УДК 004.8:004.9 DOI: 10.30837/0135-1710.2024.180.004

С.Ф. ЧАЛИЙ, В.О. ЛЕЩИНСЬКИЙ

ПОБУДОВА ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ НА ОСНОВІ ФОРМУВАННЯ КАУЗАЛЬНИХ ЗАЛЕЖНОСТЕЙ

Розглянуто задачі побудови пояснень в інтелектуальних інформаційних системах, представлених у вигляді «чорного ящику». Розроблено узагальнену каузальну модель пояснення, яка об'єднує темпоральну, причинно-наслідкову та цільову складові. Модель забезпечує побудову багатоаспектного пояснення, що може бути використано не лише після реалізації рішення, а й до початку його імплементації. Запропоновано узагальнений каузальний метод формування пояснення на основі побудови та уточнення каузальних залежностей між вхідними даними та результатом інтелектуальної системи. Метод орієнтований на автоматизацію побудови та уточнення пояснень як для розробників, так і для кінцевих користувачів.

1. Вступ

Інтелектуальні інформаційні системи призначені для збору, аналізу та обробки інформації та знань при вирішенні задач підтримки прийняття рішень. Характерна особливість сучасних інтелектуальних систем полягає у здатності адаптуватися до змін зовнішнього середовища та вчитися на минулому досвіді. Такі системи використовують підходи, засновані на знаннях, а також машинне навчання при формуванні рішень. Вони дають можливість кінцевим користувачам приймати обґрунтовані рішення при вирішенні складних проблем у сферах охорони здоров'я, фінансів, маркетингу, освіти тощо [1]. Внаслідок використання машинного навчання на даних, що відображають практичні рішення актуальних задач в заданій предметній області, алгоритми прийняття рішень в інтелектуальних системах зазвичай є непрозорими і, отже, незрозумілими для користувачів. Відсутність інформації щодо логіки формування результату в інтелектуальних системах знижує довіру користувачів до цих рішень і, відповідно, не забезпечує умов для ефективного їх використання.

Для вирішення даної проблеми використовуються пояснення [2, 3]. Мета побудови пояснень полягає у представленні процесу прийняття рішення у зрозумілому для користувача вигляді шляхом визначення причинно-наслідкових залежностей, що призвели до конкретних рішень інтелектуальної системи. Пояснення відображають зв'язки між значеннями вхідних даних, поточних даних, що виникають у процесі прийняття рішення, а також отриманим результатом [4]. Завдяки поясненням користувачі отримують можливість оцінити дії, які призвели до певного результату, та критично оцінити рішення інтелектуальної системи [5]. Ефективні пояснення мають виділяти ключові причинно-наслідкові зв'язки, з можливостями деталізації в залежності від кваліфікації користувача. Такий підхід дає можливість зробити пояснення зрозумілим та інтерактивно його адаптувати. Зазначене свідчить про актуальність проблеми побудови пояснень в інтелектуальних системах з можливостями виділення ключових каузальних залежностей на різних рівнях представлення процесу прийняття рішення в інтелектуальній системі.

2. Аналіз літературних даних і постановка проблеми дослідження

Сучасні дослідження щодо побудови пояснювального штучного інтелекту [6-9] розглядають побудову систем, алгоритми роботи яких можуть бути безпосередньо

інтерпретовані [6]. Можливість такої інтерпретації забезпечує прозорість процесу прийняття рішення [7]. Однак такі пояснення в загальному вигляді не враховують рівень знань та культурні особливості користувачів. Тому для побудови зрозумілих пояснень ключовою умовою є індивідуалізація пояснень з можливістю їх інтерактивного уточнення [8]. Така індивідуалізація формується згідно з ментальною моделлю користувача [9]. Проте побудова інтерактивних пояснень потребує великих витрат ресурсів, оскільки останні мають використовувати деталізовані знання щодо предметної області для узгодження пояснень. Пояснення має також оновлюватись при коригуванні алгоритму роботи інтелектуальної системи [8], що потребує враховувати його темпоральний аспект при визначенні причин отриманих рішень. Виконання наведених вимог дає можливість перейти від забезпечення прозорості процесу прийняття рішень до його зрозумілості [10].

Темпоральні аспекти пояснення розглянуто в [4, 11]. В [4] представлено підхід до моделювання альтернативних причин рішення на основі темпоральної упорядкованості подій процесу прийняття рішення. В [11] обґрунтовано можливість формування каузальних залежностей на основі темпоральної упорядкованості подій процесу прийняття рішення при побудові пояснення в інтелектуальній системі. Темпоральна упорядкованість щодо процесу прийняття рішення задається в рамках відносного представлення часу [12], що дає можливість представити елементи процесу прийняття рішення у вигляді множини темпоральних правил [13] та в подальшому сформувати з цих правил можливі альтернативи такого процесу [14]. Представлений в [14] підхід обумовлює можливість побудови темпорально-каузального представлення процесу прийняття рішення в інтелектуальній системі як основи для пояснення щодо отриманих в такій системі результатів.

