УДК 004. 8:004.9 **doi: 10.26906/SUNZ.2024.4.163**

С. Ф. Чалий, В. О. Лещинський

Харківський національний університет радіоелектроніки, Харків, Україна

МОДЕЛЮВАННЯ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ НА ОСНОВІ ІНТЕГРАЦІЇ ТЕМПОРАЛЬНИХ ТА КАУЗАЛЬНИХ ЗАЛЕЖНОСТЕЙ

Анотація. Предметом вивчення в статті є процеси побудови пояснень в інтелектуальних системах з використанням темпоральних та каузальних залежностей. **Метою** є розробка підходу до побудови пояснень на основі інтеграції темпоральних та каузальних залежностей щодо процесу прийняття рішення з тим, щоб забезпечити можливість формування пояснення як для зовнішніх, так і для внутрішніх користувачів інтелектуальних інформаційних систем. **Завдання:** визначення відмінностей доступу до інформації в ІІС для зовнішніх та внутрішніх користувачів; розробка ієрархічної моделі пояснення на базі темпоральних та каузальних залежностей; розробка методу побудови пояснення з використанням темпоральних та каузальних залежностей. Використовуваними **підходами** є: методи побудови пояснень, методи побудови каузальних залежностей. Отримані наступні **результати**. Виконано структуризацію відмінностей доступу до інформації користувачів інтелектуальних систем для обґрунтування необхідності багаторівневої деталізації пояснень. Запропоновано ієрархічну модель пояснення на основі темпоральних та каузальних залежностей. Запропоновано метод побудови пояснень з використанням темпоральних та каузальних залежностей між станами або діями процесу формування рішення в інтелектуальній системі. **Висновки**. Наукова новизна отриманих результатів полягає в наступному. Запропоновано ієрархічну модель пояснення, яка містить локальний, процесний та глобальний рівні пояснень згідно можливостей доступу користувачів до інформації щодо процесу формування рішення, що дає можливість враховувати неповноту інформації щодо стану інтелектуальної системи при поясненні її рішень для зовнішніх та внутрішніх користувачів. Розроблено метод побудови пояснень, який містить фази побудови пояснень з використанням темпоральних та каузальних залежностей між вхідними даними, станами процесу формування рішення та отриманим результатом та представлення пояснень, що дає можливість сформувати каузальні залежності на декількох рівнях деталізації та представити їх користувачу у вигляді упорядкованої за вагою множини альтернативних пояснень.

Ключові слова: інтелектуальна система, система штучного інтелекту, пояснення, процес прийняття рішення, темпоральні залежності, каузальні залежності, каузальність, можливість**.**

Вступ

Сучасні інтелектуальні інформаційні системи (ІІС) широко використовують алгоритми машинного навчання для вирішення складних практичних задач, пов'язаних із класифікацією, прогнозуванням, побудовою рекомендацій тощо. Однак, такі алгоритми зазвичай є непрозорими для користувачів, тобто приводять до відображення ІІС як «чорної скриньки» або «сірої скриньки» [1]. Непрозорість означає, що користувачі не мають доступу до внутрішньої логіки роботи системи і тому не мають можливості пояснити, як саме приймаються рішення.

Для забезпечення довіри користувачів до результатів роботи ІІС необхідно, щоб вони розуміли процес побудови рішення та причини отриманого результату [2]. Для того, щоб забезпечити зрозумілість для користувачів, процес прийняття рішень в ІІС має бути інтерпретованим. Для інтерпретації роботи непрозорих ІІС необхідно формувати пояснення, які розкривають логіку функціонування інтелектуальної системи [3].

Такі пояснення мають враховувати відмінності між зовнішніми та внутрішніми користувачами ІІС. Зовнішніми користувачами є кінцеві споживачі рішень системи, наприклад, клієнти рекомендаційних систем на платформах електронної комерції. Внутрішніми користувачами є аналітики та експерти, які займаються налаштуванням та підтримкою інтелектуальних інформаційних систем, наприклад коригують тип рекомендацій [2].

Відповідно, для зовнішніх користувачів важливим є розуміння, які саме вхідні дані стали причиною отриманого рішення. Тому пояснення для цієї категорії користувачів мають базуватись на каузальних залежностях [5] між значеннями вхідних змінних та результатом роботи ІІС. Це дає можливість користувачам оцінити відповідність рішення їх потребам.

