

УДК 681.3.06(075.8)

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ ЗАСОБАМИ АНАЛІЗУ Й КЕРУВАННЯ РЕЛЯЦІЙНИМИ БАЗАМИ ДАНИХ

Кудренко С.О.

(Національний авіаційний університет)

Здійснено системне дослідження підходів до аналізу даних засобами MS SQL SERVER 2014. Показано, що використання методів інтелектуального аналізу та добування даних дозволяє суттєво підвищити ефективність процесу прийняття рішень. Запропоновано алгоритм створення моделі даних, яка є доступною будь-яким клієнтським застосуванням. Сформульовано підходи до вибору алгоритму для обробки даних, створення проектів інтелектуального аналізу даних. Використання Microsoft SQL Server 2014 та інструментарію Analysis Services забезпечує автоматизоване створення моделей Data Mining, що дозволяє розв'язувати низку різноманітних задач.

Постановка проблеми. Протягом останніх десяти років в інформаційній індустрії спостерігається стійке зростання інтересу до технологій аналізу даних і заснованим на цих технологіях систем підтримки прийняття рішень [1]. За всією цією областю закріпився, що став уже звичним в англійській літературі, термін Data Mining або Knowledge Discovery. Українською мовою вживають також словосполучення "інтелектуальний аналіз даних". Наведемо класичне визначення цього терміна, запропоноване в 1996 році в роботі Fayyad, Piatetsky-Shapiro, і Smyth "Advances in Knowledge Discovery and Data Mining":

"The non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" - Нетривіальний процес виявлення нових, потенційно корисних, коректних і інтерпретованих закономірностей в даних [2]. Microsoft SQL Server- це система аналізу й керування реляційними базами даних у розв'язках електронної комерції, виробничих галузей і сховищ даних [3]. Відомі попередні версії: SQL Server 2005, SQL Server 2008, SQL Server 2008 R2, SQL Server 2012. В Microsoft SQL Server 2014 ще більш покращені критично важливі можливості, які представлені в більш ранніх версіях, за рахунок надання рекордної продуктивності, доступності й зручності керування для критично важливих застосувань.

Враховуючи різноманітність форм представлення даних, та алгоритмів, що використовуються, а також сфер застосування, інтелектуальний аналіз даних може проводитися за допомогою програмних продуктів наступних класів [4]:

- спеціалізованих "коробкових" програмних продуктів для інтелектуального аналізу;
- математичних пакетів;
- електронних таблиць і різного роду надбудов над ними;

- засобів, інтегрованих у системи керування базами даних (СКБД);
- інших програмних продуктів.

SQLServer 2014 є комплексом нових функціональних засобів роботи в оперативній пам'яті, який вбудовано в основну базу даних для операцій OLTP і зберігання даних, що доповнюють існуючі технології сховищ даних в оперативній пам'яті й функціональні можливості бізнес-аналітики для створення найбільш різнобічного розв'язку по експлуатації засобів баз даних в оперативній пам'яті на ринку [2,3,5].

Аналіз останніх досліджень і публікацій. Останні тенденції та нововведення в індустрії зберігання та аналізу даних, полягають в переході до концепції Data Warehouse (Сховища Даних). Це заснований на сучасних системах управління базами даних, "предметно-орієнтований, інтегрований комплекс, що підтримує хронологію набору даних, організований для цілей підтримки прийняття рішень" [6]. Введення цього нового шару в архітектуру інформаційно-аналітичних програм було продиктовано не тільки міркуваннями концептуальної чистоти системи, а й невдалим досвідом реалізації систем аналізу безпосередньо в середовищі СУБД. Останні орієнтовані, як правило, на вирішення завдань оперативного збору інформації та підтримки транзакцій (OLTP), що, по-перше, обумовлює вибір певної моделі зберігання даних - реляційної, а, по-друге, призводить до того, що працююча OLTP система активно використовує всі обчислювальні ресурси. Обидва ці фактори, призводять до того, що аналітичні підсистеми, які вимагають очищених і, часом надлишкових, даних, а також значних обчислювальних ресурсів, працюють в системах, суміщених з OLTP незадовільно. Нас у першу чергу будуть цікавити засоби, інтегровані із СКБД. Як приклад можна навести СКБД Microsoft SQL Server і служби Analysis Services, які входять до її складу і забезпечують користувачів засобами аналітичної обробки даних у режимі on-line (OLAP) та інтелектуального аналізу даних, які вперше з'явилися в MS SQL Server 2000. Не тільки Microsoft, але й інші провідні розроблювані СКБД мають у своєму арсеналі засоби інтелектуального аналізу даних.

