

<<数据仓库与数据挖掘技术>>

图书基本信息

书名：<<数据仓库与数据挖掘技术>>

13位ISBN编号：9787302281665

10位ISBN编号：7302281661

出版时间：2012-12

出版时间：清华大学出版社

作者：孙水华，赵钊林，刘建华 编著

页数：274

字数：424000

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技术>>

内容概要

《数据仓库与数据挖掘技术》主要介绍数据仓库和数据挖掘技术的基本概念、相关技术和应用案例及方法。

全书共分为9章，主要内容包括：

数据仓库与数据挖掘的概念和体系结构、数据仓库开发模型、etl技术、olap技术、商务智能系统、数据预处理技术、数据挖掘技术、数据仓库开发实例、报表设计等内容。

本书各章节的案例均使用microsoft

sql server

2005进行操作实践讲解。

通过对具体实例的学习和实践，使读者掌握数据仓库和数据挖掘中必要的知识点，达到学以致用目的。

《数据仓库与数据挖掘技术》适合作为高等院校本科学生的教材，也可供企业信息化管理人员、技术人员以及软件开发人员阅读参考。

<<数据仓库与数据挖掘技术>>

书籍目录

第1章 数据仓库与数据挖掘概述

1.1 数据仓库的产生与发展

1.1.1 数据仓库的产生

1.1.2 数据仓库的发展

1.1.3 数据仓库的研究与开发现状

1.1.4 数据仓库的作用

1.2 数据仓库的基本概念

1.2.1 数据仓库的定义与基本特性

1.2.2 数据仓库与数据库的区别

1.2.3 数据仓库数据的组织架构

1.3 数据仓库的体系结构

1.3.1 虚拟的数据仓库体系结构

1.3.2 单独的数据仓库体系结构

1.3.3 单独的数据集市体系结构

1.3.4 分布式数据仓库结构

1.4 数据仓库的相关概念

1.4.1 数据源

1.4.2 数据的存储层

1.4.3 olap服务器

1.4.4 前端工具

1.5 数据挖掘技术概述

1.5.1 数据挖掘技术产生的背景

1.5.2 数据挖掘的基本概念

1.5.3 数据挖掘的对象

1.5.4 数据挖掘功能

1.5.5 数据挖掘与传统分析方法的区别

1.5.6 数据仓库与数据挖掘的关系

1.5.7 数据挖掘的发展趋势

1.6 数据挖掘过程

1.6.1 fayyad过程模型

1.6.2 crisp-dm过程模型

1.6.3 其他数据挖掘过程模型

1.7 常用的数据挖掘技术

1.8 小结

1.9 习题

第2章 数据仓库开发模型

2.1 数据仓库开发模型概述

2.2 数据仓库的概念模型

2.2.1 企业模型的建立

2.2.2 规范的数据模型

2.2.3 常见的概念模型

2.3 数据仓库的逻辑模型

2.3.1 事实表模型设计

2.3.2 维度表模型设计

2.4 数据仓库的物理模型

<<数据仓库与数据挖掘技术>>

- 2.4.1 物理模型的设计要点
- 2.4.2 数据仓库物理模型的存储结构
- 2.4.3 数据仓库物理模型的索引构建
- 2.4.4 数据仓库物理模型的优化问题
- 2.5 数据仓库的元数据模型
 - 2.5.1 元数据的类型
 - 2.5.2 元数据的作用
 - 2.5.3 元数据的收集与维护
 - 2.5.4 元数据的使用
 - 2.5.5 元数据管理模型
- 2.6 数据仓库的粒度和聚集模型
 - 2.6.1 数据仓库粒度模型
 - 2.6.2 数据仓库聚集模型与数据分割
- 2.7 小结
- 2.8 习题
- 第3章 etl技术
 - 3.1 etl相关概念
 - 3.1.1 数据理解
 - 3.1.2 数据抽取
 - 3.1.3 数据清洗
 - 3.1.4 数据转换
 - 3.1.5 数据加载
 - 3.2 etl过程建模
 - 3.2.1 etl系统面临的挑战
 - 3.2.2 etl过程描述
 - 3.2.3 etl概念模型
 - 3.2.4 etl逻辑模型
 - 3.3 etl增量抽取机制
 - 3.4 etl过程数据质量控制
 - 3.4.1 数据质量问题分类
 - 3.4.2 数据质量控制技术
 - 3.5 etl并行处理技术
 - 3.6 小结
 - 3.7 习题
- 第4章 olap技术
 - 4.1 olap概述
 - 4.1.1 olap的定义
 - 4.1.2 数据仓库与数据分析的关系
 - 4.1.3 多维分析的基本概念
 - 4.1.4 olap的多维数据分析
 - 4.1.5 olap与oltp的比较
 - 4.2 多维数据库及其存储
 - 4.2.1 多维数据库
 - 4.2.2 多维数据库的数据存储
 - 4.2.3 多维数据库与数据仓库
 - 4.3 olap的类型
 - 4.3.1 多维olap

