

УДК 681.3.06 (075.8)

С.В. ГРИБКОВ, Л.Г. ЗАГОРОВСЬКА
Національний університет харчових технологій, м. Київ

ВИКОРИСТАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТА ДОБУВАННЯ ДАНИХ ПРИ УДОСКОНАЛЕННІ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

Розглянуто вимоги до сучасних систем підтримки прийняття рішень та окреслено напрямки їх удосконалення. Показано, що використання методів інтелектуального аналізу та добування даних дозволяє суттєво підвищити ефективність управлінських рішень. Запропонований підхід реалізовано на прикладі макаронного виробництва. Використання СУБД MS SQL SERVER 2005 та інструментарію Analysis Services забезпечило автоматизоване створення моделей Data Mining, що дозволило розв'язати низку задач управління підприємством.

Проблема удосконалення форм та методів управління підприємством є нагальною для більшості вітчизняних харчових підприємств. Це обумовлено необхідністю задоволення сучасних вимог до якості продукції, специфікою виробництва та прагненням виживання в умовах жорсткої ринкової конкуренції. Тому управління підприємством повинно забезпечувати на усіх рівнях вироблення та прийняття ефективних управлінських рішень щодо характеру та обсягів виробництва, встановлення зв'язків з постачальниками сировини та споживачами продукції, взаємодії з фінансово-кредитними органами, вироблення стратегії і тактики використання виробничих ресурсів та науково-технічного потенціалу. Цього можна досягти за рахунок

інтелектуалізації процесів прийняття рішень з використанням нових прогресивних інформаційних технологій.

Аналіз стану автоматизації підприємств харчової галузі показав, що останнім часом набули розвитку системи автоматизованого управління технологічними процесами, які зазвичай поставляються з технологічним обладнанням, а удосконалення та підтримка функцій управління підприємством та його підрозділами не пішло далі автоматизації документообігу та розв'язання окремих локальних задач.

Зазначені вимоги та розвиток інформаційних інтелектуальних технологій спонукають до розроблення систем підтримки прийняття рішень (СППР) нового типу, основною задачею яких є пошук та надання інформації керівництву підприємства, що дасть можливість зрозуміти не тільки що сталося, а також чому це сталося та отримати рекомендації щодо вироблення та прийняття ефективних управлінських рішень [1, 2].

Тому актуальною задачею є дослідження та розроблення СППР, яка забезпечить прийняття оптимальних стратегічних рішень на усіх рівнях управління, що в кінцевому результаті забезпечить удосконалення форм та методів управління підприємством в цілому.

При розробленні СППР, які життєво необхідні в сучасних умовах для ефективного керування підприємством, необхідно враховувати значну кількість неформальних факторів, а саме:

- необхідність розуміння контексту завдання, яке розв'язується, загального процесу й знання можливостей інших систем та осіб, що приймають участь у прийнятті рішень;
- використання знань експертів на основі попереднього досвіду;
- рівень координування прийняття рішень та планування конкретних дій;
- розширення можливостей комп'ютерних програм за рахунок використання набутого досвіду й адаптація існуючих алгоритмів на основі аналізу попередніх результатів;
- забезпечення інтерактивного зв'язку з користувачами за допомогою розуміння комп'ютером природної мови, візуалізації отриманих результатів у вигляді малюнків, багатомірних графіків, діаграм, графів, інших формальних і неформальних зображень.

Як показує досвід, розв'язання багатьох проблем управління сучасним підприємством пов'язане з обробленням неповної, неточної або суперечливої інформації, а основні критерії та залежності нечітко або неоднозначно визначені. Для реальних задач управління не існує чітких алгоритмів розв'язання, тому для їх розв'язання слід застосовувати неструктуровані процедури прийняття рішень, які використовують знання, кваліфікацію та інтуїцію особи, яка приймає рішення.

