УДК 004.8:004.9 DOi 10.30837/ bi.2022.1(98).11

С.Ф. Чалий[1], В.О. Лещинський[2]

[1]професор кафедри інформаційних управляющих систем,
Харківський національний університет радіоелектроніки, Україна,
serhii.chalyi@nure.ua

[2]доцент кафедри програмної інженерії,
Харківський національний університет радіоелектроніки, Україна,
volodymyr.leshchynskyi@nure.ua

МОЖЛИВІСНО-КАУЗАЛЬНЕ ПРЕДСТАВЛЕННЯ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНІЙ ІНФОРМАЦІЙНІЙ СИСТЕМІ

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 основі можливісного опису причинно-наслідкових зав'язків між вхідними даними та результатом роботи інтелектуальної системи, що створює умови для формування пояснення при представленні такої системи як у вигляді «чорного», так і у вигляді «сірого» ящику. Завдання: розробка узагальненої можливісно-каузальної моделі пояснення; розробка методу побудови можливісно-каузального представлення пояснення в інтелектуальній інформаційній системі. Використовуваними підходами є: методи побудови пояснень, теорія можливостей, підходи до побудови темпоральних знань. Висновки. Наукова новизна отриманих результатів полягає в наступному. Запропоновано комплекс можливісно-каузальних моделей пояснення в інтелектуальній системі, що задає причинно-наслідковий зв'язок між класом вхідних даних та класом рішення, а також між проміжними діями з процесу отримання результату та класом рішення. Можливісний аспект моделі розраховується для підмножини, що містить представників одного класу даних, або ж для окремих дій спрощеного процесу прийняття рішення. У практичному плані розроблена модель дає можливість сформувати опис процесу прийняття рішення на основі можливісних каузальних залежностей та побудувати пояснення на основі обмежених даних про процес функціонування інтелектуальної інформаційної системи. Запропоновано метод побудови пояснення в інтелектуальній інформаційній системі на основі можливісних каузальних залежностей. Метод містить етапи визначення класів вхідних даних та результату для пояснення, формування переліку можливих причинно-наслідкових залежностей, що пов'язують вхідні дані або дії процесу із рішенням інтелектуальної системи, розрахунку можливостей використання отриманих залежностей для побудови пояснення, розрахунку необхідності для отриманих залежностей та упорядкування отриманих пояснень за значенням необхідності. Метод дає можливість побудувати пояснення при представленні інтелектуальної інформаційної системи як у вигляді «чорного», так і у вигляді «сірого» ящику, відобразивши відповідно вплив вхідних даних та вплив спрощеного процесу прийняття рішення на результат інтелектуальної системи.

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

S. Chalyi, V. Leshchynskyi Possibility-causal representation of explanations in an intelligent information system. The article's subject matter is the process of constructing explanations for the obtained results and the sequence of decision-making actions in the intelligent information system. The goal is to develop an approach to the construction of explanations based on a possible description of cause-and-effect relationships between input data and the result of the work of an intelligent system, which creates conditions for the formation of an explanation when presenting such a system both in the form of "black" and in the form of "gray" box. Tasks: development of a generalized possible-causal model of explanation; development of a method of possible-causal representation of an explanation in an intellectual information system. The used approaches are: methods of constructing explanations, the theory of possibilities, approaches to the construction of temporal knowledge. Conclusions. The scientific novelty of the obtained results is as follows. A possible-causal model of explanation in an intelligent system is proposed, which specifies a cause-and-effect relationship between the class of input data and the class of decision, as well as between intermediate actions from the process of obtaining a result and the class of decision. The probabilistic aspect of the model is calculated for a subset containing representatives of the same class of data, or for individual actions of a simplified decision-making process. In practical terms, the developed model makes it possible to form a description of the decision-making process based on possible causal dependencies and to build an explanation based on limited data about the process of functioning of the intelligent information system. A method of constructing an explanation in an intelligent information system based on possible causal dependencies is proposed. The method includes the stages of determining the classes of input data and the result for explanation, forming a list of possible cause-and-effect dependencies that connect input data or process actions with the decision of an intelligent system, calculating the possibilities of using the obtained dependencies to build an explanation, calculating the need for the obtained dependencies and ordering received explanations according to necessity. The method makes it possible to build an explanation when presenting an intelligent information system both in the form of a "black" and in the form of a "gray" box, reflecting, respectively, the influence of input data and the influence of a simplified decision-making process on the result of an intelligent system.

