

# 数据挖掘技术在管理中的应用

李勇,肖智,曹国华

(重庆大学 工商管理学院,重庆 400044)

**摘要:**数据挖掘与现代管理与决策有很好的结合点。本文概述了数据挖掘技术的概念、挖掘系统典型结构、挖掘对象、挖掘方法,并探讨了数据挖掘技术在营销管理、客户关系管理、证券分析、金融分析以及政府管理与决策等典型领域的应用。

**关键词:**数据仓库;数据挖掘;金融分析;证券分析;决策支持

中图分类号:F270.7 文献标识码:A 文章编号:1008-5831(2002)06-0026-03

## Applying Data Mining Technology to Management

LI Yong, XIAO Zhi, CAO Guo-hua

(College of Business Administration, Chongqing University, Chongqing 400044, China)

**Abstract:** There is a good combination between data mining and modern management decision. This paper summarizes the concepts of data mining, architecture of a typical mining system, objects of mining and methods of mining. The paper also discusses the application of data mining about market management, client relation management, securities analysis, financial analysis, government decision, and etc.

**Key words:** data Warehouse; data Mining; financial analysis; securities analysis; decision support

### 一、数据挖掘的定义

简单地说,数据挖掘是从大量数据中提取或“挖掘”知识。也有许多学者将数据库中的知识发现或 KDD(Knowledge Discovery in Database)视为数据挖掘的同义词,还有许多类似的术语,如知识提取(Knowledge Extraction)、数据/模式分析(Data/Pattern Analysis)、数据捞取(Data Dredging)等。<sup>[1]</sup>

本文采用关于数据挖掘的广义的观点:数据挖掘是从存放在数据库、数据仓库或其他信息库中的大量数据中发现和提取有趣知识的过程,且数据挖掘是知识发现过程的一个步骤(图1)。

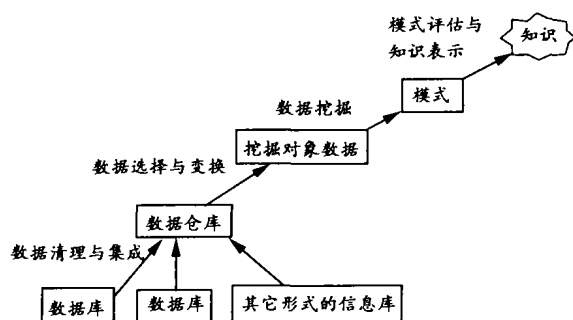


图1 数据挖掘与知识发现

### 二、数据挖掘系统结构

根据上述定义,典型的数据挖掘系统具有以下主要部分(图2):

**数据库、数据仓库或其他信息库:**这是一个或一组数据库、数据仓库、电子表格或其它类型的信息库,可以在数据上进行数据清理和数据集成;

**数据库或数据仓库服务器:**根据用户数据挖掘请求,数据库或数据仓库服务器负责提取相关数据;

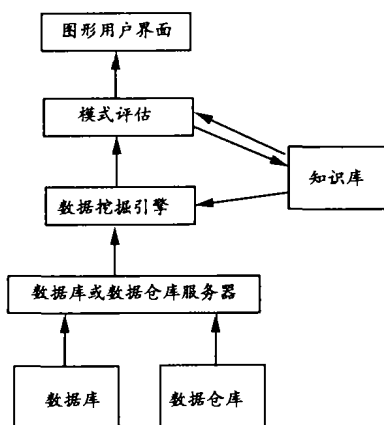


图2 典型的数据挖掘系统结构

知识库:库中存放领域知识,指导搜索或模式评估;

收稿日期:2002-06-22

作者简介:李勇(1967-),男,四川德阳人,重庆大学工商管理学院博士研究生,主要从事商务智能、信息系统与决策支持系统研究。

数据挖掘引擎:这是数据挖掘系统基本的部分,由一组功能模块构成,负责实施数据挖掘的操作;

模式评估模块:此模块通常通过兴趣度的度量,使挖掘聚焦到有趣(用户感兴趣)的模式上;

