

<<数据仓库与数据挖掘原理及应用>>

图书基本信息

书名：<<数据仓库与数据挖掘原理及应用>>

13位ISBN编号：9787302228196

10位ISBN编号：7302228191

出版时间：2011-1

出版时间：清华大学

作者：郑岩

页数：300

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<数据仓库与数据挖掘原理及应用>>

### 前言

数据仓库是将大量传统数据库数据进行抽取、清洗和转换，并按主题进行重新组织，可比喻为随时间推移不断丰富的“宝藏”；而数据挖掘是从海量数据中发现人们感兴趣的知识，这些知识是隐含的、事先未知的潜在有用信息，挖掘的知识表示形式为概念、规则、规律和模式等，可比喻为“淘宝”。随着Internet的迅速普及和广泛应用，每天都产生大量各种各样的信息，但它们背后到底隐藏着什么，这驱使人类不断探索。

工欲善其事必先利其器。

在当今信息爆炸的时代，数据挖掘堪比“利器”，让我们面对海量数据时不再感到茫然和不知所措。随着数据仓库的发展和应用，数据挖掘将展现无限的生机和活力，可以辅助、部分代替甚至拓展人的智能和决策，造福人类。

数据经整合汇总为信息，信息经挖掘抽象为知识，知识是智能的基石。

因此，信息化到知识化再到智能化将是人类社会发展的必然趋势。

数据仓库和数据挖掘正逐步渗透和深入到社会的各个领域，并不断催生新的应用。

本书主要介绍数据仓库和数据挖掘的理论、方法、技术及其应用。

此外，用较多篇幅阐述数据仓库和数据挖掘新的应用实例。

全书分为三篇。

第一篇介绍数据仓库的起源和演变过程，阐述数据仓库的定义、体系结构、组成、元数据、数据粒度和数据模型以及ETL过程，论述数据仓库设计和实现的方法，并结合具体应用详细阐述了如何构建数据仓库及其主要应用，包括OLAP和OLAM等。

第二篇介绍数据挖掘的起源和发展趋势，以及数据挖掘与web挖掘的技术和方法，包括聚类分析、分类、预测和关联分析等，详细分析了数据挖掘在电信领域的具体应用，如客户细分、重入网识别和WAP日志挖掘等。

第三篇讨论数据、信息和知识的关系，论述知识表示的主要方法和知识管理的核心技术，介绍当前研究热点——语义网和本体的核心技术和方法，分析了语义网和本体的主要应用。

## <<数据仓库与数据挖掘原理及应用>>

### 内容概要

《数据仓库与数据挖掘原理及应用》从专业角度全面介绍了数据仓库和数据挖掘的理论、方法、技术及其应用，系统地阐述了数据仓库和数据挖掘的产生、发展和应用及其主要概念、原理和算法，并结合当前数据仓库和数据挖掘中一些新的应用实例进一步加以说明，力求学以致用。

全书分为三篇。

第一篇介绍数据仓库的起源和演变过程，阐述数据仓库的定义、体系结构、组成、元数据、数据粒度和数据模型以及ETL过程，论述数据仓库设计和实现的方法。

结合具体应用详细阐述了如何构建数据仓库及其主要应用，包括OLAP和OLAM等。

第二篇介绍数据挖掘的起源和发展趋势，以及数据挖掘与web挖掘的技术和方法，包括聚类、分类、预测和关联分析等，详细分析了数据挖掘在电信领域的具体应用，如客户细分、重入网识别和WAP日志挖掘等。