INTELLIGENT SYSTEM, ARTIFICIAL INTELLIGENCE SYSTEM, EXPLANATION, DECISION-MAKING PROCESS, CAUSALITY, POSSIBILITY, CAUSALITY

Вступ

Сучасні дослідження в галузі психології пізнання свідчать про те, що людина, як правило, потребує обґрунтування нових знань з використанням відповідних пояснень [1-3].

Пояснення щодо процесу прийняття рішення, які надаються в інтелектуальних системах, формують причинно-наслідкові зв'язки між вхідними даними та результатом [2] і тому забезпечують умови для того, щоб користувачі був впевнений у правильності отриманих результатів [4, 5].

Актуальність використання пояснень в інтелектуальних системах є наслідком протиріччя, що виникає при використання алгоритмів машинного навчання у процесі формування рішень. Такі алгоритми орієнтовані на представлення таких систем у вигляді «чорного ящику», що робить їх незрозумілими для користувача. До того ж результати, отримані при використанні алгоритмів машинного навчання, можуть бути спотворені внаслідок упередженості або викидів у вхідних даних. Відповідно, отримані в інтелектуальній інформаційній системі рішення можуть не в повній мірі задовільнити користувачів таких систем. На практиці така невідповідність приводить до неефективного використання інтелектуальної інформаційної системи, зокрема до відмови від застосування запропонованих рішень.

При використанні «прозорих» алгоритмів формування рішення виникає інше обмеження, пов'язане із юридичним захистом інтелектуальної власності. В даному випадку має місце заборона щодо розкриття деталей та умов процесу прийняття рішення для користувачів. Юридичні обмеження також призводять до зниження довіри користувачів до рішень системи штучного інтелекту та відповідного неефективного використання цих рішень на практиці [6].

Тому наукові дослідження в області формування пояснень для інтелектуальних систем інтенсивно розвиваються в останні роки, особливо в рамках програми XAI, що була започаткована керівництвом агенції DARPA у 2017 році [7].

Ключові напрямки досліджень у сфері пояснюваних інтелектуальних систем базуються як на визначенні відповідності пояснень потребами користувачів, наприклад користувачів рекомендаційних систем [8], так і на визначенні явних або неявних причинно-наслідкових залежностей між вхідними факторами та рішенням інтелектуальної системи [9-13].

Однак існуючі підходи є спеціалізованими, значною мірою залежать від типу алгоритму, що використовується в інтелектуальній системі, і не приділяють достатньо уваги побудові пояснень на основі комбінованого опису процесу прийняття рішення, що враховує як зв'язок між вхідними даними та результатом інтелектуальної системи, так і залежності між ключовими діями та результатом вказаного процесу. Такий зв'язок носить можливісний характер, оскільки він враховує не лише можливість впливу вхідних даних на рішення системи, а й необхідність вибору підмножини цих вхідних даних, які є найбільш суттєвими для пояснення щодо результату роботи інтелектуальної інформаційної системи. Можливісний підхід [14] дає можливість узагальнено описати каузальні залежності для побудови пояснень, абстрагуючись від особливостей конкретного механізму прийняття рішень в інтелектуальній системі.

Зазначене свідчить про актуальність задачі побудови каузального представлення пояснень з урахуванням можливісного опису такого причинно-наслідкового зв'язку.

1. Постановка задачі

Метою статті є розробка підходу до побудови пояснень на основі можливісного опису причинно-наслідкових зав'язків між вхідними даними та результатом роботи інтелектуальної системи, що створює умови для формування пояснення при представленні такої системи як у вигляді «чорного», так і у вигляді «сірого» ящику.

Для досягнення поставленої мети вирішуються такі задачі:

– розробка узагальненої можливісно-каузальної моделі пояснення;

– розробка методу побудови можливісно-каузального представлення пояснення в інтелектуальній інформаційній системі.

