

DOI: <https://doi.org/10.36910/6775-2524-0560-2026-63-24>

УДК 004.891:004.62

Макоганюк Анастасія Олегівна, к.т.н., доцент

<https://orcid.org/0000-0002-7668-1653>

Державний університет інтелектуальних технологій і зв'язку, м. Одеса, Україна

ІНТЕЛЕКТУАЛЬНА МОДЕЛЬ АВТОМАТИЗОВАНОГО ВИЯВЛЕННЯ СЕМАНТИЧНИХ АСОЦІАТИВНИХ ПРАВИЛ У РОЗПОДІЛЕНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

Макоганюк А.О. Інтелектуальна модель автоматизованого виявлення семантичних асоціативних правил у розподілених інформаційних системах. У статті проведено комплексне дослідження актуальної науково-технічної задачі, що полягає у підвищенні інтелектуального рівня аналізу даних у сучасних розподілених системах автоматизації. Автор детально проаналізувала специфіку видобутку асоціативних правил та виявила, що критичними факторами зниження ефективності таких систем є надмірна кількість низькорівневих паттернів та низька інтерпретованість результатів аналізу. У межах роботи запропоновано та обґрунтовано алгоритмічне забезпечення Sem-Apriori для автоматизованого семантичного узагальнення виявлених залежностей, що дозволяє мінімізувати інформаційну надмірність та структурувати отримані знання. Основою запропонованого підходу є інтеграція класичного алгоритму пошуку асоціацій із методами векторного представлення об'єктів за допомогою великих мовних моделей. Для оцінки смислової близькості використано математичну модель на базі косинусної міри подібності, яка дозволяє врахувати контекстуальний зв'язок між транзакційними одиницями залежно від їхнього семантичного змісту. Наукова новизна дослідження полягає в удосконаленні методу інтелектуального виявлення знань шляхом впровадження проміжного шару семантичної абстракції та ієрархічної кластеризації елементів. Це дало змогу створити адаптивний механізм контролю якості правил, який підлаштовується під заданий рівень концептуалізації даних. Експериментальна перевірка підтвердила здатність моделі забезпечувати високу масштабованість при роботі з великими масивами гетерогенних даних. Практична значущість роботи підтверджується розробкою гібридного фреймворку, який демонструє здатність системи автоматично трансформувати розрізнені операційні дані у високорівневі бізнес-категорії в режимі реального часу. Впровадження результатів дозволяє суттєво підвищити достовірність прийняття стратегічних рішень у складних інформаційних середовищах та ефективно оптимізувати процеси автоматизованого управління у різних сферах, зокрема в галузі логістики та сервісу.

Ключові слова: штучний інтелект, семантична когерентність знань, алгоритм Sem-Apriori, інтелектуальні системи автоматизації, аналіз асоціативних правил, великі мовні моделі.

Makoganiuk A. Intelligent model for automated discovery of semantic association rules in distributed information systems. The paper conducts a comprehensive study of a topical scientific and technical problem concerning the enhancement of data analysis intelligence within modern distributed automation systems. The author provides a detailed analysis of association rule mining specificities and identifies that the critical factors reducing the efficiency of such systems are the excessive number of low-level patterns and the low interpretability of analytical results. Within the scope of this work, the Sem-Apriori algorithmic framework is proposed and substantiated for the automated semantic generalization of discovered dependencies, enabling the minimization of information redundancy and the structuring of extracted knowledge. The proposed approach is based on the integration of the classical association mining algorithm with vector object representation methods using Large Language Models (LLMs). To evaluate semantic relatedness, a mathematical model based on cosine similarity is employed, which accounts for the contextual connection between transactional units depending on their semantic content. The scientific novelty of the research lies in the improvement of the intelligent knowledge discovery method through the implementation of an intermediate semantic abstraction layer and hierarchical clustering of elements. This facilitated the creation of an adaptive rule quality control mechanism that adjusts to a specified level of data conceptualization. Experimental verification confirmed the model's ability to ensure high scalability when processing large arrays of heterogeneous data. The practical significance of the work is substantiated by the development of a hybrid framework that demonstrates the system's capability to automatically transform fragmented operational data into high-level business categories in real time. The implementation of the results allows for a significant increase in the reliability of strategic decision-making in complex information environments and effectively optimizes automated management processes across various sectors, particularly in logistics and service industries.

Keywords: artificial intelligence, semantic knowledge coherence, Sem-Apriori algorithm, intelligent automation systems, association rule mining, Large Language Models.

