

ПРИЧИННО-НАСЛІДКОВЕ МОДЕЛЮВАННЯ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ

Об'єктом дослідження є процес побудови пояснень в інтелектуальних інформаційних системах. Запропоновано причинно-наслідкову модель пояснення з ієрархічною тривірневою структурою, яка формує каузальні залежності на глобальному, процесному та локальному рівнях у темпоральному та каузальному аспектах. Запропоновано метод побудови пояснень на основі причинно-наслідкової моделі, що містить етапи визначення користувачів, побудови пояснень на кожному рівні ієрархії причинно-наслідкової моделі та формування інтегрованого тривірневого пояснення. Використання методу дозволяє деталізувати пояснення відповідно до потреб внутрішніх користувачів (розробників, адміністраторів) та зовнішніх користувачів (кінцевих клієнтів), що дозволяє підвищити прозорість рішень та довіру до отриманих рішень інтелектуальних систем.

1. Вступ

Сучасні інтелектуальні системи, які включають в себе як підсистеми штучного інтелекту, так і традиційні інформаційні підсистеми, набули поширення у таких сферах людської діяльності, як охорона здоров'я, фінансові послуги, управління підприємствами, освіта [1]. Ці інтелектуальні інформаційні системи (ІС) забезпечують підтримку прийняття рішень, що безпосередньо впливають на життя людей та функціонування підприємств та організацій [2]. Підсистеми штучного інтелекту використовують методи машинного навчання та складні репрезентації даних [3]. Тому вони, здебільшого, функціонують як «чорні скриньки», що утруднює пояснення процесу їхньої роботи [1], [4]. Традиційні інформаційні системи зазвичай використовують детерміновані процеси із зрозумілою логікою, проте ці процеси з точки зору користувача часто мають вигляд «сірих скриньок» внаслідок великої кількості параметрів та складних алгоритмів [5]. Алгоритми та програмні рішення також можуть бути захищені в рамках комерційної таємниці, що обмежує їхню прозорість для користувача. Результатом юридичного захисту є представлення ІС як «правової чорної скриньки». Проте використання складних непрозорих алгоритмів у процесі формування рішень в ІС знижує довіру користувачів до цих рішень, що суттєво обмежує ефективність їх практичного застосування [6].

З метою підвищення довіри користувачів ІС та, відповідно, ефективного використання рішень ІС проводяться дослідження у сфері пояснюваного штучного інтелекту (ХАІ) [7]. Пояснення забезпечують прозорість роботи ІС за рахунок інтерпретації причин послідовності дій процесу формування рішень.

Таким чином, проблема формування та використання пояснень в ІС з урахуванням причин рішень та процесу формування рішень в ІС є актуальною.

2. Аналіз літературних даних і постановка проблеми дослідження

Сучасні підходи до побудови пояснюваних ІС включають два напрямки. Перший напрямок орієнтований на створення інтерпретованих моделей, логіка яких побудована на правилах, графах прийняття рішень тощо [8]. Другий напрямок реалізує post-hoc пояснення для моделей, які представляються для користувача як «чорні скриньки». В рамках даної групи виконується локальна апроксимація складної функції лінійною моделлю (LIME) [6] та визначення внеску значень вхідних даних за кооперативною ігровою схемою (SHAP) [9].

Обидва напрямки забезпечують інтерпретацію конкретних рішень, проте не

враховують динаміку прийняття рішень. Користувач отримує перелік факторів впливу на результат, але не має інформації, коли і як саме вони вплинули на цей результат. Така можливість є критичною для предметних областей, де послідовність дій або послідовність подання вхідних даних впливає на отримане рішення (медична діагностика, фінансовий аналіз тощо).

Темпоральні аспекти відображаються як упорядкована послідовність для пар подій. В [10], [11] запропоновано моделювати можливі причини рішення на основі темпорального порядку подій процесу формування рішення та подальшої побудови причинно-наслідкових залежностей на основі темпоральної упорядкованості. Темпоральний порядок процесу формування рішення в роботах [12]-[15] задається за допомогою темпоральних правил.

