



С. Ф. Чалий<sup>1</sup>, В. О. Лещинський<sup>2</sup>

<sup>1</sup>професор кафедри інформаційних управляючих систем,  
Харківський національний університет радіоелектроніки, Україна,  
serhii.chalyi@nure.ua

<sup>2</sup>доцент кафедри програмної інженерії,  
Харківський національний університет радіоелектроніки, Україна

## ПОБУДОВА ПОЯСНЕНЬ НА ЛОКАЛЬНОМУ РІВНІ ПРЕДСТАВЛЕННЯ ПРОЦЕСУ ФОРМУВАННЯ РІШЕНЬ ДЛЯ ВНУТРІШНІХ КОРИСТУВАЧІВ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ

Предметом вивчення в статті є процеси побудови пояснень щодо отриманих результатів та послідовності дій з прийняття рішення в інтелектуальній інформаційній системі.

Метою є розробка підходу до побудови пояснень на локальному рівні представлення процесу формування рішень інтелектуальної системи на основі побудови темпоральних та каузальних залежностей, що створює умови для виявлення «вузьких місць» даного процесу при представленні інтелектуальної системи як «сірої» скриньки. Завдання: розробка моделі пояснення на локальному рівні представлення для внутрішніх користувачів інтелектуальної системи; розробка методу побудови пояснень на локальному рівні представлення процесу формування рішень інтелектуальної системи. Використовуваними підходами є: методи побудови пояснень, підходи до побудови темпоральних знань. Висновки. Наукова новизна отриманих результатів полягає в наступному. Розроблено модель пояснення на локальному рівні представлення процесу формування рішення для внутрішніх користувачів інтелектуальної інформаційної системи. Модель використовує поєднання темпоральних залежностей, що визначають лінійну послідовність дій процесу, та каузальних правил-обмежень, які задають обов'язкові дії, що виконуються для всіх варіантів процесу формування рішення. Запропонована модель створює умови для виявлення дій, які обмежують нормальне виконання процесу формування рішення та призводять до результатів, що не відповідають вимогам зовнішніх користувачів. Розроблено метод побудови пояснень на локальному рівні представлення процесу формування рішень інтелектуальної системи. Метод передбачає виділення темпоральних залежностей між парами послідовних станів поточного процесу формування рішення, відбір темпоральних залежностей, які виконуються для всіх варіантів процесу побудови рішення та формування на їх основі каузальних залежностей. Використання методу дає можливість представити як пояснення обмеження щодо формування рішення, що створює умови для виявлення вузьких місць у послідовності дій з отримання результату в інтелектуальній системі.

ІНТЕЛЕКТУАЛЬНА ІНФОРМАЦІЙНА СИСТЕМА, СИСТЕМА ШТУЧНОГО ІНТЕЛЕКТУ, ПОЯСНЕННЯ, ПРОЦЕС ПРИЙНЯТТЯ РІШЕНЬ, ПРОЦЕСНИЙ ПІДХІД ПРИЧИННО-НАСЛІДКОВИЙ ЗВ'ЯЗОК, МОЖЛИВІСТЬ, КАУЗАЛЬНІСТЬ

**S. Chalyi, V. Leshchynskyi. Construction of explanations at the local level of decision-making process representation for internal users of the intelligent system.** The subject of study in the article is the processes of constructing explanations for the obtained results and the sequence of decision-making actions in an intelligent information system. The goal is to develop an approach to constructing explanations at the local level of representation of the decision-making process of an intelligent system based on building temporal and causal dependencies, which creates conditions for identifying "bottlenecks" of this process when representing the intelligent system as a "gray" box. Tasks: development of an explanation model at the local level of representation for internal users of an intelligent system; development of a method for constructing explanations at the local level of representation of the decision-making process of an intelligent system. The approaches used are: methods of constructing explanations, approaches to constructing temporal knowledge. Conclusions. The scientific novelty of the obtained results is as follows. An explanation model at the local level of representation of the decision-making process for internal users of an intelligent information system has been developed. The model uses a combination of temporal dependencies that determine the linear sequence of process actions and causal rules-constraints that specify mandatory actions performed for all variants of the decision-making process. The proposed model creates conditions for identifying actions that limit the normal execution of the decision-making process and lead to results that do not meet the requirements of external users. A method for constructing explanations at the local level of representation of the decision-making process of an intelligent system has been developed. The method involves identifying temporal dependencies between pairs of consecutive states of the current decision-making process, selecting temporal dependencies that are performed for all variants of the decision-making process, and forming causal dependencies based on them. Using the method makes it possible to present decision-making constraints as an explanation, which creates conditions for identifying bottlenecks in the sequence of actions for obtaining a result in an intelligent system.