<<数据仓库与数据挖掘技术>>

- 4.3.2 关系olap
- 4.3.3 混合型olap
- 4.3.4 molap与rolap的比较
- 4.4 olap的体系结构
- 4.5 olap中的索引技术
 - 4.5.1 b-tree索引
 - 4.5.2 位图索引
 - 4.5.3 位图索引的扩展--标识符索引
 - 4.5.4 索引性能比较
 - 4.5.5 索引的选择
- 4.6 olap的评价标准
 - 4.6.1 olap的衡量标准
 - 4.6.2 olap服务器和工具的评价标准
- 4.7 olap的前端展现
 - 4.7.1 olap工具
 - 4.7.2 olap结果的展现方法
- 4.8 小结
- 4.9 习题
- 第5章 商务智能系统
 - 5.1 商务智能概述
 - 5.1.1 商务智能的概念
 - 5.1.2 商务智能的发展历程
 - 5.1.3 商务智能的商业效益
 - 5.2 商务智能系统架构
 - 5.2.1 商务智能系统的核心技术
 - 5.2.1 商务智能的体系结构
 - 5.3 商务智能系统的功能
 - 5.4 商务智能系统的应用
 - 5.4.1 商务智能系统特点
 - 5.4.2 我国商务智能系统应用现状分析
 - 5.5 小结
 - 5.6 习题
- 第6章 数据预处理技术
 - 6.1 数据预处理概述
 - 6.1.1 数据预处理的必要性
 - 6.1.2 数据预处理的基本方法
 - 6.1.3 数据预处理的研究现状
 - 6.2 数据清理
 - 6.2.1 填充缺失值
 - 6.2.2 光滑噪声数据
 - 6.2.3 数据清理过程
 - 6.3 数据集成
 - 6.4 数据变换
 - 6.5 数据归约
 - 6.5.1 数据立方体聚集
 - 6.5.2 属性子集选择
 - 6.5.3 维度归约