З іншого боку, дослідження у сфері причинно-наслідкового виведення [16], [17] дають можливість визначати причини отриманого результату, але вони не приділяють уваги темпоральному порядку дій процесу. Проте емпіричні спостереження в області когнітивної психології [18] підтверджують, що люди вважають пояснення переконливим лише тоді, коли воно враховує темпоральний аспект. Тому поєднання формування причинно-наслідкових залежностей на базі темпоральних правил дає можливість врахувати динаміку роботи ПС [19].

Таким чином, існуючі підходи до побудови пояснень фокусуються переважно на визначенні факторів, які впливають на отримане рішення, і не приділяють достатньо уваги упорядкованому в часі процесу формування рішення, що і потребує розробки комплексного багаторівневого підходу до побудови пояснень. Такий підхід має виявити причини рішення на основі вхідних даних, а також дій процесу з формування рішення, що і свідчить про актуальність теми даного дослідження.

3. Мета і задачі дослідження

Метою дослідження є розробка причинно-наслідкового підходу до побудови пояснень в інтелектуальних системах з тим, щоб відобразити у формі пояснення причини отриманого рішення, а також причини дій процесу формування рішення й обмеження на ці дії, що дає можливість представити роботу інтелектуальної системи у зрозумілій для користувачів формі.

Для досягнення мети роботи вирішуються такі задачі: розробка причинно-наслідкової моделі пояснення; розробка методу побудови пояснень на основі причинно-наслідкової моделі.

4. Причинно-наслідкова модель пояснення

Об'єктом дослідження є процес побудови пояснень в ПС.

Предметом дослідження є моделі пояснень та методи їх побудови в ПС.

Розроблена причинно-наслідкова модель пояснення M має ієрархічну структуру та включає три базових рівні L представлення множини каузальних залежностей R , що базуються на темпоральній упорядкованості вхідних даних T_{in} та дій процесу формування рішення T_d . Ці рівні визначають такі аспекти пояснення:

- глобальний L_{GI} , де пояснення задає тлумачення причин рішення ПС в цілому, без деталізації внутрішнього механізму його формування;

- процесний L_{Pr} , де пояснення відображає причинно-наслідкові залежності між діями процесу формування рішення в ПС та отриманим рішенням; ці залежності базуються на темпоральній упорядкованості дій процесу;

- локальний L_{Loc} , де описуються окремі обмеження на процес формування рішення, причому ці обмеження задаються через темпоральну упорядкованість пар дій процесу.

Наприклад, у медичній діагностичній системі модель пояснення на глобальному рівні містить причинно-наслідкові залежності із множини виду «симптоми $\rightarrow$ діагноз», на

процесному рівні модель M визначає послідовність аналізів та діагностичних дій, а на локальному рівні – обов'язкові етапи обстеження.

Формально дана модель має вигляд:

$$M = \langle L, R, T \rangle, \quad (1)$$

де $L = \{L_{Gl}, L_{Pr}, L_{Loc}\}$ – множина рівнів ієрархії пояснення; $T = \{T_{in}, T_d\}$ – темпоральна упорядкованість даних та дій процесу; R – відношення, на основі яких будується пояснення.

Така ієрархічна структура дає можливість формувати пояснення як каузальні залежності із заданим ступенем деталізації відповідно до потреб внутрішніх та зовнішніх користувачів. Внутрішні користувачі вирішують задачі із розробки та підтримки функціонування ІС і тому в рамках пояснень потребують детального розуміння внутрішньої логіки поточного процесу прийняття рішень для того, щоб виявити можливі помилки й порушення обмежень у даному процесі, удосконалити моделі роботи ІС та забезпечити відповідність регуляторним вимогам. Тому для цієї категорії користувачів пояснення мають відображати причини можливих послідовностей дій, а також порушення обмежень при функціонуванні ІС. Зовнішні користувачі використовують рішення ІС у своїй професійній діяльності, і тому пояснення для цих користувачів мають забезпечити зрозумілість причин отриманих результатів та здатність відкоригувати вхідні дані й параметри ІС для отримання бажаних результатів.