INTELLIGENT INFORMATION SYSTEM, ARTIFICIAL INTELLIGENCE SYSTEM, EXPLANATION, DECISION-MAKING PROCESS, PROCESS APPROACH, CAUSAL RELATIONSHIP, POSSIBILITY, CAUSALITY

## Вступ

Сучасні інтелектуальні інформаційні системи (ІС) використовуються для вирішення складних задач у сферах медицини, фінансів, промисловості тощо. ІС орієнтовані на обробку великих обсягів даних, виявлення закономірностей в цих даних та подальше прийняття рішень на основі алгоритмів машинного навчання. Однак, внаслідок використання таких алгоритмів, сучасні інтелектуальні системи, як правило, є «непрозорими» з точки зору користувачів. Тобто внутрішня логіка та процес формування рішень можуть бути незрозумілими для користувачів [1, 2].

Непрозорість функціонування ІС може призвести до недовіри з боку користувачів, особливо у критичних сферах, де рішення системи можуть мати значний вплив на життя та здоров'я людей. Крім того, відсутність розуміння того, як система прийшла до певного рішення, ускладнює процес виявлення та виправлення потенційних помилок у процесі формування рішень в ІС [3, 4].

Для вирішення проблеми непрозорості ІС та підвищення довіри до результатів роботи інтелектуальних систем доцільно надавати користувачам зрозумілі та інформативні пояснення щодо процесу прийняття рішень [5, 6]. Ці пояснення мають відповідати потребам як внутрішніх, так і зовнішніх користувачів.

Зовнішні користувачі є кінцевими споживачами рішень інтелектуальної системи. Вони зазвичай не володіють суттєвими знаннями про її внутрішню структуру та алгоритми роботи. Такі користувачі мають отримати зрозуміле обґрунтування рішень ІС на концептуальному рівні.

Внутрішні користувачі, зокрема розробники, аналітики, інженери з аналізу даних та експерти з предметної області, забезпечують конфігурування та налагодження роботи інтелектуальної системи й відповідають за її розробку, налаштування та підтримку. Для цих користувачів важливо в рамках пояснення отримати опис ключових послідовностей дій та ключових залежностей процесу формування рішень, щоб ефективно виявляти та усувати потенційні «вузькі місця», оптимізувати роботу системи та забезпечувати її відповідність вимогам кінцевого користувача. Зазначене свідчить про актуальність використання процесного підходу до побудови пояснень.

Процесний підхід базується на ідеї представлення процесу прийняття рішень у вигляді послідовності взаємопов'язаних дій та станів, що дозволяє відслідковувати логіку та причинно-наслідкові зв'язки в роботі інтелектуальної системи [7, 8]. Процесний підхід широко використовується в управлінні бізнес-процесами для опису, аналізу та оптимізації складних процесів в організаціях [9, 10].

Застосування процесного підходу до побудови пояснень в інтелектуальних системах створює умови для

побудови структурованого опису процесу прийняття рішень для внутрішніх користувачів. Маючи такі пояснення, розробники та аналітики можуть краще зрозуміти поведінку системи, виявити потенційні помилки та вдосконалити алгоритми прийняття рішень [11].

Дослідження у сфері пояснювального штучного інтелекту (ХАІ) активно розвиваються протягом останніх років. В рамках даного напрямку вирішуються задачі розробки методів побудови пояснень, а також створення ментальних моделей користувачів. Перша задача орієнтована на забезпечення прозорості роботи інтелектуальних систем. Друга задача орієнтована на підтримку розуміння пояснень користувачами [5, 12, 13].

