

ВИКОРИСТАННЯ ТЕХНОЛОГІЇ DATA MINING ДЛЯ ОЦІНКИ ЯКОСТІ НАВЧАННЯ У ВИЩОМУ НАВЧАЛЬНОМУ ЗАКЛАДІ

В. І. Тимофєєв, канд. техн. наук, доц.;

О. В. Лисенко, магістрант

Національний університет кораблебудування, м. Миколаїв

Анотація. Здійснено огляд стану проблеми оцінки якості освіти. Розглянуто технології інтелектуального аналізу даних, які можуть бути застосовані для вирішення задачі оцінки якості навчання. Запропоновано використання технології Data Mining як інструменту для вирішення проблеми оцінки якості навчання у вищих навчальних закладах. Проведено порівняльну оцінку існуючих інструментів інтелектуального аналізу.

Ключові слова: Data Mining, Data Warehouse, OLAP, гіперкуб, організація навчального процесу, якість освіти.

Аннотация. Осуществлен обзор состояния проблемы оценки качества образования. Рассмотрены технологии интеллектуального анализа данных, которые могут быть применены для решения задачи оценки качества обучения. Предложено использование технологии Data Mining как инструмента для решения проблемы оценки качества обучения в высших учебных заведениях. Проведена сравнительная оценка существующих инструментов интеллектуального анализа.

Ключевые слова: Data Mining, Data Warehouse, OLAP, гиперкуб, организация учебного процесса, качество образования.

Abstract. Education quality estimation problem review is carried out. Technologies of intellectual data analysis, which can be used for the decision of task of education quality estimation are considered. Application of Data Mining method as instrument for solving of education quality estimation problem is offered. Comparative estimation of the current Data Mining instruments is conducted.

Keywords: Data Mining, Data Warehouse, OLAP, hypercube, organization of educational process, quality of education.

ПОСТАНОВКА ПРОБЛЕМИ

Навчальний процес у вищих навчальних закладах (ВНЗ) України спрямований на реалізацію змісту освіти на певному освітньо-кваліфікаційному рівні відповідно до державних галузевих стандартів освіти. Державний галузевий стандарт вимагає певного рівня якості навчання.

Якість освіти можна розглядати як багатовимірне поняття, до показників якого належать: якість викладацького складу, якість студентів, якість знань, якість навчальних програм, мотивація викладацького складу і студентів, стан матеріально-технічної бази ВНЗ, якість інфраструктури, затребуваність і конкурентоспроможність випускників на ринку праці, досягнення випускників і студентів та ін. Усі ці показники по-різному впливають на інтегральну оцінку, і встановити їх ступінь впливу кількісно є складною проблемою. Для успішного менеджменту якості освіти необхідні знання про ступінь впливу різних показників на кінцевий результат.

Можливим варіантом вирішення цієї проблеми є застосування аналітичного аналізу результатів діяльності ВНЗ із використанням технологій Data Mining.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Якість навчання у ВНЗ оцінюється на сьогодні вимогами національних (ДАК МОНМС України) або міжнародних (EFMD, EQUIS, AACSB, SEEMAN, IQA)

систем акредитації. Показники якості та їх мінімальні значення в цих системах встановлені експертним шляхом і помітно відрізняються. Дослідження впливу різних факторів на якість освіти із застосуванням методів, які базуються на аналізі експертних оцінок, також дають суттєво відмінні результати в залежності від складу експертних груп [2]. Очевидно, що для кількісного встановлення ступеня впливу різних факторів на якість вищої освіти необхідно застосувати більш об'єктивні методи.

У різних ВНЗ проводяться дослідження в цьому напрямку. Запропоновано аналізувати успішність студентів засобами Data Mining [1]. Показана можливість аналізувати фактори, які впливають на успішність, методами статистичного аналізу, використовуючи пакет STATISTICA [7]. Подана теоретична основа для практичних розробок, класифіковані головні причини, які впливають на якість навчання [5]. Розроблена модель оцінки якості освіти ВНЗ за допомогою інформаційно-аналітичної системи [3]. Розглянута задача автоматичного отримання знань з баз даних [6]. Однак задача з'ясування кількісного впливу різних факторів на якість освіти в повній мірі не вирішена, хоча є дуже актуальною відносно уваги до якості вищої освіти з боку уряду України [4].

МЕТА РОБОТИ — визначення методів аналізу даних для оцінки якості навчання і покращення організації навчального процесу у вищих навчальних закладах.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Інформаційна, високотехнологічна економіка висуває нові вимоги до якості вищої освіти. Для виконання цих вимог необхідно мати критерії і засоби її оцінювання, які дозволили б визначити ступінь впливу різних факторів підготовки фахівця на кінцевий результат. Проведений аналіз показав, що ця проблема не є вирішеною.

