



ТЕМПОРАЛЬНО-ВЕКТОРНЕ УПОРЯДКУВАННЯ ДАНИХ ДЛЯ ПОБУДОВИ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ

Чалий С.Ф., д.т.н., професор, кафедра ІУС, ХНУРЕ
Кравченко Р.В., аспірант, кафедра ІУС, ХНУРЕ

Сучасні інтелектуальні інформаційні системи здатні вирішувати складні завдання, пов'язані із підтримкою прийняття рішень на основі виявлення закономірностей у великих даних, розпізнавання зображень, генерації текстів тощо. Проте алгоритми машинного навчання, що використовуються в цих системах, як правило є повністю або частково непрозорими для користувачів. З позицій користувача вони мають вигляд "чорної" або "сірої" скриньки. Така властивість інтелектуальних систем не дає можливість перевірити результат їх роботи, особливо з урахуванням можливих неточностей та упередженостей у вхідних даних, які можуть мати суттєвий вплив на формування рішення. Користувач, що отримує упереджене рішення, не буде в повній мірі довіряти результатам роботи інтелектуальної системи і перевірятиме це рішення. В результаті може суттєво знизитись ефективність застосування отриманих рішень і використання інтелектуальної системи в цілому.

Для вирішення представленої проблеми використовується концепція пояснень. Пояснення мають «розшифровувати» процес прийняття рішення в системі штучного інтелекту, надати причини отриманого результату і зробити його повністю зрозумілим для користувача [1]. Пояснення ґрунтуються на зв'язках між вхідними дані (або проміжними даними, що можуть бути видимими для користувача), а також результатом роботи системи. Ці дані відображають стани системи у процесі прийняття рішення. Зазвичай вказані зв'язки представляються у вигляді зважених причинно-наслідкових залежностей [2]. Такі залежності мають темпоральну упорядкованість, відображаючи послідовність станів системи штучного інтелекту у процесі прийняття рішення [3].

Проте при представленні причинно-наслідкових залежностей між станами інтелектуальної системи в цілому отримане на базі цих залежностей пояснення не відповідає критерію складності. Критерій складності базується на обчисленні кількості змінних, які використовуються в поясненні [4]. При зменшенні складності в поясненні мають залишитися лише ключові змінні, які мають найбільш суттєвий вплив на результат інтелектуальної системи. Для побудови зрозумілого пояснення з урахуванням лише ключових змінних необхідно узгодити семантику змінних з семантикою предметної області користувача. Однак стани системи в багатьох випадках визначаються множинами перелічуваних даних з використанням специфічних термінів предметної області. Наприклад, в рекомендаційній системі використовуються назви характеристик товарів, які пропонує така система.

Таким чином, для побудови зрозумілих пояснень доцільно враховувати не лише послідовність станів інтелектуальної системи у часі, а й відмінності

семантики змінних – властивостей станів інтелектуальної системи, що потребує представлення їх у векторній формі [5]. Зазначене свідчить про актуальність задачі темпорально-векторного упорядкування даних щодо станів інтелектуальної системи при побудові пояснень. Структуру темпорально-векторного представлення даних наведено на рисунку 1.



Рисунок 1 – Темпорально-векторне упорядкування даних

Розроблене представлення дає можливість вирішити задачі визначення причин пояснення, а також узгодження знань при формуванні пояснення. Причини визначаються на основі темпоральної упорядкованості даних. Узгодження знань реалізується на основі пошуку за схожістю векторів слів у векторній базі. Результатом пошуку є семантично близький опис даних.

Список літератури

1. Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, 40(2), 44-58. <https://doi.org/10.1609/aimag.v40i2.2850>.
2. Chalyi, S., & Leshchynskiy, V. (2023). Probabilistic counterfactual causal model for a single input variable in explainability task. *Advanced Information Systems*, 7(3), 54-59.
3. Чала, О.В. (2018). Побудова темпоральних правил для представлення знань в інформаційно-управляючих системах. *Сучасні інформаційні системи*, 2(3), 54-59.
4. Chalyi, S. & Leshchynskiy, V. (2023). Інформаційна технологія оцінки пояснень в інтелектуальній інформаційній системі. Системи управління, навігації та зв'язку. *Збірник наукових праць*, (4), 120-124. 10.26906/SUNZ.2023.4.120.
5. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.