Ключові властивості наведених рівнів представлення пояснення у причинно-наслідковій моделі наведено в табл. 1.

Таблиця 1

Властивості рівнів пояснення в причинно-наслідковій моделі

Характеристика	Глобальний рівень	Процесний рівень	Локальний рівень
Користувачі	Зовнішні користувачі (кінцеві клієнти)	Внутрішні користувачі (розробники, data scientists)	Внутрішні користувачі (системні аналітики)
Прозорість ІС для користувача	Чорна скринька	Сіра скринька	Сіра скринька або біла скринька
Об'єкт пояснень	Причини рішення ІС	Причини дій процесу формування рішення в ІС	Причини обов'язкових дій у процесі формування рішення в ІС
Вхідні дані для пояснення	Лінійно упорядковані значення вхідних змінних, рішення ІС	Логи процесу формування рішень в ІС	Логи процесу формування рішень в ІС
Представлення пояснення	Можливісні причинно-наслідкові залежності між вхідними даними та рішенням ІС	Можливісні причинно-наслідкові залежності між діями	Детерміновані або ймовірнісні обмеження щодо допустимих варіантів процесу формування рішення в ІС

Глобальний рівень причинно-наслідкової моделі пояснення містить каузальні залежності між вхідними даними та рішенням ІС. На даному рівні ІС є непрозорою для кінцевого користувача, тобто розглядається як «чорна скринька». Для побудови залежностей на даному рівні причинно-наслідкового пояснення як можлива причина використовуються або безпосередньо значення вхідних змінних, або приховані фактори

(неявні вхідні дані). Крім того, на вході можуть бути використані темпоральні правила, що відображають послідовність значень цих змінних у наборі вхідних даних. Лінійний порядок вхідних даних у часі дає можливість врахувати послідовність впливу вхідних змінних на формування рішення. Наслідком є рішення системи, яке має бути використано у відповідній предметній області.

Каузальні залежності R_{Gl} , що є основою пояснення на даному рівні, визначаються в рамках теорії можливості, тобто з використанням показників можливості Π_{Gl} та необхідності N_{Gl} . Такий підхід дозволяє враховувати невизначеність, що виникає внаслідок неповноти інформації щодо внутрішнього механізму функціонування ІС. Пояснення на глобальному рівні орієнтовані на користувачів – кінцевих клієнтів, що використовують рішення ІС у своїй професійній діяльності.

Формально пояснення на глобальному рівні L_{Gl} задають причинно-наслідкові залежності R_{Gl} між вхідними даними X та рішенням Y з урахуванням упорядкованості вхідних даних T_{in} та прихованих факторів U :

$$L_{Gl} = R_{Gl}(X, T_{in}, Y, U) : \forall R_{Gl} \exists (\Pi_{Gl}, N_{Gl}). \quad (2)$$

Множина вхідних змінних може бути упорядкована за відношенням T_{in} :

$$X = \{x_i : (\forall i, i+1) \exists T_{in}^{i,i+1}\}. \quad (3)$$

Каузальні залежності на даному рівні визначаються шляхом оцінки можливості та необхідності конкретних значень вхідних для даних $X = x$ бути причиною конкретного рішення $Y = y$.