При побудові пояснень суттєва увага приділяється розробці методів, які дозволяють генерувати пояснення на основі аналізу внутрішньої структури та процесу прийняття рішень в інтелектуальних системах. Такі методи, як LIME [14], SHAP [15] та Grad-CAM [16], дозволяють визначити важливість окремих ознак або компонентів моделі для прийняття конкретного рішення.

Інший напрямок досліджень пов'язаний побудовою пояснень з використанням темпоральних та каузальних залежностей. Каузальні моделі дозволяють описати причинно-наслідкові зв'язки між змінними та подіями в процесі прийняття рішень [17, 18]. Використання каузальних залежностей дозволяє створювати більш інтерпретовані та зрозумілі пояснення, які відображають логіку роботи інтелектуальної системи [19, 20]. В роботах [21–23] запропоновано підхід до побудови каузальних моделей пояснень для інтелектуальних систем, який дозволяє враховувати темпоральні та причинно-наслідкові зв'язки між подіями та станами системи.

В роботах [24, 25] запропоновано підхід до побудови пояснень з урахуванням темпорального аспекту.

В роботах [26, 27] запропоновано підходи до опису темпоральних залежностей між подіями та станами в процесі прийняття рішень, а також підхід до побудови процесу прийняття рішень на основі темпоральних залежностей [28]. Ці залежності дозволяють відобразити динаміку зміни станів системи та причинно-наслідкові зв'язки між ними.

Однак в наведених дослідженнях розглядаються як пояснення окремі ключові елементи процесу формування рішень в інтелектуальних системах. Представленню пояснення для внутрішніх користувачів шляхом опису процесу на основі темпоральних та каузальних залежностей не приділяється достатньо уваги, що свідчить про важливість виділення рівнів процесного представлення пояснення для внутрішніх користувачів та розробки відповідних методів побудови пояснень в інтелектуальних системах в темпоральному та каузальному аспектах. Темпоральні

залежності дозволяють описати послідовність станів та подій в процесі прийняття рішень, тоді як каузальні залежності відображають причинно-наслідкові зв'язки між цими станами та подіями. Поєднання цих двох типів залежностей дає можливість створювати детальний опис процесу формування пояснень, які відображають логіку роботи інтелектуальної системи для внутрішніх користувачів.

### 1. Постановка задачі

Метою статті є розробка підходу до побудови пояснень на локальному рівні представлення процесу формування рішень інтелектуальної системи на основі побудови темпоральних та каузальних залежностей, що створює умови для виявлення «вузьких місць» даного процесу при представленні інтелектуальної системи як «сірої» скриньки.

Для досягнення поставленої мети вирішуються такі задачі:

- розробка моделі пояснення на локальному рівні представлення для внутрішніх користувачів інтелектуальної системи;
- розробка методу побудови пояснень на локальному рівні представлення процесу формування рішень інтелектуальної системи.

### 2. Модель пояснення на локальному рівні представлення процесу формування рішення

Загальну схему представлення пояснень для зовнішніх та внутрішніх користувачів наведено на рис. 1. Пояснення містить локальний, процесний та глобальний рівні. Перші два рівні відображають процес формування рішення з різним ступенем деталізації та орієнтовані на внутрішнього користувача. Дані два рівня відображають процедурні знання і тому явно враховують темпоральний аспект. Третій рівень орієнтований на зовнішнього користувача та відображає причини отриманого в інтелектуальній системі результату. На даному рівні темпоральний аспект враховується неявно, оскільки залежність між вхідними даними та отриманим результатом передбачає, що був реалізований процес формування рішення. Глобальний рівень представляє процес формування рішення в інтелектуальній інформаційній системі як одну глобальну дію, відображаючи базові зв'язки між вхідними даними та результатом роботи системи.

Пояснення на локальному рівні представлення процесу формування рішення в ІС базується на використанні окремих темпоральних та каузальних залежностей між парами станів процесу або парами подій, що відображають виникнення відповідних станів.