В розглянутих результатах досліджень визначено необхідні умови для побудови узагальненого опису пояснення, що містить каузальні залежності з різним ступенем деталізації для урахування рівня знань та потреб користувача. Однак побудові комплексної моделі пояснення, орієнтованої не лише на зрозуміле представлення пояснення, а й на його адаптацію згідно з потребами користувача, не приділялось достатньо уваги, що й свідчить про важливість задач, які вирішуються в даному дослідженні.

3. Мета і задачі дослідження

Метою даного дослідження є розробка каузального підходу до формування пояснень в інтелектуальних інформаційних системах, що створює умови для автоматизованого уточнення пояснень з тим, щоб зробити їх зрозумілими для користувачів з урахуванням їх цілей та потреб.

Для досягнення поставленої мети у статті вирішуються такі задачі:

– структуризація задач побудови пояснень в системах пояснювального штучного інтелекту;

– розробка каузальної моделі пояснення в інтелектуальній інформаційній системі;

– розробка методу побудови пояснень в інтелектуальній інформаційній системі з використанням каузальних залежностей.

4. Задачі побудови пояснень в системах пояснювального штучного інтелекту

Пояснювальний штучний інтелект орієнтований на підтримку розуміння користувачами процесу прийняття рішення в інтелектуальній інформаційній системі з урахуванням контексту застосування цих рішень. Узагальнено, пояснювальний штучний інтелект забезпечує зрозумілість рішень та процесу їх прийняття для всіх зацікавлених сторін, що використовують результати роботи інтелектуальної системи.

Пояснення має такі базові властивості, що забезпечують зрозумілість рішень системи штучного інтелекту:

– «прозорість» алгоритму роботи інтелектуальної системи, що зазвичай має вигляд «чорного ящика» для користувачів;

– зрозумілість для користувачів результату й процесу прийняття рішення в інтелектуальній інформаційній системі;

– відповідність пояснення цілям та потребам користувачів системи штучного інтелекту.

Згідно з наведеними властивостями, при побудові пояснення необхідно вирішити такі задачі:

– представлення у прозорому вигляді алгоритму прийняття рішень в інтелектуальній системі з тим, щоб були явно відображені темпоральні та/або причинно-наслідкові зв'язки між вхідними, проміжними даними, діями процесу, а також отриманим в системі результатом;

– представлення у зрозумілому для користувачів вигляді процесу прийняття рішення таким чином, щоб пояснення користувача відповідало його знанням щодо предметної області;

– узгодження пояснення із цілями та потребами користувачів з тим, щоб останні могли ефективно використовувати рішення інтелектуальної інформаційної системи для вирішення своїх практичних задач.

Непрозорість для користувачів систем штучного інтелекту, що мають вигляд «чорного ящику», пов’язана із використанням неявних знань у процесі прийняття рішень. Неявні знання є некодифікованими знаннями, тобто вони не відображені у доступній для людини формі.

Задача представлення у прозорому вигляді алгоритму прийняття рішень в інтелектуальній системі базується на екстерналізації знань щодо вказаного процесу. Екстерналізація полягає у перетворенні неявних знань у явну форму [15].

Першим кроком екстерналізації є ідентифікація знань, тобто виділення підмножини знань, які мають бути пояснені користувачеві. Другий крок екстерналізації полягає у кодифікації знань, тобто їх перетворенні в явну форму. Знання в явній формі можуть бути безпосередньо інтерпретовані людьми. За результатами цих двох кроків знання щодо процесу прийняття рішення можуть бути представлені цільовій аудиторії.

Таким чином, для вирішення задачі формування прозорого представлення алгоритму роботи інтелектуальної системи необхідно зробити таку систему інтерпретованою. В інтерпретованій системі опис процесу прийняття рішень є самопояснюваним, оскільки модель прийняття рішень містить явні, кодифіковані залежності. Для того, щоб система штучного інтелекту була інтерпретованою, алгоритм її роботи зазвичай представляють через інтеграцію зважених темпоральних або каузальних правил. Наприклад, використовуються дерева рішень, продукційні правила тощо. Однак вимога побудови інтерпретованих систем може привести до використання простіших моделей для прийняття рішень і, відповідно, до зниження ефективності імплементації отриманих результатів. Використання інтерпретованих пояснень дає можливість подолати цей недолік.

Інша ключова особливість інтерпретованих систем штучного інтелекту полягає в тому, що така система може бути зрозумілою лише для користувачів з певним рівнем підготовки та знань щодо процесів функціонування таких систем. Тому властивість інтерпретованості системи штучного інтелекту є важливою для налагодження алгоритмів їх роботи.