图形用户界面:本模块在用户和数据挖掘模块之间交互,并以可视化的方法向用户提交挖掘结果。

### 三、数据挖掘对象

从理论上讲,数据挖掘可以在任何类型的信息库上进行,如:关系数据库、数据仓库、事务数据库、高级数据库系统、一般数据文件和 WWW。高级数据库系统包括面向对象和对象-关系数据库,面向特殊应用的数据库如空间数据库、时间序列数据库、文本数据库和多媒体数据库。对不同类型的挖掘对象,将采用不同的挖掘算法和技术,就面向管理的数据挖掘应用而言,较普遍的挖掘对象一般是关系数据库、数据仓库、事务数据库、时间序列数据库等。<sup>[1]</sup>

### 四、数据挖掘方法

数据挖掘方法是数据挖掘研究的核心内容,国内外已经产生了大量的研究成果,有的已经开发出了实用的数据挖掘系统软件或系统原型。数据挖掘方法一般可以分为描述和预测两大类,描述性挖掘数据源中数据的一般特性,预测型挖掘分析数据,建立一个或一组模型,并试图预测新数据的行为,具体介绍如下。

#### (一)概念描述(Concept Description)

概念描述由特征化和比较组成。前者对目标类的数据集进行汇总,并对目标类的一般特性进行描述(也称为数据概化);而后者对目标类数据集和对比类数据集进行汇总概化,并对二者进行比较和区别。

概念特征化有两种一般方法:基于数据立方体 OLAP 的方法和面向属性归纳的方法。面向属性归纳方法包含以下技术:数据聚焦、通过属性删除或属性概化数据、计数和聚集值累加、属性概化控制和概化数据可视化。概化数据可以用多种形式可视化,包括概化关系、交叉表、条形图、饼图、数据立方体视图、曲线和规则等。

概念比较可以用类似于概念特征化的方式,使用面向属性的归纳或数据立方体方法进行。可以量化地比较从目标类和对比类概化得到的元组。<sup>[1]</sup>

#### (二)关联规则挖掘(Mining Association Rules)

关联规则分析发现大量数据中项集之间有趣的关联或相互联系,比如在顾客购买行为分析中发现形如“90%的顾客在一次购买活动中购买商品 A 的同时购买商品 B”之类的知识即是关联规则挖掘的一个典型实例。关联规则挖掘的典型应用包括顾客购物分析、超市卖场设计、仓储规划、股票分析等。

常用的关联规则挖掘算法有 Apriori 算法、频繁模式增长(FP-增长)、多层次关联规则挖掘、多维关联规则挖掘、基于约束的关联规则挖掘等。

#### (三)分类和预测(Classification and Prediction)

分类和预测可以用于提取重要数据类的模型或预测未来的数据趋势。分类是预测类标号,而预测则是通过建立连续值函数的模型来预测数据趋势。

数据分类是一个两步过程。第一步,随机的选取一组训练样本构成训练数据集,对训练数据集应用学习模型(如分类规则、判定树或数学公式等),从而建立一个分类模型,用于描述预定义的数据类集或概念集;第二步,使用分类模型进行分类,首先使用测试数据集评估分类模型的预测准确率,如果模型的准确率是可以接受的,则可以继续对其他类标号未知的数据集(也称为分析数据集)中的元组或对象进行分类。

预测是构造和使用模型评估无标号样本类,或预测给定样本可能具有的属性值或值区间。

常用的分类方法有判定树归纳分类、贝叶斯分类、基于案例的推理、遗传算法、粗糙集等;常用的预测算法有线性回归与多元回归、非线性回归、广义线性模型等。

#### (四)聚类分析(Cluster Analysis)

将物理或抽象对象的集合分组成为由类似的对象组成的多个类的过程,称为聚类。由聚类所生成的簇(Cluster)是一组数据对象的集合,同一个簇中的对象彼此相似,与其他簇中的对象相异。在许多应用实践中,可以将一个簇中的数据对象作为一个整体来对待。聚类与分类的主要不同点在于聚类不依赖预定义的类和带有类标号的训练数据集,它要划分的类是未知的,故将聚类称为观察式学习,而不是示例式学习。

