

<<数据挖掘导论>>

图书基本信息

书名：<<数据挖掘导论>>

13位ISBN编号：9787115146984

10位ISBN编号：7115146985

出版时间：2006-5

出版时间：人民邮电出版社

作者：Pang-Ning Tan, Michael Steinbach

页数：422

译者：范明 范宏建

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>

<<数据挖掘导论>>

内容概要

《数据挖掘导论》全面介绍了数据挖掘的理论和方法，将重点放在如何用数据挖掘知识解决各种实际问题，涉及学科领域众多，适用面广。

包含大量的图表、综合示例和丰富的习题，并且使用示例，关键算法的简洁描述和习题，尽可能直接聚焦于数据挖掘的主要概念。

《数据挖掘导论》涵盖五个主题：数据、分类、关联分析、聚类和异常检测。

除异常检测外，每个主题都包含两章：前面一章讲述基本概念、代表性算法和评估技术，后面一章较深入地讨论高级概念和算法。

目的是使读者在透彻地理解数据挖掘基础的同时，还能了解更多重要的高级主题。

此外，书中还提供了大量示例、图表和习题。

作者简介

Pang-Ning Tan, 现为密歇根州立大学计算机与工程系助理教授, 主要教授数据挖掘、数据库系统等课程。

此前, 他曾是明尼苏达大学美国陆军高性能计算研究中心副研究员 (2002-2003)。

书籍目录

第1章 绪论	11.1 什么是数据挖掘	21.2 引发数据挖掘的挑战	21.3 数据挖掘的起源	31.4 数据挖掘任务	41.5 本书的内容与组织	7文献注释	7参考文献	8习题	10第2章 数据	132.1 数据类型	142.1.1 属性与度量	152.1.2 数据集的类型	182.2 数据质量	222.2.1 测量和数据收集问题	222.2.2 关于应用的问题	262.3 数据预处理	272.3.1 聚集	272.3.2 抽样	282.3.3 维归约	302.3.4 特征子集选择	312.3.5 特征创建	332.3.6 离散化和二值化	342.3.7 变量变换	382.4 相似性和相异性的度量	382.4.1 基础	392.4.2 简单属性之间的相似度和相异度	402.4.3 数据对象之间的相异度	412.4.4 数据对象之间的相似度	432.4.5 邻近性度量的例子	432.4.6 邻近度计算问题	482.4.7 选取正确的邻近性度量	50文献注释	50参考文献	52习题	53第3章 探索数据	593.1 鸢尾花数据集	593.2 汇总统计	603.2.1 频率和众数	603.2.2 百分位数	613.2.3 位置度量：均值和中位数	613.2.4 散布度量：极差和方差	623.2.5 多元汇总统计	633.2.6 汇总数据的其他方法	643.3 可视化	643.3.1 可视化的动机	643.3.2 一般概念	653.3.3 技术	673.3.4 可视化高维数据	753.3.5 注意事项	793.4 OLAP和多维数据分析	793.4.1 用多维数组表示鸢尾花数据	803.4.2 多维数据：一般情况	813.4.3 分析多维数据	823.4.4 关于多维数据分析的最后评述	84文献注释	84参考文献	85习题	86第4章 分类：基本概念、决策树与模型评估	894.1 预备知识	894.2 解决分类问题的一般方法	904.3 决策树归纳	924.3.1 决策树的工作原理	924.3.2 如何建立决策树	934.3.3 表示属性测试条件的方法	954.3.4 选择最佳划分的度量	964.3.5 决策树归纳算法	1014.3.6 例子：Web 机器人检测	1024.3.7 决策树归纳的特点	1034.4 模型的过分拟合	1064.4.1 噪声导致的过分拟合	1074.4.2 缺乏代表性样本导致的过分拟合	1094.4.3 过分拟合与多重比较过程	1094.4.4 泛化误差估计	1104.4.5 处理决策树归纳中的过分拟合	1134.5 评估分类器的性能	1144.5.1 保持方法	1144.5.2 随机二次抽样	1154.5.3 交叉验证	1154.5.4 自助法	1154.6 比较分类器的方法	1164.6.1 估计准确度的置信区间	1164.6.2 比较两个模型的性能	1174.6.3 比较两种分类器的性能	118文献注释	118参考文献	120习题	122第5章 分类：其他技术	1275.1 基于规则的分类器	1275.1.1 基于规则的分类器的工作原理	1285.1.2 规则的排序方案	1295.1.3 如何建立基于规则的分类器	1305.1.4 规则提取的直接方法	1305.1.5 规则提取的间接方法	1355.1.6 基于规则的分类器的特征	1365.2 最近邻分类器	1375.2.1 算法	1385.2.2 最近邻分类器的特征	1385.3 贝叶斯分类器	1395.3.1 贝叶斯定理	1395.3.2 贝叶斯定理在分类中的应用	1405.3.3 朴素贝叶斯分类器	1415.3.4 贝叶斯误差率	1455.3.5 贝叶斯信念网络	1475.4 人工神经网络(ANN)	1505.4.1 感知器	1515.4.2 多层人工神经网络	1535.4.3 人工神经网络的特点	1555.5 支持向量机	1565.5.1 最大边缘超平面	1565.5.2 线性支持向量机：可分情况	1575.5.3 线性支持向量机：不可分情况	1625.5.4 非线性支持向量机	1645.5.5 支持向量机的特征	1685.6 组合方法	1685.6.1 组合方法的基本原理	1685.6.2 构建组合分类器的方法	1695.6.3 偏倚—方差分解	1715.6.4 装袋	1735.6.5 提升	1755.6.6 随机森林	1785.6.7 组合方法的实验比较	1795.7 不平衡类问题	1805.7.1 可选度量	1805.7.2 接受者操作特征曲线	1825.7.3 代价敏感学习	1845.7.4 基于抽样的方法	1865.8 多类问题	187文献注释	189参考文献	190习题	193第6章 关联分析：基本概念和算法	2016.1 问题定义	2026.2 频繁项集的产生	2046.2.1 先验原理	2056.2.2 Apriori算法的频繁项集产生	2066.2.3 候选的产生与剪枝	2086.2.4 支持度计数	2106.2.5 计算复杂度	2136.3 规则产生	2156.3.1 基于置信度的剪枝	2156.3.2 Apriori算法中规则的产生	2156.3.3 例：美国国会投票记录	2176.4 频繁项集的紧凑表示	2176.4.1 最大频繁项集	2176.4.2 频繁闭项集	2196.5 产生频繁项集的其他方法	2216.6 FP增长算法	2236.6.1 FP树表示法	2246.6.2 FP增长算法的频繁项集产生	2256.7 关联模式的评估	2286.7.1 兴趣度的客观度量	2286.7.2 多个二元变量的度量	2356.7.3 辛普森悖论	2366.8 倾斜支持度分布的影响	237文献注释	240参考文献	244习题	250第7章 关联分析：高级概念	2597.1 处理分类属性	2597.2 处理连续属性	2617.2.1 基于离散化的方法	2617.2.2 基于统计学的方法	2637.2.3 非离散化方法	2657.3 处理概念分层	2667.4 序列模式	2677.4.1 问题描述	2677.4.2 序列模式发现	2697.4.3 时限约束	2717.4.4 可选计数方案	2747.5 子图模式	2757.5.1 图与子图	2767.5.2 频繁子图挖
