

## **ЗАДАЧІ ТА МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ОСВІТНІХ ДАНИХ**

*Досліджено основні задачі інтелектуального аналізу освітніх даних в рамках навчального процесу у закладах вищої освіти. Виявлено специфіку даних, що підлягають аналізу, встановлено основні цілі аналізу. Здійснено огляд та обґрунтовано актуальність застосування методів Data Mining у вищій професійній освіті.*

**Ключові слова:** інтелектуальний аналіз освітніх даних; навчальна аналітика; системи електронного навчання.

**Постановка проблеми.** Інтенсивний розвиток освітніх інформаційних технологій супроводжується накопиченням великих об'ємів цифрової інформації стосовно процесів навчання. Інноваційне оновлення засобів, методів та форм навчання супроводжується появою нових інструментальних засобів їх аналізу. Застосування методів комп'ютерної аналітики для аналізу електронних форм освітньої інформації з метою виявлення прихованих знань дозволяє розширити можливості вдосконалення вищої професійної освіти у країні.

**Аналіз останніх досліджень та публікацій.** Дослідження у сфері інтелектуального аналізу освітніх даних проводяться в рамках двох наукових напрямів: інтелектуальний аналіз освітніх даних (Educational Data Mining – EDM) та навчальна аналітика (Learning Analytics – LA). Наукові праці, присвячені різним аспектам інтелектуального аналізу освітньої інформації, друкуються в матеріалах щорічних міжнародних конференцій International Conference on Educational Data Mining (2008–2019) та Learning Analytics And Knowledge (2011–2019). Чисельні дослідження зарубіжних науковців (R. Baker [1], C. Chisholm [2], G. Cobo [3], S. Gowda [1], J. Morán [3], C. Romero [4], W. Shibin [5], M. Stewart [2], T. Feng [5], S. Ventura [4]) стосуються специфіки застосування методів інтелектуального аналізу (Data Mining – DM) в освітній сфері з метою надання допомоги студентам та викладачам, моделювання поведінки студентів, здійснення педагогічного прогнозування. Останніми роками до інтелектуального аналізу освітніх даних долучилися російські педагоги-науковці, досліджуючи ретроспективу та перспективу його розвитку (Н. Зоріна, В. Панченко [6]), застосування методів Data Mining для підтримки прийняття рішень в освітній сфері (Н. Горлушкіна, І. Коцюба, М. Хлопотов [7]), для системи оцінки якості освіти та адаптивного навчання (А. Веряев, Г. Татарнікова [8]), для підтримки сумісної мережевої діяльності суб'єктів навчального процесу (Є. Патаркін [9]).

Серед вітчизняних науковців, які проводять дослідження у сфері інтелектуального аналізу освітніх даних, можна відмітити роботи щодо дослідження сучасного стану та перспектив розвитку Educational Data Mining (Ю. Ковальчук [10]), підготовки майбут-

ніх фахівців з ІТ до здійснення навчальної аналітики (Л. Панченко [11]), оптимізації й аналізу результатів використання LMS Moodle засобами комп'ютерної аналітики (С. Петренко [12]), використання систем Data Mining для прогнозування педагогічних явищ та процесів (М. Коляда [13]).

Загалом дослідження у сфері інтелектуального аналізу освітніх даних серед науковців-педагогів України є розрізненими та поодинокими. Не вирішена проблема педагогічного обґрунтування освітніх вимірів, не достатньо дослідженими є проблеми впровадження інтелектуального аналізу в реальну освітню практику закладів вищої освіти. Що обумовлює необхідність проведення подальших наукових розробок стосовно застосування методів та розв'язання задач Data Mining у вищій професійній освіті з урахуванням особливостей освітньої системи нашої країни.

**Метою статті** є дослідження основних задач і методів інтелектуального аналізу освітніх даних та виявлення їх потенційних можливостей для підвищення ефективності навчального процесу в закладах вищої освіти України.

**Виклад основного матеріалу.** Інтелектуальний аналіз освітніх даних – це сукупність методів виявлення цих раніше невідомих, практично корисних знань про освітній процес і його учасників з метою підтримки прийняття рішень.

