of interference parameters), the components in the correlation integrals $b_{\eta=1,r_2=0}^s$ generated by its presence are fully compensated.

The noise components $n_{\text{ш1}}$ and $n_{\text{ш2}}$ remain the same as for the classical case of incoherent reception of the FSK signal.

Thus, the potential interference immunity of the incoherent demodulation algorithm (21) of the FSK signal, provided that the average power of such FSK

interference significantly exceeds the power of the useful signal and there are no errors in the estimation of its parameters, is the same as in its absence.

Conclusions

This procedure of incoherent demodulation of mutually non-orthogonal digital signals with frequency modulation has a number of advantages:

– provided that the average power of such interference significantly exceeds the power of the useful FSK signal and there are no errors in the estimation of the interference parameters, the potential (limit) immunity of the incoherent demodulation procedure (23) is the same as in the absence of interference;

– provided that the average power of such interference significantly exceeds the power of the useful FSK signal and there are no errors in the estimation of the interference parameters, the potential (limit) immunity of the incoherent demodulation procedure (23) is the same as in the absence of interference;

– this mathematical model of the procedure can be used in the implementation of frequency resource reuse programs and in the development of promising interference-protected radio communication tools.

REFERENCES

1. Bobrovskiy V. Y. Mnohopol'zovatel'skoe detektyrovaniye / V. Y. Bobrovskiy. – Ulianovsk. : Vektor – 2007. – 348 s.
2. Yerokhin V. F. Alhorytm demoduliyatsii, shcho zabezpechuie povtornе vykorystannia chastot tsyfrovoho radiomovlennia / V. F. Yerokhin, I. M. Krutofist // Zakhyst informatsii. – 2005. – № 25. – P. 42–47.
3. Burachenko D. L. Potentsyalnaia pomekhо- ustoichyvost rozdeleniya tsyfrovyykh syhnalov. Metodyka, prohrammy, rezultaty raschetov / D. L. Burachenko, V. F. Erokhyn, V. O. Rashych – L., 1987. – 122 p.
4. Yerokhin V. F. Optymalni alhorytmy rozdilennia dvokh vzaïemno neortohonalnykh syhnaliv / V. F. Yerokhin, Ye. V. Peleshok // Visnyk NTUU “KPI”. Seriia Radiotekhnika. Radioaparobuduvannia. – 2012. – V. 49. – P. 33–41.
5. Yerokhin V. F. Optymalnaia demoduliyatsiia tsyfrovoho syhnala pry addytyvnom vozdeïstvyi moshchnoi podobnoi pomekhy // V. F. Yerokhin V. N. Raevskiy // Yzv. vuzov. NTUU “KPI”. Seryia Radioelektronika. – 2009. – T. 52, № 9. – P. 17–29.
6. Yerokhin V. F. Syntez alhorytma rozdeleniya heterokhronnykh tsyfrovyykh syhnalov / V. F. Yerokhin, D. V. Liulyn // Elektronnoe modelirovanie. – 1999. – T. 21, № 5. – P. 46–54.
7. Yerokhin V. F. Demoduliyatsiia konfliktuiushchykh tsyfrovyykh syhnalov / V. F. Yerokhin. – K. : KVIUS – YK. V. M. Hlushkova, 1993. – 132 p.
8. Yerokhin V. F. Asymptotychna efektyvnist koherentnykh demoduliyatoriv tsyfrovyykh syhnaliv, shcho sposterihaiutsia na foni podibnykh potuzhnykh zavod / V. F. Yerokhin, I. M. Krutofist // Trudy akademii. NAOU. – 2005. – № 65. – P. 76–81.
9. Fink L. M. Teoriia peredachi dyskretnykh soobshcheniy / L. M. Fink. – M. : Sov. radyo, 1970. – 728 p.
10. Khvorostenko N. P. Statystycheskaia teoriia demoduliyatsiy dyskretnykh syhnalov / N. P. Khvorostenko. – M. : Sviiaz, 1968. – 336 p.
11. Prudnykov A. P. Yntehraly y riady. Tom 1. Elementarnye funktsyy / A. P. Prudnykov, Yu. A. Brychkov, O. Y. Marychev. – M. : Nauka. Hlav. red. fiz.-mat. lyt., 1981. – 798 p.