Можливість визначається з урахуванням прихованих факторів $u \in U$, які разом із поточними вхідними даними x приводять до поточного результату y , тобто $f(x, u) = y$:

$$\Pi_{Gl}(Y = y | X = x) = \max_{u \in U} \{\pi(u) : f(x, u) = y\}. \quad (4)$$

Наприклад, в банківській кредитній системі можливість $\Pi_{Gl} = 0,75$ рішення $Y = y$: *кредит схвалено*, що отримано на основі вхідних даних $X = x$: (*дохід = 80000 грн, вік = 40 років*) за заданих прихованих факторів $U = \{\text{поточна процентна ставка, ризику, сезонні коливання}\}$ та $\max_{u \in U} \pi(u) = \pi(\text{поточна процентна ставка}) = 0,75$, означає, що найбільший вплив на рішення щодо схвалення кредиту має поточна процентна ставка. Значення процентної ставки можна розглядати як основну причину виділення кредиту для клієнта віком 40 років з доходом 80000 гривень.

Оцінка необхідності є дуальною до оцінки можливості Π_{Gl} та базується на оцінці можливості альтернатив до Π_{Gl} і, відповідно, визначає ступінь неминучості рішення $Y = y$ за вхідних даних $X = x$ з урахуванням прихованих факторів $u \in U$:

$$N_{Gl}(Y = y | X = x) = 1 - \Pi_{Gl}(Y \neq y | X = x). \quad (5)$$

$\Pi_{Gl}(Y \neq y | X = x)$ відображає можливість всіх інших результатів крім Y . Наприклад $Y \neq y$: *{кредит відхилено, кредит потребує додаткових документів}*. Високе значення

необхідності означає, що альтернативні результати $Y \neq y$ мають низьку можливість. Якщо $N_{Gl}(Y = y | X = x) = 0,85$, то можливість альтернативного результату не перевищує 0,15, що свідчить про «обов'язковість» результату $Y = y | X = x$. У відповідності до прикладу кредитування, $N_{Gl} = 0,85$ означає, що для представлених даних $X = x$: (дохід = 80000 грн, вік = 40 років) кредит на практиці буде схвалено.

Таким чином, високі оцінки можливості та необхідності свідчать про те, що задані вхідні дані та приховані фактори є причиною отриманого рішення. Тобто оцінка можливості визначає максимальний ступінь довіри до каузальної залежності з урахуванням значень прихованих факторів u . Міра необхідності є дуальною до міри можливості та показує, наскільки неминучим є певний результат. Відповідно, пояснення на глобальному рівні формується за значенням оцінки можливості та при обмеженнях за значенням оцінки необхідності.

Процесний рівень причинно-наслідкової моделі пояснення призначений для опису механізму формування рішення в ІС. На даному рівні пояснюється послідовність дій процесу прийняття рішень на основі формування причинно-наслідкових залежностей між цими діями. Такі залежності встановлюються як між послідовними парами дій, так і між діями, між якими існують проміжні етапи процесу формування рішення [3], [11]. Такий підхід до формування пояснення дає можливість представити користувачам не лише причини кінцевого рішення ІС, але й логіку формування цього рішення через послідовність упорядкованих у часі дій. Причинно-наслідкові залежності на даному рівні базуються на темпоральних правилах типів «Next» та «Future», що відображають безпосередні та віддалені темпоральні залежності між діями процесу формування рішення. Правила типу «Next» задають порядок у часі для дій, які безпосередньо йдуть одна за одною. Правила типу «Future» охоплюють декілька послідовних дій і враховують опосередковані зв'язки між першою та останньою діями із цієї послідовності. Процес формування рішення у загальному випадку може мати декілька альтернативних реалізацій, які відрізняються можливими послідовностями дій. Відповідно, каузальні залежності на процесному рівні причинно-наслідкового пояснення визначаються з використанням теорії можливості на основі показників можливості та необхідності для кожної пари дій, представленої темпоральним правилом «Next» або «Future». Показник можливості дозволяє оцінити ступінь довіри до першої дії як причини другої дії із відповідної пари дій. Необхідність є оцінкою можливості альтернативної послідовності дій. Тобто необхідність обумовлює порівняння поточної причини з альтернативними або контрфактичними. Такий підхід дозволяє оцінити поточну реалізацію процесу формування рішення у порівнянні з можливими альтернативними послідовностями дій.