第三篇讨论数据、信息和知识的关系，论述知识表示的主要方法和知识管理的核心技术，介绍当前研究热点——语义网和本体的核心技术和方法，分析了语义网和本体的主要应用。

《数据仓库与数据挖掘原理及应用》可作为计算机专业研究生或高年级本科生教材，也可以作为计算机研究和开发人员以及相关专业人士的参考资料。

## 书籍目录

第1章 数据仓库基础1.1 引言1.1.1 演变过程1.1.2 定义1.2 体系结构1.2.1 两层的体系结构1.2.2 三层的体系结构1.3 组成1.4 元数据1.4.1 定义和分类1.4.2 标准化1.4.3 CWM1.4.4 UMI、MOF和XML与CWM的关系1.5 数据粒度1.6 数据模型1.7 ETI1.7.1 主要流程1.7.2 数据抽取1.7.3 数据转换1.7.4 数据加载第2章 数据仓库设计和实现2.1 数据仓库设计2.1.1 设计方法2.1.2 体系结构设计2.1.3 数据模型设计2.2 ETL设计2.3 数据仓库实现第3章 数据仓库实例3.1 实例一3.1.1 选择主题3.1.2 逻辑模型设计3.1.3 物理模型设计3.1.4 ETL设计3.2 实例二3.2.1 总体结构设计3.2.2 概念模型设计3.2.3 逻辑模型设计3.2.4 物理模型设计3.2.5 数据清洗设计3.2.6 ETL设计第4章 OLAP和OLAM4.1 OLAP4.2 OLAM4.2.1 体系结构4.2.2 特点4.2.3 基于Web的OLAM第二篇 数据挖掘第5章 数据挖掘基础5.1 概述5.1.1 定义5.1.2 功能5.1.3 模型5.1.4 展望5.2 实现5.3 工具5.3.1 概述5.3.2 比较第6章 聚类分析6.1 硬聚类6.1.1 算法种类6.1.2 相似度计算6.1.3 实现方法6.1.4 主要算法6.2 模糊聚类6.2.1 概述6.2.2 主要算法6.3 评价第7章 分类和预测7.1 神经网络7.2 决策树7.3 实现过程第8章 关联分析8.1 概述8.2 Apriori8.3 FP-Growth第9章 Web挖掘9.1 概述9.1.1 定义9.1.2 自然语言理解9.1.3 Web挖掘过程9.2 Web文档抽取和表示9.2.1 Web文档抽取9.2.2 Web文档表示9.3 特征提取9.4 Web聚类9.5 Web分类9.5.1 朴素贝叶斯9.5.2 其他方法9.5.3 评价第10章 数据挖掘实例10.1 TOM和TOM10.2 客户细分10.2.1 客户生命周期10.2.2 客户价值10.2.3 数据准备10.2.4 分析过程10.2.5 结果10.3 重入网识别10.3.1 定义10.3.2 数据准备10.3.3 分析过程10.3.4 结果10.4 WAF日志挖掘10.4.1 定义10.4.2 数据准备10.4.3 分析过程10.4.4 结果第三篇 语义网和本体第11章 知识11.1 概述11.2 知识分类11.3 知识表示11.3.1 知识表不观11.3.2 知识表示方法11.4 知识管理11.4.1 概述11.4.2 知识管理与信息管理的关系11.4.3 核心技术第12章 语义网和本体12.1 语义网12.1.1 概述12.1.2 层次结构12.1.3 元数据12.1.4 核心技术12.1.5 开发工具Jena12.1.6 Web3.012.2 本体12.2.1 哲学本源12.2.2 定义12.2.3 建模12.2.4 分类12.2.5 构建方法12.2.6 描述语言12.2.7 实例参考文献

## <<数据仓库与数据挖掘原理及应用>>

### 章节摘录

插图：进入信息时代以来，特别是近些年，数据库规模日益扩大，数据呈爆炸性增长。

图灵奖获得者吉姆·格雷提出了一个经验定律，即网络环境下每18个月产生的数据量等于有史以来的数据量之和，仅仅依靠数据库管理系统的查询检索机制和统计分析方法，已经远远不能满足实际需求，面临着“数据爆炸，知识匮乏”的严峻挑战。

例如股票经纪人需要从日积月累的大量股票行情变化的历史记录（数据）中发现其规律以预测未来的趋势；天文学家需要从获取的观测数据（其规模可达数千吉字节）中发现新的遥远天体及其运动规律；医生需要从大量病人电子病历中发现某种疾病的起因、症状等。

这些数据的共同特点是：其一数据量巨大，一般都是GB级乃至TB级；其二都以结构化的形式存储在数据库中，包含了大量潜在、有价值的知识，有的已被发现，有的还未被发现。

如何有效地管理和利用数据库中的海量数据，以及如何发现其中潜在的知识，需要一种新的、更为有效的手段对各种数据源进行整合并挖掘以发现新知识，更好地发挥这些数据的潜能。

因此，数据仓库（Data Warehouse，DW）和数据挖掘（Data Mining，DM）技术应运而生。

数据仓库是一个可更好地支持企业或组织决策，面向主题的、集成的、相对稳定的、随时间不断变化的数据集合；数据挖掘则是使用计算机对大量数据进行快速、有效地分析和处理，从中提取知识，并以一种形式化的、可以理解的方式表达，以便于决策的过程。

目前，数据仓库和数据挖掘技术已经成为计算机领域的研究热点之一，引起了数据库、机器学习、统计分析等领域专家的广泛关注。

<<数据仓库与数据挖掘原理及应用>>

编辑推荐

《数据仓库与数据挖掘原理及应用》由清华大学出版社出版。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>