2. Схема побудови пояснення в системі штучного інтелекту

Загальна схема побудови пояснення у існуючій системі штучного інтелекту, що представлена у вигляді «чорного» або «сірого» ящику, полягає у формуванні зовнішньої пояснювальної підсистеми. Дана підсистема використовує вхідні дані, вихідний результат, а також доступні проміжні дані про процес прийняття рішення у системі штучного інтелекту. Для пояснення отриманого результату використовується спрощена модель прийняття рішення.

При представленні системи штучного інтелекту у вигляді «чорного ящику» доступними для побудови пояснення є лише вхідні дані та рішення системи. Відповідну схему побудови пояснення представлено на рис. 1.

В даному випадку формуються причинно-наслідкові залежності між значеннями окремих елементів вхідних даних та рішенням. Ключова ідея пояснення полягає в тому, щоб показати залежності, що відображають вплив ключових значень вхідних даних на отримане рішення.

З метою спрощення моделі процесу прийняття рішення при побудові пояснення доцільно

визначити залежності не між окремими вхідними даними та отриманим результатом, а між класами елементів вхідного набору даних та класом рішення.

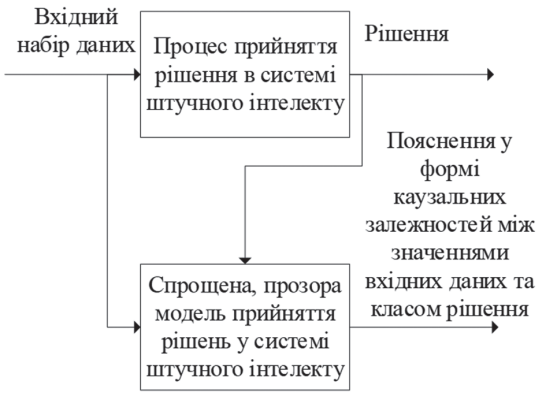

Рис. 1. Побудова пояснення на основі вхідних даних при представленні системи штучного інтелекту у вигляді «чорного ящику»

Наприклад, пояснення щодо результату класифікації зображення людини, що біжить, може відображати залежності між фрагментами зображення, які містять елементи тіла в русі, та результуючим рішенням. Або пояснення щодо пропозиції комп'ютера в рекомендаційній системі може містити зв'язки між значенням комплектуючих (модель процесора, об'єм пам'яті, тощо) та запропонованою моделлю комп'ютера.

Слід зауважити, що вказані залежності формуються для груп комп'ютерів. Наприклад, процесор типу i7 може виступати в якості причини вибору ноутбука певної фірми внаслідок того, що забезпечується найкраще співвідношення ціни та потужності.

При представленні системи штучного інтелекту у вигляді «сірого ящику» доступною є також часткова інформація про процес прийняття рішення у такій системі. Дана інформація зазвичай представлена у формі логу (журналу подій). Останній формується підсистемою моніторингу.

Відповідна схема побудови пояснення представлена на рис. 2.

Журнал подій S містить у собі набір трас $\{S_m\}$.

Кожна з трас S_m описує доступну для зовнішнього спостерігача послідовність станів процесу прийняття рішення:

$$S_m = \langle s_{m,1}, s_{m,2}, ..., s_{m,z}, ... \rangle \quad (1)$$

Стани задаються через множину значень змінних, що характеризують властивості системи. Стани є упорядкованими у часі, оскільки містять темпоральні мітки.

Тобто кожний стан описується множиною значень змінних $s_{m,z} = \{\{x^i_{m,z}\}, t^i_{m.z}\}$, серед яких $x^i_{m,z}$ задає значення таких, наприклад, властивостей, як:

– назва дії із процесу прийняття рішення, що привела до поточного стану;

– об'єкт або змінна, з якою виконувалась дія, тощо.