Для внутрішніх користувачів створити умови для аналізу та удосконалення процесу формування рішень в ІІС.

Тому пояснення для цієї категорії користувачів мають враховувати темпоральну упорядкованість дій процесу прийняття рішень, а також причинно-наслідкові зв'язки між проміжними діями та кінцевим результатом [6].

Таким чином, проблема побудови процесно-орієнтовані пояснення для внутрішніх користувачів і каузальних пояснень для зовнішніх користувачів є актуальною.

Широкі дослідження в сфері побудови пояснень для інтелектуальних систем започатковані згідно програми XAI (Explainable Artificial Intelligence) від агентства DARPA [3]. Ця програма орієнтована на створення методів побудови зрозумілих моделей формування рішень в ІІС, а також на розробку інтерфейсів представлення пояснень щодо рішень, отриманих в таких системах.

Існуючі методи побудови пояснень можна розділити на дві групи, пов'язані з прозорістю або непрозорістю інтелектуальних інформаційних систем:

методи, що базуються на інтерпретації моделі, та методи пост-обробки [7].

Перша група включає методи, що використовують моделі, які можуть бути безпосередньо інтерпретовані користувачем, наприклад, дерева рішень або лінійні моделі.

Друга група методів передбачає побудову додаткової моделі, яка апроксимує початкову модель прийняття рішень і використовується для формування пояснень.

Для пост-обробки на сьогодні широко використовуються алгоритми LIME (Local Interpretable Model-agnostic Explanations) [8] та SHAP (SHapley Additive exPlanations) [9].

Алгоритм LIME будує локальну інтерпретовану модель (наприклад, лінійну регресію) навколо конкретного рішення базової моделі. Локальна модель апроксимує поведінку базової моделі і може бути використана як пояснення, визначаючи, які ознаки найбільше вплинули на результат.

Метод SHAP базується на теорії ігор Шеплі. Він обчислює внесок кожної ознаки у отримане рішення на основі усереднення внеску ознаки по всім можливим множинам ознак.

Для побудови пояснень щодо процесу обробки зображень виконується візуалізація областей зображень, які мали найбільший вплив на рішення моделі комп'ютерного зору. Вони будуються шляхом обчислення градієнтів функції втрат відносно пікселів вхідного зображення [10].

Розглянуті методи фокусуються на поясненні впливу окремих ознак на рішення ІІС, але не враховують причинно-наслідкові зв'язки та послідовність дій у процесі прийняття рішень.

Темпоральні залежності [11] використовуються для опису послідовності дій процесу формування рішення в інтелектуальній системі в цілому [12, 13] та формування новий версій процесу на основі ймовірнісного виводу [14]. Використання темпоральних залежностей для побудови пояснень з урахуванням змін вимог користувачів запропоновано в [15].

Каузальні залежності визначають причинно-наслідкові зв'язки між окремими діями, а також між послідовністю дій та отриманим рішенням. Крім того, каузальні залежності відображають вплив значень вхідних змінних на результат роботи інтелектуальної системи [16].

Поєднання темпоральних та каузальних залежностей для побудови пояснень запропоновано в роботі [17]. Однак задача деталізації пояснень в даній роботі не розглядалась.

Таким чином, існуючі підходи до побудови пояснень орієнтовані на визначення впливу окремих факторів на отримане рішення. Однак комплексному підходу, в якому розглядались би причини рішення в цілому та причини окремих дій процесу прийняття рішення не приділялось достатньо уваги. Тому актуальною задачею є розробка підходу до побудови пояснень в інтелектуальних системах на основі інтеграції темпоральних та каузальних залежностей щодо процесу прийняття рішення для заданого рівня деталізації.

Це дає можливість сформувати пояснення, які розкривають як загальну логіку функціонування інтелектуальної системи, так і причини прийняття конкретних рішень.

Метою статті є розробка підходу до побудови пояснень на основі інтеграції темпоральних та каузальних залежностей щодо процесу прийняття рішення з тим, щоб забезпечити можливість формування пояснення як для зовнішніх, так і для внутрішніх користувачів інтелектуальних інформаційних систем.