Темпоральні залежності  $f_{n-1,n}^{(i)}$  визначають пари упорядкованих у часі послідовних станів (подій виникнення станів)  $e_{n-1}^{(i)}$  та  $e_n^{(i)}$   $i$ -реалізації процесу

формування рішення. Перехід між цими станами є наслідком виконання дії  $d_{n-1}^{(i)}$ :

$$f_{n-1,n}^{(i)} : e_{n-1}^{(i)} \xrightarrow{d_{n-1}^{(i)}} e_n^{(i)}. \tag{1}$$

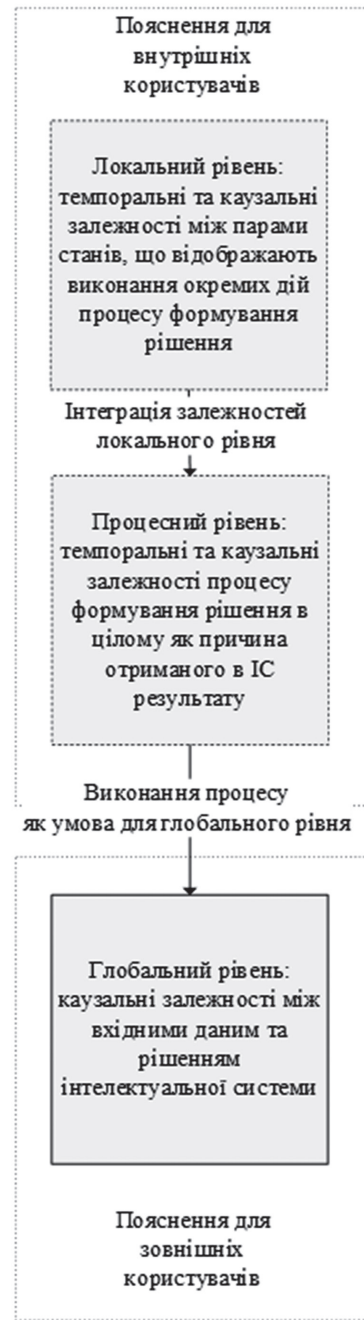


Рис. 1. Ієрархічне представлення пояснення для внутрішніх та зовнішніх користувачів

Тобто поточний процес формування рішення в ІС відображається в поясненні у вигляді лінійної послідовності упорядкованих у часі подій, причому дана лінійна послідовність однозначно задається темпоральними залежностями між парами послідовних подій.

Загальну схему локального рівня процесного опису пояснення для  $i$ -реалізації процесу формування рішення в інтелектуальній системі представлено на рис. 2.

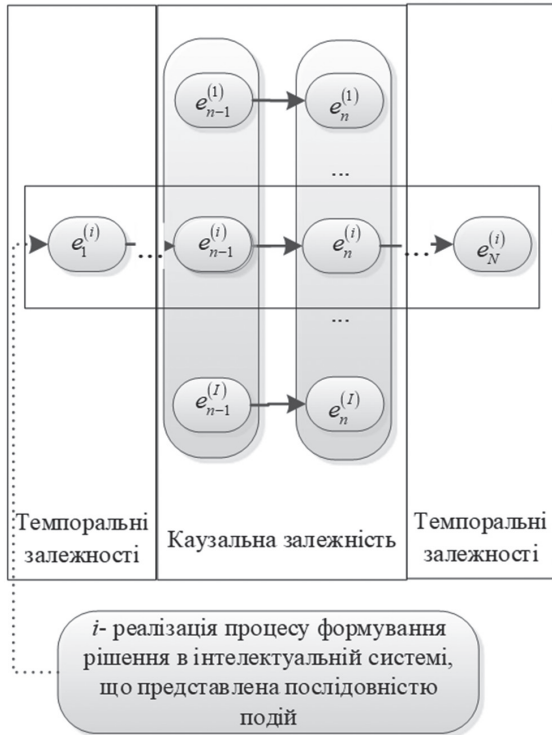


Рис. 2. Локальний рівень процесного опису пояснення

Дана схема містить послідовність подій виникнення станів  $e_1^{(i)}, \dots, e_n^{(i)}, \dots$  для даної реалізації процесу, а також темпоральні та каузальні залежності між парами станів.