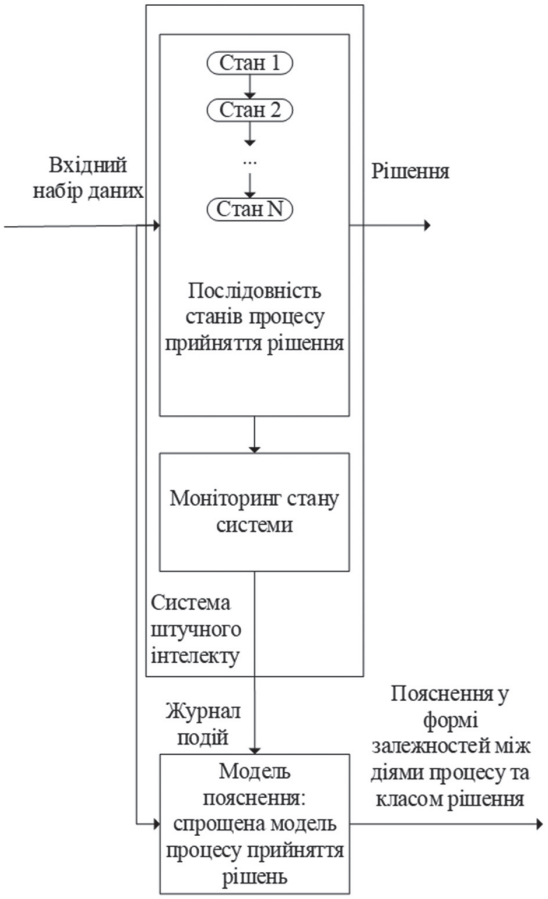

Рис. 2. Побудова пояснення на основі журналу подій при представленні системи штучного інтелекту у вигляді «сірого ящику»

Змінна $t^i_{m.z}$ є позначкою часу, яка дає можливість задати темпоральні правила.

Характерна особливість логу полягає в тому, що кожна з трас є записом про виконання одного екземпляру процесу прийняття рішення, а сукупність трас задає всі реалізовані на практиці альтернативи такого процесу.

Відповідно, зворотний інжиніринг логу дає можливість побудувати модель процесу і, на основі аналізу такої моделі, знайти причини отриманого в системі штучного інтелекту рішення. Такий зворотний інжиніринг зазвичай виконується методами інтелектуального аналізу процесів.

Однак дослідження отриманої в результаті інтелектуального аналізу процесів моделі потребує додаткової підготовки, що утруднює сприйняття такого пояснення користувачем.

Альтернативний підхід до побудови пояснення базується на виділенні темпоральних залежностей між станами системи. Такі залежності представляються у формі темпоральних правил. Темпоральні задають порядок пар станів $s_{m,z}$ у часі:.

Темпоральні правила мають вигляд:

$$fn: s_{m,z} \to s_{m,z+1}, \quad (2)$$

$$ff: s_{m,z} \to s_{m,z+q} \,|\, q > 1. \quad (3)$$

Правило (1) визначає зв'язок для двох послідовних у часі станів процесу прийняття рішення.

Правило (2) задає зв'язок для двох станів $s_{m,z}$ та $s_{m,z+q}$, між якими є проміжні у часі стани, наприклад $s_{m,q-1}$.

Наведені правила дають можливість відобразити як поточні (1), так і довгострокові (2) наслідки дій процесу прийняття рішення в інтелектуальній інформаційній системі.

Узагальнене правило f_q^m об'єднує правила (2) та (3) та упорядковує у часі довільну пару станів $s_{m,z}$ та $s_{m,z+q}$:

$$f_q^m: s_{m,z} \to s_{m,z+q} \,|\, q \geq 1. \quad (4)$$

Досягнення результату роботи системи типу R_l можна розглядати як кінцевий стан $s_{m,Q}$ процесу прийняття рішення. Відповідно, опис процесу у вигляді темпоральних правил дає можливість використати для побудови пояснення темпоральні залежності між ключовими проміжними станами системи штучного інтелекту та отриманим рішенням (5), або ж між патерном послідовності станів та відповідним рішенням (6).

$$s_{m,z} \to R_l, \quad (5)$$

$$f_q^z \to R_l. \quad (6)$$

При необхідності деталізувати стани дані залежності можуть бути зведені до зв'язку між доступними значеннями проміжних змінних системи $x_{m,z}^i$ та отриманим рішенням, або ж між послідовністю зміни значень таких змінних та результатом.

3. Можливісно-каузальна модель пояснення в інтелектуальній інформаційній системі

Задача представлення каузальних залежностей для пояснень у випадку представлення інтелектуальної системи як «чорного ящику» полягає у встановленні зв'язку між класами вхідних даних та класами рішень.

Позначимо множину вхідних даних як $X = \{x_i\}$.