Для досягнення поставленої мети вирішуються такі задачі:

– визначення відмінностей доступу до інформації в ІІС для зовнішніх та внутрішніх користувачів;

– розробка ієрархічної моделі пояснення на базі темпоральних та каузальних залежностей;

– розробка методу побудови пояснень з використанням темпоральних та зважених каузальних залежностей.

Відмінності доступу користувачів до інформації в ІІС

Вимоги користувачів до пояснень визначаються на основі особливостей задач, які вирішують внутрішні та зовнішні користувачі інтелектуальних систем. Ключова відмінність між цими користувачами полягає у рівні доступу до інформації про структуру та процес функціонування ІІС.

Внутрішні користувачі – це фахівці, які безпосередньо залучені до розробки, навчання, тестування та підтримки інтелектуальної системи. До цієї категорії належать інженери машинного навчання, фахівці з аналізу даних, команда підтримки, а також, в деяких випадках, розробники програмного забезпечення. Внутрішні користувачі мають частковий доступ до інформації про архітектуру системи, використані алгоритми та моделі, процес обробки даних та прийняття рішень. Тобто для них інтелектуальна система представлена у вигляді «сірої скриньки».

Зовнішні користувачі використовують результати роботи інтелектуальної системи для вирішення своїх задач. Вони можуть бути експертами в предметній області, для якої застосовується система, проте зазвичай не мають глибоких знань в сфері інтелектуальних технологій.

Для зовнішніх користувачів система виглядає як «чорна скринька». Вони мають можливість надавати вхідні дані та отримувати рішення ІІС, але внутрішня структура процесу обробки інформації є для них прихованою.

Таким чином, в залежності від рівня доступу до інформації в інтелектуальних інформаційних систем, внутрішні та зовнішні користувачі мають різне розуміння щодо того, як саме інтелектуальна система формує свої висновки Внутрішні користувачі можуть аналізувати проміжні результати, налаштовувати параметри алгоритмів, модифікувати навчальні вибірки тощо. Зовнішні користувачі оперують

лише вхідними даними та вихідними рішеннями, не маючи можливості безпосередньо вплинути на хід обробки інформації всередині системи.

Ієрархічна модель пояснення на базі темпоральних та каузальних залежностей

З огляду на відмінності в доступі до інформації для внутрішніх і зовнішніх користувачів, запропоновано ієрархічну модель пояснення рішень інтелектуальної інформаційної системи.

Розроблена модель враховує представлення системи як вигляді «сірої», так і у вигляді «чорної» скриньки, що створює умови для інтерпретації процесу формування рішення з різним ступенем деталізації.

Модель P містить три рівні пояснення: локальний P_l, процесний P_π та глобальний P_g, причому локальний задає обмеження на можливі реалізації процесу формування рішення.

$$P = P_g : \exists P_\pi \left| P_l \right. . \qquad (1)$$

Локальний рівень пояснення P_l призначений для внутрішніх користувачів. На даному рівні пояснення будуються з використанням темпоральних залежностей s_m^{m-n} між $m-n$ та m станами процесу формування рішення, що відображають послідовність дій із досягнення рішення в ІІС.

Обмеження у виразі (1) задаються через каузальні правила-обмеження, які виконуються для всіх відомих $i-$ реалізацій процесу прийняття рішень. Тобто такі темпоральні залежності s_m^{m-n} існують при любому відомому виконанні процесу формування рішення :

$$P_l = \left\{ s_m^{m-n} \left| (\exists m, n) : (\forall i) \exists s_m^{m-n} \right. \right\}. \qquad (2)$$

Пояснення на основі обмеження на локальному рівні дає можливість виявити «вузькі місця» в роботі інтелектуальної системи.

Процесний рівень також орієнтований на внутрішніх користувачів. Пояснення на цьому рівні містять інформацію про послідовність дій що привела до формування певного класу схожих рішень. Така послідовність дій задається через темпоральні залежності s_m^{m-n}:

$$P_\pi = \left\langle s_2^1, ..., s_m^{m-n}, ... \right\rangle \Rightarrow Y. \qquad (3)$$

На базі темпоральних формуються зважені каузальні залежності w_m^{m-n} і тоді пояснення приймає вигляд:

$$P_\pi = \left\langle w_2^1, ..., w_m^{m-n}, ... \right\rangle \Rightarrow Y. \qquad (4)$$

Пояснення процесного рівня дають можливість виявити загальні закономірності роботи інтелектуальної системи.