Темпоральні залежності мають, наприклад, вигляд  $d_1^{(i)} \rightarrow e_2^{(i)}$  для першої дії процесу формування рішення  $d_1^{(i)}$ , або  $e_{l-1}^{(i)} \rightarrow e_l^{(i)}$  для останньої дії даного процесу.

В тому випадку, якщо одна й та ж сама дія  $d_{n-1}^{(i)}$  в одному й тому ж стані  $e_{n-1}^{(i)}$  виконується для всіх можливих варіантів реалізації процесу прийняття рішення, то відповідне темпоральне правило розглядається як каузальна залежність:

$$e_{n-1}^{(i)} \Rightarrow e_n^{(i)} \mid (\forall i) \exists d_{n-1}^{(i)}. \quad (2)$$

Згідно (2), темпоральна залежність  $e_{l-1}^{(i)} \rightarrow e_l^{(i)}$  не має альтернативи для всіх відомих варіантів процесу формування рішення в ІС. Тому вираз (2) для каузальної залежності приймає вигляд каузального обмеження:

$$c_n^{(n-1)} : e_{n-1} \Rightarrow e_n. \quad (3)$$

Оскільки дане обмеження не залежить від конкретної реалізації процесу формування рішення, то його можна розглядати як детерміноване правило-обмеження.

Подія  $e_n$  фіксує завершення дії  $d_{n-1}$ , тому дану дію можна розглядати як один із атрибутів події  $e_n$ . Сукупність атрибутів даної події фіксує не лише

завершення  $d_{n-1}$ , а і властивості стану після завершення цієї дії.

Тому каузальне обмеження доцільно представити у більш загальному вигляді:

$$c_n^{(n-1)} : e_{n-1} \Rightarrow e_n. \quad (4)$$

Вираз (4) показує, що умовою події  $e_n$  є наявність попередньої події  $e_{n-1}$ , а причина визначається через атрибути цих подій, які були змінені при переході від  $e_{n-1}$  до  $e_n$ .

У випадку визначення підмножини атрибутів  $A$  вираз (4) приймає вигляд:

$$c_n^{(n-1)} : e_{n-1} \stackrel{A}{\Rightarrow} e_n. \quad (5)$$

Ключовим атрибутом зазвичай є дія, тому обмеження (4) визначає обов'язковість виконання дії  $d_{n-1}$  у стані  $e_{n-1}$ .

Відповідно, представлення цього правила як елемента пояснення на локальному рівні дає можливість внутрішньому користувачеві побачити всі обов'язкові дії процесу формування рішення і обов'язкові зміни станів, що є наслідком цих дій. У випадку, якщо такі дії накладають додаткові (необов'язкові) обмеження на процес формування рішення, то це може свідчити про «вузькі місця» у даному процесі.

Коригування наведених обмежень дає можливість усунути ці проблемні місця.

### 3. Метод побудови пояснень на локальному рівні представлення процесу формування рішень інтелектуальної інформаційної системи

На локальному рівні пояснення формується на основ каузального обмеження, що відображає допустимі зміни атрибутів пара послідовних станів процесу формування рішення в інтелектуальній системі. Такі пояснення фактично визначають детерміновані правила побудови рішення ІС.

Метод побудови пояснень на локальному рівні містить такі етапи.

Етап 1. Виділення темпоральних залежностей між парами послідовних станів поточного процесу формування рішення згідно (1).

Результатом даного етапу є множина  $F^{(i)}$  пар подій  $e_{n-1}^{(i)}, e_n^{(i)}$  щодо виникнення станів процесу, який відображає послідовність виконання  $i$ -процесу побудови рішення у часі:

$$F^{(i)} = \left\{ e_{n-1}^{(i)} \xrightarrow{d_{l-1}^{(i)}} e_n^{(i)} \right\}. \quad (6)$$

Етап 2. Відбір підмножини  $A$  атрибутів подій виникнення станів поточної  $i$ -реалізації процесу, що відображають причини переходу між станами.