--------	--------------	----------------	--------------	-------------	---------------	-------	-------	-----	----------	------------	---------------	----------------	------------	-------------------	-----------------	-------------	------------	------------	-------------	----------------	--------------	-----------------	--------------	------------------	------------	------------------------	--------------------	--------------------	------------------	-----------------	--------------------	--------	--------	------	------------	--------------	------------	---------------	--------------	---------------------	--------------------	----------------	-------------------	-----------	----------------	--------------	------------	-----------------	--------------	-------------------	----------------------	-------------------	----------------	-----------------------	--------	--------	------	------------------------	------------	-------------------	-------------	------------------	-----------------	---------------------	-------------------	-----------------	-----------------------	-------------------	----------------	--------------------	-------------------------	----------------------	-----------------	------------------------	-----------------	---------------	-----------------	---------------	--------------	-----------------	---------------------	--------------------	---------------------	---------	---------	-------	----------------	-----------------	------------------------	------------------	-----------------------	--------------------	--------------------	----------------------	---------------	-------------	--------------------	---------------	----------------	-----------------------	-------------------	-----------------	------------------	--------------------	--------------	-------------------	--------------------	--------------	------------------	-----------------------	------------------------	-------------------	-------------------	-------------	--------------------	---------------------	------------------	-------------	-------------	---------------	--------------------	---------------	---------------	--------------------	-----------------	------------------	-------------	---------	---------	-------	---------------------	-------------	----------------	---------------	---------------------------	-------------------	----------------	----------------	-------------	-------------------	--------------------------	---------------------	------------------	-----------------	----------------	--------------------	---------------	-----------------	------------------------	----------------	-------------------	--------------------	----------------	-------------------	---------	---------	-------	------------------	---------------	---------------	-------------------	-------------------	-----------------	---------------	-------------	---------------	-----------------	---------------	-----------------	-------------	---------------	----------------