Формулювання цілей статті. В проведеному дослідженні за мету ставилось ідентифікувати особливості підходів до аналізу даних засобами MS SQL SERVER 2014.

Для досягнення поставленої мети в роботі вирішувалися наступні задачі:

- ідентифікувати особливості використання методів інтелектуального аналізу та добування даних для вирішення завдань забезпечення ефективності прийняття та реалізації управлінських рішень;
- інтерпретувати зміст складових комплексу нових функціональних засобів роботи в оперативній пам'яті, який вбудовано в основну базу даних для операцій OLTP і зберігання даних;
- створити реляційну структуру інтелектуального аналізу даних та розробити алгоритм створення моделі даних, яка доступна будь-яким клієнтським застосуванням.

Виклад основного матеріалу дослідження. При проведенні цього дослідження використовувались загальнонаукові методи: аналіз та синтез для встановлення причинно-наслідкових зв'язків між елементами системи аналізу й керування реляційними базами даних у розв'язках електронної комерції, виробничих галузей і сховищ даних. Абстрагування, формалізація та аналогія для побудови моделі даних. Крім того використовувались і специфічні методи дослідження [4,7,8]: Асоціація (або відношення), класифікація, кластеризація, прогнозування використовувались для аналізу та опису типів операцій з формування, обробки та збереження і відновлення даних. Для вирішення завдань в рамках визначення критеріїв відбору та для підтримки вибору певних даних в рамках загальної структури використовували метод дерева рішень. Фактологічною базою дослідження є інформація про складові комплексу нових функціональних засобів роботи в оперативній пам'яті, який вбудовано в основну базу даних для операцій OLTP і зберігання даних.

На першому етапі дослідження ідентифікували складові зазначеного комплексу. Служби Analysis Services в SQL Server 2014 є підсистема обробки даних, яка використовується в розв'язках бізнес-аналітики й інформаційного забезпечення. Служби надають аналітичні дані, які застосовуються у звітах і клієнтських застосуваннях на зразок Excel, звітах служб Reporting Services і інших сторонніх засобах бізнес-аналітики. Служби Analysis Services використовують для створення високопродуктивних структур запитів, бізнес-логіки й ключових показників ефективності в універсальній моделі даних, яка доступна будь-яким клієнтським застосуванням, що підтримують служби Служби Analysis Servicesу якості джерела даних. Щоб створити модель, використовують SQL Server Data Tools, попередньо вибравши шаблон проекту табличної або багатомірної моделі й інтелектуального аналізу даних. Моделі заповнюються даними із зовнішніх систем обробки даних [9-10]. Звичайно це сховища даних, розміщені в системі керування реляційними базами даних SQL Server або Oracle.

До складу служб Microsoft SQL Server Analysis включено ряд наступних засобів створення рішень інтелектуального аналізу даних. Майстер інтелектуального аналізу даних, який є компонентом середовища SQL Server Data Tools (SSDL). Він спрощує створення структур і моделей інтелектуального аналізу даних або в реляційних джерелах даних, або в багатомірних даних у кубах. За допомогою цього майстра користувач вибирає потрібні дані і потім застосовує необхідні методи інтелектуального аналізу - кластеризацію, нейронні мережі або моделювання часових рядів. Засоби перегляду моделей доступні як у середовищі SQL Server Management Studio, так і в середовищі SQL Server Data Tools (SSDL) і призначені для дослідження моделей інтелектуального аналізу після їхнього створення. Можна переглядати моделі за допомогою засобів, спеціально створених для кожного алгоритму, або виконати більш глибокий аналіз за допомогою засобу перегляду змісту моделі. Конструктор прогнозуючих запитів доступний як у середовищі SQL Server Management Studio, так і в середовищі SQL Server Data Tools (SSDL) і призначений для створення прогнозуючих запитів. Щоб