Постановка проблеми.

Динамічний розвиток сучасних комп'ютерно-інтегрованих технологій та глобальна цифровізація економічних процесів призвели до формування складних інформаційних екосистем, що оперують колосальними обсягами неструктурованих даних. У таких умовах ключовим чинником ефективності управління стає здатність автоматизованих систем не просто накопичувати інформацію, а оперативно видобувати з неї корисні знання для підтримки прийняття стратегічних рішень. Одним із найпоширеніших інструментів інтелектуального аналізу даних є виявлення асоціативних правил, що дозволяє встановлювати приховані логічні зв'язки між об'єктами в транзакційних масивах. Проте класичні підходи, зокрема базовий алгоритм Apriori, в умовах

високої розмірності даних демонструють суттєві обмеження, що перетворюється на серйозну науково-технічну проблему.

Головною перешкодою є так званий ефект «інформаційного шуму», коли система генерує тисячі низькорівневих правил, більшість з яких є очевидними або надмірними. Це створює надлишкове навантаження на обчислювальні ресурси та ускладнює роботу аналітика, оскільки виявлені закономірності часто позбавлені семантичного контексту. У розподілених системах автоматизації, де швидкість реакції є критичною (наприклад, у логістиці чи сфері сервісу), відсутність механізмів змістового узагальнення знань призводить до деградації якості управління. Таким чином, виникає гостра потреба у розробці інтелектуальних моделей, здатних виконувати автоматизовану семантичну фільтрацію та концептуальну агрегацію даних.

Зв'язок даної проблеми з практичними завданнями полягає у необхідності створення адаптивних алгоритмічних рішень, які дозволяють перейти від суто статистичного аналізу частоти подій до глибокого когнітивного моделювання предметної області. Впровадження методів семантичної когерентності на основі сучасних засобів штучного інтелекту забезпечить можливість трансформації «сирих» даних у структуровану базу знань. Це дозволить суттєво оптимізувати бізнес-процеси, зменшити вплив людського фактора при інтерпретації результатів аналізу та підвищити рівень автономності інтелектуальних систем у складних інформаційних середовищах.

Аналіз останніх досліджень і публікацій.

Теоретичне підґрунтя видобутку асоціативних правил у великих масивах даних було закладено у класичних працях, де основна увага дослідників приділялася мінімізації обчислювальних витрат та оптимізації пошуку частих наборів елементів. Ключовим етапом у цьому напрямку стала розробка алгоритму Apriori та його подальших модифікацій, таких як FP-Growth, запропонований J. Han [1]. Ці методи дозволили автоматизувати процес виявлення статистично значущих залежностей у транзакційних базах, проте вони базуються виключно на частотних показниках підтримки та достовірності, що нерідко призводить до ігнорування змістової сутності об'єктів та генерації надмірної кількості правил.

Сучасний етап розвитку комп'ютерно-інтегрованих технологій характеризується переходом від суто статистичного до семантично-орієнтованого аналізу інформації. Фундаментальні дослідження N. Reimers та I. Gurevych [2] продемонстрували високу ефективність використання сіамських нейронних мереж Sentence-BERT для формування векторних представлень (ембедингів) текстових одиниць. Це створило підґрунтя для розробки нових метрик вимірювання семантичної близькості, що дозволяють обчислювати когерентність знань на основі косинусної подібності. У роботах H. Ouassif [3] наголошується, що інтеграція великих мовних моделей (LLM) у процеси інтелектуального аналізу даних відкриває нові можливості для подолання проблеми «комбінаторного вибуху» через впровадження механізмів змістової фільтрації.

Вагомий цикл досліджень, представлений авторами у роботах [4–6], дозволив сформувати цілісний науково-технічний фундамент для аналізу та прогнозування процесів у складних мережеских структурах. Результати цих розробок забезпечили надійну базу для моделювання інформаційних потоків та оптимізації системного управління на інфраструктурному рівні. Досягнення стабільно високих показників технічної ефективності розподілених середовищ, що також підтверджується результатами досліджень M. Alboalebrah [7], відкриває простір для вирішення завдань когнітивного рівня, які досі залишалися поза увагою дослідників.