Received (Надійшла) 15.03.2023

Accepted for publication (Прийнята до друку) 24.05.2023

Некогерентний прийом двох синхронних взаємно заважаючих цифрових сигналів з частотною маніпуляцією

Є. В. Пелешок, М. А. Дєдов, Б. А. Ніколаєнко

Анотація. Розглянуто синтез процедури некогерентної демодуляції двох синхронних взаємно неортогональних цифрових сигналів з частотною маніпуляцією. За відсутності завади дана процедура вироджується в процедуру класичної некогерентної демодуляції цифрового сигналу з частотною маніпуляцією. Коли миттєва потужність одного з сигналів значно перевищує миттєву потужність іншого, завадостійкість останнього наближається до завадостійкості прийому в каналі з адитивним білим гаусовим шумом без завади. Ця процедура може бути використана при розробці модемних компенсаторів, що забезпечують повторне використання радіочастотного ресурсу, а також при розробці перспективних завадозахищених пристроїв радіозв'язку.

Ключові слова: радіозв'язок, цифровий сигнал, некогерентна демодуляція, частотна модуляція.

Л. М. Сакович¹, Г. Я. Криховецький², Я. Е. Курята¹

¹ Інститут спеціального зв'язку та захисту інформації Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна

² Науково-дослідний інститут військової розвідки, Київ, Україна

ОЦІНКА СТРУКТУРНОЇ НАДІЙНОСТІ СИСТЕМ ЗВ'ЯЗКУ

Анотація. В статті запропоновано, для підвищення точності кількісної оцінки часткових і комплексних показників надійності напрямків зв'язку, вперше врахувати час роботи окремих засобів зв'язку. Сутність пропозицій полягає у використанні математичної моделі оцінки значень показників надійності об'єктів зі змінною структурою. Тобто, для організації зв'язку між абонентами, можливе використання різноманітних каналів зв'язку із множини можливих, які завантажені у різний час. Формалізовано рішення цієї задачі і приведено приклад використання результатів досліджень. Ефект від реалізації пропозицій щодо кількісної оцінки значень показників надійності напрямків зв'язку полягає у забезпеченні необхідного рівня надійності мінімальною кількістю каналів і засобів зв'язку.

Ключові слова: структурна надійність, напрямки зв'язку, оцінка часткових і комплексних показників надійності.

Вступ

Досвід військових конфліктів останніх років свідчить про необхідність забезпечення високої стійкої систем зв'язку, тобто їх живучості й надійності. Оцінка структурної надійності під час модернізації існуючих і проектуванні перспективних цифрових систем зв'язку спрямована на мінімізацію необхідної кількості каналів і засобів зв'язку при забезпеченні високих вимог до їхнього застосування за призначенням з необхідною якістю. Але існуючі методики не враховують сучасні досягнення у галузі надійності великих технічних систем.

Постановка проблеми. Не має сенсу оцінювати надійність системи зв'язку в цілому, через те, що відмова засобів зв'язку в одному напрямку не впливає на якість зв'язку в інших. Тобто необхідно кількісно оцінювати структурну надійність напрямків зв'язку між конкретними абонентами. Але в дійсний час при цьому враховують структуру системи зв'язку і статистичні дані щодо відмов окремих засобів зв'язку в певних умовах. Однак, до теперішнього часу, відомі методики оцінки надійності системи зв'язку не враховують час роботи окремих засобів, що суттєво впливає на надійність напрямків зв'язку.

Аналіз останніх досліджень і публікацій. Сучасні дослідження в галузі теорії надійності складних технічних об'єктів і систем спрямовані на отримання необхідних значень показників якості за рахунок впровадження резервування найменш надійних елементів, що має місце при організації зв'язку різноманітними засобами. При цьому традиційно використовують оцінку комплексного показника надійності – коефіцієнта готовності за умови, що всі елементи системи зв'язку працюють одночасно [1-3]. Але реально системи зв'язку використовують різний час, що впливає на показники надійності. Це показано на прикладі засобів зв'язку, які відрізняються багато режимністю, тобто відносяться до об'єктів зі змінною структурою [4, 5]. Крім того, для абонентів систем зв'язку найважливішим є не кое-

фіцієнт готовності обладнання, а ймовірність зв'язності, що не завжди враховується [6-10]. Таким чином, у відомих наукових публікаціях не розглянуто вплив властивості зміни структури системи зв'язку під час її використання за призначенням, що не дозволяє об'єктивно оцінювати реальну структурну надійність й потребує використання додаткових каналів зв'язку для досягнення необхідних значень показників надійності.

Мета статті – використання властивості зміни структури системи зв'язку під час її застосування за призначенням для отримання об'єктивної оцінки структурної надійності.

Виклад основного матеріалу

Система зв'язку – це сукупність взаємопов'язаних і узгоджених за завданням, місцем і часом дії вузлів та ліній зв'язку різноманітного призначення, що розгортаються та використовуються з метою розв'язання завдань обміну інформацією між абонентами [10-12]. Вона характеризується показниками, що приведені на рис. 1.

Згідно з [11], мобільність системи військового зв'язку – це її здатність у встановлений термін розгортатися, згортатися, переміщуватися та змінювати свою структуру відповідно до обстановки. Тобто вона відноситься до об'єктів зі змінною структурою, що не враховано у відомих методиках оцінки структурної надійності.

