

Адамик О.В., к.е.н., доцент,
Тернопільський національний економічний університет,
м. Тернопіль, Україна

ВИКОРИСТАННЯ ІНСТРУМЕНТІВ DATA MINING ДЛЯ АНАЛІЗУ КОШТОРИСУ БЮДЖЕТНИХ УСТАНОВ

Адамик, О.В., Використання інструментів Data Mining для аналізу кошторису бюджетних установ [Текст] / Оксана Василівна Адамик // Бухгалтерський облік, аналіз і аудит в управлінні економічними процесами світової і національної економіки: сучасний стан та перспективи: матеріали міжнар. наук.-практ. конф. 11 грудн. 2015 р., (ПДАТУ, м. Кам'янець-Подільський). – Тернопіль: Крок, 2015. – 373 с. – С. 233-235

У оперативних базах даних накопичується значний обсяг інформації, що характеризує фінансово-господарську діяльність суб'єктів господарювання. Вирішення задач аналізу можна здійснювати, використовуючи стандартний апарат запитів до баз даних, а окрім того застосувати методи інтелектуального аналізу даних. У західній літературі використовують термін **Data Mining (DMg)**, що дослівно перекладається як добування даних, „знаходження знань у базах даних”, „інтелектуальний аналіз даних”.

Класичне визначення технології „видобування даних” (Data Mining) звучить таким чином: це виявлення в початкових („сирих”) даних – раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань. Тобто інформація, знайдена в процесі застосування методів Data Mining, повинна бути нетривіальною і раніше невідомою, наприклад, отриманий показник середніх продажів не є таким. Знання повинні описувати нові зв'язки між властивостями, передбачати значення та характеристики одних параметрів на основі інших. Основна мета Data Mining полягає у виявленні прихованих правил і закономірностей у великих масивах даних.

Інтелектуальний аналіз даних (DMg) – це процес підтримки прийняття рішень, заснований на пошуку в даних прихованих закономірностей (шаблонів інформації). Більшість методів ІАД була спочатку розроблена в рамках теорії штучного інтелекту в 70-80-х рр. XX в. Вони набули поширення тільки останніми роками як надбудова над сховищами даних (DW).

„Сировиною” для DMg можуть бути:

- плоскі таблиці оперативних баз даних. Саме з них і починалася історія DMg;
- дані, одержані після обробки за допомогою OLAP-технологій. У багатьох випадках їх використання є більш ефективним.

Data Mining є, практично, надбудовою над OLAP-системою.

Як відомо, координування і контроль виконання кошторису здійснюється органами Державної Казначейської служби України. Тобто інформація акумулюється у „одному місці”, що створює сприятливі для застосування DMg-

інструментів в аналізі кошторису бюджетних установ. Раніше [2] нами вказано про сприятливі передумови використання OLAP-систем у бюджетних установах. Дані про їх фінансово-господарську діяльність добре піддаються формалізації, їх можна представити у вигляді гіперкуба або полікуба. Останні у свою чергу є „сировиною” для обробки інструментами OLAP-систем і надалі DMg-систем, або відразу ж засобами DMg.

Якщо у межах технології OLAP основний тягар аналізу лягає на плечі людини, то використання Data Mining перекладає тягар формулювання гіпотез і виявлення незвичних шаблонів на комп'ютер.

У загальному випадку DMg складається з трьох стадій [4]:

- 1) виявлення закономірностей (вільний пошук (Discovery)). На цій стадії здійснюються такі процеси як: а) виявлення закономірностей умовної логіки (Conditional Logic), б) виявлення закономірностей асоціативної логіки (Associations and Affinities), в) виявлення трендів і коливань (Trends and Variations);
- 2) використання виявлених закономірностей для передбачення невідомих значень (прогнозне моделювання (Predictive Modeling)) включає у себе такі дії як передбачення невідомих значень (Outcome Prediction) та прогнозування розвитку процесів (Forecasting);
- 3) аналіз винятків (Forensic Analysis), призначений для виявлення і тлумачення відхилень і аномалій у знайдених закономірностях (Deviation Detection).