Формально процесний рівень L_{Pr} причинно-наслідкової моделі пояснення представляється через множину дій D процесу формування рішень, упорядкованих за темпоральними відношеннями R_N та R_F типів «Next» та «Future» відповідно. Оцінка причин дій процесу виконується на базі можливості P_{Pr} та необхідності N_{Pr} :

$$L_{Pr} = \langle D, R_N, R_F, P_{Pr}, N_{Pr} \rangle. \quad (6)$$

Оцінка можливості виконується для поточного рішення $Y = y$ та поточної послідовності дій $D = d$ процесу формування рішення на основі темпоральних відношень R_F , оскільки останні визначають ланцюжки дій $R_F(d_i, d_j)$, упорядкованих з використанням проміжних відношень R_N . Тобто знаходиться ключовий ланцюжок дій

$(d_i, d_j): j-1 > 1$, який привів до отриманого результату:

$$\Pi_{Pr}(Y = y|D = d) = \max_{j, j \in D} \left\{ \pi(R_F(d_i, d_j)) \right\}. \quad (7)$$

Розрахунок $\max_{j, j \in D} \left\{ \pi(R_F(d_i, d_j)) \right\}$ виконується по всіх відомих альтернативних шляхах формування результату. Тобто $\pi(R_F(d_i, d_j))$ визначається через нормовану кількість повторів правила $R_F(d_i, d_j)$ на відомій множині реалізацій процесу формування рішення в ІС.

Необхідність N_{Pr} на процесному рівні розраховується аналогічно глобальному рівню як доповнення можливості Π_{Pr} :

$$N_{Pr}(Y = y|D = d) = 1 - \Pi_{Pr}(Y \neq y|D = d). \quad (8)$$

У підсумку, на даному рівні пояснення встановлюється можливість та необхідність для правил R_F , що визначають послідовність дій, яка мала найбільший вплив на отримане рішення. Наприклад, дія з призначення лікування d_j може бути визначена на основі послідовності дій після аналізу крові або ж після МРТ.

Локальний рівень причинно-наслідкового пояснення задає обмеження на множину можливих варіантів реалізації процесу формування рішення. Відповідно, локальний рівень є нижнім рівнем в ієрархії причинно-наслідкового пояснення. На даному рівні виділяються пари упорядкованих у часі дій процесу формування рішення в ІС, які мають виконуватися для всіх або переважної більшості відомих варіантів даного процесу. Такий підхід створює умови для побудови детермінованих або ймовірнісних причинно-наслідкових залежностей між парами дій.

Формально локальний рівень причинно-наслідкового пояснення задається через множину дій D та правила C_N та C_F , які базуються на правилах R_N та R_F відповідно і виконуються для всіх варіантів процесу формування рішення в ІС:

$$L_{Loc} = \langle D, C_N, C_F \rangle. \quad (9)$$

C_N визначається через темпоральне правило, яке виконується для всіх відомих послідовностей дій $(d_1, \dots, d_n) \in D$ процесу формування рішення:

$$C_N = R_N : \forall (d_1, \dots, d_n) \exists R_N. \quad (10)$$

Наприклад, при двохфакторній авторизації в банківській системі після дії d_1 із вводу пароля завжди виконується дія d_2 із підтвердження особи.

Процесний та локальний рівень причинно-наслідкової моделі пояснення орієнтовані, в першу чергу, на внутрішніх користувачів, тоді як глобальний рівень – на зовнішніх користувачів. Ці два класи користувачів відрізняються можливостями доступу до інформації щодо процесу роботи ІС, а також практичними задачами у сфері їхньої професійної діяльності.

5. Метод побудови пояснень на основі причинно-наслідкової моделі

Розроблений метод містить такі етапи.