Кожен стан процесу формування рішення характеризується множиною значень змінних, які

відображають властивості сутностей предметної області. Ці властивості можуть бути статичними або динамічними. Статичні властивості характеризують об'єкт в цілому. Динамічні властивості змінюють свої значення з часом, по мірі виконання процесу формування рішення. Такі властивості дають можливість відрізнити поточний стан сутності від інших її станів. Тому на другому етапі методу відбираються динамічні атрибути станів тобто суттєві властивості, зміна яких відображає перехід між станами процесу. Відбір атрибутів базується на вхідній інформації щодо групування вхідних даних по сутностям предметної області. Фактично, відібрані атрибути мають бути найбільш інформативними для пояснення переходу між станами.

Етап 3. Виділення темпоральних залежностей, які виконуються для всіх варіантів процесу формування рішення в інтелектуальній системі.

На даному етапі перевіряється умова (2) виконання темпоральних залежностей з урахуванням відібраних атрибутів станів для всіх відомих варіантів процесу формування рішення. Це дозволяє виявити детерміновані послідовності зміни станів процесу формування рішення.

Етап 4. Формування каузальних залежностей.

На даному етапі формуються темпоральні обмеження згідно (4) або (5).

Результатом методу є упорядкована за послідовністю дій процесу множини каузальних обмежень, що представляють собою пояснення локального рівня представлення процесу формування рішення в ПС.

Розглянемо приклад використання розробленого методу побудови пояснення на основі побудови каузальних обмежень при формуванні персоналізованого переліку товарів та послуг в рекомендаційній системі.

Вхідними даними є лог процесу формування рекомендацій в онлайн-кінотеатрі. Лог складається з трас. Кожна траса складається із послідовності подій, та є записом про одну реалізацію процесу побудови рекомендацій. Події траси упорядковані за часом виникнення. Атрибути подій містять інформацію про дії з формування персоналізованого переліку фільмів, що відображають

Розглянемо випадок, коли рекомендаційна система надає користувачеві рекомендації щодо фільмів на основі його вподобань, що відображені в історії перегляду.

Процес формування рекомендацій включає наступний узагальнений перелік дій  $D^{(1)} = \langle d_1^{(1)}, \dots, d_7^{(1)} \rangle$ :

– збір даних про користувача  $(d_1^{(1)})$ : вік, стать, місце проживання тощо.

– аналіз історії перегляду користувача  $(d_2^{(1)})$ : фільми, що переглядалися, оцінки, що були виставлені.

– визначення вподобань користувача  $(d_3^{(1)})$ : жанри, актори, режисери тощо.

– пошук схожих користувачів  $(d_4^{(1)})$ : виконується на основі вподобань та історії перегляду.

– формування списку рекомендованих фільмів  $(d_5^{(1)})$ : виконується на основі вподобань схожих користувачів.

– ранжування рекомендованих фільмів  $(d_6^{(1)})$ : виконується за відповідністю до вподобань користувача.

– надання користувачеві топ-N рекомендацій  $(d_7^{(1)})$ .

Кожна з цих дій відповідає певній події щодо виникнення стану процесу формування рекомендацій. Позначимо ці події як  $e_1^{(1)}, e_2^{(1)}, \dots, e_7^{(1)}$ .

Реалізація методу із наведеними вхідними даними має такий вигляд.

Етап 1. Виділення темпоральних залежностей між парами послідовних станів.

Виділяємо наступні темпоральні залежності між парами послідовних у часі подій:

$$e_1^{(1)} \rightarrow e_2^{(1)}, e_2^{(1)} \rightarrow e_3^{(1)}, e_3^{(1)} \rightarrow e_4^{(1)}, \\ e_4^{(1)} \rightarrow e_5^{(1)}, e_5^{(1)} \rightarrow e_6^{(1)}, e_6^{(1)} \rightarrow e_7^{(1)}.$$

Етап 2. Виділення дій, які привели до переходів між станами.