оцінити якість набору даних, можна перевірити точність моделей за набором контрольних або зовнішніх даних або виконати перехресну перевірку.

Середовище SQL Server Management Studio є інтерфейс, призначений для роботи з існуючими рішеннями інтелектуального аналізу даних, розгорнутими в екземплярі служб Analysis Services. Щоб оновити дані структур і моделей, необхідно виконати їхню повторну обробку. Служби SQL Server Служби Integration Services включають засоби, які можуть бути використані для очищення даних, для автоматизації різних задач, таких як створення прогнозів і відновлення моделей, а також для створення розв'язків в області інтелектуальної обробки тексту. Майстер інтелектуального аналізу даних у службах Microsoft SQL Server Analysis Services запускається щоразу, коли до проекту інтелектуального аналізу даних додається нова структура даних. Майстер допомагає вибрати джерело даних і створити представлення джерела даних, яке визначає дані, що використовуються для аналізу, а також створити початкову модель. На останньому етапі роботи майстра можна розділити дані на навчальний і перевірочний (тестовий) набори й включити такі функції, як деталізація. Перш ніж запускати майстер, необхідно дати відповіді на наступні питання [11]:

- Як будуть створюватися структура й модель інтелектуального аналізу даних: на основі реляційної бази даних або на основі існуючого куба в базі даних OLAP?
- Які колонки містять ключі, що однозначно ідентифікують запис варіанта?
- Які колонки або атрибути будуть використовуватися в прогнозі?
- Які колонки або атрибути підходять для використання як вхідних даних для аналізу?
- Який алгоритм використовувати? Алгоритми, використовувані службами SQLServerAnalysisServices, мають різні характеристики й дають різні результати. В цьому випадку, ми не обмежені однією моделлю для кожного набору даних, тому потрібно експериментувати, додаючи нові моделі.
- Чи буде потреба в перевірці моделей на уніфікованому наборі даних? Якщо так, то можна відкласти деякі дані для перевірки. Можна вибрати відсоток і, якщо буде потрібно, вказати кількість рядків.

Після цього слід зробити наступний важливий вибір: використати реляційне джерело даних або створити модель на основі багатомірних даних (OLAP) [12]. На даному етапі майстер інтелектуального аналізу даних розділяється на дві гілки залежно від того, утримуються дані в реляційному джерелі даних або в кубі. Усе інше, крім процесу вибору даних, залишається колишнім, можливість додати необов'язковий контрольний набір даних тощо, але вибір даних з куба є трохи складнішим, ніж використання реляційних даних.

Наступним етапом дослідження було вирішення питання створення реляційної структури інтелектуального аналізу даних. Для інтелектуального аналізу даних рекомендується використовувати реляційну таблицю або джерело даних, якщо тільки дані вже не зберігаються в кубі або якщо необхідно провести

інтелектуальний аналіз вимірів OLAP або результатів агрегатів або обчислень OLAP. Необхідно вирішити, який алгоритм використовувати для обробки даних. Це рішення може бути складним. Кожен алгоритм у службах Analysis Services має свої характеристики й дає свої результати, тому можна поекспериментувати і спробувати різні моделі, перш ніж вирішити, яка з них найкраще підходить для конкретних даних і задач бізнесу.

