

ТЕХНОЛОГІЇ ВЕЛИКИХ ДАНИХ У ФОРМУВАННІ ОСВІТНІХ СИСТЕМ

Проценко Н.М.

*кандидат економічних наук, доцент,
доцент кафедри інформаційних технологій, кібернетики
та захисту інформації
Державного біотехнологічного університету
м. Харків, Україна*

Стрімкий розвиток сучасного світу, що супроводжується розвитком науки і технологічних процесів, вимагає застосування найбільш швидких способів генерації та передачі знань в усіх сферах життєдіяльності людства, насамперед, у сфері освіти. Освітня система безперервно створює та накопичує величезні обсяги даних, а потреба доступу і можливості працювати з цими даними широкого кола суб'єктів визначає питання про системну роботу в цьому напрямку одним із пріоритетних.

В системах навчальних закладів накопичується величезна кількість інформації про різні аспекти освітнього процесу: дані студентів, їх успішність та відвідуваність, інформація про викладачів та їх науково-освітню та адміністративну діяльність, різноманітні тексти, аудіо, відео, що пов'язані з освітнім контентом. Ці дані необхідно ефективно зберігати, обробляти та аналізувати. Проте, обсягів даних настільки великі, що їх неможливо виміряти в гігабайтах і, тим більше, обробити без використання відповідних інструментів. Однією з таких технологій для обробки великих архівів та великих потоків даних може стати технологія оперування великими даними Big Data.

У науковій літературі прийнято визначати Big Data за трьома «v» [1]:
– volume – обсяг (за прогнозами Всесвітнього економічного форуму, до 2025 р. обсяг щоденного інтернет-трафіку даних по всьому світу сягне 463 ексабайтів [2]);

– velocity – швидкість (зазначені вище обсяги даних, на відміну традиційної обробки пакета даних, надходять в обробку у реальному часі, тобто дані накопичуються миттєво, причому не має значення тривалість потоку самих даних);

– variety – різноманітність (дані надходять з різних джерел та мають безліч різноманітних форматів – відеодані, фотографії, звукові записи, текстові повідомлення, коментарі, фіксація переглядів сторінок тощо).

В деяких випадках до ознак Big Data додають ще достовірність (veracity), варіативність (variability), візуалізація (visualization), цінність (value).

Слід зазначити, що технологія Big Data не є чимось принципово новим, але на сучасному етапі розвитку технологій відбуваються зміни у принципах підходу до аналізу даних. З'явилася можливість не просто збирати великі обсяги даних, а й аналізувати не пов'язані між собою фактори. Комп'ютер виявляє закономірності, які не може виявити людський мозок, представляючи зовсім несподівані кількісні взаємозв'язки. Методи, що використовуються в аналізі Big Data, базуються на розпізнаванні образів, комп'ютерному навчанні, статистики та психометрії.

Одна з найцікавіших моделей роботи з Big Data – це прогнозування. Прогнозна аналітика – це метод аналізу даних, що концентрується на прогнозуванні майбутньої поведінки об'єктів та суб'єктів з метою прийняття оптимальних рішень, тобто комбінація відомих даних дозволяє забезпечити прогноз шуканого невідомого. Множинні дані збираються із записів інтернет-сервісів, студентських систем, опитувань, соціальних мереж та різних спостережень під час експериментів. Збір та обробка подібних даних, це величезна справа, тому що потрібно знати, на які моменти дивитися, та вміти виявляти потрібну корисну інформацію. Модель може працювати для прогнозування сьогодення, використовуючи статистику з минулого, з'ясовуючи зацікавленість студентів до онлайн-курсу, чи прогнозування майбутнього (використовуючи попередні оцінки), визначаючи ступень можливості студентів вирішувати наступні завдання за такими результатами.

Важливою проблемою освіти є питання виявлення нових, часом прихованих, взаємозв'язків у великих даних, нових знань (data mining). В основу Data Mining покладено концепцію шаблонів поведінки та особистісних якостей учнів [3, с. 29]. Використання цих технологій дозволяє дізнатися про труднощі, які викликають при вивченні певних предметів, з якими тестами учні краще справляються, якій формі занять надають перевагу, які теми у них викликають найбільший інтерес і як можна оптимально побудувати навчальну програму, щоб студент набув саме тих компетенцій, які йому будуть необхідні у сфері його майбутньої професійної діяльності.