Одним зі шляхів розв'язання поставленої проблеми може бути використання аналітичного аналізу результатів діяльності ВНЗ із застосуванням технологій Data Mining.

Інформаційні системи ВНЗ містять, як правило, оперативні дані, які характеризують навчальний процес. Вони не дозволяють з'ясувати вплив різних факторів на якість кінцевого результату навчального процесу. Більше того, ці оперативні дані дуже часто не використовуються взагалі по закінченні оперативного періоду. Вже давно існує необхідність у побудові аналітичної системи для глибокого аналізу даних в їх історичному розвитку. Основним функціональним призначенням цієї системи буде визначення залежності між деякими показниками якості освіти і множиною факторів, яка впливає на ці показники, що дозволить приймати ефективні управлінські рішення.

На сьогодні загальнодоступним параметром, який характеризує рівень знань студентів, а отже, і якість навчання, є їх оцінки. Натомість причин, які впливають на даний показник, існує велика кількість, і досить важко визначити, наскільки сильно той чи інший фактор впливає на результат.

Якість підготовки фахівця у ВНЗ визначають: попередня підготовка студентів і їх здібності, мотивація студентів до успішної навчальної діяльності, рівень творчих навичок, рівень професійних знань, рівень виконавської дисципліни, рівень загальної активності й винахідливості, рівень культурного, етичного і морального виховання, рівень соціальних умов життя студентів, якість науково-педагогічного складу ВНЗ, мотивація викладачів до успішної педагогічної діяльності, рівень наукових досліджень, що проводяться у ВНЗ, якість навчальних планів і програм, забезпеченість студентів навчально-методичною літературою, затребуваність і конкурентоспроможність випускників на ринку праці, досягнення випускників та студентів, конкурс при вступі до ВНЗ, рівень матеріально-технічної бази й фінансового становища ВНЗ, якість інфраструктури, соціально-економічний стан у країні, рівень міжнародних зв'язків ВНЗ, рівень автономності в діяльності ВНЗ, наявність у ВНЗ системи управління якістю вищої освіти та ін.

Не всі з цих факторів можуть бути виражені об'єктивними кількісними показниками, деякі з них є взаємопов'язаними. Однак для більшості з цих факторів існують реальні кількісні показники, аналізуючи які, можна встановити ступінь впливу

кожного з факторів у їх взаємодії на якість вищої освіти. Дані для такого аналізу можна отримати з інформаційних систем ВНЗ і зовнішніх баз даних, накопичуючи їх з року в рік.

Результат такого аналізу може бути записаний у вигляді наступного рівняння:

$$Y=f(x_1 \dots x_n, z_1 \dots z_r, w_1 \dots w_s),$$

де $x_1 \dots x_n$ — незалежні змінні — є внутрішніми властивостями ВНЗ; $z_1 \dots z_r$ — незалежні змінні, що є зовнішніми чинниками, які впливають на процеси у ВНЗ; $w_1 \dots w_s$ — невраховані властивості або чинники; n, r, s — кількість відповідних змінних; Y — цільова функція.

Побудова вказаної математичної моделі передбачає наявність сукупності результатів спостережень. У цій сукупності цільова функція відповідає за якість освіти, для якої необхідно встановити функціональну залежність із визначеними факторами середовища.

У подальшому при перевірці значущості коефіцієнтів виключаються з розгляду другорядні фактори, вплив яких на цільову функцію буде незначним.

Для проведення аналітичного аналізу оперативні дані з інформаційних систем ВНЗ і зовнішніх баз даних треба уніфікувати, агрегувати й розмістити в сховищі даних, тобто використати технологію Data Warehousing. Сховище даних дозволить зберігати хронологічно впорядкований постійний набір даних, що підвищить якість подальшої аналітичної обробки.

Для попереднього аналізу і перевірки виникаючих гіпотез використовується OLAP-сервер, який може бути реалізований трьома способами: MOLAP (Multidimensional OLAP), ROLAP (Relational OLAP) і HOLAP (Hybrid OLAP) у залежності від способу зберігання даних. В основі концепції OLAP, або оперативної аналітичної обробки даних (On-Line Analytical Processing), лежить багатовимірне концептуальне подання даних (Multidimensional conceptual view).