主要的聚类方法有:划分方法;基于密度的方法;基于网络的方法;基于模型的方法,如统计学方法、神经网络方法等;孤立点分析,如基于距离的孤立点检测、基于偏离的孤立点检测等。

#### (五)复杂数据类型的挖掘(Mining Complex Type of Data)

通常的数据挖掘方法多用于对以结构化数据为主的关系数据库、事务数据库、数据仓库进行挖掘,而随着数据处理工具、先进数据库系统以及 INTERNET 技术的迅猛发展,大量的形式多样的复杂类型的数据不断涌现,如复杂对象、空间数据、多媒体数据、时间序列数据、文本数据和 Web 数据等。对这些类型数据的分析成为数据挖掘的一个重要领域。在企业及行政事务管理领域,对时间序列数据、文本数据以及 Web 数据的挖掘具有重要的实际作用。<sup>[1]</sup>

### 五、数据挖掘在管理中的应用

现代管理离不开信息技术,如管理信息系统、决策支持系统、电子商务系统、在线事务处理与在线分析处理等都是现代管理与信息技术紧密结合的典型实例。数据挖掘技术的产生与发展的主要动因之一就是来自于现代企业管理及现代行政管理的需要;相应地,数据挖掘以及相关技术的发展也推动了现代管理与决策科学的进步。目前,国内外大量专家学者围绕数据挖掘在管理中的应用这一主题进行了大量的研究和实践,特别是在管理决策支持、市场营销、客户管理、证券分析、金融分析、政府财政管理与分析等领域应用数据挖掘技术开展了大量的探索<sup>[2-4]</sup>。下面,对管理中应用数据挖掘技术的一些典型情况作一个概要介绍:

#### (一)数据挖掘在零售业中的应用

零售业(包括传统的超级市场和电子商务条件下的网上超级市场)是数据挖掘的主要应用领域,这是因为零售业积

累了大量的销售数据,如顾客购买记录、货物进出记录、消费与服务记录等,这为数据挖掘提供了丰富的资源。

在零售业中建立和运用数据挖掘系统,通常可以进行以下几方面的工作:(1)进行销售、顾客、产品、时间和地区的多维分析(在数据仓库的基础上);(2)促销活动的有效性分析;(3)顾客保持力(顾客忠诚度)分析;(4)购买推荐方案分析与设计;(5)卖场设计方案分析;(6)企业经济指标分析,如利润、成本分析与预测等;(7)通过分类、聚类挖掘等手段,为对会客客户进行个性化服务提供帮助,等等。

#### (二)数据挖掘在客户关系管理(CRM)中的应用

客户关系管理(CRM)是指对企业和客户之间的交互活动进行管理的过程。CRM的基本思想是:“客户是企业利润的中心,而不是成本的中心”,所以CRM的主要活动就是对市场进行分割,找出现有的和潜在的高利润的客户,然后策划和实施促销活动,设计和实施高效率、高质量的客户服务,从而达到提高企业的竞争力和增加企业利润的目标。<sup>[6]</sup>

要实现上述目标,就需要关于客户和他们购买行为的有效数据。从理论上讲,数据越多越好,但数据越多,分析人员得到信息就越难,所以,数据挖掘技术就成为CRM的有力的支撑工具。数据挖掘技术可以帮助CRM进行以下几方面的工作:(1)客户盈利能力分析;(2)潜在客户分析与新客户获取;(3)交叉营销分析与管理;(4)客户等级划分(客户的细分)与客户分级管理;(5)客户服务需求分析与客户的保持等。

