第一图书网, tushu007.com

<<数据挖掘概念与技术>>

图书基本信息

书名：<<数据挖掘概念与技术>>

13位ISBN编号：9787111205388

10位ISBN编号：7111205383

出版时间：2007-3

出版时间：机械工业

作者：(加)Jiawei Han;Micheline Kamber

页数：488

译者：范明,孟小峰

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

内容概要

本书全面地讲述数据挖掘领域的重要知识和技术创新。
在第1版内容相当全面的基础上，第2版展示了该领域的最新研究成果，例如挖掘流、时序和序列数据以及挖掘时间空间、多媒体、文本和Web数据。
本书可作为数据挖掘和知识发现领域的教师、研究人员和开发人员的一本必读书。

本书第1版曾是受读者欢迎的数据挖掘专著，是一本可读性极佳的教材。
第2版充实了数据挖掘领域研究新进展的题材，增加了讲述最新的数据挖掘方法的若干章节。
本书适合作为高等院校计算机及相关专业高年级本科生的选修课教材，特别适合作为研究生的专业课教材。

作者简介

Jiawei
Han，伊利诺伊大学厄巴纳－尚佩恩分校计算机科学系教授。
由于在数据挖掘和数据库系统领域卓有成效的研究工作，他曾多次获得各种荣誉和奖励，其中包括2004年ACM
SIGKDD颁发的创新奖。
同时，他还是ACM《Transactions on Knowledge Discovery from
Data》的主编，以及IEE

书籍目录

出版者的话
专家指导委员会
中文版序
译者序
序
前言
第1章 引言
1.1 什么激发了数据挖掘，为什么它是重要的
1.2 什么是数据挖掘
1.3 对何种数据进行数据挖掘
1.3.1 关系数据库
1.3.2 数据仓库
1.3.3 事务数据库
1.3.4 高级数据和信息系统与高级应用
1.4 数据挖掘功能—可以挖掘什么类型的模式
1.4.1 概念/类描述：特征化和区分
1.4.2 挖掘频繁模式、关联和相关
1.4.3 分类和预测
1.4.4 聚类分析
1.4.5 离群点分析
1.4.6 演变分析
1.5 所有模式都是有趣的吗
1.6 数据挖掘系统的分类
1.7 数据挖掘任务原语
1.8 数据挖掘系统与数据库系统或数据仓库系统的集成
1.9 数据挖掘的主要问题
1.10 小结
习题
文献注释
第2章 数据预处理
2.1 为什么要预处理数据
2.2 描述性数据汇总
2.2.1 度量数据的中心趋势
2.2.2 度量数据的离散程度
2.2.3 基本描述数据汇总的图形显示
2.3 数据清理
2.3.1 缺失值
2.3.2 噪声数据
2.3.3 数据清理作为一个过程
2.4 数据集成和变换
2.4.1 数据集成
2.4.2 数据变换
2.5 数据归约
2.5.1 数据立方体聚集
2.5.2 属性子集选择

2.5.3 维度归约
2.5.4 数值归约
2.6 数据离散化和概念分层产生
2.6.1 数值数据的离散化和概念分层产生
2.6.2 分类数据的概念分层产生
2.7 小结
习题
文献注释
第3章 数据仓库与OLAP技术概述
3.1 什么是数据仓库
3.1.1 操作数据库系统与数据仓库的区别
3.1.2 为什么需要分离的数据仓库
3.2 多维数据模型
3.2.1 由表和电子数据表到数据立方体
3.2.2 星形、雪花形和事实星座形模式：多维数据库模式
3.2.3 定义星形、雪花形和事实星座形模式的例子
3.2.4 度量的分类和计算
3.2.5 概念分层
3.2.6 多维数据模型中的OLAP操作
3.2.7 查询多维数据库的星形网查询模型
3.3 数据仓库的系统结构
3.3.1 数据仓库的设计和构造步骤
3.3.2 三层数据仓库的系统结构
3.3.3 数据仓库后端工具和实用程序
3.3.4 元数据储存库
3.3.5 OLAP服务器类型：ROLAP、MOLAP与HOLAP
3.4 数据仓库实现
3.4.1 数据立方体的有效计算
3.4.2 索引OLAP数据
3.4.3 OLAP查询的有效处理
3.5 从数据仓库到数据挖掘
3.5.1 数据仓库的使用
3.5.2 由联机分析处理到联机分析挖掘
3.6 小结
习题
文献注释
第4章 数据立方体计算与数据泛化
4.1 数据立方体计算的有效方法
4.1.1 不同类型立方体物化的路线图
4.1.2 完全立方体计算的多路数组聚集
4.1.3 BUC：从顶点方体向下计算冰山立方体
4.1.4 Star-Cubing：使用动态星形树结构计算冰山立方体
4.1.5 为快速高维OLAP预计算壳片段
4.1.6 计算具有复杂冰山条件的立方体
4.2 数据立方体和OLAP技术的进一步发展
4.2.1 数据立方体的发现驱动的探查
4.2.2 在多粒度的复杂聚集：多特征立方体