Цільовою аудиторією інтерпретованих систем штучного інтелекту є спеціалісти-розробники.

Задача побудови зрозумілого для кінцевого користувача опису процесу прийняття рішення полягає у формуванні множини каузальних залежностей, що визначають вплив вхідних та проміжних даних на отриманий в інтелектуальній системі результат.

Прозорість алгоритму роботи інтелектуальної системи для користувача є необхідною, але не є достатньою умовою для того, щоб користувач зрозумів процес та результат прийняття рішення в такій системі. Користувачі мають різний рівень підготовки і розуміння технологій та моделей, які використовує система штучного інтелекту.

Зрозуміле пояснення має містити ключові залежності в рамках прозорого алгоритму роботи інтелектуальної системи у відповідності до рівня знань користувача. Виділення ключових пояснень виконується з використанням показника чутливості [16]. Оскільки пояснення містить спрощену модель процесу прийняття рішень, чутливість визначає діапазон вхідних даних, для яких надаються однотипні або однакові пояснення. Використання даного показника дає можливість узагальнити пояснення для підмножин вхідних даних і таким чином виділити підмножину базових, найсуттєвіших пояснень.

Задача узгодження пояснення полягає у виборі такої підмножини причинно-наслідкових залежностей, яка відповідає вимогам користувачів.

Для вирішення цієї задачі можуть бути використані два підходи:

– орієнтований на особливості предметної області;

– орієнтований на безпосередню оцінку потреб користувачів.

Перший підхід може бути реалізований на основі онтології предметної області або з використанням векторного пошуку. Слід зазначити, що з урахуванням стрімкого розвитку великих та локальних мовних моделей використання векторного пошуку для узгодження знань має суттєві переваги. В даному випадку використовується база документів щодо запитів клієнтів, які відображають їхні цілі та вимоги. Запит перетворюється на вектор і за схожістю векторів виконується пошук відповіді (пояснення) у векторній базі даних.

Однак даний підхід має суттєвий недолік, пов'язаний зі значними витратами на створення онтологій та спеціалізованих векторних баз даних для кожної предметної області.

Другий підхід полягає у визначенні обмежень щодо відповідності пояснення цілям та потребам користувача. Такі обмеження можуть бути сформовані з використанням теорії можливостей [17] на основі показників можливості та необхідності для кожної причинно-наслідкової залежності, що входить до складу пояснення.

В цілому вирішення задач представлення за допомогою пояснення процесу прийняття рішення та його результатів у прозорому та зрозумілому вигляді у відповідності до цілей та потреб користувача інтелектуальної інформаційної системи забезпечує вирішення глобальнішої задачі підвищення довіри користувачів до рішень інтелектуальної системи.

5. Узагальнена каузальна модель пояснення в інтелектуальній інформаційній системі

Відповідно до розглянутих задач побудови пояснень, пропонується узагальнена модель пояснення, що визначає пояснення у таких аспектах: темпоральному, каузальному, цільовому.

Ключові особливості вказаних аспектів пояснення, які відображають інтеграцію темпорального опису процесу прийняття рішення та каузальних залежностей, що представляють причини прийняття отриманого рішення, наведено в табл. 1.

Пояснення у темпоральному аспекті представляється як упорядкований у часі спрощений опис процесу прийняття рішень в інтелектуальній інформаційній системі. На даному рівні

Таблиця 1

Характеристика складових узагальненої моделі пояснення

Рівень	Властивості
Темпоральний	– упорядкований у часі інтерпретований опис процесу прийняття рішень; – критерій оцінки: коректність пояснення.
Каузальний	– множина причинно-наслідкових залежностей, що відображають значення вхідних та проміжних даних як причин визначеного рішення інтелектуальної інформаційної системи; – критерій оцінки: чутливість пояснення.
Цільовий	– підмножина каузальних залежностей, що відображають найсуттєвіші для користувача причини рішення інтелектуальної системи; – критерій оцінки: складність пояснення.

пояснення P_t представляється множинами темпоральних та причинно-наслідкових залежностей, що описують послідовність дій з прийняття рішень $S = \langle S_0, S_1, ..., S_i, ..., S_I \rangle$.

Кожен $i-$ й елемент цієї послідовності характеризується набором значень змінних $v_{i,j}$, що відображають вхідні (S_0), проміжні (S_i) дані та результуюче рішення (S_I): $S_i = \{v_{i,j}\}$.