З практики відомо, що чим більше аналітик може маніпулювати даними, будувати різні моделі, оцінювати отримані параметричні залежності, тим точнішим та змістовнішим буде результат аналізу. Крім цього оброблення інформації стає більш ефективним при поєднанні у використанні засобів візуалізації, різноманітного графічного інструментарію, автоматизованих засобів формування запитів, оперативної аналітичної обробки (OLAP), що дозволяє швидше зрозуміти залежності між даними, інтерпретувати проміжні й остаточні результати.

Сучасний рівень розвитку аналітичних технологій дозволяє створювати моделі, що істотно підвищують ефективність управлінських рішень. Для полегшення розв'язання задач управління виробничими підрозділами підприємства доцільно використовувати методологію інтелектуального аналізу даних (Data Mining) [3]. Поява даної методології обумовлена накопиченням значних обсягів інформації в базах даних (БД), які стало складно зберігати та користуватися традиційними способами. Через швидкий розвиток апаратного й програмного забезпечення доводиться постійно модифікувати значні об'єми даних, що призводить до неминучих втрат й викривлення інформації. Для запобігання цьому доцільно використовувати інформаційні сховища даних (СД), які дозволяють уникнути подібних втрат та забезпечать швидкий доступ до даних.

Технологія Data Mining реалізує новий підхід видобування інформації зі сховища даних. Основне призначення Data Mining – звільнення користувача від рутинних операцій і можливих труднощів у застосуванні сучасних статистичних і евристичних методів. Якщо раніше доступ користувача до СД допускав лише одержання відповідей на поставлені питання, то технологія Data Mining при її правильному застосуванні дозволяє віднайти приховані правила, зв'язки й закономірності в наборах різних даних, які користувач не може заздалегідь передбачити й правильно оцінити. Отримані нові знання про взаємозв'язки і причини їх утворення сприяють поглибленню розуміння досліджуваної предметної області і, як наслідок, прийняттю більш якісних і обґрунтованих рішень на усіх рівнях управління.

Досягнення успіху від застосування інтелектуального аналізу даних значною мірою залежить від дотримання наступних вимог:

- чіткого і зрозумілого формулювання завдання, що підлягає розв'язанню;
- використання адекватних даних і їх групування;
- можливості і здатності аналітика висувати різні гіпотези про зв'язок між даними.

Методологія Data Mining ґрунтується на підході побудови різноманітних моделей, які забезпечують при прийнятті рішень розв'язання задач класифікації, кластеризації, регресії, прогнозування, асоціації, виявлення послідовностей. При цьому дана методологія забезпечує реалізацію багаторівневих ієрархічних систем деталізації й узагальнення даних, одержання будь-яких їх проєкцій, що дозволяють розглядати як дрібні деталі, так і побачити картину в цілому.

На основі проведеного аналізу ринку інформаційних технологій для практичної реалізації зазначеної методології обираємо СУБД MS SQL Server 2005 із вбудованим інструментарієм Analysis Services, який повною мірою забезпечує автоматизоване створення моделей Data Mining та підтримує роботу з ними при їх

подальшому використанні [4]. Перевагами даного вибору є:

- забезпечення повного набору функціональних можливостей для створення моделей інтелектуального аналізу даних та подальшого їх супроводження;
- зручність у використанні за рахунок макросів для створення елементів підсистеми, графічного відображення моделей, їх структур та результатів реалізації;
- швидке інтегрування в існуючі інформаційні системи підприємства;
- створення систем на основі технології "клієнт-сервер";
- підтримка створення засобів завантаження та перетворення даних.

Зазначені методологію та інструментарій використовуємо при удосконаленні СППР для розв'язання цілої низки задач управління виробничого рівня ВАТ „Макаронна фабрика”. В якості першочергових виділяємо наступні задачі: визначення послідовності виконання замовлень, аналізу замовників, управління ресурсами, контролю якості продукції, матеріально-технічного забезпечення, оптимізації технологічних процесів, прогнозування техніко-економічних показників тощо.