4.2.3 数据立方体中被约束的梯度分析
4.3 面向属性的归纳—另一种数据泛化和概念描述方法
4.3.1 数据特征化的面向属性的归纳
4.3.2 面向属性归纳的有效实现
4.3.3 导出泛化的表示
4.3.4 挖掘类比较：区分不同的类
4.3.5 类描述：特征化和比较的表示
4.4 小结
习题
文献注释
第5章 挖掘频繁模式、关联和相关
5.1 基本概念和路线图
5.1.1 购物篮分析：引发性例子
5.1.2 频繁项集、闭项集和关联规则
5.1.3 频繁模式挖掘：路线图
5.2 有效的和可伸缩的频繁项集挖掘方法
5.2.1 Apriori算法：使用候选产生发现频繁项集
5.2.2 由频繁项集产生关联规则
5.2.3 提高Apriori算法的效率
5.2.4 不候选产生挖掘频繁项集
5.2.5 使用垂直数据格式挖掘频繁项集
5.2.6 挖掘闭频繁项集
5.3 挖掘各种类型的关联规则
5.3.1 挖掘多层关联规则
5.3.2 从关系数据库和数据仓库挖掘多维关联规则
5.4 由关联挖掘到相关分析
5.4.1 强关联规则不一定有趣：一个例子
5.4.2 从关联分析到相关分析
5.5 基于约束的关联挖掘
5.5.1 关联规则的元规则制导挖掘
5.5.2 约束推进：规则约束制导的挖掘
5.6 小结
习题
文献注释
第6章 分类和预测
6.1 什么是分类，什么是预测
6.2 关于分类和预测的问题
6.2.1 为分类和预测准备数据
6.2.2 比较分类和预测方法
6.3 用决策树归纳分类
6.3.1 决策树归纳
6.3.2 属性选择度量
6.3.3 树剪枝
6.3.4 可伸缩性与决策树归纳
6.4 贝叶斯分类
6.4.1 贝叶斯定理
6.4.2 朴素贝叶斯分类

6.4.3 贝叶斯信念网络
6.4.4 训练贝叶斯信念网络
6.5 基于规则的分类
6.5.1 使用IF-THEN规则分类
6.5.2 从决策树提取规则
6.5.3 使用顺序覆盖算法的规则归纳
6.6 用后向传播分类
6.6.1 多层前馈神经网络
6.6.2 定义网络拓扑
6.6.3 后向传播
6.6.4 黑盒内部：后向传播和可解释性
6.7 支持向量机
6.7.1 数据线性可分的情况
6.7.2 数据非线性可分的情况
6.8 关联分类：基于关联规则分析的分类
6.9 惰性学习法（或从近邻学习）
6.9.1 k最近邻分类法
6.9.2 基于案例的推理
6.10 其他分类方法
6.10.1 遗传算法
6.10.2 粗糙集方法
6.10.3 模糊集方法
6.11 预测
6.11.1 线性回归
6.11.2 非线性回归
6.11.3 其他基于回归的方法
6.12 准确率和误差的度量
6.12.1 分类器准确率度量
6.12.2 预测器误差度量
6.13 评估分类器或预测器的准确率
6.13.1 保持方法和随机子抽样
6.13.2 交叉确认
6.13.3 自助法
6.14 系综方法—提高准确率
6.14.1 装袋
6.14.2 提升
6.15 模型选择
6.15.1 估计置信区间
6.15.2 ROC 曲线
6.16 小结
习题
文献注释
第7章 聚类分析
7.1 什么是聚类分析
7.2 聚类分析中的数据类型
7.2.1 区间标度变量
7.2.2 二元变量

7.2.3 分类、序数和比例标度变量
7.2.4 混合类型的变量
7.2.5 向量对象
7.3 主要聚类方法的分类
7.4 划分方法
7.4.1 典型的划分方法：k均值和k中心点
7.4.2 大型数据库的划分方法：从k中心点到CLARANS
7.5 层次方法
7.5.1 凝聚和分裂层次聚类
7.5.2 BIRCH：利用层次方法的平衡迭代归约和聚类
7.5.3 ROCK：分类属性的层次聚类算法
7.5.4 Chameleon：利用动态建模的层次聚类算法
7.6 基于密度的方法
7.6.1 DBSCAN：一种基于高密度连通区域的基于密度的聚类方法
7.6.2 OPTICS：通过点排序识别聚类结构
7.6.3 DENCLUE：基于密度分布函数的聚类
7.7 基于网格的方法
7.7.1 STING：统计信息网格
7.7.2 WaveCluster：利用小波变换聚类
7.8 基于模型的聚类方法
7.8.1 期望最大化方法
7.8.2 概念聚类
7.8.3 神经网络方法
7.9 聚类高维数据
7.9.1 CLIQUE：维增长子空间聚类方法
7.9.2 PROCLUS：维归约子空间聚类方法
7.9.3 基于频繁模式的聚类方法
7.10 基于约束的聚类分析
7.10.1 含有障碍物的对象聚类
7.10.2 用户约束的聚类分析
7.10.3 半监督聚类分析
7.11 离群点分析
7.11.1 基于统计分布的离群点检测
7.11.2 基于距离的离群点检测
7.11.3 基于密度的局部离群点检测
7.11.4 基于偏差的离群点检测
7.12 小结
习题
文献注释
第8章 挖掘流、时间序列和序列数据
8.1 挖掘数据流
8.1.1 流数据处理方法和流数据系统
8.1.2 流OLAP和流数据立方体
8.1.3 数据流中的频繁模式挖掘
8.1.4 动态数据流的分类
8.1.5 聚类演变数据流
8.2 时间序列数据挖掘