Для кожної темпоральної залежності визначаємо відповідну дію:

$$e_1^{(1)} \xrightarrow{d_2^{(1)}} e_2^{(1)} \text{ – аналіз історії перегляду користувача;} \\ e_2^{(1)} \xrightarrow{d_3^{(1)}} e_3^{(1)} \text{ – визначення вподобань користувача;} \\ e_3^{(1)} \xrightarrow{d_4^{(1)}} e_4^{(1)} \text{ – пошук схожих користувачів;} \\ e_4^{(1)} \xrightarrow{d_5^{(1)}} e_5^{(1)} \text{ – формування списку рекомендованих фільмів;} \\ e_5^{(1)} \xrightarrow{d_6^{(1)}} e_6^{(1)} \text{ – ранжування рекомендованих фільмів;} \\ e_6^{(1)} \xrightarrow{d_7^{(1)}} e_7^{(1)} \text{ – надання користувачеві топ-N рекомендацій.}$$

Етап 3. Узагальнення темпоральних залежностей та відповідних дій.

На даному етапі виконується аналіз та порівняння з  $D^{(1)}$  схожих трас побудови рекомендацій  $D = \{D^{(i)} : i \neq 1\}$ .

Якщо в результаті аналізу трас логу рекомендаційної системи виявлено, що послідовність дій  $d_3$  – визначення вподобань користувача,  $d_4$  – пошук схожих користувачів,  $d_5$  – формування списку рекомендацій присутня в усіх реалізаціях процесу формування рекомендацій, то на базі вказаних дій можуть бути сформовані обмеження.

Тобто для відповідних темпоральних залежностей виконується умова:

$$(\forall i) \exists d_n^{(i)} \equiv d_n \mid 2 < n < 6. \quad (7)$$

Відповідні темпоральні залежності є однаковими на всіх трасах процесу та мають вигляд:

$$e_2 \xrightarrow{d_3} e_3, e_3 \xrightarrow{d_4} e_4, e_4 \xrightarrow{d_5} e_5. \quad (8)$$

Етап 4. Формування каузальних обмежень.

На основі виявлених на попередньому етапі темпоральних залежностей та дій формуються наступні каузальні обмеження:

$e_2 \xrightarrow{d_3} e_3$ : визначення вподобань користувача  $d_3$  є причиною переходу від події  $e_2$  (проаналізована історія перегляду) до події  $e_3$  (визначені вподобання);

$e_3 \xrightarrow{d_4} e_4$ : Пошук схожих користувачів  $d_4$  є причиною переходу від події  $e_3$  (визначені вподобання) до події  $e_4$  (знайдені схожі користувачі).

$e_4 \xrightarrow{d_5} e_5$ : Формування списку рекомендованих фільмів  $d_5$  є причиною переходу від стану  $e_4$  (знайдені схожі користувачі) до стану  $e_5$  (сформовано список рекомендацій).

Сформовані таким чином каузальні обмеження є поясненнями на локальному рівні представлення. Ці пояснення показують, що визначення вподобань користувача, пошук схожих користувачів та формування списку рекомендацій на їх основі є ключовими діями, які впливають на кінцевий результат — надані користувачеві рекомендації.

Таке пояснення дає можливість команді підтримки деталізувати логіку роботи системи визначити обов'язкові дії, які вплинули на рекомендацію щодо фільмів цільовому користувачеві. Якщо рекомендації виявилися неточними, то на основі пояснення можна зробити висновок, що причиною цього могли бути некоректно визначені вподобання користувача або невдало підібрані схожі користувачі. Це дозволить зробити відповідні корективи в алгоритми рекомендаційної системи для покращення якості рекомендацій.

### Висновки

Представлено багаторівневе представлення пояснення щодо процесу формування рішення в інтелектуальній системі, яке орієнтовано на потреби внутрішніх та зовнішніх користувачів. Пояснення для внутрішніх користувачів базується на формуванні темпоральних та каузальних залежностей, що відображають процес формування рішення. Пояснення для зовнішніх користувачів відображає залежність між вхідними даними та отриманим рішенням.

Розроблено модель пояснення на локальному рівні представлення процесу формування рішення

для внутрішніх користувачів інтелектуальної інформаційної системи. Модель містить сукупність темпоральних залежностей та каузальних правил-обмежень. Темпоральні залежності визначають лінійну послідовність дій процесу, а каузальні обмеження визначають дії, які виконуються для всіх варіантів процесу формування рішення. Модель дає можливість виявити такі дії, які обмежують нормальне виконання процесу формування рішення, що приводить до результатів, які не відповідають вимогам зовнішніх користувачів.