Глобальний рівень пояснення орієнтований на зовнішніх користувачів.

На цьому рівні система представлена у вигляді "чорної скриньки", а пояснення будуються на основі зважених каузальних залежностей, що визначають причинно-наслідковий зв'язок w_Y^X між вхідними даними та фінальним рішенням:

$$P_\pi = w_Y^X . \qquad (5)$$

Такі пояснення Дають можливість обґрунтувати результат роботи системи в термінах опису вхідних даних та результату без деталізації внутрішніх процесів обробки даних в ІІС.

Між рівнями моделі існують ієрархічні зв'язки. Темпоральні залежності локального рівня представляють собою складові процесних пояснень. Процесні пояснення розкривають деталі формування глобальних каузальних залежностей, що зв'язують вхідні дані з отриманим в інтелектуальній системі рішенням.

Темпоральні залежності відіграють ключову роль в запропонованій моделі, оскільки вони безпосередньо відображають впорядкованість станів інтелектуальної системи в процесі обробки даних та формування рішень. На основі цих залежностей в подальшому формуються каузальні залежності більш високого рівня.

Каузальність в даному контексті можна визначити як причинно-наслідковий зв'язок між певними факторами (наприклад, значеннями вхідних даних або послідовністю виконання дій процесу формування рішення) та результатом роботи системи.

Наявність повторюваних темпоральних залежностей, що спостерігаються для всіх відомих реалізацій процесу прийняття рішень, дає підстави стверджувати про наявність каузального зв'язку між відповідними станами чи діями системи.

Тобто повторюваність певних патернів в упорядкованості станів є ознакою причинно-наслідкових відношень.

Ваги каузальних залежностей в запропонованій моделі визначаються з використанням різних підходів для кожного рівня пояснення. На локальному рівні ваги правил-обмежень приймаються рівними 1, оскільки ці правила виконуються для всіх відомих варіантів процесу формування рішення.

На процесному та глобальному рівнях для визначення ваг залежностей застосовується математичний апарат теорії можливостей [18]. Ця теорія, розроблена Лотфі Заде, оперує оцінками можливості та необхідності настання певних подій.

Можливість в теорії можливостей трактується як ступінь правдоподібності певної події (наприклад, причини рішення) за відсутності доказів її неможливості.

Необхідність, в свою чергу, відображає міру впевненості в настанні події з огляду на наявні вхідні дані.

З позицій оцінювання ваг каузальних залежностей можливість характеризує потенційну здатність певного фактору впливати на результат роботи ІІС, а необхідність вказує на обов'язковість врахування цього фактору для коректного формування рішення.

Спільне використання оцінок можливості та необхідності дозволяє більш гнучко описувати

причинно-наслідкові зв'язки порівняно з класичними ймовірнісними моделями, зокрема з урахуванням невизначеності, неповноти та неточності інформації щодо процесу прийняття рішення при представленні ІІС у вигляді «сірої» або «чорної» скриньки.

Метод побудови пояснень з використанням темпоральних та зважених каузальних залежностей

Метод побудови багаторівневого представлення пояснення для інтелектуальних систем на основі темпоральних та каузальних залежностей містить дві основних фази:

– формування пояснення;

– представлення пояснення.

Фаза 1. Формування пояснення.

Дана фаза передбачає створення пояснення на локальному, процесному та глобальному рівнях деталізації.

Eman 1.1. Побудова пояснень локального рівня деталізації.

На локальному рівні пояснення будуються шляхом послідовного формування темпоральних та каузальних залежностей між станами процесу прийняття рішення в інтелектуальній системі згідно представленню (2).

Крок 1.1.1. Формування темпоральних залежностей.

Темпоральні залежності задають порядок у часі для станів процесу формування рішення. При побудові таких залежностей використовується відносний час, що дає можливість порівняти темпоральну упорядкованість для декількох реалізацій процесу формування рішення.

На даному кроці формуються темпоральні залежності X та F типу [11]. Перші задають порядок для двох послідовних станів процесу, наприклад для першого та другого станів залежність позначається як s_2^1. Другі задають порядок для станів, між якими є проміжні стани.

