

А. С. Бурда, М. А. Прудіус, Я. Г. Стефанюк, О. О. Фомічов

Харківський національний технічний університет радіоелектроніки, Харків, Україна

МЕТОДИ ОБРОБКИ ТА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ З ВИКОРИСТАННЯМ ШТУЧНИХ ІМУННИХ СИСТЕМ

Анотація. Актуальність. Штучні імунні системи, завдяки своїй здатності працювати в розподіленому середовищі, можуть бути ефективно використані для виявлення та запобігання вторгненням у мережі. Штучні імунні системи здатні адаптуватися до змінюваних умов та нових загроз. Це особливо актуально в умовах швидких технологічних змін, де традиційні методи часто вимагають тривалого перенавчання. Вони активно використовуються для виявлення та запобігання кіберзагрозам. Вони можуть аналізувати мережевий трафік, виявляти аномальні дії та забезпечувати захист у режимі реального часу. Таким чином, актуальність використання штучних імунних систем для обробки та класифікації даних полягає у їх унікальних властивостях адаптивності, здатності до розпізнавання складних патернів та аномалій, а також ефективній роботі в розподілених системах. **Метою даної роботи** є дослідження існуючих методів обробки та штучного аналізу даних з використанням штучних імунних систем. **Об'єктом дослідження** є інтелектуальний аналіз даних штучними імунними системами. **Предметом дослідження** є методи інтелектуального аналізу даних штучними імунними системами. **Результати.** Проведено аналіз існуючих методів обробки та інтелектуального аналізу даних з використанням штучних імунних систем. Імунні алгоритми кластеризації є потужним інструментом для аналізу та обробки даних. Вони дозволяють ефективно групувати дані, виявляти аномалії та адаптуватися до змін у середовищі даних, що робить їх цінними для широкого спектра застосувань, від маркетингу та медицини до фінансів і промисловості. Артилерійні алгоритми є потужними інструментами для інтелектуального аналізу даних, які пропонують ефективні, точні та продуктивні методи оптимізації та аналізу. Вони знаходять застосування у багатьох галузях, від фінансів і логістики до медицини та кібербезпеки, забезпечуючи вирішення складних задач обробки даних і прийняття рішень. Дендритні клітинні алгоритми є потужним інструментом для інтелектуального аналізу даних, зокрема для виявлення аномалій і класифікації даних. Їх здатність агрегувати різні типи сигналів і приймати рішення на основі загального рівня небезпеки робить їх особливо ефективними в складних і динамічних середовищах, таких як кібербезпека та біоінформатика. Вони забезпечують високу чутливість і адаптивність, що дозволяє їх успішно застосовувати у різних сферах для вирішення задач обробки та аналізу даних.

Ключові слова: штучна імунна система, алгоритми негативного відбору, імунні алгоритми кластеризації, артилерійні алгоритми, дендритні клітинні алгоритми, кібербезпека, афінність, клональний відбір, патерн.

Вступ

Штучні імунні системи (ШИС) – це обчислювальні моделі, натхнені біологічною імунною системою, які використовуються для вирішення різноманітних задач, включаючи обробку та класифікацію даних. Основними компонентами ШИС є імунні алгоритми [1], які імітують процеси розпізнавання і адаптації, що відбуваються в імунній системі живих організмів. Методами ШИС є імунні алгоритми негативного відбору (ІАНВ), імунні алгоритми кластеризації (ІАК), артилерійні алгоритми (АА), дендритні клітинні алгоритми (ДКА) [2].

ІАНВ імітують процес видалення незрілих Т-лімфоцитів, які розпізнають власні клітини організму (self-cells). У контексті обробки даних використовуються для виявлення аномалій та невідомих даних. ІАНВ є однією з ключових технік штучних імунних систем, що моделюють механізм негативного відбору в біологічній імунній системі. Вони призначені для виявлення аномалій або невідомих патернів у даних шляхом навчання системи розрізняти "свої" (нормальні) дані від "чужих" (аномальних) даних.