Як основу інформаційного забезпечення СППР використовуємо єдине СД, структуру якого запропоновано та обґрунтовано в роботах [5, 6]. Використання єдиного СД обумовлено тим, що для повноти аналізу та розв'язання задач прийняття рішень необхідно мати швидкий доступ до хронологічних предметно-орієнтованих наборів даних. Запропонована структура забезпечує узгодження даних із електронних джерел та заповнення ними стандартного архіву даних й тематичних областей. Стандартний архів даних має нормалізовану структуру, що забезпечує зберігання детальної інформації про функціонування підприємства в цілому. Тематичні області складаються з вітрин та окремих сховищ даних, які мають денормалізовані структури та містять агреговані дані. Кожна тематична область орієнтується на розв'язання певного класу задач ланки управління.

Моделі вітрин та сховищ даних створюємо CASE-інструментарієм автоматизованого моделювання й проектування AllFusion ERwin Data Modeler 7.1. Фрагмент моделі СД для розв'язання задач аналізу та підтримки прийняття рішень в контурі управління виробництвом наведено на рис. 1.

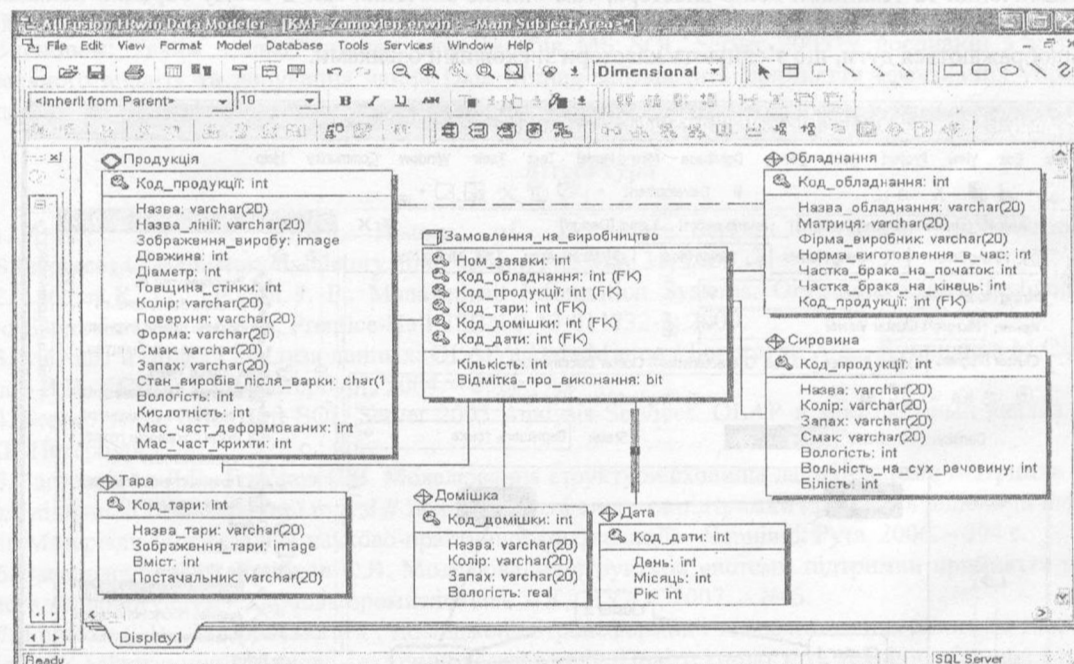


Рис. 1. Фрагмент моделі СД

Для ефективного проектування моделей СД використовуємо методологію багатовимірного проектування (Dimensional). Моделі вітрин та СД будуюмо за схемою „сніжинка” з денормалізованою структурою даних згідно до зазначеної методології. Такі моделі складаються з таблиць фактів, розмірностей та консольних таблиць, кожна з яких визначається своєю піктограмою. У наведеному фрагменті моделі СД (рис. 1) таблицю фактів є „Замовлення на виробництво”, таблицями розмірностей є „Обладнання”, „Сировина”, „Тара”, „Домішка” й „Дата”, а консольною таблицею є „Продукція”. Дана модель призначена для вітрини даних щоб відображувати стан замовлень на виробництво продукції.