Алгоритм інтелектуального аналізу даних - це набір евристик і обчислень, які створює модель інтелектуального аналізу даних з даних. Щоб створити модель, алгоритм спочатку аналізує надані дані, здійснюючи пошук певних закономірностей і тенденцій. Алгоритм використовує результати цього аналізу для вибору оптимальних параметрів створення моделі інтелектуального аналізу даних. Потім ці параметри застосовуються до всього набору даних, щоб виявити придатні до використання закономірності й отримати докладну статистику.

Модель інтелектуального аналізу даних, створювана алгоритмом з наданих даних, може мати різні форми, включаючи наступні [13]:

- набір кластерів, які описують зв'язки варіантів у наборі даних;
- дерево рішень, яке передбачає результат і описує, який вплив на цей результат виявляють різні критерії;
- математичну модель, яка прогнозує продажі;
- набір правил, які описують групування продуктів у транзакції, а також імовірності одночасної покупки продуктів.

У службах Microsoft SQL Server Analysis Services реалізовано декілька алгоритмів для використання в розв'язаннях інтелектуального аналізу даних. Ці алгоритми є реалізаціями деяких з найбільш популярних методів, що використовуються в інтелектуальному аналізі даних. Усі алгоритми інтелектуального аналізу даних Майкрософт налаштовуються, вони програмуються через API-інтерфейси або компоненти інтелектуального аналізу даних служб SQL Server Integration Services.

Крім того, підтримується використання сторонніх алгоритмів, відповідних до специфікації OLEDB для інтелектуального аналізу даних. Є також можливість розробляти власні алгоритми, які можна зареєструвати в якості служб, а потім використовувати в платформі інтелектуального аналізу даних SQL Server.

Вибір типу алгоритму здійснюється з урахуванням специфіки задач інтелектуального аналізу даних які вирішуються. Служби Analysis Services включають наступні типи алгоритмів:

- алгоритми класифікації здійснюють прогнозування однієї або декількох дискретних змінних на основі інших атрибутів у наборі даних;
- регресивні алгоритми здійснюють прогнозування однієї або декількох безперервних змінних, наприклад прибутку або збитків, на основі інших атрибутів у наборі даних;
- алгоритми сегментації ділять дані на групи або кластери елементів, що мають схожі властивості;
- алгоритми взаємозв'язків здійснюють пошук кореляції між різними

атрибути в наборі даних; Найбільш частим застосуванням цього типу алгоритму є створення правил взаємозв'язків, які можуть використовуватися для аналізу споживчого кошика.

- алгоритми аналізу послідовностей узагальнюють послідовності, що часто зустрічаються, у даних, наприклад потік даних в Інтернеті.

Однак ніщо не змушує користувача обмежуватися одним алгоритмом у своїх розв'язках. Досвідчені аналітики часто використовують один алгоритм для виявлення найбільш ефективних вхідних даних (тобто змінних), після чого застосовують інший алгоритм для прогнозування певного результату на основі цих даних. Інтелектуальний аналіз даних SQL Server дозволяє на базі однієї структури інтелектуального аналізу побудувати багато моделей таким чином, що в рамках одного розв'язку для інтелектуального аналізу даних можна було використовувати алгоритм кластеризації, модель дерева рішень, а також модель спрощеного алгоритму Байеса для отримання різних представлень даних. Кілька алгоритмів в одному розв'язку також можна використовувати для виконання окремих задач. Наприклад, за допомогою регресії можна отримувати фінансові прогнози, а за допомогою алгоритму нейронної мережі виконувати аналіз факторів, що впливають на обсяг продажів.

Щоб полегшити вибір алгоритмів для розв'язання певної задачі, у наступній таблиці наведені типи задач, для розв'язання яких звичайно використовується кожен алгоритм (табл.). У кожному випадку можна створювати різні моделі з різними алгоритмами або змінювати параметри алгоритмів для створення різних моделей. Користувач не обмежений у виборі алгоритму, і рекомендується створити декілька моделей з тими самими даними. Розглянемо особливості визначення даних для моделі. Крім вибору даних з певного джерела, необхідно вказати, яка таблиця у представленні джерела даних містить дані варіантів. Таблиця варіантів буде використовуватися для навчання моделі інтелектуального аналізу даних, тому вона повинна містити сутності, які потрібно аналізувати, наприклад, відомості про замовників і відповідні демографічні відомості. Кожний варіант повинен бути унікальним і ідентифікуватися по ключу варіанта. Крім вказівки таблиці варіантів можна включити в дані вкладені таблиці. Вкладена таблиця звичайно містить відомості, що застосовуються про сутності таблиці варіантів.