ІАК базуються на процесі клонального відбору та афінної зрілості В-лімфоцитів. Ці алгоритми створюють та зберігають популяцію антитіл, що розпізнають специфічні антигени (патогени). ІАК також є частиною штучних імунних систем, які моделюють процес клонального відбору та афінної зрілості в біологічній імунній системі. Ці алгоритми викорис-

товуються для кластеризації даних, оптимізації та інших задач, що вимагають розпізнавання патернів та адаптації.

АА моделюють взаємодію між антитілами у мережі [3]. АІН використовуються для кластеризації даних, де мережа антитіл саморегулюється та адаптується до вхідних даних. АА, або штучні імунні мережі (Artificial Immune Networks, АІН), моделюють взаємодію між антитілами в біологічній імунній системі, створюючи мережу антитіл, яка здатна до самоорганізації та адаптації. Ці алгоритми використовуються для кластеризації даних, виявлення схожих структур і патернів у даних, а також для оптимізації та прогнозування [4, 5].

ДКА моделюють роботу дендритних клітин [6], які відповідають за презентацію антигенів Т-лімфоцитам. ДКА використовуються для виявлення аномалій та класифікації даних. ДКА [7] – це інноваційний підхід у сфері штучних імунних систем, який моделює роботу дендритних клітин у біологічній імунній системі. Ці алгоритми розроблені для виявлення аномалій і класифікації даних, зокрема у сфері кібербезпеки та біоінформатики [8–10].

Отже, **метою цієї роботи** є дослідження існуючих методів обробки та штучного аналізу даних з використанням штучних імунних систем.

Основна частина

Штучні імунні системи – це інноваційна методологія, що базується на принципах біологічної

імунної системи для вирішення складних задач обробки даних. Вони використовуються для класифікації, кластеризації, виявлення аномалій, оптимізації та прогнозування. Основні типи алгоритмів ШІС включають негативний відбір, клональний відбір, штучні імунні мережі та дендритні клітинні алгоритми [11].

ІАНВ є важливою складовою ШІС і моделюють процеси, що відбуваються у біологічній імунній системі. Основна мета цих алгоритмів – виявлення аномалій шляхом навчання на нормальних даних і подальшої перевірки нових даних на предмет їх відповідності.

У біологічній імунній системі негативний відбір відбувається в тимусі, де Т-клітини, які реагують на власні антигени організму, видаляються. Аналогічно, в NSA створюється набір детекторів, які навчаються на нормальних даних і не реагують на них. Ці детектори потім використовуються для виявлення аномалій у нових даних.

На першому етапі створюється набір детекторів, які представляють випадкові можливі стани системи. Ці детектори проходять процес навчання, під час якого вони перевіряються на відповідність "своїм" даним (нормальним зразкам). Детектори, які реагують на нормальні дані, відсіюються (негативний відбір). Це означає, що лише ті детектори, які не реагують на нормальні дані, залишаються в системі.

У біологічній імунній системі негативний відбір відбувається в тимусі, де Т-клітини, які реагують на власні антигени організму, видаляються. Аналогічно, в ІАНВ створюється набір детекторів, які навчаються на нормальних даних і не реагують на них. Ці детектори потім використовуються для виявлення аномалій у нових даних. Після процесу навчання детектори використовуються для моніторингу нових даних. Якщо детектор реагує на нові дані, ці дані вважаються аномальними або підозрілими. Процес моніторингу може відбуватися в режимі реального часу, що дозволяє швидко виявляти аномалії.

Алгоритм ІАНВ передбачає ініціалізацію (створення початкового набору випадкових детекторів), негативний відбір (перевірка кожного детектора на відповідність нормальним даним та відсіювання детекторів, які розпізнають нормальні дані), а також моніторинг (застосування детекторів до нових даних та ідентифікація аномальних даних на основі реакції детекторів). Таким чином, ІАНВ є порівняно простим у реалізації, що робить їх доступними для широкого кола застосувань. Також вони можуть бути адаптовані для роботи з великими обсягами даних і в умовах розподілених систем. Алгоритм може навчатися новим патернам і адаптуватися до змін у даних.