Використовуючи інструментарій AllFusion ERwin Data Modeler 7.1 та створені моделі, генеруємо вітрини та СД у СУБД MS SQL Server 2005. Після чого реалізуємо задачу інтеграції створеного єдиного СД з діючим інформаційно-обчислювальним комплексом підприємства. Така інтеграція забезпечує автоматизоване заповнення єдиного СД відповідною інформацією. Для цього створюємо пакети перетворення та завантаження даних ETL (Extract, Transform, Load) засобами MS SQL Server 2005 Integration Services (SSIS), які забезпечують гнучку та потужну архітектуру для ефективної інтеграції даних. Побудова ETL-пакетів відбувається в

автоматизованому режимі. Ці пакети реалізують: всебічний доступ до інформації, представленої в різних електронних джерелах даних; перетворення, розбиття та агрегацію даних; завантаження інформації у єдине СД в автоматизованому режимі за заданим розкладом. Обрана технологія SSIS забезпечує глибоку інтеграцію пакетів з методами Data Mining Analysis Services. Крім цього вона забезпечує підтримку систематизації та аналіз текстової інформації без попередньої обробки, а також висунення припущень про її корисність на основі заданих бізнес-правил.

Пакети програм розробляємо в інтегрованому середовищі Visual Studio засобами MS SQL Server Business Intelligence (BI) Development Studio. Обрані засоби підтримують створення проектів з використанням інструментарію MS SQL Server Integration Services, Analysis Services, Reporting Services та сучасних мов програмування (Visual C#, Visual Basic.NET тощо). Зазначений вибір обумовлено його спроможністю поєднувати в собі різні напрямки розроблення, інтеграції, аналізу, формування звітів, створення додатків на рівні метаданих та можливості відлагоджування пакетів під час їх виконання.

Практичну реалізацію моделей Data Mining та проектів з їх використанням здійснюємо у BI Development Studio. Дане інтегроване середовище забезпечує їх побудову, випробування та застосування для розв'язання задач управління. Моделі являють собою сукупність метаданих, що відображають бізнес-правила й закономірності в джерелах даних, та обрані алгоритми Analysis Services.

Для розв'язання задачі аналізу попередніх та шойно сформованих замовлень з метою виявлення споріднених будуємо модель Data Mining. При побудові цієї моделі в якості джерела даних обираємо заздалегідь створену тематичну вітрину даних „Замовлення”, в основі якої є метадані, що описують властивості замовлення, а в якості алгоритму аналізу даних – алгоритм кластеризації. Провівши ініціалізацію створеної моделі, отримуємо 5 кластерів замовлень за значеннями своїх атрибутів.

На рис. 2 подано діаграму взаємодії кластерів за показником „Продукція” та його значенням „Вермішель тонка”. Ця діаграма дозволяє в автоматизованому режимі здійснити аналіз та перегляд варіантів поєднання кластерів та відношення належності за значенням певного атрибуту аналізу. Насичення кольору визначає належність кластера до входження в групу, яка поєднується за вказаним значенням обраного атрибуту. Чим більш насичений та темніший колір кластера, тим більше значення має в ньому обраний показник. Дана діаграма відображає частку продукції „Вермішель тонка” у кожному кластері замовлень. За аналогічним правилом відображаються дуги, що з'єднують кластери зі схожими ознаками.