Це можуть бути транзакції, здійснювані замовником, або атрибути, що мають зв'язок із сутністю "багато до одного". Наприклад, вкладена таблиця, приєднана до таблиці варіантів Customers, може містити список покупок кожного замовника. У моделі, яка аналізує трафік Web-сайту, вкладена таблиця може містити послідовності сторінок, які відвідав користувач.

У службах SQL Server Analysis Services дані повинні подаватися на вхід алгоритму інтелектуального аналізу даних у вигляді послідовності варіантів, які містяться в таблиці варіантів. Але не всі варіанти можна описати одним рядком даних. Наприклад, деякий варіант може бути виведено на підставі двох таблиць: таблиця, яка містить дані про замовника, і таблиця, яка містить відомості про покупки замовника.

Одному замовникові з таблиці замовників може відповідати декілька елементів у таблиці покупок, що ускладнює опис даних в одному рядку. Сервіси Analysis Services надають унікальний метод обробки таких випадків за допомогою вкладених таблиць. Поняття вкладеної таблиці демонструється на наступній схемі (рис.1).

Таблиця 1 - Вибір релевантного алгоритму для розв'язання задачі

Приклади задач	Релевантні алгоритми
Прогнозування дискретного атрибута Відмітка клієнтів зі списку потенційних покупців як гарних і поганих кандидатів. Обчислення ймовірності відмови сервера протягом наступних шести місяців. Класифікація варіантів розвитку хвороб пацієнтів і дослідження пов'язаних факторів.	Алгоритм дерева прийняття рішень Спрощений алгоритм Байеса Алгоритм кластеризації Алгоритм нейронної мережі
Прогнозування безперервного атрибута. Прогноз продажів на наступний рік. Прогноз кількості відвідувачів сайту з урахуванням минулих років і сезонних тенденцій. Формування оцінки ризику з урахуванням демографії.	Алгоритм дерева прийняття рішень Алгоритм часових рядів Алгоритм лінійної регресії
Прогнозування послідовності Аналіз маршруту переміщення по Веб-сайту компанії. Аналіз факторів, які ведуть до відмови сервера. Відслідковування й аналіз послідовностей дій під час відвідування поліклініки з метою формулювання рекомендацій із загальних дій.	Алгоритм кластеризації послідовностей
Знаходження груп спільних елементів у транзакціях. Використання аналізу споживчого кошика для : визначення місць розміщення продуктів. Виявлення застосовуваних продуктів, які можна запропонувати купити клієнтові. Аналіз даних опитування, проведеного серед відвідувачів події, з метою виявлення того, які дії та стенди були поєднані, щоб планувати майбутні дії	Алгоритм взаємозв'язків Алгоритм дерева прийняття рішень

Майстер інтелектуального аналізу даних містить наступні функції, що допомагають вибрати дані й правильно налаштувати джерела даних: Автоматичні визначення типів даних; Пропозиції щодо змінних; Вибір компонентів. Більшість алгоритмів автоматично виявляє стовпці, найбільш відповідні для прогнозування, і використовують у першу чергу ці стовпці. У стовпцях з великою кількістю значень використовується функція вибору компонентів для зменшення кількості елементів зв'язку даних і підвищення ймовірності одержання значимої закономірності. На вибір компонентів можна вплинути за допомогою параметрів моделі.

