

<<数据挖掘>>

图书基本信息

书名：<<数据挖掘>>

13位ISBN编号：9787312022449

10位ISBN编号：7312022448

出版时间：2008-11

出版时间：中国科学技术大学出版社

作者：朱明

页数：491

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>

## 前言

2008年是中国科学技术大学建校五十周年。

为了反映五十年来办学理念和特色，集中展示教材建设的成果，学校决定组织编写出版代表中国科学技术大学教学水平的精品教材系列。

在各方的共同努力下，共组织选题281种，经过多轮、严格的评审，最后确定50种入选精品教材系列。

1958年学校成立之时，教员大部分都来自中国科学院的各个研究所。

作为各个研究所的科研人员，他们到学校后保持了教学的同时又作研究的传统。

同时，根据“全院办校，所系结合”的原则，科学院各个研究所在科研第一线工作的杰出科学家也参与学校的教学，为本科生授课，将最新的科研成果融入到教学中。

五十年来，外界环境和内在条件都发生了很大变化，但学校以教学为主、教学与科研相结合的方针没有变。

正因为坚持了科学与技术相结合、理论与实践相结合、教学与科研相结合的方针，并形成了优良的传统，才培养出了一批又一批高质量的人才。

学校非常重视基础课和专业基础课教学的传统，也是她特别成功的原因之一。

当今社会，科技发展突飞猛进、科技成果日新月异，没有扎实的基础知识，很难在科学技术研究中作出重大贡献。

建校之初，华罗庚、吴有训、严济慈等老一辈科学家、教育家就身体力行，亲自为本科生讲授基础课。

他们以渊博的学识、精湛的讲课艺术、高尚的师德，带出一批又一批杰出的年轻教员，培养了一届又一届优秀学生。

这次入选校庆精品教材的绝大部分是本科生基础课或专业基础课的教材，其作者大多直接或间接受到过这些老一辈科学家、教育家的教诲和影响，因此在教材中也贯穿着这些先辈的教育教学理念与科学探索精神。

## <<数据挖掘>>

### 内容概要

数据挖掘技术，又称为数据库知识发现，是20世纪90年代在信息技术领域开始迅速发展起来的计算机技术。

作者结合自己近20年从事人工智能、机器学习、数据挖掘等方面的科研工作积累与教学经验，编著此书。

本书较全面系统地介绍了数据挖掘中常用和常见的数据挖掘方法，以及文本与视频数据挖掘方法。

本书的主要内容包括：数据挖掘基本知识、数据挖掘预处理方法、决策树分类及其他分类方法、关联知识挖掘方法、各种聚类分析方法，以及文本挖掘所涉及表示、分类和聚类等方法，还包括视频挖掘所涉及的视频镜头检测、字幕提取、视频摘要和视频检索等主要分析方法。

本书作为学习、掌握和应用数据挖掘方法和技术的综合指导书，是从事数据挖掘研究与应用人员，以及希望了解数据挖掘主要方法和技术的IT技术人员的良师益友；同时也是一本可用于大学高年级或研究生相关课程的教材和参考文献。

## 书籍目录

总序前言第1章 数据挖掘导论 1.1 数据挖掘的发展背景 1.2 数据挖掘定义 1.3 数据挖掘过程 1.4 数据挖掘功能 1.5 数据挖掘应用 1.6 数据挖掘发展 1.7 本章小结第2章 数据预处理 2.1 数据描述 2.1.1 数据集类型 2.1.2 数据质量 2.2 数据清理 2.2.1 缺失值处理 2.2.2 噪声数据处理 2.2.3 数据清理过程 2.3 数据集成和变换 2.3.1 数据集成 2.3.2 数据变换 2.3.3 维度归约 2.4 数据归约 2.4.1 数据立方体聚集 2.4.2 属性子集选择 2.5 本章小结第3章 分类挖掘：决策树 3.1 决策树方法 3.2 决策树深入 3.2.1 信息熵基础 3.2.2 C4.5方法 3.2.3 CART方法 3.2.4 SLIQ方法 3.2.5 SPRINT方法 3.2.6 其他决策树方法 3.3 决策树的简化 3.4 决策树的改进 3.4.1 属性选择 3.4.2 连续属性离散化 3.5 决策树的讨论 3.5.1 决策树优化问题 3.5.2 决策树优化方法 3.6 分类模型的评估 3.7 本章小结第4章 分类挖掘 4.1 贝叶斯方法 4.1.1 贝叶斯方法概述 4.1.2 朴素贝叶斯分类 4.2 k-近邻方法 4.3 人工神经网络方法 4.4 遗传进化方法 4.5 支持向量机方法 4.5.1 SVM分类方法 4.6 粗糙集方法 4.7 集成学习方法 .....第5章 关联挖掘第6章 聚类挖掘第7章 异类挖掘第8章 文本挖掘第9章 视频挖掘第10章 视频分析

## &lt;&lt;数据挖掘&gt;&gt;

## 章节摘录

第1章 数据挖掘导论数据挖掘是20世纪80年代末开始逐步发展起来的一个新的研究领域，它是多个学科和技术相结合的产物。

本章将简要介绍数据挖掘的发展背景、概念定义、主要方法及应用案例等内容。

1.1 数据挖掘的发展背景随着数据库技术的迅速发展以及数据库管理系统的广泛应用，人们利用信息技术生产和搜集数据的能力大幅度提高，无数个数据库被用于商业管理、政府办公、科学研究和工程开发等领域，超级市场中的交易数据、加油站里的汽油销售数据、旅行社的旅游信息等等，均构成了数据库系统的信息来源。

近年来，数据库所管理的数据量急剧增大，人们积累的数据越来越多。

例如：美国NASA的地球观测系统（EoS）每小时向地面发回约50 GB的图像数据；美国沃尔玛零售系统每天会产生约2亿条交易数据。

人们希望能够对其进行更高层次的分析，以便更好地利用这些数据。

激增的数据背后隐藏着许多重要的信息，目前的数据库系统可以高效地实现数据的录入、查询、统计等功能，但无法发现数据中存在的关系和规则，无法根据现有的数据预测未来的发展趋势。

缺乏挖掘数据背后隐藏的知识的手段，导致了“数据富有但知识贫乏”的现象。

于是，一个新的挑战被提了出来：在这被称之为信息爆炸的时代，信息过量几乎成为人人需要面对的问题。

如何才能不被信息的汪洋大海所淹没，从中及时发现有用的知识，提高信息利用率呢？

要想使数据真正成为一个企业的资源，只有充分利用它为企业自身的业务决策和战略发展服务才行，否则大量的数据可能成为包袱，甚至成为垃圾。

## <<数据挖掘>>

### 编辑推荐

《数据挖掘(第2版)》作为学习、掌握和应用数据挖掘方法和技术的综合指导书，是从事数据挖掘研究与应用人员，以及希望了解数据挖掘主要方法和技术的IT技术人员的良师益友；同时也是一本可用于大学高年级或研究生相关课程的教材和参考文献。

#### 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>