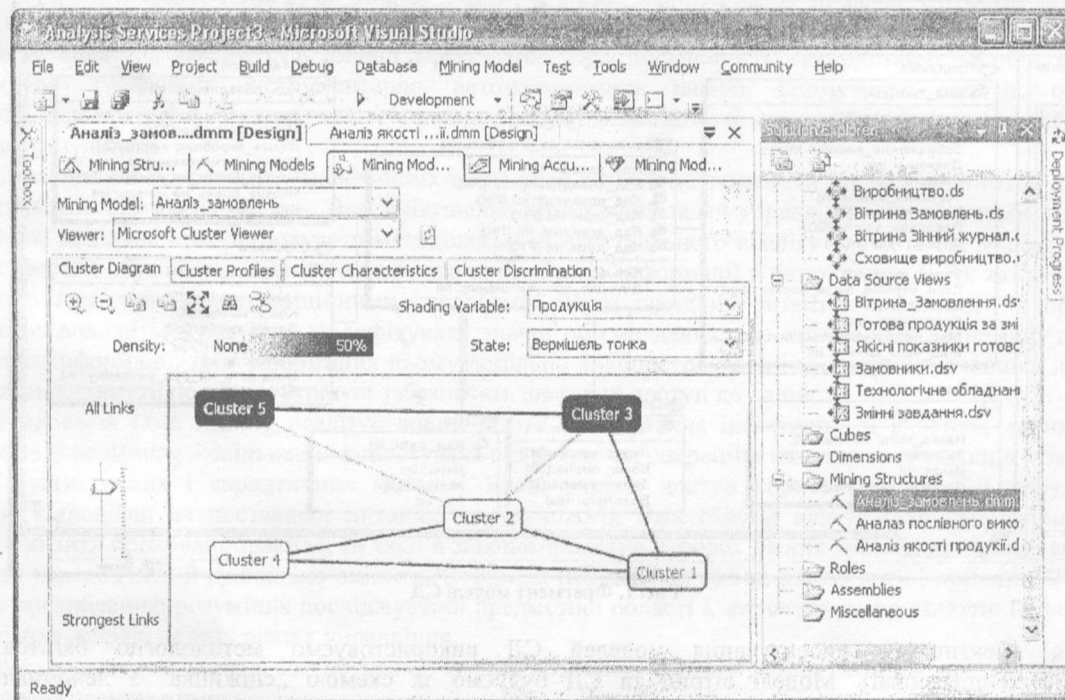


Рис. 2. Діаграма кластерів

Вибравши закладку „Cluster Characteristics” на панелі інструментів, отримуємо діаграму характеристик кластерів (рис. 3).

Дана діаграма відображає характеристики продукції, що належить до третього кластера. Аналізуючи ці показники, бачимо, що деякі з них притаманні усім видам продукції. Наприклад, вся продукція пакуватиметься у пакети з плівки „Золота фортуна”, виготовлятиметься з борошна хлібопекарського та міститиме В-каротин. Інші характеристики відносяться лише до певних видів продукції („Ріжки виток” та „Вермішель тонка”).

Використовуючи інші можливості обраного інструментарію та створену модель, отримуємо діаграми зрізів різних характеристик продукції та діаграми порівнянь.