Можливе застосування ІАНВ в кібербезпеці (виявлення вторгнень і аномалій у мережевому трафіку, моніторинг активності користувачів для виявлення підозрілих дій), моніторингу систем (виявлення несправностей у промисловому обладнанні, аналіз логів систем для виявлення аномалій у роботі програмного забезпечення) та медичній діагностиці (виявлення, наприклад, аномалій у кардіограмах або ракових клітин у зображеннях). Імунні алгоритми кластеризації використовують принципи біологічної імун-

ної системи для групування подібних даних. Ці алгоритми особливо корисні для обробки великих обсягів даних, виявлення патернів та аномалій, і є потужним інструментом для аналізу даних у різних сферах. ШІС для кластеризації використовують механізми самоорганізації та адаптації для створення мережі антитіл (або агентів), що представляють кластери даних. Основні принципи роботи включають ініціалізацію мережі антитіл, клонування та мутацію, взаємодію між антитілами, і самоорганізацію мережі.

Алгоритм ІАК містить ініціалізацію мережі антитіл, коли ініціалізується початкова популяція антитіл (або агентів), кожна з яких представляє потенційний кластер, обчислення афінності (афінність антитіла до антигену (даних) визначається на основі відстані або подібності між ними), клонування та мутації (антитіла з високою афінністю клонуються і мутують для створення нових варіантів, що можуть краще відповідати даним), взаємодію та самоорганізацію мережі (антитіла взаємодіють між собою та з даними, утворюючи мережу, яка самоорганізується та адаптується до нових даних), а також оновлення мережі (мережа антитіл постійно оновлюється на основі нових даних та взаємодій у мережі).

Під час дії цього алгоритму відбувається генерація початкової популяції антитіл, кожне з яких представляє можливе рішення або центр кластера. Потім визначається ступінь відповідності (афінності) між антитілами та антигенами (даними). Висока афінність вказує на добру відповідність даним. Антитіла з високою афінністю проходять процес клонування, створюючи копії з невеликими мутаціями. Це дозволяє досліджувати простір можливих рішень навколо антитіла з високою афінністю. Також проходить відбір кращих клонів на основі їх афінності до антигенів. Цей процес повторюється до досягнення оптимального розподілу клонів, що представляють центри кластерів. Дані кластеризуються на основі близькості до обраних центрів кластерів (антитіл).

Перевагами використання ІАК є адаптивність, коли алгоритми можуть адаптуватися до нових даних та змін у даних; гнучкість, при якій ІАК можуть використовуватися для різних задач кластеризації та оптимізації та пошук глобальних оптимумів, що, в свою чергу, дозволяє ІАК ефективно знаходити глобальні оптимуми в просторі рішень. Також слід відзначити, що можливе застосування ІАК для кластеризації різних типів даних, включаючи текстові, числові та зображення; вирішення задач оптимізації яких як оптимізація розкладів, маршрутів та ресурсів; а також для виявлення схожих структур у великих обсягах даних, що корисно для аналізу трендів та патернів. ІАК є потужним інструментом для кластеризації та оптимізації даних завдяки своїй здатності до адаптації, гнучкості та ефективному пошуку глобальних оптимумів. Вони знаходять застосування у різних сферах, включаючи аналіз даних, медицину, промисловість та маркетинг, забезпечуючи ефективні рішення для задач кластеризації та розпізнавання патернів.