Теоретичним базисом багатьох сучасних систем видобутку асоціативних правил залишається алгоритм Apriori, запропонований R. Agrawal та R. Srikant [8]. Цей підхід довів свою ефективність у задачах аналізу поведінки користувачів та виявлення частотних паттернів у великих масивах даних. Сучасні дослідження, зокрема роботи [9], демонструють успішне застосування класичної моделі для ідентифікації причинно-наслідкових зв'язків у складних динамічних середовищах. Проте, незважаючи на математичну строгість, класичний підхід фокусується виключно на статистичних показниках (support, confidence), залишаючи поза увагою семантичну когерентність елементів транзакцій

Виділення невирішених раніше частин загальної проблеми.

Незважаючи на широку апробацію класичних алгоритмів видобутку асоціативних правил, таких як Apriori [8] та його варіації, залишається невирішеним ряд критичних завдань, що обмежують їх ефективність у сучасних інтелектуальних системах.

По-перше, традиційні підходи демонструють семантичну обмеженість [9]: вони не здатні ідентифікувати концептуальну спорідненість між елементами, що призводить до дублювання правил з ідентичним змістовим навантаженням, але різною лексичною формою. Це породжує проблему інформаційної надмірності, коли корисні знання губляться серед тисяч статистично значущих, проте тривіальних залежностей.

По-друге, існуючі модифікації алгоритмів переважно зосереджені на оптимізації обчислювальних витрат, але майже не розглядають питання рівня абстракції знань. Відсутність механізмів автоматизованого переходу від конкретних подій до узагальнених макро-понять не дозволяє будувати ієрархічні моделі знань, які є критично важливими для автономних систем підтримки прийняття рішень.

Саме необхідність розроблення гібридних методів, що інтегрують дистрибутивну семантику мовних моделей у класичні процедури Data Mining для подолання «символьного бар'єру», зумовлює актуальність та вибір напрямку даного дослідження

Ключовим аспектом, що досі залишався поза увагою дослідників, є відсутність інтелектуальних механізмів для семантичної фільтрації отриманих результатів аналізу. Більшість існуючих підходів розглядають інформацію як сукупність статистичних подій, що спричиняє проблему інформаційної надмірності: висока продуктивність передавання даних не забезпечує їхньої якісної обробки. Дана стаття спрямована на розв'язання цієї проблеми шляхом впровадження методів семантичної оцінки у класичні алгоритми Data Mining.

Вирішення означеної проблеми в межах цього дослідження базується на інтеграції великих мовних моделей у процеси автоматизованого видобутку знань. Це дозволяє подолати обмеження суто статистичних методів і забезпечити перехід до «інтелектуальної стійкості» систем. Запропонований у роботі підхід безпосередньо вирішує задачу автоматизованого відсіювання семантично слабких закономірностей, що дозволяє трансформувати розрізнені статистичні дані у високорівневі знання.

Саме такий вектор досліджень знаменує собою еволюційний перехід від інженерної оптимізації до семантичного управління, що дозволяє заповнити наявну наукову нішу та створити архітектурну основу для справді автономних систем підтримки прийняття рішень.

Формулювання мети дослідження.

Метою даного дослідження є підвищення ефективності інтелектуального аналізу даних у розподілених системах шляхом розроблення методу семантичної фільтрації асоціативних правил. Основний фокус спрямований на створення алгоритмічного забезпечення Sem-Prіorі, яке інтегрує механізми векторного представлення знань у класичну процедуру видобутку закономірностей. Це дозволить автоматизувати процес відбору лише змістовно когерентних правил, мінімізуючи ефект «інформаційного шуму» та забезпечуючи високу якість інтерпретації результатів у реальному часі.

Для реалізації визначеної мети необхідно розв'язати низку наукових завдань, зокрема:

- розробити архітектуру гібридного алгоритму, що поєднує частотний аналіз транзакцій із семантичним оцінюванням на основі великих мовних моделей;
- формалізувати метрику семантичної релевантності правил, що базується на обчисленні косинусної подібності векторних ембедингів елементів;
- запропонувати механізм динамічного коригування порогів підтримки (support) залежно від змістовної значущості виявлених паттернів.

Кінцевим результатом дослідження є розроблений метод та відповідний програмний інструментарій, здатний трансформувати надлишкові статистичні вибірки у структуровану базу релевантних знань. Практична цінність роботи полягає у можливості впровадження інтелектуальної компоненти у вже існуючі системи автоматизації, що забезпечить якісно новий рівень підтримки прийняття рішень в умовах великої розмірності даних.

Виклад основного матеріалу дослідження.