<<数据挖掘导论>>

掘 2777.5.3 类Apriori方法 2787.5.4 候选产生 2797.5.5 候选剪枝 2827.5.6 支持度计数
 2857.6 非频繁模式 2857.6.1 负模式 2857.6.2 负相关模式 2867.6.3 非频繁模式、负模式和负
 相关模式比较 2877.6.4 挖掘有趣的非频繁模式的技术 2887.6.5 基于挖掘负模式的技术 2887.6.6
 基于支持度期望的技术 290文献注释 292参考文献 293习题 295第8章 聚类分析：基本概念和
 算法 3058.1 概述 3068.1.1 什么是聚类分析 3068.1.2 不同的聚类类型 3078.1.3 不同的簇类型
 3088.2 K均值 3108.2.1 基本K均值算法 3108.2.2 K均值：附加的问题 3158.2.3 二分K均值
 3168.2.4 K均值和不同的簇类型 3178.2.5 优点与缺点 3188.2.6 K均值作为优化问题 3198.3
 凝聚层次聚类 3208.3.1 基本凝聚层次聚类算法 3218.3.2 特殊技术 3228.3.3 簇邻近度
 的Lance-Williams公式 3258.3.4 层次聚类的主要问题 3268.3.5 优点与缺点 3278.4 DBSCAN
 3278.4.1 传统的密度：基于中心的方法 3278.4.2 DBSCAN算法 3288.4.3 优点与缺点 3298.5
 簇评估 3308.5.1 概述 3328.5.2 非监督簇评估：使用凝聚度和分离度 3328.5.3 非监督簇评估：
 使用邻近度矩阵 3368.5.4 层次聚类的非监督评估 3388.5.5 确定正确的簇个数 3398.5.6 聚类趋
 势 3398.5.7 簇有效性的监督度量 3408.5.8 评估簇有效性度量的显著性 343文献注释 344参考文
 献 345习题 347第9章 聚类分析：附加的问题与算法 3559.1 数据、簇和聚类算法的特性
 3559.1.1 例子：比较K均值和DBSCAN 3559.1.2 数据特性 3569.1.3 簇特性 3579.1.4 聚类算
 法的一般特性 3589.2 基于原型的聚类 3599.2.1 模糊聚类 3599.2.2 使用混合模型的聚类
 3629.2.3 自组织映射 3699.3 基于密度的聚类 3729.3.1 基于网格的聚类 3729.3.2 子空间聚类
 3749.3.3 DENCLUE：基于密度聚类的一种基于核的方案 3779.4 基于图的聚类 3799.4.1 稀疏
 化 3799.4.2 最小生成树聚类 3809.4.3 OPOSSUM：使用METIS的稀疏相似度最优划分 3819.4.4
 Chameleon：使用动态建模的层次聚类 3819.4.5 共享最近邻相似度 3859.4.6 Jarvis-Patrick聚类算
 法 3879.4.7 SNN密度 3889.4.8 基于SNN密度的聚类 3899.5 可伸缩的聚类算法 3909.5.1 可伸
 缩：一般问题和方法 3919.5.2 BIRCH 3929.5.3 CURE 3939.6 使用哪种聚类算法 395文献注释
 397参考文献 398习题 400第10章 异常检测 40310.1 预备知识 40410.1.1 异常的成因
 40410.1.2 异常检测方法 40410.1.3 类标号的使用 40510.1.4 问题 40510.2 统计方法
 40610.2.1 检测一元正态分布中的离群点 40710.2.2 多元正态分布的离群点 40810.2.3 异常检测
 的混合模型方法 41010.2.4 优点与缺点 41110.3 基于邻近度的离群点检测 41110.4 基于密度的
 离群点检测 41210.4.1 使用相对密度的离群点检测 41310.4.2 优点与缺点 41410.5 基于聚类
 的技术 41410.5.1 评估对象属于簇的程度 41510.5.2 离群点对初始聚类的影响 41610.5.3 使用簇的
 个数 41610.5.4 优点与缺点 416文献注释 417参考文献 418习题 420

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>