Розроблено метод побудови пояснень на локальному рівні представлення процесу формування рішення інтелектуальної системи. Метод містить етапи виділення темпоральних залежностей між парами послідовних станів поточного процесу формування рішення, виділення темпоральних залежностей, які виконуються для всіх варіантів процесу побудови рішення, формування каузальних залежностей. Метод дає можливість представити як пояснення обмеження щодо формування рішення, що створює умови для виявлення вузьких місць у послідовності дій з отримання результату у інтелектуальній системі.

### Список літератури:

- [1] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160.
- [2] 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 (CSUR)*, 51(5), 1-42.
- [3] Baier, L., Kühl, N., & Satzger, G. (2019, July). How to cope with change? Preserving validity of predictive services over time. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- [4] Biran, O., & Cotton, C. (2017, August). Explanation and justification in machine learning: A survey. In *IJCAI-17 workshop on explainable AI (XAI) (Vol. 8, No. 1, pp. 8-13)*.
- [5] Чалий, С., & Лещинська, І. Концептуальна ментальна модель пояснення в системі штучного інтелекту // *Вісник Національного технічного університету «ХПІ»*. Серія: Системний аналіз, управління та інформаційні технології — 2023. — Vol. 1(9). — P. 70–75 <https://doi.org/10.20998/2079-0023.2023.01.11>
- [6] Tintarev, N., & Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5), 399-439. <https://doi.org/10.1007/s11257-011-9117-5>
- [7] Van Der Aalst, W. (2016). *Process mining: data science in action*. Springer.
- [8] Rehse, J. R., Mehdiyev, N., & Fettke, P. (2019). Towards explainable process predictions for industry 4.0 in the dfki-smart-lego-factory. *KI-Künstliche Intelligenz*, 33(2), 181-187.

- [9] Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of business process management* (Vol. 1, p. 2). Heidelberg: Springer.
- [10] Weske, M. (2019). *Business process management: concepts, languages, architectures*. Springer.
- [11] Mehdiyev, N., Evermann, J., & Fettke, P. (2020). A novel business process prediction model using a deep learning method. *Business & information systems engineering*, 62(2), 143-157.
- [12] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2019). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
- [13] Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*.
- [14] 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. <https://doi.org/10.1145/2939672.2939778>SHAP:
- [15] Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774. Grad-CAM:
- [16] 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. *Proceedings of the IEEE International Conference on Computer Vision*, 618-626. <https://doi.org/10.1109/ICCV.2017.74>
- [17] Pearl, J. (2009). *Causality*. Cambridge university press.
- [18] Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.
- [19] Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4), e1312.
- [20] Moraffah, R., Karami, M., Guo, R., Raglin, A., & Liu, H. (2020). Causal interpretability for machine learning-problems, methods and evaluation. *ACM SIGKDD Explorations Newsletter*, 22(1), 18-33.
- [21] Chalyi, S., & Leshchynskiy, V. (2023). Probabilistic counterfactual causal model for a single input variable in explainability task. *Advanced Information Systems*, 7(3), 54-59.
- [22] Чалий, С. Ф., & Лещинський, В. О. (2021). Можливісна модель каузального зв'язку по вхідній змінній для побудови пояснення в інтелектуальній системі. *Вісник Національного технічного університету "ХПІ". Серія: Системний аналіз, управління та інформаційні технології*, (2), 97-103. <https://doi.org/10.20998/2079-0023.2021.02.17>
- [23] 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. <https://doi.org/10.21303/2461-4262.2020.001228>
- [24] Chalyi, S., & Leshchynskiy, V. (2020). Method of constructing explanations for recommender systems based on the temporal dimension of user choice. *EUREKA: Physics and Engineering*, 3, 43-50.
- [25] 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*, (2(6)), 41-46.
- [26] Chala O. (2018) Models of temporal dependencies for a probabilistic knowledge base. *Econtechmod. An International Quarterly Journal*. Vol. 7, No. 3. P. 53 – 58.
- [27] 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.
- [28] 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.

*Надійшла до редколегії 16.05.2024*