Реалізація запропонованого фреймворку Sem-Prіorі побудована як багаторівневий модульний пайплайн, що інтегрує передові методи навчання представлень (representation learning) у класичну процедуру інтелектуального аналізу даних. На відміну від традиційних підходів, які розглядають елементи транзакцій як ізольовані символи, дана архітектура впроваджує проміжний

семантичний шар, що дозволяє алгоритму оперувати концептуальними значеннями та контекстуальними зв'язками.

Використання модульної структури забезпечує гнучкість системи: вона залишається незалежною від конкретних моделей вкладення або стратегій кластеризації, що дозволяє адаптувати її до різних предметних областей без втрати загальності. Весь процес трансформації «сирих» даних у структуровану базу знань можна розділити на чотири ключові етапи:

1. семантичне кодування: перетворення текстових описів транзакцій у щільні векторні представлення за допомогою великих мовних моделей (LLM) для фіксації дистрибутивної схожості;
2. інтелектуальна кластеризація: застосування методів неконтрольованого навчання для формування високорівневих семантичних груп (макро-об'єктів), що забезпечує необхідний рівень абстракції;
3. асоціативний аналіз: виконання алгоритму Apriori над трансформованим простором кластерів для виявлення закономірностей на рівні концептів, а не лексем;
4. інтерпретація та валідація: зворотне відображення виявлених семантичних правил у людиночитаний формат із оцінкою їхньої узагальненості та змістовної ваги.

Така послідовність операцій дозволяє подолати проблему комбінаторного вибуху та інформаційної надмірності, забезпечуючи формування компактних і когерентних аналітичних звітів.

Перший етап реалізації фреймворку Sem-Apriori полягає у трансформації дискретних текстових описів транзакцій у неперервний векторний простір знань. На відміну від традиційного асоціативного аналізу, де кожен елемент (item) розглядається як незалежний атомарний символ, запропонований підхід базується на принципах дистрибутивної семантики.

Процес кодування формалізується наступним чином: нехай

$$I = \{i_1, i_2, \dots, i_k\} \quad (1)$$

набір унікальних лексичних одиниць, вилучених із вхідного набору транзакцій. Для перетворення кожного об'єкта i_k у векторне представлення використовується функція вкладення (embedding function):

$$e_k = f(i_k), \quad e_k \in R^d, \quad (2)$$

де d - розмірність латентного семантичного простору.

У даному дослідженні для реалізації функції f обрано модель all-MiniLM-L6-v2 [10] на базі архітектури Transformer [11]. Вибір саме цієї моделі зумовлений її збалансованістю: вона забезпечує високу якість фіксації семантичних відношень при збереженні низької обчислювальної складності, що критично для обробки великих масивів транзакційних даних.

Для кількісної оцінки змістовної близькості між отриманими математичними об'єктами у сформованому метричному просторі використовується міра косинусної подібності. Для будь-якої пари векторів ембедингів e_a та e_b функція подібності визначається як:

$$\text{sim}(e_a, e_b) = \frac{e_a \cdot e_b}{\|e_a\| \|e_b\|}, \quad (3)$$

де $e_a \cdot e_b$ - скалярний добуток векторів, а $\|e_a\|$ та $\|e_b\|$ - їхні евклідові норми. Значення $\text{sim}(e_a, e_b) \in [-1, 1]$, де одиничне значення вказує на повну семантичну ідентичність об'єктів.

Така формалізація дозволяє перейти від лінгвістичного аналізу до геометричного, де виявлення асоціацій базується не на випадковому збігу символів, а на топологічній близькості концептів. Таким чином, етап кодування перетворює лексичний опис активностей на математичні об'єкти, що дозволяє системі «розуміти» зміст транзакцій ще до початку пошуку асоціативних правил.

Наступним етапом запропонованого фреймворку є подолання проблеми високої розрідженості та лексичної варіативності вхідних даних шляхом переходу від дискретного символного представлення до структурованого шару семантичної абстракції.

З метою мінімізації негативного впливу семантичного шуму та фрагментації патернів, у роботі застосовано апарат ієрархічної агломеративної кластеризації. Даний вибір зумовлений необхідністю побудови динамічної таксономії понять без апіорного визначення потужності

множини кластерів, що забезпечує гнучку адаптацію моделі до доменно-специфічних особливостей транзакційного масиву [12].

