

ДОСЛІДЖЕННЯ МЕТОДІВ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ НАВЧАННЯ

Гайдаєнко О.В., к.т.н., доцент¹; Бандурко К.Ю. магістр²; Світенко І.В. магістр³

^{1,2,3} Національний університет кораблебудування імені адмірала Макарова

^{1,2,3} Україна, Миколаїв

¹ ORCID 0000-0002-6614-5443; ² ekaterina-homenk@ukr.net; ³ igorsvitenko5@gmail.com

Анотація. Бурхливий розвиток систем дистанційного навчання, впровадження цифрових технологій у шкільний навчальний процес призвело до появи великих масивів освітніх даних. Розглянуто методи аналізу освітніх даних та їх ексклюзивність у порівнянні з методами інтелектуального аналізу даних.

Ключові слова: освітні дані; аналіз; масив даних.

Вступна частина. Широке поширення систем дистанційного навчання надає дослідникам значно більший обсяг інформації порівняно із традиційним освітнім процесом. Пов'язано це як з активним використанням в електронному навчанні різних технологій збору даних, так і з більшим масштабом аудиторії електронних освітніх середовищ. Зростання обсягу даних сприяло появі в початку 2000-х років нового напрямку в галузі штучного інтелекту – аналізу освітніх даних [1]. Якість освіти розглядається як комплексний процес на який впливає велика кількість факторів. Отримання якісної освіти безпосередньо залежить від якості самих вимог (цілей, стандартів і норм), якості ресурсів (програми, кадровий потенціал, контингент абітурієнтів, матеріально-технічне забезпечення, фінанси тощо) та якості освітніх процесів (наукова та навчальна діяльність, управління, освітні технології), які безпосередньо забезпечують підготовку фахівців. Методи аналізу корисні для проектування освітнього середовища, в організації матеріалів навчальних курсів та управління освітніми ресурсами.

Метою роботи є підвищення ефективності навчання за рахунок прийняття управлінських рішень, які ґрунтуються на більш глибокому розумінні процесів і закономірностей присутніх в освітньому процесі.

Основна частина. В освітньому аналізі даних широко використовуються методи традиційного аналізу, основні із яких є: класифікація, кластеризація, пошук зв'язуючих правил, пошук послідовних шаблонів, а також інтелектуальний аналіз текстів (текстовий аналіз). В [2] показано який метод і в яких саме завданнях частіше всього використовується.

Перевага кластерного аналізу в тому, що він дозволяє проводити розбиття об'єктів не по одному параметру, а по цілому набору ознак. Крім того, кластерний аналіз, на відміну від більшості математико-статистичних методів, не накладає ніяких обмежень на вид розглянутих об'єктів і дозволяє розглядати безліч вихідних даних практично довільної природи.

Кластерний аналіз дозволяє розглядати досить великий обсяг інформації і різко скорочувати, стискати великі масиви інформації, робити їх компактними і наочними.

Формально задача кластеризації описується наступним чином.

Для множини об'єктів даних I , кожний з яких представлено набором атрибутів. Необхідно побудувати множину кластерів C і відображення F множини I на множину C , тобто $F : I \rightarrow C$. Відображення F задає модель даних, яка є рішенням задачі. Якість рішення задачі визначається кількістю правильно класифікованих об'єктів даних.

Множину I визначаємо наступним чином:

$$I = \{i_1, i_2, \dots, i_j, \dots, i_n\},$$

де i_j – досліджуємий об'єкт.

Прикладом такої множини може служити набір даних про оцінки по дисциплінам.

Кожний з об'єктів характеризується набором параметрів:

$$i_j = \{x_1, x_2, \dots, x_h, \dots, x_m\}.$$

В нашому прикладі такими параметрами є питома вага лекцій, практичних, індивідуальних і лабораторних занять у загальній кількості кредитних годин виділених на дисципліну.

Кожна змінна x_h може приймати значення з деякої множини:

$$x_h = \{v_h^1, v_h^2, \dots\}.$$

В даному прикладі значеннями є дійсні числа.

Задача кластеризації полягає в побудові множини:

$$C = \{c_1, c_2, \dots, c_k, \dots, c_g\}.$$

Тут c_k – кластер, містить схожі один на одного об'єкти з множини I :

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

де σ – величина, яка визначає міру близькості для включення об'єктів в один кластер; $d(i_j, i_p)$ – міра близькості між об'єктами, яка називається відстанню.

Невідомі значення $d(i_j, i_p)$ називаються відстанню між елементами i_j і i_p , якщо виконуються наступні умови:

- a) $d(i_j, i_p) \geq 0$, для всіх i_j і i_p ;
- b) $d(i_j, i_p) = 0$, тоді і тільки тоді, коли $i_j = i_p$;
- c) $d(i_j, i_p) = d(i_p, i_j)$;
- d) $d(i_j, i_p) \leq d(i_j, i_r) + d(i_r, i_p)$.

В останній час ведуться активні розробки нових алгоритмів кластеризації, які поєднують в собі методи ієрархічної кластеризації з іншими методами. Одним з таких алгоритмів є алгоритм BIRCH (Balanced Interactive Reducing and Clustering using Hierarchies) [3].

В ньому реалізовано двоетапний процес кластеризації.

В ході першого етапу формується попередній набір кластерів. На другому етапі до виявлених кластерам застосовуються інші алгоритми кластеризації – придатні для роботи в оперативній пам'яті.

Висновки. В результаті дослідження було прийнято рішення використовувати методи кластеризації та дерев рішень для подальшого їх використання у системі аналізу освітніх даних.

Реалізація даної системи дозволить підвищити ефективність навчання за рахунок прийняття управлінських рішень, які ґрунтуються на більш глибокому розумінні процесів і закономірностей присутніх в освітньому процесі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Михалко, М.І., Гайдаєнко, О.В. (2021). Автоматизована інформаційна система контролю та адміністрування відвідуваності університету. *Інформаційні технології: моделі, алгоритми, системи* :

II Всеукраїнської науково-практичної інтернет конференції (ITMAS – 2021) (С. 63-65). Миколаїв, Україна. <http://itconf.nuos.edu.ua/2021/proceedings/>

[2] Romero, C., Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*. 010, V. 40, № 6. P. 601-618. doi: 10.1109/TSMCC.2010.2053532

[3] Gupta, A. (2021, Jun 2). *Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) Algorithm in Machine Learning*. <https://medium.com/geekculture/balanced-iterative-reducing-and-clustering-using-hierarchies-birch-1428bb06bb38>

Haydaenko O.V., Bandurko K.Yu., Svitenko I.V.

Research of methods of increasing the efficiency of learning

Abstract. *The rapid development of distance learning systems, the introduction of digital technologies into the school educational process has led to the emergence of large arrays of educational data. Methods of educational data analysis and their exclusivity in comparison with methods of intellectual data analysis are considered.*

Keywords: *educational data; analysis; array of data.*