Інформаційний напрямок зв'язку – це сукупність вузлів та ліній зв'язку, яка забезпечує зв'язок між двома пунктами управління. Види зв'язку розрізняють по засобам зв'язку: каналотворювальні, комутаційні, кінцеві, автоматизації, спеціальні. До каналотворювальних засобів відносяться провідний зв'язок, оптико-електронний зв'язок, радіозв'язок, тропосферний радіозв'язок, космічний радіозв'язок, радіорелейний радіозв'язок та інші [11, 12].

Надійність системи зв'язку – це здатність забезпечувати зв'язок, зберігаючи в часі значення експлуатаційних показників у межах, передбачених нормативною і технічною документацією [3, 9, 10].



Рис. 1. Показники якості системи зв'язку

Структурна надійність системи зв'язку – об'єктивна властивість забезпечувати обмін інформацією користувачами мережі, з якістю не нижчою від заданої [3, 6, 10]. Її основні показники – коефіцієнт готовності та ймовірність зв'язності (рис. 1). Коефіцієнт готовності (A) – ймовірність того, що об'єкт виявиться працездатним у довільний момент часу, крім запланованих періодів, протягом яких його використання за призначенням не передбачене [13]:

$$A_j = \frac{T}{(T + T_b)}$$

де T – середній наробіток на відмову, T_b – середня тривалість відновлення.

Коефіцієнт готовності являє собою ймовірність технічної справності всіх елементів системи зв'язку і готовності всіх напрямків зв'язку до функціонування в повному обсязі. Проте для систем зв'язку більш об'єктивною характеристикою є ймовірність зв'язності абонентів P_3 , тобто ймовірність технічної готовності до обміну інформацією принаймні по одному каналу зв'язку [6-8, 10]. Вочевидь, що завжди $A \leq P_3$, причому рівність досягається коли кожний елемент системи використовується тільки в одному напрямі й тільки один раз. У системах із сильним взаємозв'язком елементів, що відповідає реальним системам зв'язку, задача зводиться до визначення ймовірності справного стану хоча б одного шляху з множини можливих (m) [5-10]:

$$P_3 = 1 - \prod_{j=1}^m (1 - A_j),$$

де A_j – коефіцієнт готовності шляху j між абонентами.

Під час вибору каналів зв'язку необхідно враховувати не тільки надійність засобів зв'язку, а й ерлангові втрати p , зумовлені зайнятістю каналів

за час t , що відводиться на передавання повідомлення [5-10]:

$$W = 1 - \prod_{j=1}^m (1 - A_j E_j),$$

$$E_j = (1 - p) \exp\left(-\frac{t}{T_j}\right),$$

де W – комплексний показник надійності напрямку зв'язку, T_j – наробіток на відмову засобів зв'язку шляху j , E_j – коефіцієнт врахування надійності засобів зв'язку та ерлангових втрат.

Завжди $W < P_3$, тому що кількісно оцінює потенційну структурну надійність мережі зв'язку без врахування її навантаження.

Значення кількісних показників надійності засобів зв'язку отримують з їхнього технічного опису, керівних документів з надійності, але найбільш об'єктивну оцінку дає аналіз статистичних даних щодо наробітку на відмову T_i , середнього часу відновлення T_{vi} і коефіцієнту готовності A_i засобу зв'язку виду $i = \overline{1, L}$, де L – кількість видів засобів зв'язку у системі.

В системі зв'язку її елементи працюють різний час, тому для аналізу структурної надійності окремих напрямків зв'язку доцільно використовувати результати, отримані для об'єктів зі змінною структурою [4, 5].

Розглянемо результати оцінки структурної надійності при традиційному підході [1-3, 6-10] в порівнянні з врахуванням часу роботи окремих елементів в системі зв'язку, як об'єктів зі змінною структурою [4, 5] на конкретному прикладі фрагмента системи зв'язку приведеному на рис. 2.

Показники надійності елементів за результатами експлуатації приведені в табл. 1.