Багатовимірна модель візуально зображується за допомогою куба (або у разі більше трьох вимірів — гіперкуба). Розглянемо приклад. Нехай оцінки, отримані у ВНЗ — функція від змінних «Семестр», «Відсоток аудиторних занять від загальної кількості занять» і «Предмет». Тоді вимірами виступатимуть «Час», «Кількість аудиторних занять» і «Предмет». На рис. 1 наведений багатовимірний куб даних для подання даної функції.

Вищезазвані виміри взяті в навчальних цілях, і предметна галузь не обмежується тільки ними. Наприклад, вимірами, пов'язаними з діяльністю викладачів вузу, можуть бути факультет, кафедра, викладач, посада, вчене звання, дата підведення підсумків діяльності тощо. Вимірами, пов'язаними з навчанням студентів, можуть виступати група, студент, форма навчання, початковий рівень знань, матеріальне становище. Вимірами, пов'язаними з технологією та умовами навчання, технічним

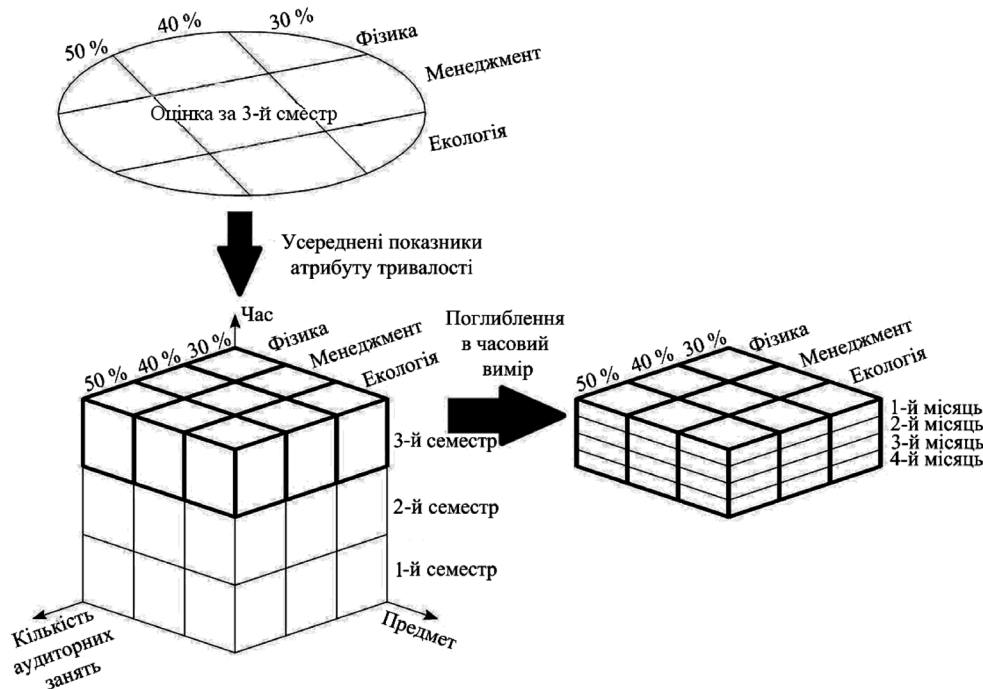


Рис. 1. Гіперкуб OLAP

і методичним оснащенням, можуть бути методи навчання, методи контролю процесу навчання, обладнання, середовище.

Побудова багатовимірної моделі даних дозволяє виконувати над гіперкубом операції зрізу (slice), обертання (rotate), консолідації (drill up) і деталізації (drill down).

У нашому прикладі (див. рис. 1) операція зрізу буде виконана в разі вибору значення «3-й семестр» виміру «Час». Зрізом даних буде підкуб — двовимірна проекція куба з вимірами «Кількість аудиторних занять» і «Предмет». Якщо при аналізі даних про стан оціночних відомостей виконати операцію консолідації для виміру «Час», то результатом будуть агреговані показники оцінок за 3-й семестр.

Для більш глибокого аналізу використовуються алгоритми Data Mining. Виділяють статистичні методи і методи штучного інтелекту. Арсенал статистичних методів Data Mining класифікований на чотири групи методів: описовий аналіз, аналіз зв'язків

(кореляційний і регресійний аналіз), багатовимірний статистичний аналіз (компонентний аналіз, канонічні кореляції та ін.), аналіз тимчасових рядів (динамічні моделі й прогнозування). До другої групи належать такі методи: штучні нейронні мережі (розпізнавання, кластеризація), еволюційне програмування, генетичні алгоритми (оптимізація), асоціативна пам'ять (пошук аналогів, прототипів), нечітка логіка, дерева рішень, системи обробки експертних знань [9]. Для поданої вище математичної моделі найбільш придатними з них є аналіз зв'язків, багатовимірний статистичний аналіз і нечітка логіка, хоча інші методи можуть також дати цікаві результати.