Завершення роботи майстра. На останньому кроці виконання майстра вказуються імена структури інтелектуального аналізу даних і пов'язаної з нею моделі інтелектуального аналізу даних. Залежно від створеного типу моделі можуть надаватися наступні можливості:

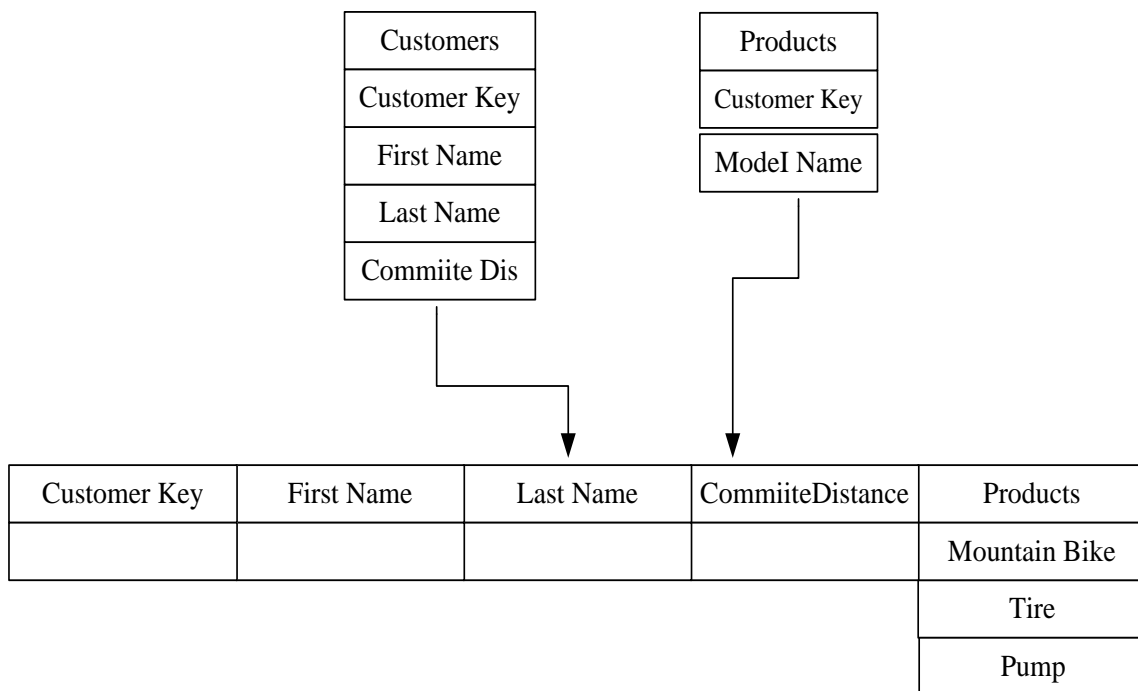


Рисунок 1 - Формування даних для розв'язання задачі на основі пов'язаних таблиць

Якщо обрано параметр «Дозволити деталізацію», у моделі включається функція деталізації. Якщо деталізація включена, користувачі з відповідними дозволами можуть переглядати вихідні дані, які використовувалися для побудови моделі. Після завершення роботи майстра інтелектуального аналізу даних за допомогою конструктора інтелектуального аналізу даних можна змінити структуру й моделі інтелектуального аналізу даних, переглянути точність моделі, переглянути характеристики структури й моделей або виконати прогнозування з використанням цих моделей. Після створення структури й моделі інтелектуального аналізу за допомогою майстра інтелектуального аналізу даних працювати з існуючими моделями й структурами можна за допомогою конструктора інтелектуального аналізу даних застосувань у середовищі SQL Server Data Tools (SSDL) або середовищі SQL Server Management Studio.

Конструктор містить засоби для розв'язання наступних задач:

Зміна властивостей структур інтелектуального аналізу даних, додавання стовпців і створення їх псевдонімів, зміна методу прив'язки або очікуваного розподілу значень.

Додавання нових моделей в існуючу структуру. копіювання моделей, зміна властивостей або метаданих моделі або визначення фільтрів на моделі інтелектуального аналізу даних.