<<数据仓库与数据挖掘技术>>

6.5.4 数值归约

6.5.5 数据离散化与概念分层

6.6 小结

6.7 习题

第7章 数据挖掘技术

7.1 概念描述

7.1.1 概念描述的生成过程

7.1.2 概念分层与数据泛化

7.1.3 概念分层方法

7.1.4 数据泛化方法

7.1.5 泛化的表示

7.1.6 属性相关分析

7.1.7 区别性描述

7.2 关联规则

7.2.1 关联规则相关概念

7.2.2 关联规则挖掘步骤

7.2.3 关联规则分类

7.2.4 关联规则的算法

7.3 数据分类

7.3.1 数据分类的基本步骤与评价准则

7.3.2 决策树

7.3.3 贝叶斯分类

7.3.4 神经网络方法

7.3.5 近邻分类方法

7.4 数据聚类

7.4.1 聚类分析概述

7.4.2 聚类算法的分类及其典型算法

7.4.3 聚类分析中的相似度量方法

7.4.4 聚类分析中的聚类准则函数

7.4.5 k-means聚类算法

7.5 遗传算法

7.5.1 遗传算法的基本术语

7.5.2 遗传算法的执行过程

7.5.3 遗传算法应用举例

7.5.4 遗传算法的基本要素

7.5.5 遗传算法的特点及应用领域

7.6 粗糙集

7.6.1 粗糙集理论的相关概念

7.6.2 粗糙集的应用举例

7.6.3 粗糙集理论研究的对象及特点

7.7 小结

7.8 习题

第8章 数据仓库开发实例

8.1 sql server 2005所提供的数据库功能

8.1.1 sql server 2005 integration services

8.1.2 sql server 2005 analysis services

8.1.3 sql server 2005 dw工具

<<数据仓库与数据挖掘技术>>

8.2 福马特商店销售分析数据仓库系统的分析与设计

8.3 数据仓库的实现

8.3.1 sql server的数据仓库创建

8.3.2 olap的实施

8.3.3 数据仓库中的数据挖掘

8.4 数据仓库的应用与管理

8.4.1 数据仓库的用户

8.4.2 数据仓库应用案例

8.4.3 数据仓库的运行技术管理

8.4.4 数据仓库应用中的法律问题

8.4.5 数据仓库的成本与效益分析

8.5 小结

8.6 习题

第9章 报表设计

9.1 报表概述

9.1.1 报表结构

9.1.2 传递报表

9.1.3 report server功能结构

9.1.4 report services的组成部分

9.2 报表向导制作报表

9.2.1 向导制作报表

9.2.2 报表设计器

9.2.3 部署报表

9.3 编辑制作报表

9.3.1 新建报表项目

9.3.2 新建数据集

9.3.3 报表格式设计

9.3.4 分组

9.3.5 钻取功能

9.3.6 文档结构图

9.4 矩阵式报表

9.4.1 数据集建立

9.4.2 矩阵布局

9.4.3 矩形布局

9.4.4 折叠结构

9.5 统计图表

9.5.1 图表元素

9.5.2 柱形图

9.5.3 折线图

9.5.4 饼图

9.5.5 圆环图

9.6 主体的多列

9.7 小结

9.8 实验

参考文献

<<数据仓库与数据挖掘技术>>

章节摘录

版权页：插图：4.复杂类型数据库 复杂类型的数据库是指非单纯文本的数据库或能够表示动态的序列数据的数据库，主要有如下几类。

(1) 空间数据库。

主要指存储空间信息的数据库，其中数据可能以光栅格式提供，也可能用矢量图形数据表示。

例如，地理信息数据库、卫星图像数据库、城市地下管道、下水道及各类地下建筑分布数据库等。

对空间数据库的挖掘可以为城市规划、生态规划、道路修建提供决策支持。

(2) 时序数据库。

主要用于存放与时间相关的数据，它可用来反映随时间变化的即时数据或不同时间发生的不同事件。

例如，连续的存放即时的股票交易信息、卫星轨道信息等。

对时序数据的挖掘可以发现事件的发展趋势、事物的演变过程和隐藏特征，这些信息将对事件的计划、决策和预警是非常有用的。

(3) 多媒体数据库。

用于存放图像、声音和视频信息的数据库。

由于多媒体技术的发展，以及相关研究（如可视化信息检索、虚拟现实技术）的成就，多媒体数据库也逐渐普及，并应用于许多重要研究领域。

目前，多媒体数据的挖掘主要放在对图像数据的检索与匹配上，随着研究的深入将会拓展到对声音、视频信息的挖掘处理。

1.5.4 数据挖掘功能 如果以数据挖掘任务为标准来划分，数据挖掘功能有如下的几类：概念描述、关联分析、分类和预测、聚类分析、孤立点分析、演变分析等。

1. 概念描述 概念描述（concept description）指的是对某种对象的内涵特征进行概括，概括就是概念描述的本质。

一般情况下，对一个数据集，其包含大量数据，做一个总体状态的概述就是一个概念。

例如，将某一公司所有卖出的汽车的基本情况进行概述与总结，就会了解到所有汽车基本情况的一个整体概念。

传统的也是最简单的数据总结方法是计算出数据库的各个字段上的求各值、平均值、方差值等统计值，或者用直方图、饼状图等图形方式表示。

而数据挖掘中的概念描述主要关心从数据泛化的角度来讨论数据总结。

数据泛化是一种把数据库中的有关数据从低层次抽象到高层次上的过程。

由于数据库上的数据或对象所包含的信息总是最原始、基本的信息（这是为了不遗漏任何可能有用的数据信息），人们有时希望能从较高层次的视图上处理或浏览数据，因此，需要对数据进行不同层次上的泛化以适应各种查询要求。

描述可以分为特征性的描述与区别性的描述。

前者描述某类对象的共同特征，生成一个类的特征性描述，该描述只涉及该类对象中所有个体的共性。

其输出可以采用多种形式，包括饼图、柱状图、曲线、多维数据立方体、含交叉表的多维表，且描述结果也可以用概化关系或规则形式表示。

后者描述异类对象之间的区别，将目标类对象的一般特性与一个或多个对比类对象的一般特性比较，而这种比较必须具备可比性的两个或多个类之间进行的。

数据区分的输出类似于数据特征化，但它应该包括比较度量，帮助区分目标类和对比类。

<<数据仓库与数据挖掘技术>>

编辑推荐

《高等院校信息管理与信息系统专业系列教材:数据仓库与数据挖掘技术》适合作为高等院校本科学生的教材,也可供企业信息化管理人员、技术人员以及软件开发人员阅读参考。

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>