Таким чином, для реалізації наведеної технології необхідно створити інформаційну систему, яку пропонується побудувати з трьох складових: Data Warehouse, OLAP, Data Mining. Концептуальна архітектура системи інтелектуального аналізу зображена на рис. 2.

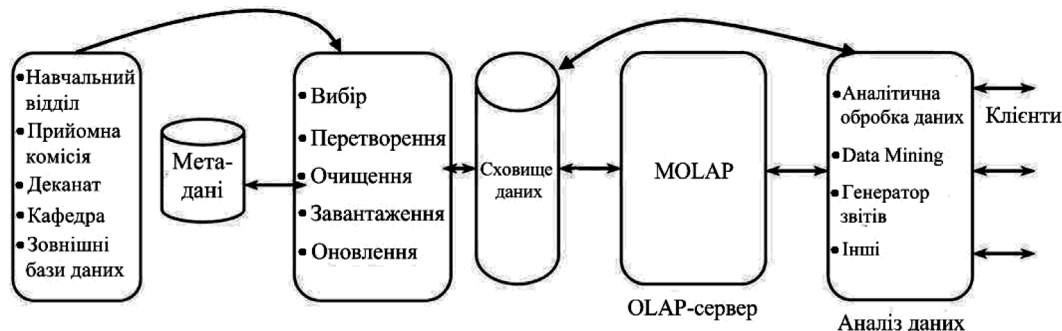


Рис. 2. Архітектура аналітичної системи

Прикладом технології, яка забезпечує подібну функціональність, є Business Intelligence (BI). Цей термін був запропонований для опису різних концепцій і методів, які покращують бізнес-рішення шляхом використання систем підтримки рішень.

Питання вибору платформи для розробки BI-рішення непросте. На січень 2011 року 75% ринку BI контролює п'ятірка лідерів: IBM, Microsoft, SAP, SAS, Oracle [8].

Компанія SAS лідирує в сфері прогресивної аналітики (Advanced Analytic Solutions). Пакет SAS Enterprise Miner надає набір інструментів і алгоритмів прогностичного й описового моделювання, що включає у себе дерева рішень, нейронні мережі, методи міркування, засновані на механізмах пошуку в пам'яті, лінійну і логістичну регресію, кластеризацію, асоціації, тимчасові ряди тощо.

Сильними сторонами компанії SAP у першу чергу є генерація форматуваних звітів і аналіз даних у різних ракурсах та з різним ступенем деталізації за допомогою технології OLAP.

Інструмент компанії IBM — IBM Intelligent Miner for Data — пропонує найсучасніші методи Data Mining, підтримує повний Data Mining-процес (від підготовки даних до презентації результатів) та підтримує мови XML і PMML.

Продукти Oracle мають широкі можливості рішення щодо організації колективної роботи, одні з найкращих засобів візуалізації. Версія Oracle Data Mining 11gR2 підтримує наступний спектр алгоритмів: класифікаційні моделі (Naive Bayes, Adaptive Bayes Network), класифікації і регресійні моделі (Support Vector Machine), пошук істотних атрибутів (Minimal Descriptor Length), кластеризація (Enhanced K-means, O-cluster), пошук асоціацій (Apriory Algorithm), виділення ознак (Non-Negative Matrix Factorization).

Розробки Microsoft представлені SQL Server Analysis Services 2008, що є не тільки комплексною BI-платформою, але й тісно інтегрований з офісними рішеннями (Microsoft Office System). У цьому інструменті крім OLAP-технології застосовуються алгоритми Data Mining: Microsoft Decision Trees, Naive Bayes, Microsoft Clustering, Sequence Clustering, Правила асоціацій, Microsoft Neural Networks, Time Series, Microsoft Linear Regression, Microsoft Logistic Regression. Інтерфейс BI-інструментів Microsoft

інтуїтивно зрозумілий, у цьому він нагадує знайомий усім інтерфейс системи Microsoft Office. До переваг, загальних для всього BI-інструментарію Microsoft, належать також гнучкість, простота використання і розгортання, особливо якщо оперативні інформаційні системи ВНЗ базуються на платформах Microsoft.