Наприклад для першого та третього станів формується залежність s_3^1.

Результатом даного кроку є множина темпоральних залежностей виду:

$$P_l = \left\{ s_2^1, s_3^1, s_4^1, ..., s_3^2, s_4^2, ... \right\}. \quad (6)$$

Крок 1.1.2. Формування каузальних залежностей. Каузальні залежності на даному рівні деталізації визначаються як темпоральні залежності, які виконуються для всіх варіантів (реалізацій) процесу прийняття рішення. На даному кроці порівнюються всі відомі $i-$ варіанти виконання процесу. Якщо така умова задовольняється, то вага залежності w_m^{m-n} встановлюється рівною одиниці:

$$\left(\forall i\right) \exists s_m^{m-n} \Rightarrow w_m^{m-n} = 1. \quad (7)$$

Результатом етапу є множина темпоральних залежностей s_m^{m-n} та підмножина каузальних залежностей, що мають одиничну вагу.

Перші виступають як елементи побудови процесного рівня пояснення, а другі – як обмеження даного рівня.

Eman 1.2. Побудова пояснень процесного рівня деталізації.

На процесному рівні аналізується сукупність темпоральних залежностей, які відображають порядок станів у поточному процесі формування рішення в інтелектуальній системі.

Каузальні залежності на цьому рівні формуються з темпоральних з використанням математичного апарату теорії можливостей, що дозволяє оцінити ступінь впевненості у наявності причинно-наслідкового зв'язку між станами на основі частоти спостереження відповідної темпоральної залежності у різних реалізаціях процесу прийняття рішення.

Eman 1.3. Побудова пояснень глобального рівня деталізації.

На глобальному рівні визначаються каузальні залежності між вхідними даними інтелектуальної системи та отриманим рішенням згідно (5).

Ваги цих залежностей також обчислюються з використанням оцінок можливості та необхідності в рамках теорії можливостей [18].

Це дозволяє врахувати невизначеність щодо впливу окремих вхідних факторів на кінцевий результат.

Фаза 2. Формування представлення пояснення.

Фаза формування представлення пояснення складається з п'яти етапів, які забезпечують вибір та деталізацію пояснень для користувача інтелектуальної системи.

Eman 2.1. Формування множини альтернативних пояснень.

На першому етапі формується множина альтернативних пояснень на основі каузальних залежностей, отриманих на попередній фазі.

Ці пояснення впорядковуються за вагою, яка відображає ступінь впевненості у причинно-наслідкових зв'язках.

Eman 2.2. Оцінка пояснень за критеріями якості.

На другому етапі виконується оцінка сформованих пояснень.

Зокрема, на даному етапі перевіряється складність пояснень. Також виконується оцінка коректності та чутливості пояснень. Перша відображає ступінь відповідності пояснення реальному процесу прийняття рішення в інтелектуальній системі. Друга відображає ступінь зміни пояснення при зміні вхідних даних або параметрів моделі.

Eman 2.3. Відбір пояснень.

На третьому етапі здійснюється відбір коректних пояснень, які мають максимальну вагу та відповідають встановленим критеріям чутливості та складності. Це дозволяє обрати релевантні та зрозумілі для користувача пояснення.

Eman 2.4. Деталізація пояснень за об'єктами, з якими оперує процес формування рішення в ІІС.

На четвертому етапі відібрані пояснення, за потреби, деталізуються шляхом встановлення зв'язків

між каузальними залежностями та об'єктами, які використовує процес формування рішення [19].

Це дозволяє користувачу краще зрозуміти вплив конкретних сутностей предметної області на логіку формування рішення.

Eman 2.5. Доповнення пояснень вхідними та проміжними даними.

На п'ятому етапі каузальні залежності, що утворюють пояснення, доповнюються вхідними даними, які були використані при формуванні відповідних станів процесу прийняття рішення.

Такий підхід забезпечує можливість інтерпретації пояснень та дозволяє користувачу встановити зв'язок між вхідними даними, діями та отриманим результатом.

Розглянемо реалізацію фази 1 даного методу для побудови пояснення щодо рекомендацій в системі електронної комерції.