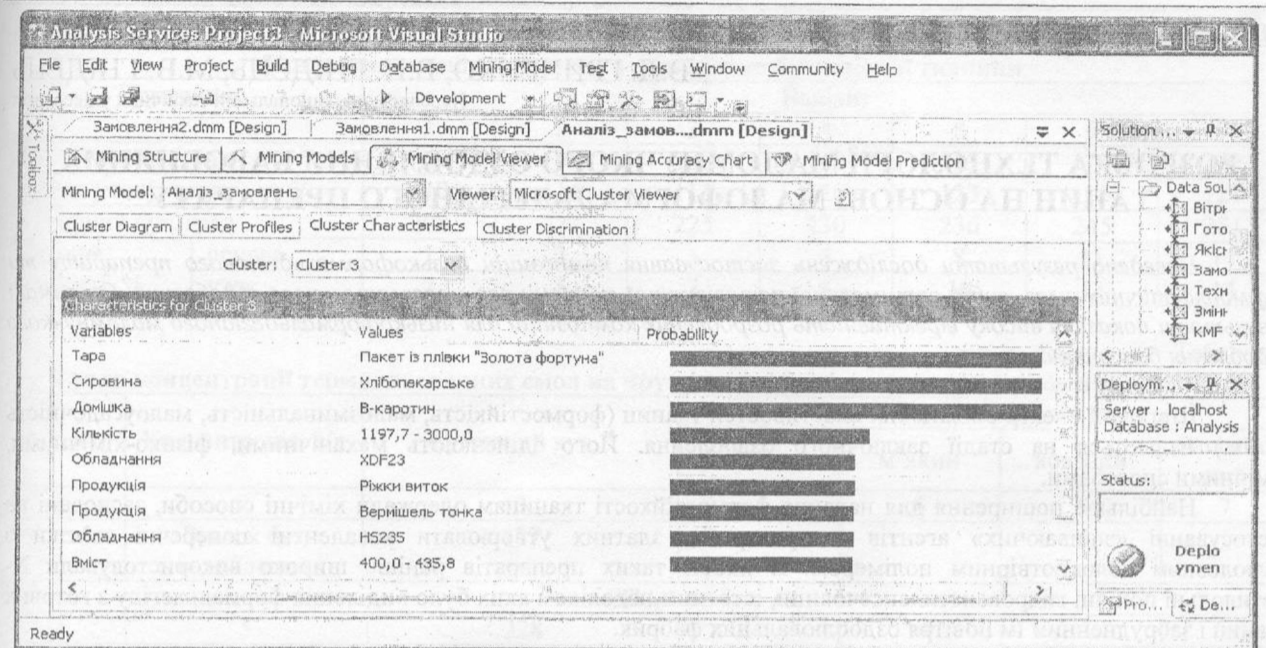


Рис. 3. Діаграма характеристик кластерів

На основі всебічного аналізу замовлень за допомогою побудованої моделі та різновидів діаграм поєднуємо окремі замовлення за набором споріднених характеристик та формуємо групи замовлень для виконання. Розв'язання даної задачі сприяє ефективному використанню технологічного обладнання та призводить до зменшення собівартості готової продукції.

Висновок: Таким чином, використання засобів MS SQL Server 2005 у поєднанні з методологією інтелектуального аналізу та добування даних Data Mining забезпечить прийняття ефективних управлінських рішень, що підвищить якість системи управління підприємством в цілому.

Література

1. Power, D.J. A Brief History of Decision Support Systems, DSSResources.COM, <http://DSSResources.COM/history/dsshistory.html>, version 2.8, May 31, 2003.
2. Laudon K. C., Laudon J. P., Management Information Systems: Organization and Technology in the Networked Enterprises, 6th Edition, Prentice-Hall, ISBN 0-13-011732-3, 2000.
3. Методы и модели анализа данных: OLAP и Data Mining / Берсегян А.А., Куриянов М.С., Степанов В.В., Холод И.И. – СПб.: БХВ-Петербург, 2004. – 336 с.: ил.
4. Бергер А. В. Microsoft SQL Server 2005 Analysis Services. OLAP и многомерный анализ данных. – СПб.: БХВ-Петербург, 2007. – 928 с.: ил.
5. Загорюська Л.Г., Грибков С.В. Моделювання структури сховища даних системи підтримки прийняття рішень для підприємств макаронної галузі // Інтелектуальні системи підтримки прийняття рішень та інформаційні технології: Матеріали Міжнародної науково-практичної конференції. – Чернівці: Рута, 2006. – 304 с.
6. Загорюська Л.Г., Грибков С.В. Моделювання структури системи підтримки прийняття рішень для підприємств харчової галузі / Харчова промисловість. – К.: НУХТ, 2007. – № 5.
7. Грибков С.В., Загорюська Л.Г. Дослідження трансформації та завантаження даних до сховища даних // Матеріали ІХ Міжнародної науково-технічної конференції „Нові технології та технічні рішення в харчовій та переробній промисловості: сьогодні і перспективи”, 17 – 19 жовтня 2005 р: У 2 ч. – Ч. II. – К.: НУХТ, 2005 – 77 с.

Надійшла 23.9.2008 р.