Між цими даними існують темпоральні залежності f_{j+n}^{i}, що задають попарну упорядкованість наборів даних $\langle S_i, ..., S_{i+n} \rangle$ у часі:

$$f_{i+n}^{i} : S_i \to S_{i+n}, 1 < n \le I - 1. \quad (1)$$

Згідно з (1), темпоральні залежності задають зв'язок як між послідовними у часі кроками з прийняття рішень $\langle S_i, S_{i+1} \rangle$, так і між віддаленими кроками, між якими є n проміжних кроків. Кількість проміжних кроків у загальному випадку залежить від рівня деталізації процесу прийняття рішення. На верхньому рівні, коли інтелектуальна система є повністю непрозорою для пояснення, $I = 1$ і ми отримуємо темпоральну залежність f_1^0 між вхідними даними та результуючим рішенням системи.

Кожна причинно-наслідкова залежність $r_{i+n}^{i,j}$ базується на відповідній темпоральній залежності f_{i+n}^{i} та має задовольняти критерію коректності. Оцінка коректності пояснення виконується на основі виявлення необхідності використання цільового значення кожної вхідної змінної з урахуванням вибору альтернативних значень змінних [18].

Порівняння альтернативних значень дає можливість визначити пояснення з урахуванням контрфактів [19]. Необхідність вибраного для пояснення значення змінної у поясненні має бути вище заданого порогу ε :

$$r_{j+n}^{i,j} : v_{i,j} \to S_{i+n} \left| N\left(v_{i,j}\right) > \varepsilon. \right. \quad (2)$$

Зокрема, перевищення порогу $\varepsilon > 0,5$ свідчить, що змінна зі значенням $v_{i,j}$ має більший вплив на результат вибраної дії процесу у порівнянні з альтернативними значеннями змінних. Тобто значення $v_{i,j}$ є необхідним для отримання результату, і тому пояснення на основі залежності $r_{j+n}^{i,j}$ є коректним.

Залежності для підмножини вхідних значень $v_{i,j}$ визначаються аналогічно виразу (2). Такі залежності із підмножиною $V_k = \{v_{i,j}\}, V_k \subseteq S_i$ значень вхідних змінних позначимо r_{j+n}^{i,V_k}.

Таким чином, модель пояснення у темпоральному аспекті складається з причинно-наслідкових залежностей $r_{i+n}^{i,j}$, кожна з яких базується на темпоральній залежності f_{j+n}^{i} та задовольняє обмеженню щодо необхідності вибору саме значення $v_{i,j}$ вхідної змінної як причини на поточному кроці результату:

$$P_t = \left\{ r_{i+n}^{i,j} : (\forall i \forall n) \exists f_{i+n}^{i}, N\left(v_{i,j}\right) > \varepsilon \right\}. \tag{3}$$

Пояснення у каузальному аспекті P_c представляється як сукупність причинно-наслідкових залежностей, що визначають вплив вхідних і проміжних даних на отриманий в інтелектуальний інформаційній системі результат. Ці залежності становлять підмножину залежностей (3), елементи якої задовольняють критерію чутливості. Чутливість дає можливість оцінити зміни пояснення при відхиленнях у вхідних даних. Ключова ідея використання даного критерія полягає в тому, щоб об'єднати однотипні пояснення для незначних змін у вхідних або проміжних даних. Як однотипні розглядаються такі залежності $r_{i+n}^{i,j}$ та $r_{i+n}^{i,l}$, які мають близькі можливості використання $\Pi\left(r_{i+n}^{i,j}\right)$ та $\Pi\left(r_{i+n}^{i,l}\right)$ для пояснення одного й того ж результату S_{i+n}, тобто :

$$\Pi\left(r_{i+n}^{i,j}\right) \approx \Pi\left(r_{i+n}^{i,l}\right). \tag{4}$$

Тому пояснення у каузальному аспекті містить підмножину правил $r_{i+n}^{i,j}$, для яких виконується умова $\Pi\left(r_{i+n}^{i,j}\right) \neq \Pi\left(r_{i+n}^{i,l}\right)$:

$$P_c = \left\{ r_{i+n}^{i,j} : (\forall j \forall l) \Pi\left(r_{i+n}^{i,j}\right) \neq \Pi\left(r_{i+n}^{i,l}\right) \right\}. \tag{5}$$

Такий підхід дає можливість упорядкувати пояснення щодо результату S_{i+n} за значенням можливості. В подальшому пояснення можна ітеративно уточнювати для представлення користувачеві, на кожній наступній ітерації представляючи йому пояснення із все меншим значенням можливості причинно-наслідкового зв'язку $r_{i+n}^{i,j}$ для обґрунтування поточного або фінального результату у процесі прийняття рішення в інтелектуальній інформаційній системі.