Задачі DMg класифікують перш за все за типами інформації, що витягується, тобто за видами знайдених закономірностей. Виділяють наступні п'ять видів задач:

1. *Класифікація* – дозволяє виявити ознаки, що характеризують однотипні групи об'єктів (класи), для того, щоб за відомими значеннями цих характеристик можна було віднести новий об'єкт до того або іншого класу. Передбачається, що характеристики класів відомі наперед (до аналізу). Як методи рішення задачі класифікації застосовують алгоритм найближчого сусіда (Nearest Neighbor), індукцію дерев рішень, «навчені вчителем» нейронні мережі.

2. *Кластеризація* – розповсюдження ідеї класифікації на складніший випадок, коли самі класи не призначені. В результаті виконання процедури кластеризації початкові дані розбиваються на однорідні групи (кластери). Це дозволяє установити виробити по відношенню до кожної з груп (наприклад, до груп покупців) певну політику. Задача кластеризації значно складніша за задачу класифікації. У даний час найбільш часто задачі кластеризації вирішуються методом «навчання без вчителя» спеціального виду нейронних мереж.

3. *Виявлення асоціації*. Асоціація – закономірність в даних, що фіксує настання двох (або більш) одночасних подій. Типовий приклад асоціації досліджується в задачі визначення пар виписаних одночасно документів (наприклад, накладна і податкова накладна, платіжне доручення про

виплату заробітної плати на дебетові картки працівників і платіжне доручення про сплату податків і зборів, пов'язаних з оплатою праці, тощо).

3. *Виявлення послідовностей.* Послідовність – закономірність у даних, що фіксує настання подій не одночасно, а з деяким певним розривом в часі. Таким чином, асоціація є окремий випадок послідовності з нульовим тимчасовим лагом. Наприклад, виписування відрядного посвідчення працівникові й авансовий звіт про витрачені кошти; податкова декларація і платіжне доручення про сплату податків.

5. *Прогнозування* – формалізована процедура прогнозу, яка на основі дослідження поточних і минулих даних дозволяє оцінити майбутні значення числових показників.

Існують два способи впровадження нової інформаційної технології в локальні інформаційні структури:

- 1) пристосування її до організаційної структури мережі розпорядників бюджетних коштів;
- 2) модернізація організаційної структури з метою найбільш ефективного використання нової інформаційної технології.

Перший спосіб є дешевшим і не вимагає великих змін в організації діяльності мережі розпорядників бюджетних коштів. Проте, ефект від його впровадження може бути незначним. Другий спосіб вимагає більших капіталовкладень, але забезпечує якісно новий рівень діяльності суб'єктів державного сектора економіки. Зокрема, нами вказувалась у [1] доцільність запровадження для бюджетних установах хмарних обчислень.

Основною перевагою методів Data Mining є здатність виявляти нові знань, які неможливо отримати методами статистичного, регресивного аналізу або економетрики. Засоби інтелектуального аналізу входять до складу найрозвиненіших ІАС, тому у зв'язку з складністю виконуваних задач інтелектуальний аналіз частіше реалізується автономними програмними системами.

Список використаних літературних джерел:

1. Адамик, О. В. Інструменти “хмарних обчислень” як фактор удосконалення обліку в бюджетних установах // Економічний аналіз: зб. наук. праць. – Тернопіль: “Економічна думка”, 2015. – Том 19. – № 2. – С. 179-184.
2. Адамик, О.В. Використання інструментів технологій OLTP та OLAP для обліку й аналізу виконання кошторису бюджетних установ // Наука молода, 2015. – № 23. – 225 с. – С. 195-201.
3. Діденко Д.Г. Бази даних та інформаційні системи: Лекція 25. Системи аналітичної обробки даних OLAP – [Електронний ресурс]. – Режим доступу: <http://www.simulation.kiev.ua/dbis/lecture25.html>
4. Щавелёв Л.В. Оперативная аналитическая обработка данных: концепции и технологии – [Електронний ресурс]. – Режим доступу: http://www.olap.ru/basic/olap_and_ida.asp