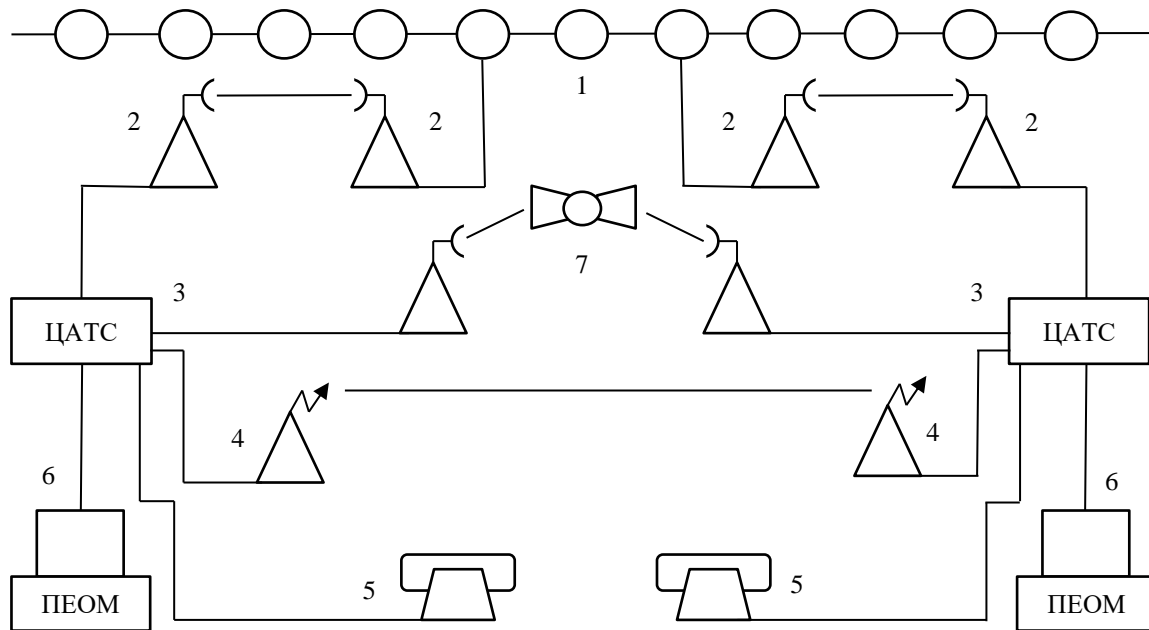


Рис. 2. Фрагмент системи зв'язку

Таблиця 1 – Показники надійності елементів напрямку зв'язку

<i>i</i>	Засоби зв'язку	T_i , год	T_{vi} , год	A_i
1.	Мережа урядового зв'язку (оренда каналів)	8000	0,40	0,99995
2.	Радіорелейні станції	5000	0,33	0,999934
3.	Цифрова АТС	2300	3,00	0,99869
4.	Радіостанція	2000	0,33	0,99983
5.	Телефонний апарат	800	1,00	0,99875
6.	Переносна ЕОМ	1500	0,50	0,99967
7.	Комплекс космічного зв'язку (оренда каналів)	12000	1,17	0,9999

Нижню межу значення структурної надійності напрямку зв'язку отримуємо за умови, що засоби зв'язку використовують цілодобово протягом року. Задача аналогічна орієнтовній оцінці надійності радіоелектронних засобів під час їхнього проектування, коли відома кількість елементів усіх типів у виробі та їхній параметр потоку відмов $Z_i = 1/T_i$. При використанні n_i адаптивних елементів в напрямку зв'язку, його параметр потоку відмов дорівнює

$$Z_H = \sum_{i=1}^L n_i Z_i.$$

В такому разі наробіток напрямку зв'язку на відмову дорівнює $T = 1/Z_H$, а прогнозована кількість відмов за рік експлуатації $T_p = 8760$ год елементів виду i

$$N_i = n_i T_p Z_i = n_i T_p / T_i.$$

Тоді прогнозована кількість відмов напрямку зв'язку за рік експлуатації

$$N = \sum_{i=1}^L N_i = T_p \sum_{i=1}^L n_i / T_i.$$

Наробіток на відмову напрямку зв'язку за рік експлуатації $T = T_p / N$, а середній час відновлення після відмови

$$T_B = \frac{1}{N} \sum_{i=1}^L N_i n_i T_{vi}.$$

Це найгірший випадок і якщо результати задовольняють, то далі оцінюють реальні значення показників надійності з врахуванням відносного часу роботи окремих елементів u_i протягом року:

$$T_{pi} = u_i T_p, \quad N_i = n_i T_{pi} / T_i, \quad N = \sum_{i=1}^L N_i;$$

$$T = \frac{T_p}{N}, \quad T_B = \frac{1}{N} \sum_{i=1}^L N_i T_{vi}, \quad A = T / (T + T_B).$$

Результати розрахунків для прикладу, що розглядається, приведено в табл. 2 і 3, відповідно. За даними табл. 2 отримуємо:

$$Z_H = 0,00671; \quad T = 149 \text{ год}; \quad N = 58,8; \quad T_B = 2 \text{ год}.$$

Аналіз отриманих результатів показує, що найменш надійним елементом є телефонний апарат, а найбільший час відновлення за рік безперервної експлуатації потребує цифрова АТС. Нижня межа коефіцієнту готовності напрямку зв'язку дорівнює $A_H = T / (T + T_B) = 0,987$; $U_H = 0,013$, що є неприйнятним.