Артилерійні алгоритми – це специфічний підхід до інтелектуального аналізу даних, який може використовуватися для оптимізації та покращення продук-

тивності в завданнях обробки великих даних. Цей термін може відноситися до різних алгоритмічних стратегій, але в основному має на увазі точні, ефективні та добре структуровані методи, які застосовуються до розв'язання складних задач у галузі штучного інтелекту та машинного навчання. АА зазвичай включають методи оптимізації, такі як генетичні алгоритми, рій частинок, алгоритми мурашиних колоній, які дозволяють ефективно знаходити найкращі рішення серед великого простору можливих варіантів. Ці алгоритми мають високу продуктивність, що досягається через паралельні обчислення, ефективні структури даних і спеціалізовані методи обробки. Точність і надійність є ключовими характеристиками артилерійних алгоритмів, що дозволяє їм успішно застосовуватися в критичних додатках, таких як фінансове моделювання, кібербезпека та медичний аналіз.

Генетичні алгоритми (ГА) використовують принципи природного відбору та генетики для оптимізації і пошуку рішень. Процес роботи ГА наступний: ініціалізація популяції випадкових рішень, оцінка кожного рішення за допомогою функції придатності, відбір кращих рішень для створення нового покоління через кросовер і мутацію, повторення процесу до досягнення оптимального рішення. Застосовуються ГА для оптимізації торгових стратегій на фінансових ринках та розв'язання задач комівояжера.

Алгоритми рою частинок моделюють поведінку зграї птахів або риби для пошуку оптимальних рішень. Процес роботи включає ініціалізацію популяції частинок з випадковими положеннями та швидкостями, оцінку кожної частинки на основі її положення, оновлення швидкостей і положень частинок на основі їхньої історії і найкращих положень сусідів, повторення процесу до досягнення оптимального рішення. Застосовуються ці алгоритми для оптимізації параметрів машинного навчання, а також розв'язання задач оптимізації у виробництві.

Алгоритми мурашиних колоній використовують принципи поведінки мурах для знаходження найкоротших шляхів і оптимізації. Спочатку відбувається ініціалізація мурах у випадкових позиціях, потім моделюється рух мурах по графу на основі феромонових слідів і евристичної інформації, далі оновлення феромонових слідів на основі якості знайдених рішень. Застосовується даний тип алгоритмів для розв'язання задач маршрутизації, а також для оптимізації логістичних ланцюгів.

Артилерійні алгоритми базуються на ідеях створення та підтримки мережі антитіл, яка може змінювати свою структуру відповідно до вхідних даних. Основні процеси включають ініціалізацію мережі антитіл, тобто створення початкової популяції антитіл, кожне з яких представляє можливе рішення або кластер; афінну зрілість і вибір (антитіла взаємодіють з антигенами (даними) та з іншими антитілами в мережі, що сприяє їх афінній зрілості. Антитіла з високою афінністю до антигенів залишаються в мережі, а антитіла з низькою афінністю видаляються); клонування та мутацію (клонування антитіл з високою афінністю та їх мутація для створення нових антитіл, які можуть краще розпізнавати антигени); само-

організацію мережі. В даному випадку антитіла утворюють мережу, яка самоорганізується на основі взаємодій між антитілами та антигенами. Ця мережа адаптується до змін у вхідних даних.

Перевагами використання АА є: самоорганізація, масштабованість та гнучкість. АА є потужним інструментом для кластеризації, виявлення аномалій, оптимізації та прогнозування завдяки своїм властивостям самоорганізації та адаптивності. Вони знаходять застосування в різних сферах, включаючи аналіз даних, кібербезпеку, медицину та промисловість, забезпечуючи ефективні рішення для складних задач у сучасному інформаційному середовищі.