Пояснення у цільовому аспекті P_a представляється у вигляді упорядкованої підмножини

каузальних залежностей r_{j+n}^{i,V_k} з мінімальною складністю, тобто таких залежностей, які містять лише найсуттєвіші для формування рішення значення вхідних та проміжних змінних. Критерій мінімальної складності задається як кількість змінних, що є необхідними для формування пояснення, тобто V_k. Необхідність включення значень вхідних змінних у поясненні визначається через показник можливості використання пояснення $\Pi\left(r_{j+n}^{i,V_k}\right)$. Тобто несуттєве зниження можливості використання залежності $\Pi\left(r_{j+n}^{i,V_k}\right) \approx \Pi\left(r_{j+n}^{i,V_m}\right)$ у складі пояснення для підмножини V_k з видаленою змінною у порівнянні з аналогічними підмножинами V_m, що містять цю змінну, свідчить про те, що видалене значення змінної є несуттєвим для обґрунтування отриманого поточного або цільового результату.

Згідно з наведеним обґрунтуванням, формальне представлення пояснення у цільовому аспекті має вигляд:

$$P_a = \left\{ r_{j+n}^{i,V_k} : (\forall k) |V_k| = \min_m \left(|V_m| \right) \middle| \Pi\left(r_{j+n}^{i,V_k}\right) \approx \Pi\left(r_{j+n}^{i,V_m}\right) \right\}. \quad (6)$$

Використання можливостей $\Pi\left(r_{j+n}^{i,V_k}\right)$ для виділення найважливіших змінних у каузальній залежності для пояснення обумовлює відповідність пояснення задачам користувача.

Дійсно, можливість в теорії можливостей обумовлює найбільшу вірогідність виникнення певної події. Можливість розраховується на основі відомої інформації щодо використання рішення інтелектуальної системи, яке було сформовано на основі відомих вхідних даних.

Прикладом можуть бути вибір та покупка споживачем рекомендованого товару із заданими технічними характеристиками в системі електронної комерції. Висока вірогідність покупки товару свідчить про узгодженість його рекомендованих властивостей з потребами споживача.

Зазначена причина обумовлює високе значення можливості для залежності r_{j+n}^{i,V_k} бути включеною у пояснення для користувача.

Представлена узагальнена модель P пояснення охоплює всі три наведені аспекти і має вигляд:

$$P = \{ P_t, P_c, P_a : P_t \subset P_c, P_c \subseteq P_a \}. \quad (7)$$

Модель (7) у темпоральному аспекті призначена для представлення пояснень розробникам інтелектуальної системи з метою підвищити ефективність удосконалення алгоритму прийняття рішень в інтелектуальній системі.

Узагальнена модель пояснення у каузальному аспекті орієнтована на кваліфікованого користувача і призначена для підтримки адаптації процесу прийняття рішення згідно з вимогами користувача.

Модель у цільовому аспекті призначена для підтримки вибору рішення кінцевим користувачем.

6. Метод побудови пояснень в інтелектуальній інформаційній системі з використанням каузальних залежностей

6.1. Основні етапи методу

Запропонований узагальнений метод побудови пояснень з урахуванням темпорального, каузального та цільового аспектів процесу прийняття рішення враховує розглянуті особливості задач представлення даного процесу у прозорому та зрозумілому вигляді, а також задачі узгодження пояснення із потребами користувача.

Вхідна інформація для метода представлена такими множинами значень даних, що використовуються у процесі прийняття рішення:

– вхідні дані, що зазвичай є доступними для користувача;

– проміжні дані, що відображають виконання дій процесу прийняття рішення, або дій користувача; такі дані зазвичай є частково доступними;

– результуючі дані, що складають опис отриманого рішення.

Метод містить етапи побудови темпорального, каузального та цільового представлення пояснення:

Етап 1. Побудова темпорального представлення пояснення P_t.

Крок 1.1. Формування набору із темпоральних правил f_{i+n}^{i}, що описують упорядкованість у часі (1) даних щодо процесу прийняття рішення.

Крок 1.2. Розрахунок значення необхідності $N\left(v_{i,j}\right)$ для вхідних змінних отриманих темпоральних залежностей.

Крок 1.3. Відбір множини каузальних залежностей. Дана множина включає правила $r_{j+n}^{i,j}$, для яких значення необхідності перевищує заданий поріг.

Результатом даного етапу є множина правил $r_{j+n}^{i,j}$, що відображають ключові закономірності процесу прийняття рішення. Сукупність правил визначає спрощену модель даного процесу, зокрема, з тієї причини, що проміжні дані зазвичай є лише частково доступними.

Побудова моделі процесу прийняття рішення може бути виконана методом [14], оскільки каузальні правила $r_{j+n}^{i,j}$ були сформовані на основі відповідних темпоральних правил f_{i+n}^{i}.

Етап 2. Побудова каузального представлення пояснення P_c.

Крок 2.1. Розрахунок показника чутливості для елементів множини P_t.

Крок 2.2. Відбір підмножин каузальних залежностей, для яких виконується умова (4).