Результати розрахунків значень показників надійності напрямку зв'язку з врахуванням відносного часу роботи окремих елементів приведено в табл. 3, при цьому $N = 24,8$; $T = 353,2$ год; $T_B = 1,28$ год.

Таблиця 2 – Розрахунки нижньої межі значень показників надійності напрямку зв'язку при $T_{pi} = T_p = 8760$ год

i	n_i	Z_i	$n_i Z_i$	N_i	$N_i n_i T_{vi}$
1	1	0,000125	0,000125	1,095	0,438
2	4	0,0002	0,0008	7,008	9,250
3	2	0,000435	0,00087	7,621	45,726
4	2	0,0005	0,0010	8,760	5,782
5	2	0,00125	0,0025	21,900	43,800
6	2	0,000667	0,00133	11,651	11,651
7	1	0,000083	0,000083	0,7271	0,851

Таблиця 3 – Розрахунок значень показників надійності напрямку зв'язку із врахуванням відносного часу роботи елементів

i	u_i	T_{pi} , год	N_i	$N_i T_{vi}$
1	0,95	8322	1,04	0,416
2	0,95	8322	3,33	1,099
3	1,00	8760	7,62	22,86
4	0,01	87,6	0,09	0,030
5	0,10	876	2,19	2,190
6	0,90	7884	10,51	5,100
7	0,04	350	0,03	0,035

Коефіцієнт готовності напрямку зв'язку з урахуванням відносного часу роботи елементів

$$A_n = 0,996; U_n = 0,004.$$

Значення T_i , T_{vi} й A_i не залежать від відносного часу роботи елементів, тому для оцінки значень показників надійності напрямку зв'язку відповідно рекомендацій [1-3, 6-10] перетворюємо схему зв'язку (рис. 2) у модель з надійності (рис. 3).

При паралельно-последовному з'єднанні елементів отримуємо

$$A = A_{56}^2 A_3^2 (1 - (1 - A_2^4 A_1)(1 - A_7)(1 - A_4^2));$$

$$A_{56}^2 = (1 - (1 - A_5)(1 - A_6))^2.$$

Після підстановки A_i з табл. 1 отримуємо $A = 0,9974; U = 0,0026$.

При цьому значення коефіцієнту готовності напрямку зв'язку відрізняється від отриманого з урахуванням відносного часу роботи елементів всього на 0,14%.

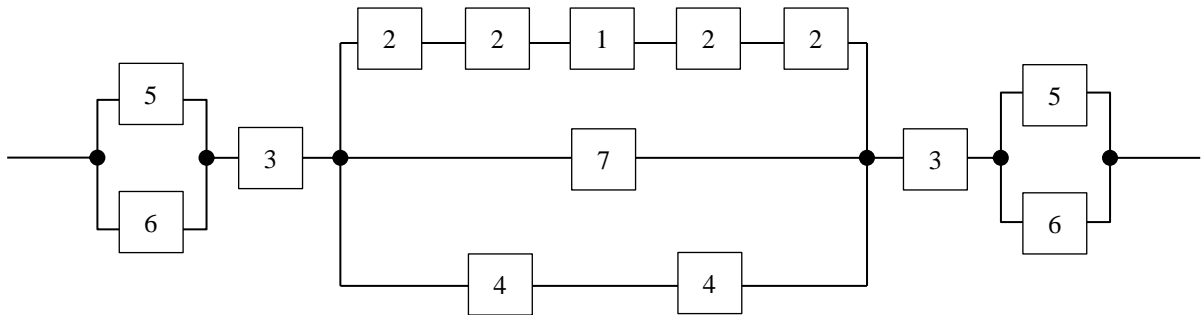


Рис. 3. Модель з надійності напрямку зв'язку

Якщо схема з'єднання елементів відрізняється від паралельно-последовної, задача розрахунку суттєво ускладнюється [1-3]. Наприклад, для схеми, приведеної на рис.4., отримуємо:

$$A_c = A_3 \times ((1 - (1 - A_1)(1 - A_2))(1 - (1 - A_4))(1 - A_5)) + (1 - A_3)(1 - A_2 A_5)(1 - A_1 A_4).$$

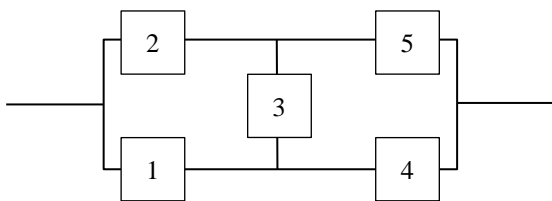


Рис. 4. Схема з'єднання елементів, що відрізняється від паралельно-последовної

Розрахунок спрощується, якщо $A_i = A$:

$$A_c = A(1 - (1 - A)^2)^2 + (1 - A)(1 - (1 - A^2)^2).$$

Порівняльний аналіз варіантів розрахунку показників надійності інформаційного напрямку зв'язку показує доцільність врахування відносного часу роботи окремих елементів.