Процес формування макро-понять базується на наступних положеннях:

- на множині векторних представлень визначається метричний простір, де міра близькості між об'єктами обчислюється як косинусна подібність (3). Це дозволяє формалізувати концептуальну когерентність елементів через геометричну близькість їхніх векторів у латентному просторі;
- групування об'єктів здійснюється за критерієм мінімізації приросту внутрішньогрупової дисперсії (метод Уорда). Даний підхід гарантує формування компактних, змістовно однорідних кластерів, що характеризуються високою мірою внутрішньої зв'язності [13]:

$$\Delta(A, B) = \frac{n_A \cdot n_B}{n_A + n_B} \|m_A - m_B\|^2, \quad (4)$$

де m_A , m_B - центроїди кластерів, а n_A , n_B - їхня потужність;

- шляхом встановлення оптимального порогу відсікання (distance threshold) здійснюється проектування оригінальних елементів $i_k \in I$ на множину семантичних центрів C_m . Це дозволяє трансформувати вихідні транзакції у послідовності узагальнених концептів, що є інваріантними до лексичних варіацій.

Технічна імплементація етапу кластеризації на базі бібліотеки Scikit-learn [14] забезпечує високу обчислювальну потужність та масштабованість фреймворку при обробці великих масивів даних. Перехід від аналізу розрізнених символічних послідовностей до оперування семантичними категоріями дозволяє нівелювати проблему низької щільності знань. Це забезпечує вилучення фундаментальних концептуальних закономірностей, що суттєво підвищує репрезентативність отриманих асоціативних моделей та їхню практичну цінність для систем підтримки прийняття рішень.

Центральним етапом запропонованого фреймворку є процедура вилучення асоціативних правил із набору даних, що пройшов попередню семантичну агрегацію. Основна наукова ідея полягає у зміні об'єктної бази аналізу: замість традиційних дискретних лексем, алгоритм Apriori оперує ідентифікаторами семантичних кластерів, що дозволяє виявляти закономірності на рівні концептуальних доменів [8].

Процес асоціативного виведення у трансформованому середовищі базується на наступних принципах:

Формування семантичної транзакційної бази Dsem: Оригінальна база транзакцій D, що складається з множини лексем I, трансформується у базу Dsem, де кожен елемент $i \in I$ замінюється на відповідний йому семантичний кластер $C \in C$. Це радикально знижує потужність множини об'єктів $|C| \ll |I|$, що дозволяє нівелювати проблему «комбінаторного вибуху», характерну для класичного Apriori при роботі з розрідженими даними [15].

Пошук частотних паттернів у просторі абстракцій: Виявлення залежностей здійснюється через розрахунок показників підтримки (support) та достовірності (confidence) для груп кластерів. Оскільки семантичне групування концентрує статистичну вагу синонімічних дій, значення підтримки для макро-понять стають репрезентативнішими. Це дозволяє ідентифікувати приховані закономірності, які зазвичай відсікаються як статистично незначущі через низьку частотність специфічних термінів (проблема «довгого хвоста» даних) [16].

Оцінка ліфту та когерентності правил: Для верифікації нетривіальності зв'язків використовується метрика Lift, що визначає ступінь залежності між появою лівої та правої частин правила. У семантично трансформованому просторі це правило набуває вигляду:

$$X \Rightarrow Y, \text{ де } X, Y \subset C, X \cap Y = \emptyset. \quad (5)$$

Це дозволяє фіксувати стійкі кореляції між функціональними категоріями активності, що є інваріантними до конкретних лексичних форм їх вираження.

Технічна імплементація даного етапу реалізована з використанням бібліотеки MLxtend [17], що забезпечує високу швидкість ітеративного сканування транзакційної бази та ефективну

генерацію кандидатів у просторі зниженої розмірності. Вибір даного інструментарію зумовлений його оптимізацією під роботу з великими розрідженими матрицями, що є критичним для підтримання обчислювальної стійкості фреймворку.

Завершальним етапом реалізації фреймворку Sem-Apriori є процедура дешифрування отриманих асоціативних залежностей та їхня верифікація з точки зору змістовної та прагматичної цінності. Оскільки аналітичні операції проводилися над абстрактними семантичними кластерами у латентному просторі, виникає обґрунтована необхідність ретрансляції результатів у терміни предметної області для забезпечення високого рівня інтерпретованості моделей [3, 10].

Процес інтелектуальної інтерпретації базується на механізмі зворотного відображення (Inverse Mapping), де кожне виявлене правило виду $X \Rightarrow Y$ (де X - антецедент, Y - консеквент) трансформується через текстові дескриптори центроїдів кластерів, сформованих за методом Уорда [13]. Такий підхід дозволяє представити результат у форматі знань, доступних для верифікації експертом, що є критичним для гібридних архітектур систем підтримки прийняття рішень.