Вхідні дані підрозділяються на класи, які описуються підмножинами X_j:

$$X_j \subset X, \left(\bigcap X_j\right) = \varnothing. \quad (7)$$

Аналогічно, позначимо множину рішень інтелектуальної інформаційної системи $R = \{r_i\}$. Класи рішень визначаються підмножинами R_l, що не перетинаються:

$$R_l \subset R, \left(\bigcap R_l\right) = \varnothing \quad (8)$$

Для кожного з рішень зазвичай є відомою ймовірність використання вхідних даних. Наприклад, якщо було продано ноутбук в системі електронної комерції, то відомими є комплектуючі цього ноутбука. Відповідно, можна підрахувати ймовірність вибору ноутбуку з певним процесором, жорстким диском або пам'яттю. Таку множину ймовірностей для елементів X_j позначимо P_j.

Тобто $P_j = \{p_{k,j}\}$, причому $p_{k,j}$ – це ймовірність вибору рішення R_l із вхідними даними x_i.

Тоді можливісно-каузальна модель пояснення при представленні інтелектуальної системи у вигляді «чорного ящику» для однієї вхідної змінної має вигляд:

$$\begin{aligned} &\textit{Вхідні дані}: \{P_j\}, \{R_l\}, \\ &\textit{Можливісно – каузальна залежність}: \\ &x_{k,j} \Rightarrow R_l \,|\, p_{k,j} = \Pi(X_j), \end{aligned} \quad (9)$$

де $x_{k,j}$ – значення k – вхідної змінної із підмножини X_j, яка є можливою причиною рішення класу R_l; $\Pi(X_j)$ – можливість використання одного із елементів підмножини X_j в якості причини для рішення класу R_l.

Сенс виразу (2) полягає в наступному: значення $x_{k,j}$ X_j змінної із підмножини можливих значень X_j є найбільш можливою причиною отриманого рішення.

Модель пояснення для декількох вхідних змінних має такий вигляд:

$$\begin{aligned} &\textit{Вхідні дані}: \{P_j\}, \{R_l\}, \\ &\textit{Можливісно – каузальна залежність}: \\ &\langle x_{k,j}, x_{k,j+1}, \ldots \rangle \Rightarrow R_l \\ &|\, p_{k,j} = \Pi(X_j), p_{k,j+1} = \Pi(X_{j+1}), \\ &N(X_j) \geq N(X_{j+1}) \end{aligned} \quad (10)$$

В даному випадку вхідні змінні упорядковуються за необхідністю, оскільки необхідність фактично визначає можливість використання альтернативних до $x_{k,j}$ значень вхідних даних для пояснення.

У випадку представлення інтелектуальної системи як «сірий ящик» доступною є додаткова інформація у вигляді логів або журналів подій, що дає можливість побудувати опис відомої частини процесу прийняття рішення у вигляді темпоральних правил. Тоді модель пояснення може базуватись на трьох підходах:

– на основі обмежень на виконання дій процесу прийняття рішення;

– на основі вибору можливої дії/стану процесу прийняття рішення серед підмножини альтернатив;

– на основі вибору можливого темпорального правила.

Згідно першого підходу, в якості можливої причини рішення розглядаються обмеження на виконання процесу прийняття рішення в інтелектуальній системі. В якості обмежень розглядаються дії, які були виконані на всіх трасах процесу. Тобто для обмежень

виконується умова $p_{m,q}=1$. Відповідна можливісно-каузальна залежність задає зв'язок між станами $s_{m,q}$ та класом рішення R_l.

Згідно другого підходу в якості можливої причини рішення інтелектуальної системи розглядається один із альтернативних станів процесу прийняття рішення. Тобто із станів на різних трасах формується множина альтернатив і як причина рішення вибирається стан із найбільшою ймовірністю. Представлення такої залежності має вигляд:

$$\begin{aligned} &\textit{Вхідні дані}: \{P_j\}, \{R_l\}, \\ &\textit{Можливісно – каузальна залежність}: \\ &\langle s_{m,z}, s_{m,q}, \ldots \rangle \Rightarrow R_l \\ &\left| p_{m,z} = \Pi(S_z), p_{m,q} = \Pi(S_q), \right. \\ &N(S_z)) \geq N(S_q), \end{aligned} \tag{11}$$

де S_z, S_q – підмножини альтернативних станів процесу прийняття рішення.