8.2.1 趋势分析
8.2.2 时间序列分析中的相似性搜索
8.3 挖掘事务数据库中的序列模式
8.3.1 序列模式挖掘：概念和原语
8.3.2 挖掘序列模式的可伸缩方法
8.3.3 基于约束的序列模式挖掘
8.3.4 时间相关序列数据的周期性分析
8.4 挖掘生物学数据中的序列模式
8.4.1 生物学序列比对
8.4.2 生物学序列分析的隐马尔可夫模型
8.5 小结
习题
文献注释
第9章 图挖掘、社会网络分析和多关系数据挖掘
9.1 图挖掘
9.1.1 挖掘频繁子图的方法
9.1.2 挖掘变体和约束子结构的模式
9.1.3 应用：图索引、相似性搜索、分类和聚类
9.2 社会网络分析
9.2.1 什么是社会网络
9.2.2 社会网络的特征
9.2.3 链接挖掘：任务和挑战
9.2.4 挖掘社会网络
9.3 多关系数据挖掘
9.3.1 什么是多关系数据挖掘
9.3.2 多关系分类的ILP方法
9.3.3 元组ID传播
9.3.4 利用元组ID传播进行多关系分类
9.3.5 用户指导的多关系聚类
9.4 小结
习题
文献注释
第10章 挖掘对象、空间、多媒体、文本和Web数据
10.1 复杂数据对象的多维分析和描述性挖掘
10.1.1 结构化数据的泛化
10.1.2 空间和多媒体数据泛化中的聚集和近似
10.1.3 对象标识和类/子类层次的泛化
10.1.4 类复合层次泛化
10.1.5 对象立方体的构造与挖掘
10.1.6 用分治法对规划数据库进行基于泛化的挖掘
10.2 空间数据挖掘
10.2.1 空间数据立方体构造和空间OLAP
10.2.2 挖掘空间关联和并置模式
10.2.3 空间聚类方法
10.2.4 空间分类和空间趋势分析
10.2.5 挖掘光栅数据库
10.3 多媒体数据挖掘

10.3.1 多媒体数据的相似性搜索
10.3.2 多媒体数据的多维分析
10.3.3 多媒体数据的分类和预测分析
10.3.4 挖掘多媒体数据中的关联
10.3.5 音频和视频数据挖掘
10.4 文本挖掘
10.4.1 文本数据分析和信息检索
10.4.2 文本的维度归约
10.4.3 文本挖掘方法
10.5 挖掘万维网
10.5.1 挖掘Web页面布局结构
10.5.2 挖掘Web链接结构识别权威Web页面
10.5.3 挖掘Web上的多媒体数据
10.5.4 Web文档的自动分类
10.5.5 Web使用挖掘
10.6 小结
习题
文献注释
第11章 数据挖掘的应用和发展趋势
11.1 数据挖掘的应用
11.1.1 金融数据分析的数据挖掘
11.1.2 零售业的数据挖掘
11.1.3 电信业的数据挖掘
11.1.4 生物学数据分析的数据挖掘
11.1.5 其他科技应用的数据挖掘
11.1.6 入侵检测的数据挖掘
11.2 数据挖掘系统产品和研究原型
11.2.1 如何选择数据挖掘系统
11.2.2 商用数据挖掘系统的实例
11.3 数据挖掘的其他主题
11.3.1 数据挖掘的理论基础
11.3.2 统计学数据挖掘
11.3.3 可视数据和音频数据挖掘
11.3.4 数据挖掘和协同过滤
11.4 数据挖掘的社会影响
11.4.1 无处不在和无形的数据挖掘
11.4.2 数据挖掘、隐私和数据安全
11.5 数据挖掘的发展趋势
11.6 小结
习题
文献注释
附录 微软用于数据挖掘的OLE DB简介
A.1 模型创建
A.2 模型训练
A.3 模型预测和浏览
索引

第一图书网, tushu007.com

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