Крок 2.3. Формування множини P_c шляхом об'єднання залежностей, для яких виконується умова (4).

Результатом даного етапу є множина каузальних правил, які відображують суттєві причини отриманого рішення згідно з критерієм чутливості. Іншими словами, при використанні отриманих на поточному етапі правил значення пояснення буде змінено лише при суттєвих (можливісних) змінах у вхідних даних.

Етап 3. Формування цільового представлення пояснення P_a.

Крок 3.1. Розрахунок складності пояснень за показником $\left|V_k\right|$.

Крок 3.2. Відбір підмножини значень вхідних змінних згідно з (6) із ітеративним уточненням показника складності.

Крок 3.3. Формування підмножини P_a із каузальних залежностей, для яких виконано

умову (6).

Крок 3.4. Упорядкування підмножини P_a за значенням можливості каузальних правил.

Особливість даного етапу полягає в тому, що узгодження пояснення виконується без урахування специфіки предметної області.

Розроблений метод дає можливість сформувати пояснення у відповідності до потреб користувачів у заданій предметній області з урахуванням наявної інформації про вибір цих користувачів у минулому.

6.2. Приклад використання методу

Розглянемо приклад застосування методу побудови пояснень для рекомендаційної системи. Як вхідні дані виступає інформація про вибір користувачів, що є схожими за інтересами із цільовим користувачем. Ці користувачі вибирали товари певної групи, наприклад, комп'ютери з певними технічними характеристиками процесора, пам'яті, екрану, жорсткого диску тощо. Рекомендаційна система на основі інформації про вибір схожих користувачів запропонує цільовому користувачеві комп'ютери із аналогічними характеристиками. Як пояснення користувач має отримати інформацію про те, які характеристики комп'ютера були ключовими при формуванні рекомендацій. В даному прикладі інтелектуальна система має вигляд «чорного ящика», тобто проміжні дані відсутні.

На першому етапі методу формуються необхідні каузальні залежності виду «модель процесора – марка комп'ютера», «об'єм пам'яті – марка комп'ютера» тощо. Тобто представлені в цих залежностях значення вхідних змінних (конкретна модель процесора, конкретне значення об'єму пам'яті) є необхідними при побудові рекомендацій щодо вибору комп'ютера. На другому етапі із множини необхідних каузальних залежностей відбираються унікальні правила. Розглянемо випадок, коли значення об'єму пам'яті та об'єму жорстокого диску, які розглядаються як причина для рекомендації комп'ютера, мають близьке значення показника можливості. В такому випадку відбирається одне зі значень цих змінних. Відбір може бути виконано на основі порівняння значень можливості, а також з урахуванням додаткових критеріїв, зокрема, критерію складності або ж специфічних для предметної області показників. Результатом етапу є множина правил, що визначають вірогідні й необхідні причини рекомендації певної моделі комп'ютера в рекомендаційній системі. На третьому етапі методу порівнюється складність правил. Якщо видалення однієї з причин (одного зі значень змінної) для правила не зменшує можливість вибору пояснення, то вказана змінна є неважливою для потреб користувача і правило спрощується. В результаті кожне каузальне правило має містити мінімальну кількість значень вхідних змінних, які вказують на найвірогідніші причини отримано рекомендації з урахуванням потреб користувача.

Детальніші приклади використання розрахунку коректності, чутливості та складності на трьох етапах методу наведено в публікаціях авторів [16, 18, 20].

7. Обговорення результатів дослідження

В рамках експериментальної перевірки запропонованих теоретичних результатів розглядається побудова пояснень для задачі аналізу поведінки покупців – користувачів сайту електронної комерції. Сутність даної задачі полягає в тому, щоб на основі інформації про послідовність дій користувача на сторінках сайту оцінити його намір здійснити покупку, або ж покинути інтернет-магазин. Як вхідні дані для прогнозування намірів користувача використовується інформація про переглянуті сторінки, а також про сеанси взаємодії з системою електронної комерції. Задача вирішується з використанням штучних нейронних

мереж – багатошарового персептрону та LSTM (long short-term memory) [21].

Набір даних для експериментальної перевірки містить інформацію про підмножину сеансів взаємодії користувачів із системою електронної комерції. Кожен сеанс містить інформацію лише про одного користувача та характеризується значеннями таких змінних:

– клас сеансу, який визначається наявністю або відсутністю покупки (змінна «Revenue»);

– кількість різнотипних сторінок, які відвідав користувач, а також час перебування на цих сторінках (змінні «Administrative», «Administrative Duration», «Informational», «Informational Duration», «Product Related» and «Product Related Duration»);

– оцінки поведінки користувача на сторінках сайту, виміряні з використанням «Google Analytics» (змінні «Bounce Rate», «Exit Rate» and «Page Value»);

– технічні дані щодо операційної системи та браузера на комп'ютері відвідувача, інформація про регіон, з якого направив запит користувач, а також тип трафіку, тип відвідувача (новий чи вже зареєстрований користувач) та інформація про відповідність дати відвідування святковим або вихідним дням (змінні «Operating system», «Browser», «Region», «Visitor type», «Weekend»).