В даному випадку фактично конструюються темпоральні правила $s_{m,z} \to R_l$, $s_{m,q} \to R_l$, на базі яких визначаються можливісні каузальні залежності результату від відповідних дій процесу прийняття рішення. Зв'язок між станами та результатом у наведених правилах виділяється тому, що проміжні стани відображають результати окремих дій у процесі формування рішення.

Згідно третього підходу, формується зв'язок між темпоральними правилами та результатом. Тобто задана правилами послідовність дій розглядається як можлива причина отриманого в інтелектуальній системі результату.

Розглянемо приклад залежності між проміжними станами процесу прийняття рішення та результатом для рекомендаційної системи. В даному випадку в онлайн-підсистемі побудови рекомендацій аналізується послідовність дій користувача системи (тобто послідовність кліків по екранній формі). Користувачі по різному взаємодіють з однією й тією ж екранною формою, вибираючи або пропускаючи запропоновані на екрані товари. Тому аналіз трас логу для різних користувачів дає можливість порівняти траєкторії їх руху та надати один із варіантів переміщення по екрану як пояснення щодо запропонованої рекомендації товарів або послуг. Зокрема, якщо користувач розглядає (вибирає мишою) властивості найбільш популярних товарів, то відповідна залежність буде використана у якості пояснення.

4. Метод побудови можливісно-каузального представлення пояснення в інтелектуальній інформаційній системі

Побудова неведеного можливісно-каузального представлення передбачає побудову опису вхідних даних, формування можливих залежностей та подальше їх упорядкування з урахуванням розрахованої необхідності впливу конкретних значень вхідних змінних на отримане в системі штучного інтелекту рішення.

Метод містить таку узагальнену послідовність етапів.

Етап 1. Визначення класів вхідних та вихідних даних для пояснення.

Крок 1.1 Визначення класів вхідних даних.

Реалізація даного етапу залежить від особливостей та структури вхідних даних. Однак в цілому має бути отримана однорівнева (по групам) або багаторівнева (ієрархічна) класифікація вхідних даних.

Результатом даного етапу є множина класів X_j, що описується підмножинами X_j. Для однорівневого розбиття на класи виконується умова (7). При побудові ієрархії класів вхідних даних вираз (7) доповнюється рекурсивно, тобто виділяються підмножини $X_{j,g}$, для яких виконується умова:

$$\bigcup_g X_{j,g} = X_j, \left(\bigcap X_{j,g}\right) = \varnothing. \tag{12}$$

Виділення підмножин інших рівнів ієрархії виконується аналогічно.

Крок 1.2 Формування класів рішень.

На даному етапі формуються класи рішень таким чином, щоб для кожного рішення з класу можна було сформувати ідентичне або схоже пояснення.

Класи рішень R_l визначаються так, щоб любе рішення даного класу мало множину значень вхідних даних із відповідних класів.

Наприклад, якщо рішенням є пропозиція ноутбука в рекомендаційній системі, то клас рішення містить всі ноутбуки з однаковим типом процесору, об'ємом пам'яті та жорсткого диску, тощо. Тобто клас рішення R_l визначається одними й тими ж підмножинами X_j, для яких ми можемо розрахувати можливісно-каузальні залежності:

$$X^l = \left\{ X_j : \forall x_i \in X_j \; \exists r_n \in R_l \right\} \tag{13}$$

Крок 1.3 Визначення класів темпоральних правил. Даний крок виконується за умови, що інтелектуальна система представлена як «сірий ящик».

На даному кроці темпоральні правила формуються та розподіляються на класи в залежності від значень атрибутів $x^i_{m,z}$ станів $s_{m,z}$.

Тоді клас темпорального правила визначається, наприклад, класом $x^i_{m,z}$. Зокрема, ми можемо виділити дії за їх функціональною ознакою: фільтрація, перевірка умов, виявлення латентних факторів, тощо.

Етап 2. Формування можливих причинно-наслідкових залежностей.

На даному етапі виділяються залежності виду $x_{k,j} \Rightarrow R_l$ для вхідних змінних та $s_{m,z} \Rightarrow R_l$ й $f^z_q \Rightarrow R_l$ для темпоральних правил.

Етап 3. Розрахунок можливостей використання отриманих залежностей для побудови пояснення.