Дендритні клітини у біологічній імунній системі відіграють ключову роль у виявленні патогенів та ініціації імунної відповіді. Вони збирають інформацію з навколишнього середовища і, на основі отриманих сигналів, активуються або залишаються неактивними. ДКА моделюють цей процес, використовуючи три типи сигналів: сигнали, що свідчать про наявність патогенів (патоген-асоційовані молекулярні патерни), сигнали, що свідчать про нормальні, "безпечні" умови (безпечно-асоційовані молекулярні патерни, сигнали), що свідчать про стрес або пошкодження клітин (небезпечно-асоційовані молекулярні патерни). Алгоритм ДКА включає в себе ініціалізацію, коли створюється популяція дендритних клітин; збір сигналів, при якому кожна клітина збирає сигнали з навколишнього середовища даних; агрегацію сигналів, при якій дендритні клітини агрегують отримані сигнали, визначаючи рівень "небезпеки"; розподіл сигналів, при якому клітини розподіляють зібрані сигнали і приймають рішення про свою активацію на основі загального рівня безпеки; активацію, якщо рівень безпеки перевищує певний поріг, клітини активуються і сигналізують про наявність аномалій, та якщо рівень безпеки низький, клітини залишаються неактивними, що свідчить про нормальні умови; класифікацію активованих клітин, які використовуються для класифікації даних як аномальні або нормальні.

Перевагами ДКА є здатність до виявлення навіть незначних аномалій завдяки багаторівневій агрегації сигналів, можливість адаптуватися до змінюваних умов у середовищі даних, ефективна робота з великими обсягами даних та у розподілених системах. Застосовують ДКА в кібербезпеці, моніторингу систем та біоінформатиці.

Висновки

Проведено аналіз існуючих методів обробки та інтелектуального аналізу даних з використанням штучних імунних систем. Імунні алгоритми негативного відбору є потужним інструментом для виявлення аномалій та класифікації даних. Завдяки своїй здатності адаптуватися до змін та працювати у розподілених середовищах, вони залишаються актуальними в сучасних умовах, забезпечуючи ефективні рішення для моніторингу та аналізу даних у різних сферах, від кібербезпеки до медичної діагностики. Імунні алгоритми кластеризації є потужним інструментом для аналізу та обробки даних. Вони дозволяють ефективно групувати дані, виявляти аномалії та адап-

туватися до змін у середовищі даних, що робить їх цінними для широкого спектра застосувань, від маркетингу та медицини до фінансів і промисловості.

Артилерійні алгоритми є потужними інструментами для інтелектуального аналізу даних, які пропонують ефективні, точні та продуктивні методи оптимізації та аналізу. Вони знаходять застосування у багатьох галузях, від фінансів і логістики до медицини та кібербезпеки, забезпечуючи вирішення складних задач обробки даних і прийняття рішень.

Дендритні клітинні алгоритми є потужним інструментом для інтелектуального аналізу даних, зокрема для виявлення аномалій і класифікації даних. Їх здатність агрегувати різні типи сигналів і приймати рішення на основі загального рівня небезпеки робить їх особливо ефективними в складних і динамічних середовищах, таких як кібербезпека та біоінформатика. ДКА забезпечують високу чутливість і адаптивність, що дозволяє їх успішно застосовувати у різних сферах для вирішення задач обробки та аналізу даних.