Етап 1. Визначення внутрішніх та зовнішніх користувачів.

Визначаються кінцеві клієнти, які будуть використовувати пояснення для вирішення

свої практичних задач, а також розробники, аналітики даних тощо, які будуть використовувати пояснення для удосконалення ІС.

В залежності від результатів етапу будуть сформовані пояснення на одному, двох, або трьох рівнях.

Етап 2. Побудова пояснення на глобальному рівні.

Крок 2.1. Визначення підмножини значень вхідних змінних $X = x$, властивостей рішення $Y = y$ та внутрішніх факторів $u \in U$ для побудови пояснення.

Крок 2.2. Розрахунок можливості та необхідності для встановлення причин рішення згідно з (4) та (5).

Результатом етапу є пояснення причини рішення на основі прихованих факторів за умови поточних вхідних даних з оцінками його можливості та необхідності.

Етап 3. Побудова пояснення на процесному рівні.

Крок 3.1. Визначення множини дій процесу D .

Крок 3.2. Формування темпоральних правил R_N та R_F .

Крок 3.3. Розрахунок можливості та необхідності для ланцюжка дій – причини отриманого рішення згідно з (7) та (8).

Результатом етапу є пояснення причини рішення як ключової послідовності дій процесу формування рішення з оцінками його можливості та необхідності.

Етап 4. Побудова пояснення на локальному рівні.

Крок 4.1. Знаходження обмежень C_N на основі темпоральних правил R_N , які виконуються для всіх варіантів процесу формування рішення.

Крок 4.2. Знаходження обмежень C_F на основі темпоральних правил R_F , які виконуються для всіх варіантів процесу формування рішення.

Крок 4.3. Перевірка виконання обмежень для поточного процесу формування рішення.

Результатом етапу є пояснення у формі виконання/невиконання обмежень поточного процесу формування рішення.

Етап 5. Побудова об'єднаного трирівневого пояснення, що містить як причини рішення значення вхідних даних та дії процесу формування рішення, а також інформацію про задоволення обмежень на процес формування рішення.

6. Експериментальна перевірка методу побудови пояснень на основі причинно-наслідкової моделі

Експериментальну перевірку розробленого методу виконано з використанням логу фірми Volvo IT, який містить записи про невіршені проблеми управління сервісними процесами у інформаційній системі компанії.

Вхідний лог містить 819 трас, що відображають можливі варіанти реалізації процесу сервісного обслуговування у інформаційній системі компанії Volvo IT. Ключові дії d_i даного процесу – Accepted, Queued, Completed.

Аналіз логу на етапі 1 методу дає можливість виділити таких ключових внутрішніх користувачів: менеджери процесів, аналітики. Зовнішніми користувачами є клієнти, які надають інформацію про невіршені проблеми, та бізнес-користувачі.

За результатами етапу 2 методу виділено набір вхідних даних $X = \{Тип продукту (product)\}$, а також набір факторів $U = \{Робоча група (org : group), Країна (resource country), Вплив (impact)\}$, які впливають на рішення системи. Рішення містить статус процесу: $Y = \{Статус завершення : Completed / Not Completed\}$.

Розрахунок причин виконано для робочих груп. Результати для ключових робочих груп (з максимальним значенням можливості) наведено в табл. 2.

Таблиця 2

Можливості впливу робочих груп на результат Y процесу		
Робоча група	Можливість $\pi(u)$	Необхідність $n(u)$
A2	0,85	0,62
C	0,78	0,45
G4	0,72	0,38
B	0,69	0,34

Аналіз табл. 2 дає можливість зробити висновок, що найвищу можливість успішного виконання процесу у Volvo IT має робоча група A2. Тобто якщо поточний процес обслуговує інша група, то поясненням невирішеної проблеми є недостатня кваліфікація цієї робочої групи.

На етапі 3 побудовано темпоральні правила R_N та R_F та виконано розрахунок можливості та необхідності для цих правил.