Перегляд шаблонів і правил у рамках моделі. Дослідження зв'язків або дерев прийняття рішень. Є спеціалізовані засоби перегляду для кожної моделі, які допомагають аналізувати дані й досліджувати закономірності, виявлені в процесі інтелектуального аналізу даних.

Тестування моделі шляхом створення діаграми точності прогнозів або

аналіз кривої прибутку для моделей. Порівняння моделей за допомогою класифікаційних матриць або перевірка набору даних і його моделей шляхом перехресної перевірки.

Створення прогнозів і запити змісту існуючої моделі інтелектуального аналізу даних. Побудова запитів з невідомим компонентом, а також настроювання запитів для формування прогнозів для повних таблиць зовнішніх даних. Розглянемо проекти інтелектуального аналізу даних в Microsoft SQL Server Analysis. Проект інтелектуального аналізу даних - це частина розв'язання служб Analysis Services. Під час проектування об'єкти, створювані в цьому проекті, доступні для тестування й запитів у складі бази даних робочої області. Щоб дати користувачам можливість переглядати об'єкти в проекті й встановлювати до них запити, необхідно розгорнути проект в екземплярі служб Analysis Services, яка працює в багатомірному режимі.

Модель інтелектуального аналізу даних визначає алгоритм, тобто метод аналізу, який буде застосовуватися до даних. У кожен структуру інтелектуального аналізу даних додається одна або декілька моделей інтелектуального аналізу даних. Залежно від задач можна об'єднати декілька моделей в одному проекті або створити окремі проекти для кожного типу моделі або аналітичної задачі. Після створення структури й моделей виконується обробка кожної моделі. При цьому дані з представлення джерел даних проходять через алгоритм, який створює математичну модель даних. Цей процес також називається навчанням моделі. Після обробки стає доступним перегляд моделі інтелектуального аналізу даних і створення прогнозуючих запитів до неї. Якщо дані, отримані в процесі навчання, перебувають у кеші, то запити деталізації дають можливість отримати докладні відомості про варіанти, використовувані в моделі.

Якщо потрібно використовувати модель у робочому середовищі (наприклад, для складання прогнозів або для перегляду звичайними користувачами), то її можна розгорнути на іншому сервері. Якщо надалі знадобиться повторна обробка моделі, разом з моделлю можна експортувати визначення базової структури інтелектуального аналізу даних (при цьому також необхідно визначення джерела даних і представлення джерела даних). Під час розгортання моделі також необхідно перевірити правильність завдання параметрів обробки структури й моделі, а також забезпечити наявність у потенційних користувачів необхідних дозволів на виконання запитів, перегляд моделей і деталізацію даних структури або моделі.

При вирішенні передбачувальних задач інтелектуального аналізу якості прогнозу, який видає модель можна оцінити на перевірочному наборі даних, для якого відоме значення прогнозованого параметра. В MSSQL Server 2014 сервіси Analysis Services надають засоби, котрі спрощують розподіл даних на навчальний і перевірочний набори. Таке секціонування можна виконати автоматично під час побудови моделі інтелектуального аналізу даних. Точність прогнозів, створюваних моделями, можна перевірити за допомогою таких засобів, як діаграма точності прогнозів і матриця класифікації.

Інший підхід, який називається перехресною перевіркою, полягає в тім, що створюються підмножини даних і порівнюються результати роботи моделі на кожній підмножині. Такий підхід може використовуватися як при вирішенні передбачувальних (предиктивних), так і описових задач. Засоби автоматизації перехресної перевірки доступні при використанні MSSQLServer2014 версії Enterprise або Developer.

Висновки з даного дослідження.