Це забезпечує достатню точність результатів розрахунку при спрощенні їх реалізації, особливо при наявності складної схеми з'єднання елементів яка відрізняється від паралельно-последовного з'єднання.

Алгоритм реалізації запропонованого підходу щодо кількісної оцінки показників надійності інформаційного напрямку зв'язку формалізовано на рис. 5, де A_n і P_{zn} – припустимі значення коефіцієнту готовності та ймовірності зв'язності абонентів, відповідно.

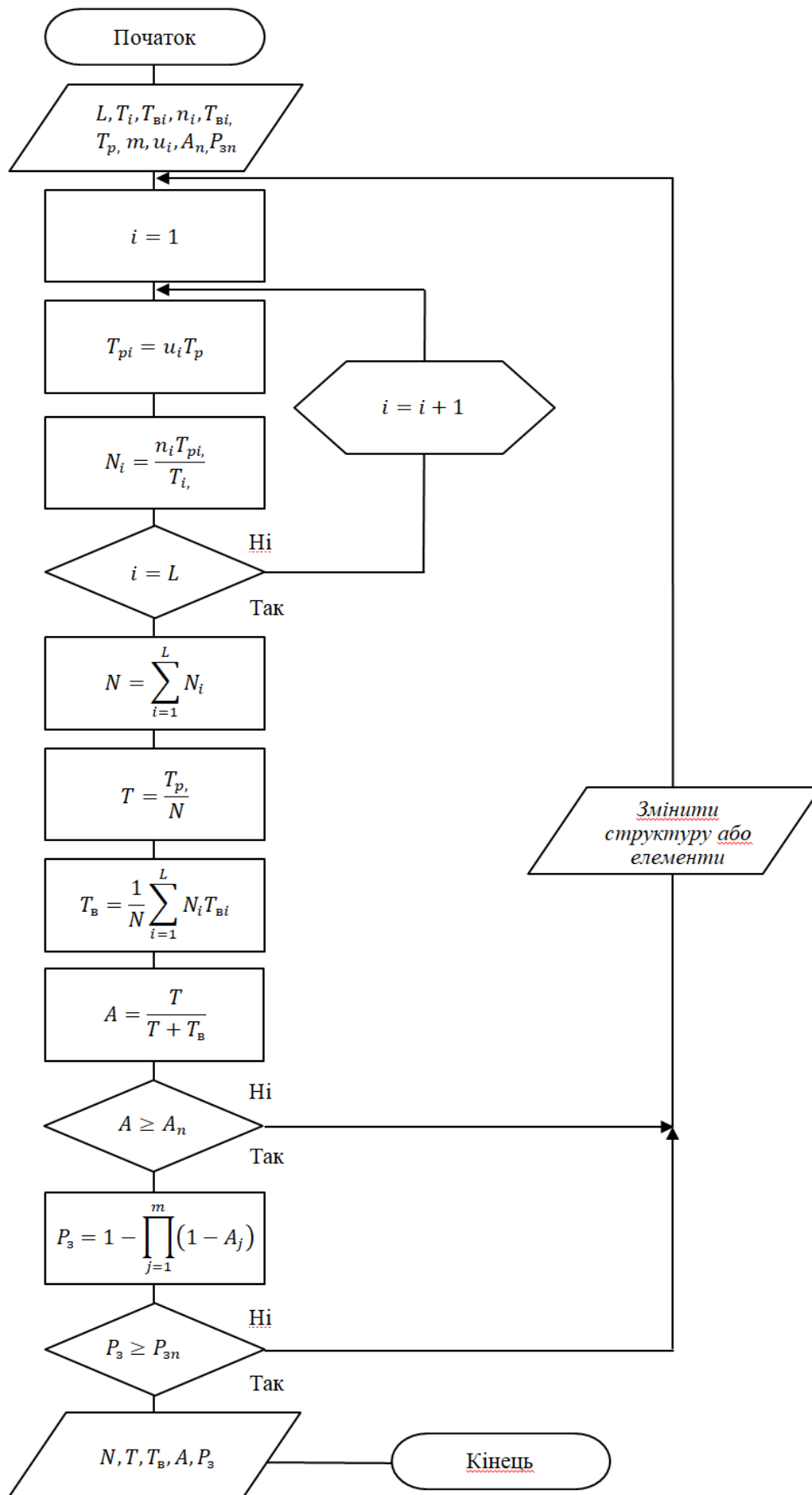


Рис. 5. Блок-схема алгоритму оцінки показників напрямку зв'язку

Висновки

В результаті порівняльного аналізу можливих варіантів кількісної оцінки показників надійності інформаційних напрямків зв'язку показано, що врахування відносного часу роботи окремих елементів дозволяє з необхідною точністю оцінити часткові (наробіток на відмову, середній час відновлення) і комплексні (коефіцієнт готовності і неготовності, ймовірність зв'язності абонентів) значення показників надійності.