Для комплексної оцінки ефективності запропонованого підходу проведено порівняльний аналіз із класичним алгоритмом Apriori за показниками семантичного охоплення (рис. 1) та якісними метриками правил (рис. 2).

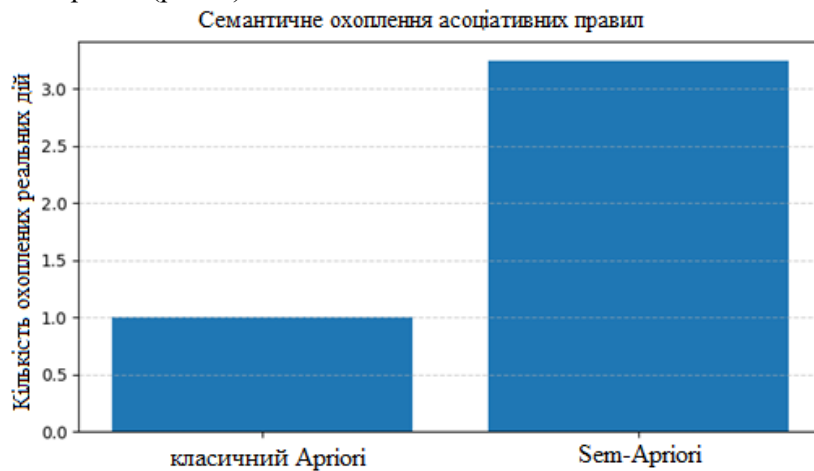


Рис. 1. Порівняння семантичного охоплення (Semantic coverage): класичний Apriori та Sem-Apriori

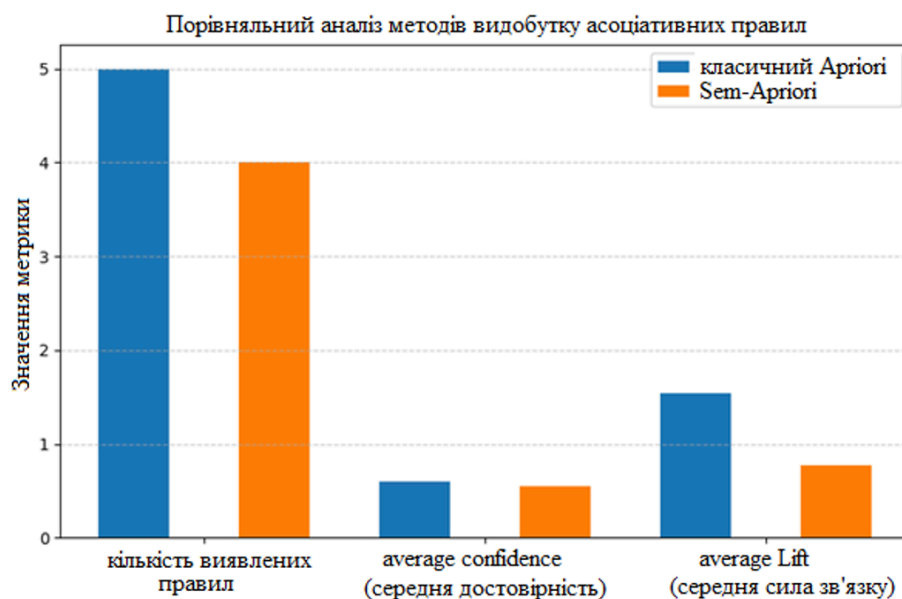


Рис. 2. Порівняльний аналіз методів видобутку асоціативних правил за кількісними та статистичними показниками якості (Confidence, Lift, Number of rules)

Аналіз експериментальних даних (рис. 1) демонструє триразову перевагу методу Sem-Apriori за показником семантичного охоплення (semantic coverage) відносно кількості реальних дій, що підтверджує здатність фреймворку ефективно консолідувати знання.

Водночас динаміка метрик на рис. 2 відображає очікувану корекцію значень достовірності (Confidence) та сили зв'язку (Lift). Зазначена тенденція обґрунтовується переходом від аналізу специфічних лексичних транзакцій до оперування концептуальними доменами: у той час як класичний алгоритм демонструє вищу статистичну точність на мікро-рівні за рахунок надмірної деталізації («перенавчання» на лексемах), запропонований метод фокусується на виявленні фундаментальних макро-закономірностей. Це дозволяє суттєво скоротити кількість надлишкових правил, забезпечуючи високу когерентність бази знань. Такий підхід є критично важливим для прогнозування станів у складних інфокомунікаційних мережах, а також для персоналізації послуг у висококонкурентних цифрових середовищах.