Мета, задачі та методи досліджень, які проводяться в рамках напрямів інтелектуального аналізу освітніх даних (Education Data Mining) та навчальної аналітики (Learning Analytics) багато в чому співпадають. Дослідники розрізняють їх у тому, що Education Data Mining більше зосереджений на автоматизації виявлення закономірностей в освітніх даних, тоді як Learning analytics більше націлена на підготовку даних у вигляді, придатному для їх аналізу людиною [14; 15].

Для дослідження у сфері інтелектуального аналізу освітніх даних використовують традиційні методи Data Mining: класифікацію, кластеризацію, виявлення взаємозв'язків, моделювання, пошук асоціативних правил та послідовних шаблонів, Text Mining (інтелектуальний аналіз текстів), Visual Mining. Однак реалізація цих методів в освітній сфері для аналізу цифро-

вих даних стосовно процесів навчання має свої особливості, обумовлені цілями аналізу та специфікою даних, які аналізуються.

Науковці зазначають, що основною метою інтелектуального аналізу освітніх даних є удосконалення освітнього процесу шляхом [16]:

- розвитку знань про педагогічні контексти електронного навчання;
- адаптивного вироблення рекомендацій студентам з вивчення матеріалу та забезпеченням зворотного зв'язку на основі аналізу ефектів взаємодії «викладач-студент» / «система-студент»;
- прогнозування поведінки студентів у процесі навчання на основі створення моделі студента, яка включає інформацію про його пізнавальні можливості, знання, мотиви;
- розробки нових моделей, форм, способів подання навчального матеріалу, які б відповідали різним стилям навчання та рівням пізнавальних можливостей, виявленим у студентів за допомогою методів інтелектуального аналізу даних.

Освітні дані, що підлягають аналізу, зазвичай мають складну структуру, представлені в різних системах навчання та є не звичними для працівників сфери освіти, оскільки вони є цифровими слідами, залишеними у логах та базах даних і стосуються дій студентів у середовищі електронного навчання [17]:

- сторінка, через яку студент заходить на сайт освітньої системи, і через яку залишає сайт;
- сторінки, які студенти відвідують найчастіше;
- браузер, який використовує студент, працюючи з електронною системою навчання;
- кількість відвідувань і кількість відвідувачів сторінки електронної системи навчання;
- частота відвідувань у часі (у вигляді часового ряду) – для сайту загалом та для окремої сторінки;
- географічне розташування місця, звідки студент входить до електронної системи навчання та час з'єднання;
- число відвідувань та їх тривалість для окремого студента за певний період часу;
- найбільш популярні ключові слова для пошуку інформації в системі;
- число переглядів/скачувань навчальних матеріалів;
- число різних елементів з навчальним контентом, переглянутих (відвіданих, прочитаних, скачаних) студентом за сеанс роботи або за більш тривалий період часу;
- статистичні індикатори спілкування на форумі освітньої системи (число постів, популярність обговорюваної теми, кількості звернень з питаннями до інших студентів та до викладача);
- бали, отримані студентом за виконання певного навчального завдання;
- обсяг навчального матеріалу, який студент вивчає перед виконанням завдання.

Застосування методів Data Mining для аналізу такого роду інформації дозволяє виявляти приховані закономірності та знання, які традиційними методами аналізу отримати бути не можуть. Кожен з методів призначений для розв'язання певної задачі, серед яких можна виділити задачі традиційного Data Mining

та задачі, специфічні саме для інтелектуального аналізу освітніх даних.

До основних задач традиційного Data Mining відносять наступні:

1. Класифікація – встановлення залежності між вхідними і дискретними вихідними змінними, які характеризують процес навчання, що дає можливість відносити об'єкти до одного зі заздалегідь відомих класів. Це дозволяє здійснювати класифікацію студентів залежно від їх успішності чи активності у системі електронного навчання. Класифікація засобів навчання залежно від частоти їх використання дозволяє викладачеві визначати оптимальну структуру навчального електронного курсу. Для розв'язання задачі класифікації застосовують методи дискримінантного аналізу, Naive Bayes, k-ближніх сусідів, дерева рішень, нейронні мережі.