Нехай у процесі формування рекомендацій користувачеві спостерігаються наступні стани: c_1 – користувач переглядає сторінку товару «Смартфон Samsung Galaxy S21»; c_2 – користувач додає товар «Смартфон Samsung Galaxy S21» до кошика; c_3 – : Користувач оформлює замовлення на товар «Смартфон Samsung Galaxy S21».

На першому етапі формуються темпоральні залежності s_2^1, s_3^1, s_3^2 між цими станами. Оскільки ці залежності є характерними для всіх користувачів, які купують даний товар, то $w_2^1 = w_3^1 = w_3^2 = 1$, тобто дані залежності є каузальними правилами-обмеженнями.

На другому етапі формується процес із послідовності каузальних залежностей s_2^1, s_3^2, що визначають послідовний перехід від першого до другого, а потім до третього стану, а також залежності s_3^1, що визначає обов'язкову наявність c_3 в майбутньому після c_1. Цей фрагмент процесу є детермінованим, оскільки правила s_2^1, s_3^2 та s_3^1 виступають в якості обмежень.

На другій фазі визначаються атрибути станів $c_1 - c_3$ і пояснення деталізується з урахуванням об'єктів, до яких належать ці атрибути.

На глобальному рівні деталізації пояснення визначаються каузальні залежності між характеристиками користувача (вік, стать, місто проживання) та фактом купівлі товару «Смартфон Samsung Galaxy S21».

Розраховується значення можливості і відбирається пояснення з найбільшим значенням можливості.

Зокрема, пояснення виду «Користувачі чоловічої статі віком 25-40 років часто купують смартфон Samsung Galaxy S21 після його перегляду та додавання у кошик».

Висновки

З метою обґрунтування необхідності багаторівневої деталізації пояснень виконано структуризацію відмінностей доступу до інформації користувачів інтелектуальних систем.

Показано, що внутрішні користувачі мають частковий доступ до внутрішньої інформації щодо процесу прийняття рішення, що потребує відповідної деталізації пояснень.

Запропонована ієрархічну модель пояснення на основі темпоральних та каузальних залежностей, яка містить локальний, процесний та глобальний рівні пояснень згідно можливостей доступу користувачів до інформації щодо процесу формування рішення в інтелектуальній системі.

Локальний та процесний рівні орієнтовані на внутрішніх користувачів і передбачають більш детальний аналіз роботи системи як "сірої скриньки". Глобальний рівень призначений для зовнішніх користувачів і надає узагальнені пояснення при представленні системи у вигляді "чорної скриньки". Темпоральні залежності визначають порядок дій процесу формування рішення. На їх основі формуються зважені каузальні залежності.

Модель дозволяє врахувати невизначеність та неповноту інформації при поясненні рішень інтелектуальних систем для зовнішніх та внутрішніх користувачів.

Запропоновано метод побудови пояснень з використанням темпоральних та каузальних залежностей між станами або діями процесу формування рішення в інтелектуальній системі.

Метод містить фази формування та представлення пояснень. Метод дозволяє сформувати каузальні залежності на декількох рівнях деталізації та представити їх користувачу у вигляді упорядкованої за вагою множини альтернативних пояснень.

Деталізація пояснень з урахуванням властивостей об'єктів, що використовуються у процесі формування рішення, а також доповнення пояснень на основі каузальних залежностей вхідними даними сприяє кращому розумінню користувачем логіки роботи інтелектуальної системи.

REFERENCES

1. Castelvecchi, D. (2016). Can we open the black box of AI? Nature, 538(7623), 20-23. https://doi.org/10.1038/538020a.
2. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38. https://doi.org/10.1016/j.artint.2018.07.007.
3. Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. AI Magazine, 40(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850.
4. Tintarev, N., & Masthoff, J. (2007). A survey of explanations in recommender systems. In IEEE 23rd International Conference on Data Engineering Workshop (pp. 801-810). https://doi.org/10.1109/ICDEW.2007.4401070.
5. Pearl, J. (2009). Causality: Models, Reasoning and Inference (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511803161.