СПИСОК ЛІТЕРАТУРИ

1. Lauren M. Sompayrac How the Immune System Works. 4th Edition // Wiley-blackwell. 2012. – 153 p.
2. Kephart, J. O. A biologically inspired immune system for computers". Proceedings of Artificial Life IV: The Fourth International Workshop on the Synthesis and Simulation of Living Systems. MIT Press. 1994. – pp. 130–139.
3. Andrews and Timmis "A Computational Model of Degeneracy in a Lymph Node". Artificial Immune Systems. Lecture Notes in Computer Science. Vol. 4163. 2006. – pp. 164–177. doi:10.1007/11823940_13.
4. Kuchuk, H., Kovalenko, A., Ibrahim, B.F. and Ruban, I. (2019), "Adaptive compression method for video information", International Journal of Advanced Trends in Computer Science and Engineering, vol. 8(1), pp. 66–69, doi: <http://dx.doi.org/10.30534/ijatcse/2019/1181.22019>
5. Худов В.Г. Аналіз відомих методів сегментування зображень, що отримані з бортових систем оптикоелектронного спостереження / В.Г. Худов, Г.А. Кучук, О.М. Маковейчук, А.В. Крижний // Системи обробки інформації, 2016. – Вип. 9 (146). – С. 77-80.
6. Mendao; et al. "The Immune System in Pieces: Computational Lessons from Degeneracy in the Immune System". 2007 IEEE Symposium on Foundations of Computational Intelligence. 2007. – pp. 394–400. doi:10.1109/FOCI.2007.371502.
7. Greensmith, J.; Aickelin, U. "Artificial Dendritic Cells: Multi-faceted Perspectives". Human-Centric Information Processing Through Granular Modelling (PDF). Studies in Computational Intelligence. Vol. 182.2009. – pp. 375–395. doi:10.1007/978-3-540-92916-1_16.
8. Kuchuk G.A., Mohammad A.S. and Kovalenko, A.A. (2011), "Method for reducing data transmission time in a wireless network", *Control, navigation and communication systems*, No. 3 (19), pp. 209-213.
9. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", *2021 International Conference on Information and Digital Technologies (IDT)*, Zilina, Slovakia, doi: <https://doi.org/10.1109/IDT52577.2021.9497534>
10. Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Yu., Yevstrat, D., Chyrva, Yu. & Kuchuk, H. (2022), "Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples", *Eastern-European Journal of Enterprise Technologies*, 2022, 6 (4(120)), pp. 40–49, doi: <https://doi.org/10.15587/1729-4061.2022.269128>
11. Yaloveha V., Podorozhniak A., Kuchuk H. Convolutional neural network hyperparameter optimization applied to land cover classification. *Radioelectronic and computer systems*. 2022. No. 1. С. 115–128. DOI: <https://doi.org/10.32620/reks.2022.1.09>

Received (Надійшла) 22.05.2024

Accepted for publication (Прийнята до друку) 31.07.2024

Methods of data processing and intellectual analysis using artificial immune systems

Anatoly Burda, Nikita Prudius, Yaroslav Stefanyuk, Oleksandr Fomichev

Abstract. Relevance. Artificial immune systems, especially algorithms of negative selection and dendritic cells, demonstrate high efficiency in detecting rare and unknown anomalies. Distributed computing systems and networks require effective monitoring and protection methods. Their ability to recognize specific patterns and classify medical images makes them indispensable in these fields. They are also actively used to detect and prevent cyber threats. They can analyze network traffic, detect anomalous activity, and provide real-time protection. Thus, the relevance of using artificial immune systems for data processing and classification lies in their unique properties of adaptability, ability to recognize complex patterns and anomalies, as well as effective work in distributed systems. **The purpose** is to research existing methods of processing and artificial analysis of data using artificial immune systems. **The object** of the study is the intellectual analysis of data by artificial immune systems. **The subject** are the methods of intellectual analysis of data by artificial immune systems. **Results** an analysis of existing methods of data processing and intellectual analysis using artificial immune systems was carried out. Immune clustering algorithms are a powerful tool for data analysis and processing. They allow you to efficiently group data, detect anomalies, and adapt to changes in the data environment, making them valuable for a wide range of applications, from marketing and medicine to finance and industry. Artillery algorithms are powerful tools for intelligent data analysis, offering efficient, accurate and productive methods of optimization and analysis. They find applications in many industries, from finance and logistics to medicine and cyber security, providing solutions to complex data processing and decision-making tasks. Dendritic cell algorithms are a powerful tool for intelligent data analysis, particularly for anomaly detection and data classification. Their ability to aggregate different types of signals and make decisions based on the overall level of danger makes them particularly effective in complex and dynamic environments such as cyber security and bioinformatics. they provide high sensitivity and adaptability, which allows them to be successfully used in various fields to solve data processing and analysis problems.

Keywords: artificial immune system, negative selection algorithms, immune clustering algorithms, artillery algorithms, dendritic cell algorithms, cyber security, affinity, clonal selection, pattern.