

<<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：9787302197102

10位ISBN编号：7302197105

出版时间：2009-5

出版时间：清华大学出版社

作者：陈志泊

页数：249

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>

## &lt;&lt;数据仓库与数据挖掘&gt;&gt;

## 前言

随着计算机和信息时代的迅猛发展，人类收集、存储和访问数据的能力大大增强，快速增长的海量数据集已经远远超出了人类的理解能力，传统的数据分析工具也显得力不从心。

如何不被这些海量数据淹没，而是有效地组织这些数据，并且从中找出有价值的知识，从而帮助人类制定正确的决策？

针对这一问题，数据仓库和数据挖掘技术应运而生，并且显示出强大的生命力。

要将海量数据转换成为有用的信息和知识，首先要有效地收集和组织数据。

数据仓库是良好的数据收集和工具，它的任务是搜集来自各个业务系统的有用数据，存放在一个集成的储存区内。

在数据仓库的丰富完整的数据基础上，数据挖掘技术可以从中挖掘出有价值的知识，从而帮助决策者做出正确决策。

本书主要介绍数据仓库和数据挖掘技术的基本原理和应用方法，全书共分为12章，主要内容包括数据仓库的概念和体系结构、数据仓库的数据存储和处理、数据仓库系统的设计与开发、关联规则、数据分类、数据聚类、贝叶斯网络、粗糙集、神经网络、遗传算法、统计分析、文本和Web挖掘。

其中，前3章主要介绍数据仓库的基本原理和数据仓库系统的组建方法，后面的章节介绍当前流行的数据挖掘算法的主要思想和理论基础，并且给出丰富的应用实例。

本书紧跟数据仓库和数据挖掘技术的发展和人才培养的目标，有以下几个特点：（1）可读性强，文字叙述深入浅出，易读易用，即使是初学者，阅读起来也比较容易。

（2）概念清晰，条理清楚，内容取舍合理。

（3）本书强调基础，重视实例。

各章节都以经典算法为主，介绍其主要思想和基本原理，并且给出恰当和丰富的实例。

（4）书中实例和课后习题实用、丰富，通过练习，读者可以对各个知识点从不同角度得到训练，巩固和掌握所学知识。

（5）教学资源丰富，本书提供多媒体教学课件和习题参考答案，方便教学。

对于上述资源，读者可到清华大学出版社的网站<http://www.tup.com.cn/>下载。

（6）对于数据仓库的组建方法和多数章节中的数据挖掘算法，本书都使用Microsoft SQL Server 2005进行了操作实现，这种做法与市场主流开发工具和技术同步，有利于读者走向社会。

本书各章节之间衔接自然，同时各章节又有一定的独立性，读者可按教材的自然顺序学习，也可以根据实际情况挑选需要的章节学习。

本书可以作为高等学校计算机及相关专业本科、研究生学习数据仓库和数据挖掘的教材，也可供相关领域的广大科技工作人员和高校师生参考。

本书由陈志泊担任主编，第1~第3章由聂耿青编写，第5、第6和第11章由韩慧编写，第4和第10章由孙俏编写，第7~第9和第12章由王建新编写。

由于时间仓促，加之编者水平有限，书中不足之处敬请批评指正。

## <<数据仓库与数据挖掘>>

### 内容概要

本书主要介绍数据仓库和数据挖掘技术的基本原理和应用方法，全书共分为12章，主要内容包括数据仓库的概念和体系结构、数据仓库的数据存储和处理、数据仓库系统的设计与开发、关联规则、数据分类、数据聚类、贝叶斯网络、粗糙集、神经网络、遗传算法、统计分析、文本和Web挖掘。

本书既重视理论知识的讲解，又强调应用技能的培养。

每章首先介绍算法的主要思想和理论基础，之后利用算法去解决实例中给出的任务，而且对于数据仓库的组建方法和多数章节中的数据挖掘算法，本书都使用Microsoft SQL Server 2005进行了操作实现。本书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。

本书每章均配有习题，习题形式为选择题、简答题和操作题，可以帮助读者进一步巩固和掌握所学知识。

此外，本书提供多媒体教学课件和习题参考答案，读者可到清华大学出版社网站（<http://www.tup.com.cn/>）下载。

本书可以作为高等学校计算机及相关专业本科、研究生的数据仓库和数据挖掘教材，也可供相关领域的广大科技工作人员和高校师生参考。

书籍目录

第1章 数据仓库的概念与体系结构 1.1 数据仓库的概念、特点与组成 1.2 数据挖掘的概念与方法 1.3 数据仓库的技术、方法与产品 1.4 数据仓库系统的体系结构 1.5 数据仓库的产生、发展与未来 1.6 小结 1.7 习题第2章 数据仓库的数据存储与处理 2.1 数据仓库的数据结构 2.2 数据仓库的数据特征 2.3 数据仓库的数据ETL过程 2.4 多维数据模型 2.5 小结 2.6 习题第3章 数据仓库系统的设计与开发 3.1 数据仓库系统的设计与开发概述 3.2 基于SQL Server 2005的数据仓库数据库设计 3.3 使用SQL Server 2005建立多维数据模型 3.4 小结 3.5 习题第4章 关联规则 4.1 概述 4.2 引例 4.3 经典算法 4.4 相关研究与应用 4.5 小结 4.6 习题第5章 数据分类 5.1 引例 5.2 分类问题概述 5.3 决策树 5.4 支持向量机 5.5 近邻分类方法 5.6 小结 5.7 习题第6章 数据聚类 6.1 引例 6.2 聚类分析概述 6.3 聚类分析中相似度的计算方法 6.4 kmeans 聚类算法 6.5 层次聚类方法 6.6 小结 6.7 习题第7章 贝叶斯网络第8章 粗糙集第9章 神经网络第10章 遗传算法第11章 统计分析第12章 文本和Web挖掘参考文献

## 章节摘录

插图：(3) 分区维。

分区维以同一结构生成两个或多个维时，这些维结构相同，只是数值不同。

例如，对于时间维，每一年都有相同的季度：相同的月和相同的天（除了闰年以外）。

假定把度量事实表分割为2007年的数据和2008年的数据，那么在OLAP分析中将频繁使用时间分区维来分割数据仓库中的数据。

其中一个时间维是针对2007年的数据，而另一个时间维针对2008年的数据。

(4) 分类维。

分类维是通过对一个维的属性值分组而创建的。

例如，客户表中有家庭收入属性，如果希望查看客户根据收入的购物方式，就可以生成一个含有家庭收入的分类维。

(5) 退化维。

当维表中的主键在事实表中没有与外键关联时，这样的维称为退化维。

退化维与事实表并无关系，但有时在查询限制条件（如订单号码、出货单编号等）中需要用到退化维。

以销售分析为例，通常是把出货日期作为事实时间，而把订单日期或需求日期等作为查询条件，这里，订单日期或需求日期就是退化维。

(6) 一致维。

当有好几个数据集市要合并成一个企业级数据仓库时，可以使用一致维来集成数据集市，以便确保数据仓库可以使用每个数据集市的事实。

## <<数据仓库与数据挖掘>>

### 编辑推荐

《数据仓库与数据挖掘》为普通高等教育“十一五”国家级规划教材之一。

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>