

МЕТОД ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ АКАДЕМІЧНИХ ДАНИХ НА ОСНОВІ ЛОГІЧНОЇ КЛАСИФІКАЦІЇ

Козирев А. Д.

Науковий керівник – к.т.н., доц., проф. Шубін І. Ю.

Харківський національний університет радіоелектроніки, каф. ПІ

м. Харків, Україна

e-mail: andrii.kozyriev@nure.ua.

The presented work studies in detail the application of the method of logical classification, which is a powerful tool for grouping students based on their academic needs and characteristics. This method allows you to distinguish groups of students with similar characteristics, based on the analysis of a wide range of academic indicators, such as average score, course, specialization, active participation in academic activities and other criteria that reflect their educational needs and preferences. By identifying student groups with similar needs, universities and colleges can develop specialized support programs that better meet the specific requirements and expectations of students.

Інтеграція технологій в освітній процес, у тому числі в аспекти персоналізованого навчання, просувається стрімко. Прогрес у сфері інформаційно-комунікаційних технологій дозволив ефективно обробляти великі обсяги даних у секторі освіти. Це, у свою чергу, дало змогу освітнім інституціям аналізувати зібрану інформацію про взаємодії зі студентами для вдосконалення освітнього процесу.

Термін "великі дані" описує інтеграцію структурованих даних з баз даних та неструктурованих даних з новітніх джерел, включно з соціальними мережами, мобільними пристроями, датчиками, інтелектуальними вимірювальними приладами та фінансовими системами [1]. Сучасні методології надають можливість організаціям аналізувати інформацію будь-якого роду, розміру або швидкості її надходження, щоб приймати більш виважені рішення на їх основі.

Для обробки значного масиву інформації все частіше використовуються методи видобування даних та аналізу великих даних. Метод видобування даних, відомий також як добування знань з баз даних, служить для виявлення непомітної інформації у великих масивах даних. Цей метод аналізує частини даних, щоб виявити повторювані патерни поведінки або прогнозовані моделі, які можуть бути отримані з аналізованої інформації.

Методика видобування даних фокусується на виявленні неявних знань з сирих даних через використання алгоритмів машинного навчання та інструментарію штучного інтелекту, розкриваючи раніше невідомі, незвичайні та важливі для практики дані, доступні для людського

розуміння. Спочатку задіяний з метою економічного аналізу, його можливості були успішно адаптовані для потреб освіти [2].

Одним із ключових напрямків застосування видобування даних є групування інформації за спільними атрибутами для ідентифікації загальних шаблонів, особливо серед студентських груп. Також, методи логічного групування ефективно використовуються у багатьох сферах, включаючи, але не обмежуючись, біологією, фізикою, метеорологією. Важливим аспектом цих методів є стадія розробки математичних моделей, з окремим акцентом на особливості оброблюваних даних. Зазвичай, для аналізу використовується логіка висловлювань, проте в цьому дослідженні пропонується підхід, базований на алгебрі скінченних предикатів.

Розглянемо структуру завдань, заснованих на застосуванні рівнянь з предикатами. Змінні, такі як y_1, y_2, \dots, y_l , представляють специфічні атрибути об'єктів. Це може охоплювати, наприклад, структуроване визначення критеріїв для аналізу певних ситуацій, таких як критерії оцінювання якості або класифікація суб'єктів згідно з їх значенням. На відміну від булевих змінних, що приймають лише два можливих значення, предикативні змінні мають змогу набувати широкий спектр значень зі своїх дискретних доменів значень.

У ситуаціях, коли дискретні змінні x_1, x_2, \dots, x_n відображають атрибути, які дозволяють оцінити потенційні властивості об'єктів, ці атрибути та їх особливості можуть бути пов'язані між собою складними логічними зв'язками. Це дає можливість виражати зв'язки між атрибутами за допомогою предикатних рівнянь типу:

$$P(y_1, y_2, \dots, y_l; x_1, x_2, \dots, x_n) = 1. \quad (1)$$

Ключові параметри, що описують виконання студентами практичних робіт, охоплюють:

- унікальний ідентифікатор студента, рівень освіти (курс або показник завершення навчального закладу);
- загальна кількість завдань, які студент виконав, що дозволяє виміряти активність студента та його залученість у навчальний процес;
- відсоток завдань, виконаних правильно з усіх спроб, що є ключовим показником якості засвоєння матеріалу та ефективності навчання;
- середня кількість спроб, необхідних студенту для успішного розв'язання завдання, що вказує на ступінь складності завдань для студента та його здатність до вирішення проблем;
- середній рівень складності завдань, обраних студентом та успішно вирішених ним, який відображає здібності студента до виконання більш складних завдань та його академічні інтереси.

Таку інформацію можна отримати з бази даних освітнього порталу за допомогою базових SQL-запитів [3].

Задача класифікації зводиться до створення набору

$$C = \{c_1, c_2, \dots, c_k, \dots, c_g\}, \quad (2)$$

де c_k – клас, що містить схожі один на одного об'єкти з множини I ,

$$c_k = \{i_j, i_p \mid i_j \in I, i_p \in I, d(i_j, i_p) < \sigma\}, \quad (3)$$

де $d(i_j, i_p)$ – міра близькості між об'єктами, звана відстанню, σ – величина, яка визначає міру близькості для включення об'єктів в один клас (в ході експерименту її значення було підібрано емпірично).

Отже, мета полягає в класифікації студентської аудиторії за академічними показниками для наступного детального аналізу. Визначення числа груп (m), де можливим рішенням є поділ студентів на, наприклад, три групи: студенти з високими академічними досягненнями (високий середній бал), студенти середнього рівня (середній бал) та студенти, яким необхідна додаткова підтримка (низький середній бал). Вибір критеріїв для класифікації: для цього можна застосувати такі показники, як середній бал та кількість зарахованих кредитів. Реалізація методу класифікації: застосовуючи метод k -середніх, ініціюємо процес з трьома визначеними центрами груп, послідовно розподіляючи кожного студента до найбільш відповідної групи відповідно до обраних критеріїв. Оцінка ефективності класифікації: ефективність класифікації можливо оцінити через аналіз внутрішньогрупової відстані (сума відстаней від об'єктів до центру їх групи) та міжгрупової відстані (відстань між центрами груп).

Список використаних джерел:

1. John Walker S. Big data: a revolution that will transform how we live, work, and think. *International journal of advertising*. 2014. Vol. 33, no. 1. P. 181–183. URL: <https://doi.org/10.2501/ija-33-1-181-183> (date of access: 04.03.2024).

2. Kash B. R. P., Thappa D. M. H., Kavitha V. Big data in educational data mining and learning analytics. *International journal of innovative research in computer and communication engineering*. 2014. Vol. 02, no. 12. P. 7515–7520. URL: <https://doi.org/10.15680/ijrcce.2014.0212044> (date of access: 04.03.2024).

3. Kozyriev A., Shubin I. Method for solving quantifier linear equations for formation of optimal queries to databases. *CEUR workshop proceedings*. 2023. Vol. 3396. P. 449–459.