Можливість розраховується як найбільша ймовірність $p_{k,j}$ для відповідної підмножини вхідних даних, станів або правил.

Етап 4. Розрахунок необхідності для отриманих залежностей.

Необхідність розраховується через можливість всіх інших значень змінних, крім змінних із підмножини X_j.

Етап 5. Упорядкування отриманих пояснень за значенням необхідності.

Результатом даного етапу є набір змінних, які є можливими причинами отриманого рішення. Ці змінні упорядковуються за ступенем довіри до них. Остання визначається згідно необхідності використання цих змінних.

В тому випадку, якщо інтелектуальна система представлена у вигляді «сірого ящику», тоді результатом є набір темпоральних правил, які є можливими причинами рішення із класу R_l.

Особливість використання темпоральних правил полягає в тому, що вони визначають не лише проміжні стани, які є можливими причинами для отриманого результату, але й дії, які привели до цих станів, а також зв'язок між діями, тобто упорядкованість дії-причини та попередньої до неї дії в часі.

В результаті можна отримати комплексну причину результату: ключова дія (або дії), які привели до даного рішення, а також передумови для виконання цих ключових дій.

Так, якщо ми розглядаємо людино-машинний процес прийняття рішення, при якому людина використовує проміжне рішення інтелектуальної системи і відповідно змінює алгоритм роботи, то в якості причини ми можемо з'ясувати, чи використала ця людина-оператор проміжні результати роботи системи на подальших етапах процесу прийняття рішення.

Наприклад, чи були використані результати автоматизованої діагностики у процесі сервісного обслуговування клієнтів. Або ж навпаки, з метою спрощення роботи виконавець проігнорував отриману інформацію.

Розглянемо приклад формування можливісно-каузальної залежності для пояснення щодо пропозиції комп'ютера в рекомендаційній системі.

Використовуються вхідні дані про продажі ноутбуків у системі електронної комерції. Ці містять інформацію про модель комп'ютера.

На базі інформації про моделі ноутбуків формується множина класів рішень $R = \{R_l\}$, що містить моделі ноутбуків на базі процесорів $i7, i9$, тощо.

Множина класів вхідних даних, для, наприклад, процесора, формується у вигляді набору $X = \{i3, i5, i7, i9\}$. Кожна із підмножин $i3, i5, i7, i9$ містить назви моделей відповідного типу, наприклад $i7 = \{i7-1185, i7-1165, ...\}$.

На базі інформації про продажі формується множина ймовірностей продажу ноутбуків із відповідними процесорами: $p(i7) = \{p(i7-1185), p(i7-1165), ...\}$.

На наступних етапах розраховується можливість $\Pi(i7)$:

$$\Pi(i7) = \max\left(p(i7-1185), p(i7-1165), ...\right), \quad (14)$$

та формується каузальна залежність виду

$$i7 - 1185 \Rightarrow \text{Модель ноутбука} \quad (15)$$

як основа пояснення щодо запропонованої моделі.

Аналогічні розрахунки виконуються для інших вхідних змінних, пов'язаних із пам'яттю, жорстким диском, параметрами екрану, зовнішніх портів, тощо.

Залежності виду (15) в подальшому упорядковуютья за значенням необхідності.

Висновки

Розроблено комплекс можливісно-каузальних моделей пояснення в інтелектуальній системі, що задають причинно-наслідковий зв'язок між класами вхідних даних, станів, темпоральних правил та класом рішення.

При представленні інтелектуальної системи у вигляді «чорного ящику» пояснення визначається через можливість впливу вхідних даних на рішення інтелектуальної системи. Можливість розраховується для підмножини, що містить представників одного класу даних.

Модель в даному випадку забезпечує пояснення щодо рішення інтелектуальної системи через вплив вхідних даних на результат, представляючи такий вплив як можливісний каузальний зв'язок міх входом та виходом інтелектуальної системи.

При представленні інтелектуальної системи у вигляді «сірого ящику» пояснення визначається через причинно-наслідкові зв'язки між діями спрощеного процесу прийняття рішення.

Безпосередньо спрощений процес відображає послідовність отримання результату, сформовану на основі доступних даних про функціонування інтелектуальної системи, наприклад, на основі журналів подій. Така послідовність представляється темпоральними залежностями.