#### (三)数据挖掘在证券分析中的应用

作为市场经济重要特征的证券市场,在国民经济生活中有着重要的地位和影响。证券是市场的内部规律非常复杂,影响证券市场的因素众多,如政治、经济形势、投资者心理、上市公司情况等等,所以按照一般的手段很难摸清证券市场的变化规律。对证券市场进行有效分细的途径之一是对海量交易数据、客户数据、上市公司经营情况数据以及国家宏观政治经济信息进行有效的管理和深度的分析。以我国上海和深圳两个交易所为例,上市股票达1500只左右、投资者达数百万、每天的交易笔数都平均在百万条以上。这是一个信息的宝库,在现有的技术条件下,我们可以运用大型数据库系统、数据仓库系统、在线分析处理系统(OLAP)和数据挖掘技术对海量证券数据进行有效管理和深度分析,为证券管理机构、证券交易所、证券公司、证券咨询公司以及广大上市公司和投资者的管理与决策提供有力的帮助。数据仓库和数据挖掘在证券分析中的应用一般有以下几个方面:(1)对证券交易数据进行在线多维分析(以数据仓库为基础);(2)对上市公司进行分类与聚类挖掘,为投资者和证券公司提供参考;(3)对个股、或某类股票或大盘进行时序分析(如趋势分析、序列模式挖掘和周期分析等);(4)对投资者进行分类与聚类,为证券公司和证券咨询公司进行有效的投资者关系管理提供帮助(5)对投资者进行投资偏好分析;(6)对投资者的异常交易行为进行挖掘;(7)对证券市场中的犯罪行为进行挖掘分析,等等。

#### (四)数据挖掘在金融管理及分析中的应用

绝大多数银行和金融机构都提供丰富多样的储蓄服务(如支票、存款和商业及个人用户交易)、信用服务(如商业和消费贷款等)、投资服务,有些还提供保险服务和股票投资服务等。在银行及其它金融机构中产生的金融数据通常比较完整、可靠和高质量,这为金融数据分析与挖掘奠定了良好的基础<sup>[8]</sup>。在金融数据中进行挖掘分析通常可以进行以下一些工作:(1)为多维数据分析和数据挖掘设计和建立数据仓库;(2)贷款偿还预测和客户信用政策分析;(3)对目标客户群(准客户)进行分类与聚类;(4)洗黑钱和其他金融犯罪的侦破,等等。

#### (五)数据挖掘在政府行政管理及决策分析中的应用

信息技术在政府管理中应用非常广泛,诸如政府上网工程、电子政务建设、政府办公自动化系统等等,已成为政府管理现代化的重要内容。决策是政府管理中最重要活动,正确、及时的决策需要全面、准确、及时的信息和正确的决策方法。各种政府信息系统的建立和运行为政府决策信息的获取奠定了数据基础,如工商管理部的企业注册及年度生审查管理系统积累了大量的企业基本数据以及年度审查数据,税收部门的税务征管信息系统则积累了各级各类企业缴税情况的庞大数据资源,海关和对外经贸管理部门的进出口贸易管理信息系统收集和存储了企业详细的进出口贸易数据,公安部门的户籍管理信息系统以及罪犯管理信息系统拥有巨大的公民户籍资料数据库和详尽的罪犯资料数据库等等。这些宝贵的数据资源保障了政府部门的日常管理工作高效、准确地进行。但根据我们调研的情况表明,这些数据资源在支持各级领导,尤其是高层领导的决策方面没有起到应有的作用,主要原因是如何从海量的数据获取信息的问题没有得到解决。从信息技术的角度考虑,在各类政府的政务信息管理系统基础上建立各个层次的数据仓库系统,在此基础上进行常规的在线分析(OLAP)处理和更深入的数据挖掘处理,是一个可行的解决方案。当然,如何在政府管理信息数据仓库上进行在线分析、分析什么,采用什么挖掘方法、挖掘什么等等,需要作更详细的研究。

数据挖掘技术及其为核心内容之一的知识科学在现代管理技术,尤其在管理智能化、决策智能化方面具有广阔的应用前景,需要人们在理论探讨和应用实践方面不断地做出更多新的工作。

#### 参考文献:

- [1] Jiawei Han, Micheline Kamber. DATA MINING — Concept and Techniques(Canada)[M]. 北京:高等教育出版社,2001.
- [2] Anonymous. Data mining/CRM: Searching for an ROI, Chain Store Age [M]. New York: 2001.
- [3] Owen P Hall Jr. Mining the store[J]. The Journal of Business Strategy, 2001, (3).
- [4] Larry Adams. Mining the world of quality data[J]. 2001.
- [5] Alex Berson, Stephen Smith, 等. 构造面向CRM的数据挖掘应用[M]. 贺奇,等译. 北京:人民邮电出版社,2001.