Для кращого сприйняття інформації необхідно візуально представляти результати аналізу даних. Правильно підібраний метод візуалізації даних може принести низку переваг: збільшити швидкість прийняття рішень, покращити розуміння даних, зменшиться кількість помилок та даних з неповною інформацією тощо. За статистичними

даними Центру політичних досліджень та аналізу «Ейдос» ефективність роботи людей, які обробляють візуальну інформацію, підвищується на 17 %; 4,5 % краще згадують візуальну інформацію та інші деталі [4]. Тому розробники програмного забезпечення разом із вченими винаходять нові методи візуалізації. Найбільш прогресивними у цьому напрямі є хмара тегів, графіки та діаграми (геометричні перетворення), історичний потік та просторовий потік. Розглянемо деякі з них.

Візуалізація даних за допомогою хмари тегів є невід'ємним аспектом текстової аналітики. Залежно від того, як часто слова або речення з'являються у певному текстовому блоці, вони відображаються за різними розмірами. Якщо частіше вживаються, то виділяються більшим шрифтом на фоні слів/речень із нижчими показниками вживання. Цей онлайн-сервіс, наприклад, сприяє більш інтуїтивному уявленню термінології за певною темою або привертає увагу суб'єкта і змушує зосередитися на певній категорії понять.

Метод геометричних перетворень спрямований на трансформацію багатовимірних наборів даних з метою відображення їх у декартовому та недекартовому геометричних просторах. Прикладами окремих видів застосування цього методу можуть бути: точкові діаграми, які часто використовуються для візуалізації взаємозв'язків між даними у двох вимірах, допомагаючи знайти взаємозв'язок між двома показниками; за допомогою бульбачкової діаграма можна порівняти два параметри за третім. Розміри бульбашок визначаються значеннями третього ряду даних і дозволяють візуально виділити конкретні значення.

Векторні діаграми. Широко застосовується в різних випадках – починаючи від електротехнічних розрахунків і закінчуючи вивченням психології особистості. Може бути застосована як для побудови графіків при обробці тестів, так й для кращого розуміння студентів та умов, у яких вони навчаються, позначати поточні знання студента, мотивацію та стосунки.

Діаграма Сенкей – графік, який показує потоки та його напрями. При цьому лінії відображають взаємозв'язки між об'єктами, а товщина графів, природно, пропорційна даним. Такий вид візуалізації корисний, скажімо, для зручного аналізу потоку відвідувачів порталу (курсів з дисциплін) з кореляцією за джерелами трафіку.

Тобто, подання даних в графічному представленні – це прагнення донести повідомлення якомога простіше і більш інформативне, зберігаючи при цьому цілісність даних.

Слід зазначити, що великі дані (Big Data) мають потужний потенціал для розширення значно ширшого набору інструментів для аналізу важливих питань у системі освіти. Технологія інтелектуального аналізу даних (Data Mining) урізноманітнює методики навчання, дозволяє

впливати на успішність учнів та розуміти, як краще організувати навчальний процес. Аналізу освітніх даних дозволяє спеціалістам відстежувати навчальні події, цифрові сліди основних суб'єктів освітньої діяльності у нелінійному середовищі, не порушуючи сам процес навчання або його навігації за освітнім контентом. Дані, отримані в режимі реального часу, дозволяють віртуально відтворити або змодельовати прогноз дій студентів, визначити, як вони конструюють знання та відслідковують свій фактичний вибір, а також виявити методи, які вони використовують для вираження вибору за допомогою саморегульованого навчання в конкретному контексті.

Використання методів аналізу великих даних та отриманих результатів дозволить сформуванню стратегію прийняття організаційно-педагогічних та управлінських рішень в освіті на підставі даних та способів їх систематизації (big data management in education). Особливістю такої стратегії є те, що вона буде заснована на математично та статистично вимірних кількісних показниках, тобто самих оброблених освітніх даних, що дозволить автоматизувати процес керування навчанням та зробити його персоналізованим.

Список використаних джерел:

1. Letouzé E. Big Data for Development: Opportunities and Challenges. UN Global Pulse. 2012. URL: <https://unstats.un.org/unsd/trade/events/2014/Beijing/documents> (дата звернення: 30.11.2023).
2. Oussous A., Benjelloun F.Z., Lahcen A.A., Belfkih S. Big Data Technologies: A Survey. Journal of King Saud University – Computer and Information Sciences, 2017. DOI: <http://dx.doi.org/10.1016/j.jksuci/2017.06.001> (дата звернення: 1.12.2023).
3. Ian H. Witten, Eibe Frank. Data Mining: Practical machine learning tools and techniques. «Morgan Kaufmann», 2015. P. 229.
4. Як і для чого використовувати візуалізацію даних? Центр політичних студій та аналітики «Ейдос». URL: <http://eidos.org.ua/novyny/yak-i-dlya-choho-vykorystovuvaty-vizualizatsiyu-danyh/> (дата звернення: 29.11.2023).