6. Chalyi, S., & Leshchynskyi, V. (2020). Temporal representation of causality in the construction of explanations in intelligent systems. Advanced Information Systems, 4(3), 113-117. https://doi.org/10.20998/2522-9052.2020.3.16.
7. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052.
8. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778.
9. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774). https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
10. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 618-626). https://doi.org/10.1109/ICCV.2017.74
11. Chala, O. (2018). Models of temporal dependencies for a probabilistic knowledge base. Econtechmod. An International Quarterly Journal, 7(3), 53-58.
12. Chala, O. (2019). Development of information technology for the automated construction and expansion of the temporal knowledge base in the tasks of supporting management decisions. Technology Audit and Production Reserves, 1/2(45), 9-14. https://doi.org/10.15587/2312-8372.2019.160205.
13. Levykin V., Chala O. Development of a method of probabilistic inference of sequences of business process activities to support business process management. Eastern-European Journal of Enterprise Technologies. 2018. № 5/3(95). P. 16-24. DOI: 10.15587/1729-4061.2018.142664.
14. Чала О. В. (2020) Модель узагальненого представлення темпоральних знань для задач підтримки управлінських рішень. Вісник Національного технічного університету «ХПІ». Системний аналіз, управління та інформаційні технології. № 1(3). С. 14-18. DOI: 10.20998/2079-0023.2020.01.03.
15. Chalyi, S., & Leshchynskyi, V. (2020). Method of constructing explanations for recommender systems based on the temporal dimension of user choice. EUREKA: Physics and Engineering, (3), 43-50. https://doi.org/10.21303/2461-4262.2020.001228.
16. Чалий, С. Ф., & Лещинський, В. О. (2023). Можливісна модель каузального зв'язку по вхідній змінній для побудови пояснення в інтелектуальній системі. Системи управління, навігації та зв'язку, (1(71), 102-106. https://journals.nupp.edu.ua/sunz/article/view/3066.
17. Chalyi, S., & Leshchynskyi, V. (2020). Temporal representation of causality in the construction of explanations in intelligent systems. Advanced Information Systems, 4(3), 113-117. https://doi.org/10.20998/2522-9052.2020.3.16.
18. Dubois, D., & Prade, H. (2011). Possibility theory and its applications: Where do we stand? In Springer handbook of computational intelligence (pp. 31-60). Springer, Berlin, Heidelberg. https://doi.org/10.20998/2522-9052.2020.3.16.
19. Chalyi, S., & Leshchynskyi, V. (2022). Temporal representation of the essences of the subject area for the construction of explanations in intelligent systems. Intelligent Information Systems, 1-10.

Received (Надійшла) 10.09.2024
Accepted for publication (Прийнята до друку) 20.11.2024

Modeling explanations in intelligent systems based on the integration of temporal and causal dependencies

S. Chalyi, V. Leshchynskyi

Abstract. The **article's subject matter** is the processes of constructing explanations in intelligent systems using temporal and causal dependencies. The aim is to develop an approach to constructing explanations based on the integration of temporal and causal dependencies regarding the decision-making process in order to provide the possibility of forming explanations for both external and internal users of intelligent information systems. **Tasks:** determining the differences in access to information in IIS for external and internal users; developing a hierarchical model of explanation based on temporal and causal dependencies; developing a method for constructing explanations using temporal and causal dependencies. **The approaches used are**: methods of constructing explanations, methods of constructing causal dependencies. The following **results** were obtained. The structuring of the differences in access to information of users of intelligent systems was performed to justify the need for multilevel detailing of explanations. A hierarchical model of explanation based on temporal and causal dependencies is proposed. A method for constructing explanations using temporal and causal dependencies between states or actions of the decision-making process in an intelligent system is proposed. **Conclusions**. The scientific novelty of the obtained results is as follows. A hierarchical model of explanation is proposed, which contains local, process, and global levels of explanations according to the possibilities of user access to information about the decision-making process, which makes it possible to take into account the incompleteness of information about the state of the intelligent system when explaining its decisions for external and internal users. A method for constructing explanations has been developed, which contains the phases of constructing explanations using temporal and causal dependencies between input data, states of the decision-making process, and the obtained result, and presenting explanations, which makes it possible to form causal dependencies at several levels of detail and present them to the user in the form of an ordered by weight set of alternative explanations.

Keywords: intelligent system, artificial intelligence system, explanation, decision-making process, temporal dependencies, causal dependencies, causality, possibility.