Отримано такі ключові залежності:

- можливість: $Accepted \rightarrow In Progress(0,95)$;
- необхідність: $Queued \rightarrow Accepted(0,78)$.

Тобто залежність $Accepted \rightarrow In Progress$ має найбільший вплив на успішне завершення процесу, а залежність $Queued \rightarrow Accepted$ у більшості випадків є необхідною для успішного завершення процесу.

На етапі 4 виконано формування обмежень на послідовність дій процесу. Виконано аналіз всіх трас процесу. Обмеженнями в рамках експерименту вважаються ті правила, які виконані на більш ніж 95 % трас.

Результати етапу наведено у табл. 3.

Таблиця 3

Обмеження на виконання процесу		
Правило-обмеження	Виконання, %	Невиконання
$Queued \rightarrow Accepted$	96.8%	26 випадків
$Accepted \rightarrow Completed$	98.2%	14 випадків

На етапі 5 сформовано трирівневе пояснення для поточного варіанту процесу.

Глобальний рівень: робоча група A2 забезпечує можливість 0,85 успішного завершення процесу.

Процесний рівень: послідовність $Accepted \rightarrow In Progress$ забезпечує можливість 0,95 успішного завершення поточного процесу; послідовність $Queued \rightarrow Accepted$ є необхідною у поточному процесі і має значення необхідності 0,78.

Локальний рівень: наведені обмеження у поточному процесі повністю виконуються.

7. Обговорення результатів дослідження

Розроблена причинно-наслідкова модель пояснення та метод її побудови в цілому формують трирівневе представлення пояснень, придатне як для зовнішніх, так і для внутрішніх користувачів. Ієрархічне представлення пояснень на глобальному, процесному

та локальному рівнях дало змогу поєднати різні типи причин для отриманого рішення, включаючи значення вхідних змінних та обов'язкові операції процесу. Відмінність запропонованого підходу полягає в такому. На глобальному рівні враховано нечіткий характер впливу явних та неявних вхідних змінних через значення можливості та необхідності. На процесному рівні причини рішення встановлюються на основі темпоральних правил «Next»/«Future», які враховують як безпосередні, так і віддалені залежності між діями процесу формування рішення. На локальному рівні визначено детерміновані обмеження, які визначають множину допустимих варіантів виконання процесу формування рішення в ІС.

Розроблений метод побудови причинно-наслідкових пояснень послідовно формує пояснення на кожному рівні моделі, а потім об'єднує їх у єдину структуру.

Переваги запропонованого підходу полягають у можливості порівняти значення вхідних даних та дії процесу формування рішення як причини отриманого рішення, а також встановити виконання обмежень для поточного процесу.

Проведений на даних із журналу Volvo IT експеримент дозволив виявити робочі групи з максимальною можливістю успішного вирішення проблемних ситуацій, критичні послідовності дій процесу з найбільшими значеннями можливості та необхідності, а також підтвердити виконання обмежень на процес. Аналіз всіх альтернативних варіантів виконання процесу дозволив виявити порушення обмежень у окремих альтернативах процесу, що дає змогу локалізувати та в подальшому удосконалити ці варіанти.

Обмеження запропонованого підходу пов'язані з потребою у великій кількості трас альтернативних варіантів виконання процесу з репрезентативним розподілом подій для обчислення оцінок можливості/необхідності. Крім того, лог не містить окремого маркування для неявних вхідних даних (прихованих факторів), що ускладнює побудову пояснення.

Подальші дослідження спрямовані на проведення класифікації прихованих факторів з метою підвищити точність оцінювання можливості та необхідності. Крім того, доцільно інтегрувати у запропонований підхід графові нейронні мережі для оперативного виявлення нових темпоральних патернів для великих журналів подій.