На відміну від існуючих методик оцінки значень показників, запропонований підхід щодо оцін-

ки значень показників надійності відрізняється спрощенням розрахунків при необхідній точності результатів, використанням вперше властивості змін структури з'єднання елементів напрямку зв'язку, тобто врахуванням відносного часу роботи окремих елементів.

Реалізація запропонованих пропозицій щодо оцінки значень показників надійності інформаційних напрямків зв'язку існуючих та перспективних систем спеціального зв'язку дозволяє забезпечити необхідне значення часткових і комплексних показників з мінімальною необхідною кількістю каналів і засобів зв'язку.

СПИСОК ЛІТЕРАТУРИ

1. Зайцева Е.Н. Исследование надежности информационных систем // Электросвязь. – 2003. – № 6. – С. 37-39.
2. Романов О.І. Оцінка впливу та структури напрямків зв'язку на живучість телекомунікаційних мереж / О.І. Романов, В.Д. Голь, О.В. Волков // Збірник наукових праць ВІПІ НТУУ “КПІ”. – 2005. – Вип. № 1, – С. 142-150.
3. Харьбин А.В. О подходе к решению задачи выбора методологии оценки структурной надежности и живучести информационных систем критического применения / А.В. Харьбин, О.Н. Одарущенко // Радиоэлектронні і комп'ютерні системи. – 2006. – № 6. – С. 61-70.
4. Yevhen Ryzhov, Lev Sakovych, Oleksandr Puchkov, Yana Nebesna. Evaluation of reliability of radioelectronic devices with variable structure. Radio Electronics, Computer Science, Control. No. 3(54), 2020 pp. 31–39. DOI: <https://doi.org/10.15588/1607-3274-2020-3-3>.
5. Sergii Gnatiuk. Method of Estimating the Values of Reliability Indicators of Objects with Variable Structure / S. Gnatiuk, L. Sakovich, Y. Kuryata, R. Odarchenko, V. Gnatiuk // CPITS-2022: Cybersecurity Providing in Information and Telecommunication Systems, October 13, 2022, Kyiv, Ukraine, P. 33-43.
6. Рижак В.А., Сакович Л.М. Кількісне оцінювання структурної надійності систем зв'язку // Зв'язок, 2004. - № 4. - С. 53-57.
7. Гнатюк С. С. Аналіз показників надійності систем спеціального зв'язку / С.С. Гнатюк // Восточно-Европейский журнал передовых технологий. – 2013. – № 6(12). – С. 45-50.
8. Гнатюк С.С. Кількісне оцінювання значень показників надійності систем зв'язку / Л.М.Сакович, С.С.Гнатюк // Зв'язок. – 2013. – № 6. – С. 10-15.
9. Поповский В. В. Методы априорной оценки сетевой надёжности / В. В. Поповский, В. С. Волотка // Радиотехника. – 2014. – Вып. 178. – С. 20-23.
10. Гнатюк С.С. Методика оцінки показників надійності систем спеціального зв'язку / С.С.Гнатюк, Л.М.Сакович // Озброєння та військова техніка. – №1 (5). – 2015, С. 26-28.
11. Зв'язок військовий. Терміни та визначення: ДСТУ В. 3265–95. – К.: Держстандарт України, 1995. – 78 с.
12. Основи теорії систем зв'язку: Математичні моделі телекомунікаційних систем / В. Ф. Олійник. – К. : Техніка, 2000. – 152 с. – ISBN 966-575-028-3.
13. Надійність техніки. Терміни та визначення: ДСТУ 2860-94. – Київ: Держстандарт України, 1995. – 89 с.

Received (Надійшла) 17.03.2023

Accepted for publication (Прийнята до друку) 24.05.2023

Evaluation of structural reliability of communication systems

L. Sakovich, G. Krykhovetskyi, Ya. Kuryata

Abstract. In order to increase the accuracy of the quantitative assessment of partial and complex indicators of the reliability of communication directions, the article proposes to take into account the operating time of individual means of communication for the first time. The essence of the proposals consists in the use of a mathematical model for estimating the values of the reliability indicators of objects with a variable structure. That is, to organize communication between subscribers, it is possible to use various communication channels from a set of possible ones that are loaded at different times. The solution to this problem is formalized and an example of the use of research results is given. The effect of the implementation of the proposals regarding the quantitative assessment of the values of the reliability indicators of the communication directions is to ensure the required level of reliability with the minimum number of channels and means of communication.

