

<<数据仓库与数据挖掘>>

图书基本信息

书名：<<数据仓库与数据挖掘>>

13位ISBN编号：9787308038317

10位ISBN编号：7308038319

出版时间：2004-8

出版时间：浙江大学出版社

作者：周根贵 编

页数：327

字数：374000

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>>

### 内容概要

本书侧重于信息管理中信息组织与处理的技术，全面而系统地介绍了数据仓库与数据挖掘的基本概念、基本方法和基本技术，以及数据仓库与数据挖掘的应用领域与最新进展。

全书共分10章，包括概论，数据仓库的技术与开发，数据仓库的管理，联机分析处理，SQL Server数据仓库的应用与开发，数据挖掘与知识发现，统计类数据挖掘技术，知识类数据挖掘技术，21世纪的数据挖掘技术，数据仓库与数据挖掘的综合应用等内容。

为了便于学习，每一章都有提要和小结，并配有一定数量的习题，以帮助读者对基本内容的理解和掌握。

本书深入浅出，阐述清晰、理论与实际并重，可作为高等院校信息管理与信息系统等专业本科生、研究生的教材。

## <<数据仓库与数据挖掘>>

### 书籍目录

#### 第1章 概论

- 1.1 决策支持技术与数据库的发展
  - 1.1.1 决策支持技术的发展
  - 1.1.2 数据库技术的发展
- 1.2 数据仓库概述
  - 1.2.1 数据仓库概念的提出
  - 1.2.2 数据仓库的定义
  - 1.2.3 数据仓库的特征
  - 1.2.4 数据仓库的应用和发展
- 1.3 数据挖掘概述
  - 1.3.1 数据挖掘的定义
  - 1.3.2 数据挖掘与数据仓库的关系
  - 1.3.3 数据挖掘的应用和发展

本章小结

习题

#### 第2章 数据仓库的技术与开发

- 2.1 数据仓库的体系结构
  - 2.1.1 用户眼中的数据仓库结构
  - 2.1.2 数据仓库系统的体系结构
  - 2.1.3 数据集市
- 2.2 元数据
  - 2.2.1 元数据的定义
  - 2.2.2 元数据的主要作用
  - 2.2.3 元数据分类
- 2.3 数据仓库的数据模型
  - 2.3.1 概念模型
  - 2.3.2 逻辑模型
  - 2.3.3 物理模型
- 2.4 粒度和分割
  - 2.4.1 粒度的确定
  - 2.4.2 粒度划分实例
  - 2.4.3 数据分割
- 2.5 数据仓库和开发流程
- 2.6 总线型结构的数据仓库
  - 2.6.1 统一的维
  - 2.6.2 统一的事实
  - 2.6.3 数据仓库总线

本章小结

习题

#### 第3章 数据仓库管理技术

- 3.1 数据仓库管理的基本内容
- 3.2 休眠数据管理
  - 3.2.1 休眠数据的定义与理解
  - 3.2.2 休眠数据的处理
- 3.3 元数据的管理

## <<数据仓库与数据挖掘>>

3.3.1 传统的元数据管理方法

3.3.2 企业级中心知识库的管理方法

3.4 数据清理

3.4.1 脏数据的来源和清理

3.4.2 过期数据的清理

本章小结

习题

第4章 联机分析处理

4.1 概述

4.1.1 OLAP的定义

4.1.2 OLAP的基本概念

4.1.3 OLAP的基本分析操作

4.1.4 OLAP和OLTP的比较

4.2 多维OLAP和关系OLAP

4.2.1 数据存储

4.2.2 MOLAP和ROLAP的特征

4.2.3 星型模式

4.3 OLAP的新发展——OLAM

.....

第5章 SQL Server数据仓库的应用与开发

第6章 数据挖掘与知识发现

第7章 统计类数据挖掘技术

第8章 知识类数据挖掘技术

第9章 非结构化数据挖掘技术

第10章 数据仓库与数据挖掘的综合应用

第11章 数据挖掘的语言与工具

第12章 数据仓库与数据挖掘的综合应用

参考文献

## &lt;&lt;数据仓库与数据挖掘&gt;&gt;

## 章节摘录

版权页：插图：7.5.2统计分析工具应用在统计类数据挖掘技术的应用中，需要商业分析人员给予必要的辅助指导，因此统计类数据挖掘技术的应用的成败往往取决于商业分析人员的专业水平。

7.5.2.1统计分析类工具应用中的问题在现实世界中的数据仓库极易受噪声、空缺数据和不一致性数据的影响。

因为数据仓库太大，存在不完整的、含噪声的和不一致的数据是大型的、现实数据库或数据仓库的共同特点。

不完整数据的出现可能有多种原因。

有些感兴趣的属性，并非总是可用的。

数据含噪声（具有不正确的属性值）可能有多种原因：如收集数据的设备可能出故障，人为的或计算机的错误可能在数据输入时出现，数据传输中的错误也可能出现。

这些或许是由于技术的限制（如用于数据传输同步的缓冲区大小的限制），不正确的数据可能由命名或所用的数据代码不一致而造成的。

重复元组也会造成数据噪声，对此也需要清理。

数据清理例程通过填写空缺的值、平滑噪声数据，识别、删除孤立点，并且解决不一致来“清理”数据。

脏数据能使挖掘过程陷入混乱，导致不可靠的输出。

尽管大部分挖掘例程都有一些过程，处理不完整或噪声数据，但它们并非总是强壮的。

相反，它们更致力于避免数据过分适合所建的模型。

这样，需要一个预处理步骤清理数据中的各种问题。

（1）空缺值处理如果一个数据库中许多元组的一些属性值没有记录值，可以采用以下的方法为该属性添上空缺的值。

忽略元组：如果挖掘任务涉及分类或描述，但是缺少类标号时可以忽略元组。

该方法应用时，要求元组有多个属性缺少值，否则该方法不是很有效。

当每个属性缺少值的百分比变化很大时，它的性能就非常差。

<<数据仓库与数据挖掘>>

编辑推荐

《数据仓库与数据挖掘(第2版)》为普通高等教育“十一五”国家级规划教材,信息管理与信息系统专业核心课程精品教材系列之一。

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>