

DOI <https://doi.org/10.30525/978-9934-26-172-5-7>

АНАЛІЗ ОСОБЛИВОСТЕЙ СИСТЕМ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Одарущенко О. Б.

*кандидат технічних наук, доцент,
доцент кафедри інформаційних систем і технологій
Полтавський державний аграрний університет*

Протас Н. М.

*кандидат сільськогосподарських наук, доцент,
доцент кафедри інформаційних систем і технологій
Полтавський державний аграрний університет*

Дегтярьова Л. М.

*кандидат технічних наук, доцент,
доцент кафедри інформаційних систем і технологій
Полтавський державний аграрний університет
м. Полтава, Україна*

У зв'язку з високою обчислювальною складністю отримання знань за допомогою звичайної технології інженерії знань, протягом останніх кількох років інтенсивно розробляються методи автоматичного отримання знань з накопичених фактів. Актуальність пошуку нових підходів з аналізу даних обумовлена наступними чинниками: дані мають необмежений обсяг; дані є різномірними (кількісними, якісними та текстовими); результати обробки даних мають бути конкретні і зрозумілі; інструменти для обробки сирих даних повинні бути прості у використанні [1–5].

Мета дослідження полягає у проведенні аналізу особливостей систем інтелектуального аналізу даних.

Інтелектуальний аналіз даних є коротким позначенням досить широкого спектру процедур автоматичного аналізу даних. Загалом процес інтелектуального аналізу даних складається з трьох стадій:

1. Виявлення закономірностей (вільний пошук);
2. Використання виявлених закономірностей для передбачення невідомих значень (прогностичне моделювання);
3. Аналіз винятків, призначений для виявлення та тлумачення аномалій у знайдених закономірностях.

Іноді у явному вигляді виділяють проміжну стадію перевірки достовірності знайдених закономірностей між їх знаходженням та використанням (стадія валідації).

Усі методи інтелектуального аналізу поділяються на дві великі групи по принципу роботи з вихідними навчальними даними:

1. Методи міркувань на основі аналізу прецедентів – вихідні дані можуть зберігатися в явному деталізованому вигляді та можуть безпосередньо використовуватися для прогностичного моделювання та/або аналізу винятків.

Головною проблемою цієї групи методів є проблематичність їх використання на великих обсягах даних, хоча саме під час аналізу інформації з великих сховищ даних методи інтелектуального аналізу є найбільш ефективними.

2. Методи, у яких інформація спочатку добувається з первинних даних і перетворюється на деякі формальні конструкції (їх вид залежить від конкретного способу).

Відповідно до попередньої класифікації, цей етап виконується на стадії вільного пошуку, яка у методів першої групи в принципі відсутня. Таким чином, для прогностичного моделювання та аналізу винятків використовуються результати цієї стадії, які більш компактні, ніж самі масиви вихідних даних. При цьому отримані конструкції можуть бути або “прозорими” (інтерпретованими), чи “чорними ящиками” (не трактованими).

Розрізняють такі підходи до відбору параметрів для пошуку функціональних залежностей та виду (або вибору) класу самих залежностей:

1. Користувач сам висуває гіпотези щодо залежностей між даними. Фактично традиційні технології аналізу розвивали цей підхід. Це справедливо в тому випадку, коли процес пошуку повністю контролюється людиною. У багатьох системах інтелектуального аналізу даних у цьому процесі автоматизовано перевірку достовірності гіпотез, що дозволяє оцінити ймовірність тих чи інших залежностей у базі даних.

2. Залежності між даними знаходяться автоматично.

Процес пошуку залежностей розпадається на три етапи:

1. Виявлення залежностей (discovery) полягає у перегляді бази даних із метою автоматичного виявлення залежностей. Проблема тут полягає у відборі дійсно важливих залежностей з величезної кількості існуючих у базі даних.

2. Прогнозування (predictive modelling) передбачає, що користувач може пред'явити системі запис із незаповненими нулями і запитати

відсутні значення. Система сама аналізує вміст бази та робить правдоподібним передбачення щодо цих значень.

3. Аналіз аномалій (forensic analysis). Відбувається процес пошуку підозрілих даних, що сильно відхиляються від стійких залежностей.

В роботі наведено особливості систем інтелектуального аналізу даних, наведені основні принципи їх роботи. Наведені переваги та недоліки відомих підходів з інтелектуального аналізу даних, наведені принципи та основні процедури роботи відомих підходів з інтелектуального аналізу даних.

Література:

1. Макаренко С. И. Интеллектуальные информационные системы: учебное пособие. – Ставрополь: СФ МГГУ им. М. А. Шолохова, 2009. – 206 с.: ил..

2. Pievtsov, H., Turinskiy, O., Zhyvotovskiy, R., Sova, O., Zvieriev, O., Lanetskii, B., and Shyshatskiy, A. Development of an advanced method of finding solutions for neuro-fuzzy expert systems of analysis of the radioelectronic situation. EUREKA: Physics and Engineering. 2020, No. (4), pp. 78-89. <https://doi.org/10.21303/2461-4262.2020.001353>.

3. P. Zuiev, R. Zhyvotovskiy, O. Zvieriev, S. Hatsenko, V. Kuprii, O. Nakonechnyi, M. Adamenko, A. Shyshatskiy, Y. Neroznak, V. Velychko. Development of complex methodology of processing heterogeneous data in intelligent decision support systems. Eastern-European Journal of Enterprise Technologies. 2020, Vol. 4, No. 9 (106), pp. 14–23. DOI: <https://doi.org/10.15587/1729-4061.2020.208554>.

4. Шишацький А. В, Налапко О. Л., Одарущенко О. Б. Основні біоінспіровані алгоритми обробки різнотипних даних. Інтеграція інформаційних систем і інтелектуальних технологій в умовах трансформації інформаційного суспільства: тези доповідей IV Міжнародної науково-практичної конференції, що присвячена 50-ій річниці кафедри інформаційних систем та технологій. Полтава: ПДАУ, 2021. С. 109-114. <https://doi.org/10.32782/978-966-289-562-9>.

5. Одарущенко О. Б., Протас Н. М., Дегтярьова Л. М., Шишацький А. В. Розробка комплексної методики оцінки та прогнозування обстановки в інтелектуальних системах підтримки прийняття рішень. Theoretical and empirical scientific research: concept and trends: Collection of scientific papers «ЛОГОΣ» with Proceedings of the II International Scientific and Practical Conference (Vol. 1), Oxford, May 28, 2021. Oxford-Vinnitsia: P.C. Publishing House & European Scientific Platform, 2021. С. 209-212.