Keywords: structural reliability, directions of communication, assessment of partial and complex indicators of reliability.

АЛФАВІТНИЙ ПОКАЖЧИК

Бараненко Р.	39	Капітон А. М.	59	Покора К. В.	106
Барсуков А. І.	80	Караєва Н. В.	54	Пустовойтов П.	39
Бевзюк Р.	68	Каргін А. О.	113	Резнік Д. В.	190
Бурдейна Н. Б.	186	Коваленко А. А.	106	Рибальченко А. О.	135
Бурцев В. С.	158	Коваленко А. С.	170	Романенко І. О.	63
Вишнівський Д.	75	Коваленко О. В.	170	Росінський Д. М.	142
Вінтенко Б. Ю.	170	Козлітін О. О.	179	Сакович Л. М.	210
Волк М. О.	80	Коровко В. А.	35	Самадова С.	146
Гавриленко С. Ю.	87	Кочетов Г. М.	183	Самченко Д. М.	183
Гаптельманов А. В.	92	Криховецький Г. Я.	210	Свірцевський К. О.	142
Глива В. А.	179	Куліш Р. В.	20	Сергєєв Д. А.	113
Голець В. В.	80	Курята Я. Е.	210	Синицький Ю. В.	199
Головко Г. В.	98	Кучук Г. А.	131	Сілін Є. Л.	113
Гора М. В.	80	Кучук Н. Г.	119	Скакаліна О. В.	59
Горбачов П. Ф.	15	Лабазов В. Г.	80	Смірнов О. А.	170
Демиденко М. І.	101	Ластівка О. В.	183	Смірнов С. А.	170
Дерюгін О. В.	190	Левченко Л. О.	54	Сокол Г.	39
Дєдов М. А.	204	Лещинський В. О.	165	Соломоненко Ю. С.	63
Джеря Т. Е.	42	Любий Є. В.	15	Сукач С. В.	190
Дяченко В. О.	106	Ляшенко О.	68, 75	Суліма С. В.	151
Євсеєнко О. М.	45	Маловічко Є. В.	125	Тихенко О. М.	179
Єрмолаєв О. Д.	151	Марюшко М. В.	5	Ткаченко Т. М.	186
Єр'оміна Н.	75	Матвєєв М. І.	131	Фомічов О. О.	158
Жатканбаєва Е. А.	35	Матющенко О. Г.	20	Харченко В. С.	125
Жуков С. В.	113	Минко П. Є.	63	Худов Г. В.	63
Захаров Р.	39	Міхаль О. П.	92	Цопа В. А.	190
Здоренко Ю. М.	119	Міщенко А. В.	80	Чалий С. Ф.	165
Зозуля В. Д.	87	Мордюк В. І.	199	Чеберячко С. І.	190
Зозуля С. В.	179	Мураховська О. А.	26	Ченчева О. О.	186, 190
Зуєв А. О.	45	Нагієва П. А.	31	Шевчук В. В.	42
Іванюк В. О.	63	Ніколаєнко Б. А.	204	Шенцюань Ван	119
Ієвлєв Д.	98	Омельченко В. В.	87	Шурпик О. В.	49
Ільяшов О. А.	106	Очеретенко С. В.	35	Щепка О. О.	92
Ічанська Н. В.	49	Пащенко Р. Е.	5	Яворська О. О.	190
Кальницький М. А.	101	Пелешок Є. В.	204		

Наукове видання

СИСТЕМИ УПРАВЛІННЯ, НАВІГАЦІЇ ТА ЗВ'ЯЗКУ

Збірник наукових праць

Випуск 2 (72)

Відповідальна за випуск *К. С. Нестеренко*

Технічний редактор *Т. В. Уварова*

Коректор *О. В. Морозова*

Комп'ютерна верстка *Н. Г. Кучук, І. Ю. Петровська*

Оформлення обкладинки *І. В. Льїна*

Свідоцтво про державну реєстрацію КВ № 19512-93/2ПР від 16.11.2012 р.

Підписано до друку 16.06.2023. Формат 60×84/8. Ум.-друк. арк. 27,0. Тираж 120 прим. Зам. 616-23

Адреса редакції: Україна, 36011, м. Полтава, Першотравневий проспект, 24, тел. (050) 302-20-71

Національний університет «Полтавська політехніка імені Юрія Кондратюка»

Віддруковано з готових оригінал-макетів у цифровій друкарні Impress

61002, м. Харків, вул. Пушкінська, 56, тел. + 38 (057) 714-52-11

e-mail: irina@impress.biz.ua