Можливісний аспект моделі у даному випадку визначається для окремих дій або послідовностей дій спрощеного процесу прийняття рішення.

У практичному плані розроблена модель орієнтована на формування опису процесу прийняття рішення з використанням можливісного підходу, що створює умови для побудови зрозумілого опису процесу прийняття рішення у вигляді множини взаємопов'язаних можливісних каузальних залежностей на основі обмежених даних про процес

функціонування інтелектуальної інформаційної системи.

Розроблено метод побудови пояснення в інтелектуальній інформаційній системі на основі можливісних каузальних залежностей.

Метод містить етапи визначення класів вхідних даних для пояснення, формування можливих причинно-наслідкових залежностей, розрахунку можливостей використання отриманих залежностей для побудови пояснення, розрахунку необхідності для отриманих залежностей та упорядкування отриманих пояснень за значенням необхідності.

Метод дає можливість побудувати пояснення при представленні інтелектуальної інформаційної системи як у вигляді «чорного», так і у вигляді «сірого» ящику, відобразивши відповідно вплив вхідних даних та вплив спрощеного процесу прийняття рішення на результат інтелектуальної системи.

Список літератури:

[1] *Chi, M., de Leeuw, N., Chiu, M., & LaVancher, C.* Eliciting self-explanations improves understanding. Cognitive Science. – 1994. – Vol.18. P. 439–477.

[2] *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

[3] *Carey, S.* The origin of concepts. New York, NY: Oxford University Press. 2009. 608 p.

[4] *Adadi, A., Berrada, M.* (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. – 2018. – Vol.6. P. 52138– 52160.

[5] *Castelvecchi D.* Can we open the black box of AI? Nature. – 2016. – Vol. 538 (7623), pp. 20-23.

[6] *Tintarev N., Masthoff J.* Evaluating the effectiveness of explanations for recommender systems, User Model User-Adap Inter. – 2012. – Vol. 22, pp. 399– 439, https://doi.org/10.1007/s11257-011-9117-5

[7] *Gunning I. D. Aha* DARPA's Explainable Artificial Intelligence (XAI) Program // AI Magazine. – 2019. – Vol. 40(2), pp.44-58, doi: 10.1609/ aimag.v40i2.2850.

[8] *Adomavicius G. et al.* Incorporating contextual information in recommender systems using a multidimensional approach // ACM Transactions on Information Systems. – 2005. – Vol. 23(1). – P. 103–145.

[9] *Chalyi, S., Leshchynskyi, V.* Method of constructing explanations for recommender systems based on the temporal dynamics of user preferences. EUREKA: Physics and Engineering. – 2020. – Vol. 3, pp. 43-50. doi: 10.21303/2461-4262.2020.001228. Available at: http://journal.eu-jr.eu/engineering/article/view/14.

[10] *Chalyi S.* Probabilistic counterfactual causal model for a single input variable in explainability task / S. Chalyi, V. Leshchynskyi // Сучасні інформаційні системи = Advanced Information Systems. – 2023. – Т. 7, № 3. – С. 54-59.

[11] *Chalyi Serhii* Можливісна модель каузального зв'язку по вхідній змінній для побудови пояснення в інтелектуальній системі / Serhii Chalyi, Volodymyr Leshchynskyi // Системи управління, навігації та зв'язку. Збірник наукових праць. – Полтава: ПНТУ, 2023. – Т. 3 (73). – С. 138-143. – doi:https://doi.org/10.26906/SUNZ.2023.3.138.

[12] *Akula, Arjun R., et al.* "CX-ToM: Counterfactual Explanations with Theory-of-Mind for Enhancing Human Trust in Image Recognition Models // arXiv preprint arXiv: 2109.01401 . – 2021. (accepted to iScience 2021).

[13] *A. Akula, S. Wang, and S.-C. Zhu, "*CoCoX: Generating Conceptual and Counterfactual Explanations via Fault-Lines", AAAI, vol. 34, no. 03, pp. 2594-2601, Apr. 2020.

[14] *Dubois, Didier and Prade, Henri,* Possibility Theory, Probability Theory and Multiple-valued Logics: A Clarification // Annals of Mathematics and Artificial Intelligence . – 2002. – Vol. 32, pp.35–66.

Надійшла до редколегії 14.11.2022