2. Прогнозування – задача передбачення значення величини, яка цікавить дослідника (вихідної змінної), на основі відомих значень, що характеризують процес навчання та суб'єктів навчання. Для її розв'язання застосовуються методи математичної статистики, нейронні мережі, часові ряди. Зазвичай прогнозування здійснюється шляхом побудови моделі, яка дозволяє розраховувати чи передбачати значення вихідної змінної. Якщо вихідна змінна приймає неперервні значення, то ми маємо справу з регресією – пошуком залежності між змінними. Якщо вихідна змінна має кінцевий дискретний набір значень, то мова йде про задачу класифікації – віднесення об'єкту до того або іншого класу. У процесі аналізу освітніх даних це дозволяє моделювати та передбачати поведінку студентів у процесі навчання та встановлювати залежність між іншими величинами, що стосуються електронного навчання.

3. Кластеризація – групування об'єктів на основі даних, що описують їх сутність. Результатом кластеризації є поділ об'єктів, які стосуються навчання, на групи споріднених, схожих об'єктів – кластери. Задача кластеризації є логічним продовженням ідеї класифікації, однак при проведенні кластеризації кількість кластерів заздалегідь невідома і визначається у процесі аналізу. Розв'язок задачі кластерного аналізу допомагає краще зрозуміти дані та виявити їх структуру. Це дозволяє здійснювати більш глибокий аналіз освітніх даних з використанням методів ієрархічного кластерного аналізу, методів k-середніх та с-середніх.

4. Пошук асоціативних правил та послідовностей дозволяє виявляти взаємозв'язки між пов'язаними подіями у наборі освітніх даних. Розв'язання цієї задачі дозволяє виявляти правила виду «якщо умова, то наслідок», де «умова» та «наслідок» є подіями, які відбуваються у середовищі електронного навчання й мають високу ймовірність одночасної або послідовної появи. Пошук асоціативних правил здійснюється найчастіше з використанням алгоритму (Апріорі). Під час аналізу освітніх даних виявлення асоціативних правил дає можливість, на основі виявленого зв'язку, наприклад, між діями студента у середовищі електронного навчання, моделювати та пропонувати студентам більш оптимальні стратегії вивчення курсу.

5. Візуалізація, Visual Mining – створення графічного образу аналізованих даних шляхом перетворення

великих масивів цифрових даних, накопичених у системах електронного навчання, у доступну для розуміння та сприйняття інформацію. Для вирішення задачі візуалізації використовуються графічні методи відображення складної, багатомірної інформації, що показують наявність закономірностей в освітніх даних та використовують спеціальні засоби, легкі для сприйняття людиною їх сутності.

Науковці, які проводять дослідження у сфері інтелектуального аналізу освітніх даних, виділяють також специфічні задачі Data Mining [14]: 1) відкриття за допомогою моделей (Discovery with Models) 2) перегонка даних для прийняття рішень людиною (Distillation of Data for Human Judgment).

Відкриття за допомогою моделі припускає, що створення прогнозованої моделі є проміжною ланкою у вирішенні іншої проблеми. Наприклад, знання студента можуть бути оцінені побічно за непрямими ознаками (за частотою й тривалістю звертання до навчальних матеріалів і т. п.). На основі цього може бути побудована модель, яка прогнозує майбутню підсумкову оцінку. Це дає можливість виявити спроби студента приховати недостатні знання з певної теми, використовуючи недосконалість в роботі системи. Перегонка освітніх даних для прийняття рішень означає, що дані мають бути перетворені й представлені у вигляді, легкому для сприйняття їх людиною. Ця група методів націлена на використання переваг візуального способу подання інформації, попередньо обробленої методами аналізу даних (наприклад, для скорочення розмірності даних).

Розв'язання задач із застосуванням методів інтелектуального аналізу освітніх даних складається з наступних етапів: попередня обробка даних; виявлення закономірностей у даних; перевірка (валідація) виявлених закономірностей; застосування знайдених закономірностей для прогнозування майбутніх подій у навчальному середовищі; використання побудованих прогнозів для підтримки прийняття рішень і вироблення освітньої політики. Перші три етапи відносяться до традиційного аналізу даних, а останні два етапи характерні для аналізу саме освітніх даних.