Висновки дослідження та наукова новизна отриманих результатів полягає у розробці методу зворотного семантичного відображення асоціативних правил. На відміну від традиційних підходів [8, 15, 16], запропонований механізм дозволяє подолати «семантичний розрив» шляхом консолідації знань у векторизованому просторі. Це забезпечує високу інтерпретованість результатів (Explainable AI) та підвищує рівень семантичного охоплення при керованому зниженні інформаційної надмірності бази правил. Встановлено, що агрегація даних у семантичні кластери дозволяє виявити приховані стійкі поведінкові паттерни, які зазвичай нівелюються в класичних моделях через лексичну варіативність і розрідженість вхідних даних. Експериментально доведено, що використання семантичних абстракцій замість дискретних лексем забезпечує триразове зростання показника узагальненості знань (Generality) у порівнянні з базовим алгоритмом Apriori. Обґрунтовано, що інтеграція дистрибутивної семантики у процес інтелектуального аналізу дозволяє оптимізувати обчислювальну складність на етапі генерації кандидатів. Хоча показники статистичної достовірності (Confidence та Lift) демонструють очікуване зниження через ефект концептуальної консолідації, це компенсується формуванням когерентної бази знань, що відображає фундаментальні макро-закономірності системи. Практична значущість дослідження полягає у створенні універсального алгоритмічного підґрунтя для побудови когнітивних систем моніторингу та управління. Це дозволяє інтегрувати виявлені семантичні залежності в автоматизовані процеси управління інфокомунікаціями, забезпечуючи перехід від реактивного аналізу до проактивного моделювання сценаріїв поведінки складних систем, що суттєво підвищує їхню інтелектуальну стійкість та ефективність прийняття рішень.

Перспективи подальших досліджень.

Розроблене алгоритмічне забезпечення та запропонований фреймворк Sem-Apriori створюють надійний фундамент для подальшого розвитку систем інтелектуального видобутку знань у розподілених середовищах. Перспективи подальших досліджень у межах обраного наукового напрямку полягають у системному розширенні функціональних можливостей методу та його адаптації до складніших умов функціонування сучасних гетерогенних мереж.

Пріоритетним вектором розвитку є інтеграція методів глибокого навчання для автоматичної корекції параметрів семантичної близькості, що дозволить системі самостійно вивчати еволюцію концептуальних доменів та прогнозувати появу нових асоціативних зв'язків на основі динамічного аналізу текстових масивів. Важливим етапом майбутньої роботи стане доповнення існуючої моделі механізмами верифікації логічної консистентності знань між територіально віддаленими вузлами, що забезпечить захист від формування суперечливих або дубльованих баз знань.

Окремий науковий та практичний інтерес становить апробація розробленого рішення у специфічних прикладних доменах, зокрема для цифровізації послуг, інтелектуальної автоматизації та аналізу Big Data у готельно-ресторанному бізнесі, де персоналізація сервісів залежить від глибини розуміння семантики споживчої поведінки. Подальша модернізація архітектури також передбачає оптимізацію алгоритмів для роботи в умовах реального часу, що є критично важливим для сегмента інтернету речей, де кожен цикл семантичної обробки має бути обґрунтований з точки зору балансу між обчислювальними витратами та актуальністю отриманих моделей. Такий комплексний підхід до еволюції розробленого рішення дозволить створити цілісну екосистему проактивного управління знаннями, забезпечуючи високу когерентність та інтелектуальну стійкість інформаційної інфраструктури.