Задача експерименту полягає в тому, щоб встановити змінні, які визначають ключові причини вибору користувача, тобто покупки або завершення сеансу роботи в системі електронної комерції. В подальшому встановлені змінні порівнювались з відфільтрованими для прогнозування поведінки змінними [21]. Результати порівняння для 5 ключових змінних наведено в табл. 2.

Таблиця 2

Змінні, що відображають поведінку користувача системи електронної комерції

Оцінка поведінки користувача		Пояснення	
Рейтинг	Змінна	Можливість	Необхідність
1	Page Value	0,198	0,114
2	Exit Rate	0,087	0,068
3	Product Related	0,193	0,09
4	Product Related Duration	0,187	0,098
5	Bounce Rate	0,141	0,101

Наведені змінні мають найбільші значення показника можливості і відображають основні причини рішення щодо прогнозованих намірів користувача, що відповідає оцінці змінних в [21]. Ці змінні можуть бути використані для побудови каузальних залежностей виду «значення вхідної змінної – рішення системи» при формуванні пояснення. Проте слід врахувати, що змінні «Page Value», «Product Related», «Product Related Duration» мають близькі значення можливості. Максимальне значення необхідності має змінна «Page Value». У відповідності до (4), із трьох змінних для побудови пояснення доцільно використовувати в першу чергу «Page Value». Змінні «Exit Rate» та «Bounce Rate» мають менші значення можливості і тому можуть бути використані лише як додаткові залежності при деталізації пояснення.

Запропоновані модель та метод побудови пояснень орієнтовані на підтримку задач удосконалення, адаптації під потреби користувача та ефективного використання рішення інтелектуальної інформаційної системи, представленої у вигляді «чорного ящику». Пояснення у формі темпоральних та каузальних правил дає можливість побудувати спрощену інтерпретовану модель процесу прийняття рішення в інтелектуальній інформаційній системі, що

створює умови підтримки та удосконалення алгоритмів роботи розробником такої системи. Пояснення у формі найвірогідніших каузальних залежностей забезпечує умови для представлення кваліфікованому користувачеві множини можливих причин прийнятого рішення. Аналіз цих причини дає можливість адаптувати процес прийняття рішення, змінивши, наприклад, підмножину вхідних даних. Пояснення у формі найпростіших каузальних залежностей дає можливість виділити ключові причини отриманого рішення і представити таке спрощене пояснення кінцевому користувачеві. Обмеження на використання запропонованого підходу пов'язані із необхідністю отримати вхідні, проміжні та результуючі дані за суттєвий проміжок часу для обрахунку показників можливості та необхідності. Подальший розвиток даного підходу пов'язаний із комбінуванням можливісних каузальних залежностей із існуючими підходами до побудови спрощених моделей прийняття рішень в системах штучного інтелекту.

8. Висновки

В рамках дослідження виділено та структуровано ключові задачі побудови пояснень: представлення алгоритму формування рішення в інтелектуальній системі у інтерпретованій формі; представлення зрозумілих для користувача каузальних залежностей щодо причин прийнятого рішення як основи для пояснень; узгодження каузальних залежностей у складі пояснення із потребами користувача.

Розроблено узагальнену каузальну модель пояснення, яка об'єднує темпоральну, причинно-наслідкову та цільову складові. Темпоральна складова орієнтована на представлення пояснення як інтерпретованого опису алгоритму прийняття рішення в інтелектуальній інформаційній системі. Причинно-наслідкова складова орієнтована на формування зрозумілого опису причин дій процесу та результуючого рішення. Цільова складова орієнтована на формування пояснення у вигляді набору ключових причин рішення згідно з потребами користувача. Модель забезпечує побудову багатоаспектного пояснення, що може бути використано не лише після реалізації рішення, а й до початку його імплементації.

Запропоновано узагальнений метод побудови пояснення на основі каузальних залежностей, що містить етапи формування темпорального, причинно-наслідкового та цільового опису пояснення, які передбачають послідовне виділення найсуттєвіших причин прийнятого в інтелектуальній інформаційній системі рішення. Метод забезпечує можливість автоматизованої побудови та ітеративного уточнення пояснень як для розробників, так і для кінцевих користувачів з тим, щоб підвищити ефективність використання рішень інтелектуальних інформаційних систем.