Дослідники у сфері Educational Data Mining та Learning Analytics виділяють наступні напрями застосування методів інтелектуального аналізу освітніх даних [17]:

1. Забезпечення зворотного зв'язку для підтримки роботи викладачів: дає можливість виявляти й оптимально організувати структуру навчального контенту, робити його більш доступним на основі аналізу поведінки студентів, досліджуваної за отриманими балами, результатів опитувань, тестів, текстів студентських робіт.

2. Аналіз і візуалізація даних: дозволяє перетворювати великі масиви розподіленої у просторі та часі

інформації у легкі та зрозумілі людині структури з метою підтримки прийняття рішень стосовно процесу навчання.

3. Рекомендації студентам: з вибору навчального контенту на навчальному ресурсі на основі виявлених стилів навчання та звичок використання ресурсу.

4. Прогнозування успішності: на основі аналізу поведінки студента у системі електронного навчання.

5. Моделювання студентів: розробка моделі студента на основі виявлених у результаті проведеного аналізу мотивації, задоволеності, стилю навчання, афективного статусу, навичок.

6. Виявлення небажаної поведінки студентів: здійснюється на основі побудованої моделі студента й дозволяє заздалегідь виявляти тих студентів, які мають ризик отримання негативної оцінки з курсу.

7. Поділ студентів на групи відповідно до їх індивідуальних особливостей та особистісних характеристик.

8. Аналіз соціальних мереж для виявлення закономірностей в академічному співробітництві з метою надання допомоги щодо організації й підтримки досліджень, заснованих на певних моделях співробітництва. Дослідження поведінки користувачів, що виникає як результат дій окремих користувачів за деякими правилами, може проводитися методами агентного моделювання [18]. Проте аналіз соціальних мереж зіштовхується з проблемою неповноти даних, що збирають.

9. Розробка концептуальних карт: структурний аналіз взаємозв'язків між окремими поняттями навчального курсу з допомогою графів. Концептуальна карта студента допомагає виявити можливі прогалини в освіті, а її порівняння з концептуальною картою курсу дозволяє оцінити рівень освоєння предмета.

10. Організація освітнього контенту курсу: автоматизація організації освітнього контенту дозволяє створювати з набору навчальних матеріалів курс, настроєний під потреби тієї або іншої групи студентів.

11. Планування й складання розкладів: рекомендації з вибору майбутніх курсів, які необхідні студентів для заповнення прогалини у знаннях або подальшому розвитку за обраною спеціальністю.

Як бачимо, інтелектуальний аналіз освітніх даних є потужним інструментом для видобутку знань з метою підтримки діяльності суб'єктів навчального процесу.

**Висновки.** Таким чином, розв'язання задач з використанням методів інтелектуального аналізу освітніх даних націлене на вдосконалення процесу навчання, підвищення його ефективності шляхом оптимізації освітнього контенту курсу, моделювання поведінки студентів, виявлення зв'язків та закономірностей, прогнозування, візуалізації даних та установлення зворотного зв'язку між суб'єктами навчального процесу. Подальші напрями вдосконалення освітньої системи потребують проведення досліджень реалізації вказаних методів в умовах нашої країни.

### Список використаних джерел

1. Baker R. S., Gowda S. M. Towards automatically detecting whether student learning is shallow // Intelligent Tutoring, 2012.
2. Stewart M. F., Chisholm C. U. Comparative analysis of emotional competency within distinct student cohorts // Global J. of Engng. Educ. 2012. V. 14, № 2. – P. 163–169.
3. Cobo G., García-Solórzano D., Morán J. A. Using agglomerative hierarchical clustering to model learner participation profiles in online discussion forums // Proc. 2nd International Conference on Learning Analytics and Knowledge, 2012.