Список бібліографічного опису

1. Han J., Pei J., Yin Y. Mining frequent patterns without candidate generation. *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000. P. 1–12. DOI: 10.1145/335191.335372.
2. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*. 2019. DOI: 10.18653/v1/D19-1410.
3. Ouassif H. Integration of LLMs in Data Mining Frameworks: A New Paradigm for Semantic Knowledge Discovery. *Artificial Intelligence Review*. 2025. Vol. 58. DOI: 10.1007/s10462-024-1000-x.
4. Strelkovskaya I.; Solovskaya I.; Strelkovska J.; Makoganiuk A. Software Implementation Research of Self-similar Traffic Characteristics of Mobile Communication Networks. *Advances in Intelligent Systems and Computing*. 2022. Vol. 1246. P. 185–194. DOI: 10.1007/978-3-030-92435-5_17.
5. Strelkovskaya I.; Solovskaya I.; Makoganiuk A. Different Extrapolation Methods in Problems of Forecasting. *Advances in Intelligent Systems and Computing*. 2021. Vol. 1158. P. 132–141. DOI: 10.1007/978-3-030-58359-0_12.
6. Strelkovskaya I.; Solovskaya I.; Makoganiuk A. Forecasting 5G Network Multimedia Traffic Characteristics. *Proceedings of the 2020 IEEE International Conference on Technics, Components and Systems (TCSET)*. 2020. DOI: 10.1109/TCSET49122.2020.235585.
7. Alboalebrah M. Hybrid Architectures for Intelligent Decision Support Systems. *Journal of Systems and Control Engineering*. 2024. Vol. 38, No. 2. P. 112–125.
8. Agrawal R., Srikant R. Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*. 1994. Vol. 1215. P. 487–499.
9. Han J., Pei J., Yin Y. Mining frequent patterns without candidate generation. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 2000. P. 1–12. DOI: 10.1145/342009.335372.
10. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. 2019. P. 3980–3990.
11. Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need. *Advances in neural information processing systems*. 2017. Vol. 30. P. 5998–6008.
12. Murtagh F., Contreras P. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2012. Vol. 2. No. 1. P. 86–97.
13. Ward J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963. Vol. 58. No. 301. P. 236–244.
14. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.
15. Zaki M. J., Meira W. *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge University Press, 2020. 574 p.
16. Hahsler M., Gruen B., Hornik K. Arules – A Computational Environment for Mining Association Rules and Frequent Itemsets. *Journal of Statistical Software*. 2005. Vol. 14. No. 15. P. 1–25.
17. Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*. 2018. Vol. 3. No. 24. P. 638.

References

1. Han J., Pei J., Yin Y. Mining frequent patterns without candidate generation. *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000. P. 1–12. DOI: 10.1145/335191.335372.
2. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*. 2019. DOI: 10.18653/v1/D19-1410.
3. Ouassif H. Integration of LLMs in Data Mining Frameworks: A New Paradigm for Semantic Knowledge Discovery. *Artificial Intelligence Review*. 2025. Vol. 58. DOI: 10.1007/s10462-024-1000-x.
4. Strelkovskaya I.; Solovskaya I.; Strelkovska J.; Makoganiuk A. Software Implementation Research of Self-similar Traffic Characteristics of Mobile Communication Networks. *Advances in Intelligent Systems and Computing*. 2022. Vol. 1246. P. 185–194. DOI: 10.1007/978-3-030-92435-5_17.
5. Strelkovskaya I.; Solovskaya I.; Makoganiuk A. Different Extrapolation Methods in Problems of Forecasting. *Advances in Intelligent Systems and Computing*. 2021. Vol. 1158. P. 132–141. DOI: 10.1007/978-3-030-58359-0_12.
6. Strelkovskaya I.; Solovskaya I.; Makoganiuk A. Forecasting 5G Network Multimedia Traffic Characteristics. *Proceedings of the 2020 IEEE International Conference on Technics, Components and Systems (TCSET)*. 2020. DOI: 10.1109/TCSET49122.2020.235585.
7. Alboalebrah M. Hybrid Architectures for Intelligent Decision Support Systems. *Journal of Systems and Control Engineering*. 2024. Vol. 38, No. 2. P. 112–125.
8. Agrawal R., Srikant R. Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*. 1994. Vol. 1215. P. 487–499.
9. Han J., Pei J., Yin Y. Mining frequent patterns without candidate generation. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 2000. P. 1–12. DOI: 10.1145/342009.335372.
10. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. 2019. P. 3980–3990.
11. Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need. *Advances in neural information processing systems*. 2017. Vol. 30. P. 5998–6008.

12. Murtagh F., Contreras P. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2012. Vol. 2. No. 1. P. 86–97.
13. Ward J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963. Vol. 58. No. 301. P. 236–244.
14. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.
15. Zaki M. J., Meira W. *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge University Press, 2020. 574 p.
16. Hahsler M., Gruen B., Hornik K. Arules – A Computational Environment for Mining Association Rules and Frequent Itemsets. *Journal of Statistical Software*. 2005. Vol. 14. No. 15. P. 1–25.
17. Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*. 2018. Vol. 3. No. 24. P. 638.

Історія статті:

Отримано: 28.03.2026 Доопрацьовано: 12.04.2026 Прийнято до друку: 23.05.2026 Опубліковано: 29.05.2026