Перелік посилань:

1. Kordon A. (2016). Intelligent Systems in Industry. *Innovative Issues in Intelligent Systems. Studies in Computational Intelligence*/ Sgurev, V., Yager, R., Kacprzyk, J., Jotsov, V. (eds). Springer, Cham. 2016. Vol 623. P. 1-31 https://doi.org/10.1007/978-3-319-27267-2_1.

2. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 2019. Vol. 267. P. 1-38. https://doi.org/10.1016/j.artint.2018.07.007

3. Bodria F., Giannotti F., Guidotti R., Naretto F., Pedreschi D., Rinzivillo S. (2021) Benchmarking and survey of explanation methods for black box models. *arXiv*. https://arxiv.org/abs/2102.13076.

4. Чалий С. Ф., Лещинський В. О., Лещинська І. О. 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». 2021. № 2 (6). С. 41-46.

5. Gunning D., Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*. 2019. Vol. 40 (2). P. 44-58. https://doi.org/10.1609/aimag.v40i2.2850.

6. Tjoa E., Guan C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems.* 2021. Vol. 32 (11). P. 4793-4813. https://doi.org/10.1109/tnnls.2020.3027314. PMID: 33079674.

7. Hanif A. et al. A Comprehensive Survey of Explainable Artificial Intelligence (XAI) Methods: Exploring Transparency and Interpretability. *Web Information Systems Engineering – WISE 2023. Lecture Notes in Computer Science*/ Zhang F., Wang H., Barhamgi M., Chen L., Zhou R. (eds). Springer, Singapore. 2023. Vol. 14306. https://doi.org/10.1007/978-981-99-7254-8_71

8. Yang W., Wei, Y., Wei H. et al. Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. *Human-Centric Intelligent Systems.* 2023. Vol. 3. P. 161–188. https://doi.org/10.1007/s44230-023-00038-y.

9. 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. № 1 (9). С. 70–75. https://doi.org/10.20998/2079-0023.2023.01.

10. Arrieta A. B., Díaz-Rodríguez N., Del Ser J., Bennetot A., Tabik S., Barbado A., Garcia S., Gil-Lopez S., Molina D., Benjamins R., Chatila R., Herrera F. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. 2020. Vol. 58. P. 82-115.

11. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. 2020. Vol. 4(3). P 113–117. https://doi.org/10.20998/2522-9052.2020.3.16.

12. Chala O. Models of temporal dependencies for a probabilistic knowledge base. *Econtechmod. An International Quarterly Journal*. 2018. Vol. 7, No. 3. P. 53 – 58.

13. Чала О. В. Модель узагальненого представлення темпоральних знань для задач підтримки управлінських рішень. *Вісник Національного технічного університету «ХПІ». Системний аналіз, управління та інформаційні технології.* 2020. № 1(3). С. 14-18. https://doi.org/10.20998/2079-0023.2020.01.03.

14. 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. https://doi.org/10.15587/1729-4061.2018.142664.

15. Nonaka I., Hirotaka T.. The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford University Press, 1995. https://doi.org/10.1093/oso/9780195092691.001.0001

16. Chalyi S., Leshchynskyi V. Оцінка чутливості пояснень в інтелектуальній інформаційній системі. *Системи управління, навігації та зв'язку. Збірник наукових праць.* 2023. Т 2. С. 165-169. https://doi.org/10.26906/SUNZ.2023.2.165.

17. Dubois D., Prade H. Possibility Theory. *The Palgrave Encyclopedia of the Possible* / Glăveanu, V.P. (eds). Palgrave Macmillan, Cham. 2022. https://doi.org/10.1007/978-3-030-90913-0_175.

18. Chalyi S., Leshchynskyi V. Інформаційна технологія оцінки пояснень в інтелектуальній інформаційній системі. *Системи управління, навігації та зв'язку. Збірник наукових праць*.2023. Т. 4. С. 120-124. https://doi.org/10.26906/SUNZ.2023.4.120.

19. Byrne R.M.J. Counterfactuals in explainable artificial intelligence (XAI): evidence from human reasoning. *Proceedings of the twenty-eighth international joint conference on artificial intelligence, IJCAI 2019*, Macao, China, August 10–16, 2019. Survey track. P. 6276-6282. https://doi.org/10.24963/ijcai.2019/876.

20. Чалий С.Ф., Лещинський В.О. Метод можливісного оцінювання пояснення в системі штучного інтелекту/ *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", 2023. № 2 (10). С. 95-101.

21. Baati, K., & Mohsil, M. Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest. *Artificial Intelligence Applications and Innovations.* 2020. Vol. 583. P. 43 - 51.

Надійшла до редколегії 25.03.2024 р.

Чалий Сергій Федорович, доктор технічних наук, професор, професор кафедри ІУС ХНУРЕ, м. Харків, Україна, e-mail: serhii.chalyi@nure.ua; ORCID: 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: 0000-0002-8690-5702