4. Romero C., Ventura S. Data mining in education // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. – 2013. – v. 3 (1), – P.12–27.
5. Feng T., Shubin W., Cheng Z., Qinghua Z. Research on e-learner personality grouping based on fuzzy clustering analysis // Proc. 12th Int. Conf. Computer Supported Cooperative Work in Design, 2008. p. 1035–1040.
6. Зорина Н. В., Панченко В. М., Зорин А. Л. Интеллектуальный анализ образовательных данных : ретроспектива и перспектива развития // Научная дискуссия : вопросы математики, физики, химии, биологии // Сборник статей по материалам LIII–LIV международной научно-практической конференции, № 5–6 (38), – М. Изд. «Интернаука», 2017. – С. 42–53.
7. Горлушкина Н. Н., Коцюба И. Ю., Хлопотов М. В. Задачи и методы интеллектуального анализа данных // Образовательные технологии и общество, № 1, т. 18, 2015. – С. 472–482.
8. Веряев А. А., Татарникова Г. В. Educational Data Mining and Learning Analytics – направления развития образовательной квалитологии // Преподаватель XXI век, № 2, т. 1, 2016. – С. 150–160.
9. Патаркин Е. Д. Использование учебной компьютерной аналитики для поддержки совместной сетевой деятельности субъектов образования // Образовательные технологии и общество, № 2, т. 17, 2014. – С. 538–554.
10. Ковальчук Ю. О. Пошук, отримання й аналіз даних в освіті: сучасний стан і перспективи розвитку // Інформаційні технології і засоби навчання, 2015, т. 50, № 6. – С. 152–164.
11. Панченко Л. Ф. Підготовка майбутніх фахівців з інформаційних технологій до здійснення навчальної аналітики // Вісник КрНУ ім. М. Островського, серія «Педагогічні науки», вип. 1 (2), 2015. – С. 80–88.
12. Петренко С. В. Оптимізація й аналіз результатів використання LMS Moodle у системі змішаного навчання в університеті // Інформаційні технології і засоби навчання, 2017, т. 61, № 5. – С. 140–150.
13. Коляда М. Г. Використання систем видобутку знань Data Mining для прогнозування педагогічних процесів та явищ / М. Г. Коляда // Вісник післядипломної освіти. – 2014. – вип. 11. – С. 80–88.
14. Baker R., Siemens G. Educational Data Mining and Learning Analytics [Електронний ресурс]. – Режим доступу : <http://www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf>.
15. Siemens G., Baker R. Learning Analytics and Education Data Mining: Towards Communication and Collaboration // Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 2012 [Електронний ресурс]. – Режим доступу : <http://www.columbia.edu/~rsb2162/LAKs%20reformatting%20v2.pdf>.
16. Baker R., Yacef K. The state of Educational data mining in 2009 : A review and future vision [Електронний ресурс] // Journal of Educational Data Mining. 2009. V. 1, № 1. – P. 3–17. Режим доступу : <http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/download/8/2>.
17. Romero C., Ventura S. Educational Data Mining: A Review of the State of the Art // IEEE Transactions on Systems Man and Cybernetics. Part C (Applications and Reviews). – 2010. – v. 40, № 6. – P. 601–618.
18. Artemenko V. Agent-based modeling of the e-learning users' behavior // International Journal of Computing. 2014. V. 13, № 1. – P. 61–69.

**Н. Н. Болюбаш,**

*ЧНУ ім. Петра Могили, г. Николаев, Україна*

#### **ЗАДАЧИ И МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ДАННЫХ**

*Исследованы основные задачи интеллектуального анализа образовательных данных в рамках учебного процесса в высших учебных заведениях. Выявлена специфика данных, подлежащих анализу, установлены основные цели анализа. Осуществлен обзор и обоснована актуальность применения методов Data Mining в высшем профессиональном образовании.*

**Ключевые слова:** интеллектуальный анализ образовательных данных; учебная аналитика; системы электронного обучения.

**N. Bolyubash,**

*Petro Mohyla Black Sea National University, Mykolaiv, Ukraine*

#### **EDUCATIONAL DATA MINING TASKS AND METHODS**

*The main tasks of the Educational Data Mining in the educational process in higher educational institutions are studied. The specifics of the data to be analyzed are revealed, the main objectives of the analysis are set. The review and substantiation of the relevance of Data Mining in higher professional education was substantiated.*

**Key words:** educational data mining; learning analytics; e-learning systems.

**Рецензенти:** Мещанинов О.П., д-р пед. наук, професор;

Гришкова Р. О., д-р пед. наук, професор.