8. Висновки

Запропоновано причинно-наслідкову модель пояснення в інтелектуальній інформаційній системі, яка на глобальному рівні містить можливісні причинно-наслідкові залежності між явними й неявними вхідними даними та рішенням системи, на процесному рівні містить можливісні причинно-наслідкові залежності між діями процесу формування рішення і отриманим рішенням, на локальному рівні містить детерміновані обмеження щодо допустимих послідовностей дій із процесу формування рішення. У практичному плані модель дає можливість в рамках пояснення визначити вхідні дані як причини отриманого рішення для зовнішніх користувачів та дії процесу формування рішення як причини цього рішення для внутрішніх користувачів з тим, щоб підвищити довіру користувачів до отриманих рішень.

Запропоновано метод побудови пояснень на основі причинно-наслідкової моделі, який містить етапи визначення користувачів, побудови пояснень на глобальному, локальному та процесному рівнях, а також формування комплексного пояснення, що дає можливість підвищити ефективність функціонування інтелектуальної інформаційної системи, а також ефективність використання її рішень при вирішенні практичних задач користувачів.

Перелік посилань:

1. 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>
2. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1-35.
3. 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>
4. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 1-42.
5. Chalyi, S., Leshchynskyi, V., & Leshchynska, I. (2019). Method of forming recommendations using temporal constraints in a situation of cyclic cold start of the recommender system. *EUREKA: Physics and Engineering*, 4, 34-40. <https://doi.org/10.21303/2461-4262.2019.00952>
6. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). «Why should I trust you?» Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.
7. D. Gunning i D. Aha, (2019) «DARPA's Explainable Artificial Intelligence (XAI) Program», *AI Magazine*, 40 (2), 44-58, <https://doi.org/10.1609/aimag.v40i2.2850>.
8. Arrieta, A. B., Díez-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
9. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 4765-4774.
10. 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>
11. Чалий С. Ф., Лещинський В. О., Лещинська І. О. (2021) Контрфактуальна темпоральна модель причинно-наслідкових зв'язків для побудови пояснень в інтелектуальних системах, / Вісник Національного технічного університету «ХПІ». Сер. : Системний аналіз, управління та інформаційні технології = Bulletin of the National Technical University «KhPI». Ser.: System analysis, control and information technology: зб. наук. пр. – Харків : НТУ «ХПІ», 2 (6), С. 41-46.
12. Chala O. (2018) Models of temporal dependencies for a probabilistic knowledge base. *Econtechmod. An International Quarterly Journal*. 7 (3), 53–58.
13. Levykin V., Chala O. (2018). Development of a method of probabilistic inference of sequences of business process activities to support business process management. *Eastern-European Journal of Enterprise Technologies*. 5/3(95), 16-24. <https://doi.org/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), 14-18. <https://doi.org/10.20998/2079-0023.2020.01.03>.
15. Chala O. (2018) Models of temporal dependencies for a probabilistic knowledge base. *Econtechmod. An International Quarterly Journal*. 7, 3, 53–58.
16. Pearl, J., & Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic books.
17. Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press..
18. Sloman, S. A. (2005). Causal models: How people think about the world and its alternatives. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195183115.001.0001>
19. Chalyi, S., & Leshchynskyi, V. (2020). Causal reasoning in the construction of explanations in intelligent systems. *Advanced Information Systems*, 4(4), 105-110. <https://doi.org/10.20998/2522-9052.2020.4.14>

Надійшла до редколегії 25.05.2025 р.

Чалий Сергій Федорович, доктор технічних наук, професор, професор кафедри ІУС ХНУРЕ, м. Харків, Україна, e-mail: serhii.chalyi@nure.ua; ORCID: <https://orcid.org/0000-0002-9982-9091>
Лещинський Володимир Олександрович, кандидат технічних наук, доцент, доцент кафедри програмної інженерії ХНУРЕ, м. Харків, Україна, e-mail: volodymyr.leshchynskyi@nure.ua; ORCID: <https://orcid.org/0000-0002-8690-5702>