Крім зарубіжних існують і російськомовні розробки. Найбільшого поширення отримали два продукти класу Data Mining: PolyAnalyst і Deductor. Обидві системи підходять для організації робочого місця аналітика, який має справу з даними великого обсягу. Відмінна особливість PolyAnalyst — використання власних алгоритмів аналізу, заснованих на принципі еволюційного програмування. Система Deductor компанії BaseGroup Labs спочатку являла собою сукупність самонавчальних алгоритмів для вирішення задач KDD (виявлення знань у базах даних з можливістю початкової обробки інформації), потім була доповнена системами тиражування знань, а на сьогодні перетворилася в аналітичну платформу, яка дозволяє поставити аналіз даних на конвеєр.

Усі розглянуті BI-інструменти мають достатньо засобів для вирішення поставленої задачі, тому вибір програмного інструменту буде залежати від фінансових можливостей ВНЗ, програмної сумісності з елементами їх оперативних інформаційних систем і набору реалізованих алгоритмів Data Mining.

Методи Data Mining суттєво відрізняються за ступенем інформативності й рівнем достовірності отриманих результатів. Для визначення конкретного алгоритму Data Mining, який надавав би достовірні кількісні результати, необхідно провести порівняльні дослідження. При цьому ряд алгоритмів, які дозволяють отримати не тільки цікаві, але й якісні результати, в контексті даної задачі можна априорі виключити з розгляду.

ВИСНОВКИ

Запропоновано використання технології Data Mining як інструменту для вирішення проблеми оцінки якості навчання у ВНЗ. Проведено порівняльну характеристику аналітичних систем, придатних для вирішення цієї задачі.

Установлено необхідність проведення досліджень придатності різних методів Data Mining для знаходження залежності якості вищої освіти від множини факторів, що впливає на неї.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] **Воронин, Т. В.** Анализ успеваемости студентов при помощи средств Data Mining. Технологии Microsoft в теории и практике программирования [Текст] / Т. В. Воронин, В. Н. Разумов // Сб. тр. VIII Всероссийской науч.-практ. конф. студентов, аспирантов и молодых ученых (Томск, 23–24 марта 2011 г.). — Томск : Изд-во Нац. исслед. Томского политехн. ун-та, 2011. — С. 191–193.
- [2] **Гончаров, С. М.** Науково-методичне забезпечення кредитно-модульної системи організації навчального процесу [Текст] : монографія / С. М. Гончаров. — Рівне : НУВГП, 2005. — 266 с.

- [3] **Пампуха, І. В.** Модель оцінки якості освіти ВВНЗ за допомогою інформаційно-аналітичної системи [Текст] / І. В. Пампуха, А. В. Мальога // Зб. наук. пр. ВІТІ НТУУ «КПІ». — К. : КПІ, 2011. — №1. — С. 150–159.
- [4] Про затвердження Порядку проведення моніторингу якості освіти [Електронний ресурс] : [Постанова Кабінету Міністрів України № 1283 від 14 груд. 2011 р.]. — Режим доступу: <http://zakon.rada.gov.ua>.
- [5] **Сергушичева, А. П.** Анализ факторов, влияющих на качество обучения. Управление и экономика: опыт, традиции, инновации [Текст] / А. П. Сергушичева, Д. В. Жаров // Материалы науч.-практ. конф. (Вологда, 9–10 апр. 2010 г.). — Вологда : Легия, 2010. — С. 275–287.
- [6] **Федяев, И. О.** Автоматизированное извлечение знаний из баз данных [Текст] / О. И. Федяев, И. А. Чернов, О. А. Мандрикова // Сб. тр. Первой Междунар. студ. науч.-техн. конф. «Информатика и компьютерные технологии — 2005». — Донецк : ДонНТУ, 2005. — Ч. 7. — С. 257–258.
- [7] **Филь, Н.** Анализ факторов, влияющих на успеваемость, методами статистического анализа [Текст] / Н. Филь, М. Плисс // Тр. межвуз. науч.-практ. и учеб.-метод. конф. «Актуальные проблемы образования». — Рига : Институт транспорта и связи, 2011. — С. 54–67.
- [8] **Чубукова, И. А.** Волна Forrester и квадраты Gartner. Взгляд на ВІ 2011 [Электронный ресурс] / И. А. Чубукова. — Режим доступа: <http://irina-chubukova.blogspot.com>.
- [9] Encyclopedia of Data Warehousing and Mining [Text] / John Wang, editor. — 2nd ed. p. cm. — Hershey ; London : Information science reference (IGI Glodal), 2009. — 1248 p.

© В. І. Тимофеев, О. В. Лисенко

Надійшла до редколегії 30.11.2011

Статтю рекомендує до друку член редколегії Вісника НУК
д-р техн. наук, проф. *К. В. Кошкін*

Статтю розміщено у Віснику НУК № 1, 2012