В роботі здійснено системне дослідження підходів до аналізу даних засобами MS SQL SERVER 2014. Сформовано алгоритм створення моделі даних, яка доступна будь-яким клієнтським застосуванням, сформульовано підходи до вибору алгоритму для обробки даних, створення проектів інтелектуального аналізу даних. Доведено, що найбільш ефективні моделі розгортаються у виробничому середовищі. При цьому, можливі сценарії інтеграції засобів інтелектуального аналізу даних і користувацьких застосувань. І кінцевий користувач, у відповідь на сформований запит, буде отримувати результати аналізу у вигляді звіту. При формуванні звітів про результати проведеного аналізу можуть використовуватися можливості сервісу SQL Server Reporting Services. Встановлено, що з часом характеристики предметної області можуть змінюватися, що зажадає й зміни шаблонів інтелектуального аналізу даних. Для цього слід здійснити перенавчання існуючих моделей або створення нових. У ряді випадків SQL Server дозволяє автоматизувати процес відновлення моделей за рахунок використання сервісів.

Список літератури:

1. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И.Холод. –СПб : БХВ-Петербург, 2007. –384 с.
2. Data Mining в сфере бизнеса и финансов: Фондовые рынки [Электронный ресурс]. –Режим доступа: <http://www.franklin-grant.ru>
3. Data Mining. Применение. [Электронный ресурс]. –Режим доступа: <http://www.zsoft.ru/>
4. Джулій В.М. Структура, функції системи інтелектуальної обробки даних / В.М. Джулій, О.М. Горбатюк // Інформаційно-вимірювальні та обчислювальні системи і комплекси св технологічних процесах. – К. : 2014. – С.106-110.
5. Валькман Ю.Р. Графическое представление информации в проблемах принятия решений [Электронный ресурс]. –Режим доступа: <http://dialog-21.ru>
6. Дюк В. Data mining- интеллектуальный анализ данных [Электронный ресурс]. –Режим доступа: <http://www.bizoffice.ru/>
7. Большакова Е.И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб.пособие / Е.И. Большакова, Э.С. Клышинский, Д.В. Ландэ, А.А. Носков, О.В. Пескова, Е.В. Ягунова. –М. :

МИЭМ. 2011. –272 с.

8. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. –Спб. : Питер, 2001. – 182 с.

9. Джексон П. Введение в экспертные системы / П. Джексон. - М.: Издательский дом «Вильямс», 2001.

10. Додонов А.Г. Конкурентная разведка в компьютерных сетях / А.Г. Додонов. - К.: ИПРИ НАН Украины, 2013. - 248 с.

11. Дюк В. Data Mining : Учебный курс / В. Дюк. – СПб : Литер, 2001. – 386 с.

12. Зиновьев А.Ю. Визуализация многомерных данных / А.Ю. Зиновьев. – Красноярск : Изд-во Красноярского государственного технического университета, 2000. –180 с.

13. Лорьер Ж.-Л. Системы искусственного интеллекта / Ж.-Л. Лорьер. –М. : Мир, 1991. –568 с.

Аннотация

Интеллектуальный анализ данных средствами анализа и управления реляционными базами данных

Кудренко С.А.

Осуществлено системное исследование подходов к анализу данных средствами MS SQL SERVER 2014. Показано, что использование методов интеллектуального анализа и извлечения данных существенно повышает эффективность процесса принятия решений. Предложен алгоритм создания модели данных, доступной любым клиентским приложениям. Сформулированы подходы к выбору алгоритма для обработки данных и создания проектов интеллектуального анализа данных. Использование Microsoft SQL Server 2014 и инструментария Analysis Services обеспечивает автоматизированное создание моделей Data Mining, для решения различных задач.

Abstract

Data mining tools to analyze and manage relational databases

Kudrenko S.A.

The systematic study of approaches to data analysis by means of MS SQL SERVER 2014. It is shown that the use of predictive analysis and data mining allows to significantly increasing the effectiveness of decision-making. The algorithm for creating a data model, which is available to any client application. Formulated approaches to the choice of algorithm for data processing, creation of data mining projects. Using Microsoft SQL Server 2014 and Analysis Services provides tools for automated creation of models of Data Mining, which allows to solve a number of different tasks.