

<<数据挖掘教程>>

图书基本信息

书名：<<数据挖掘教程>>

13位ISBN编号：9787302105336

10位ISBN编号：7302105332

出版时间：2005-5-1

出版时间：清华大学出版社

作者：郭崇慧,田凤占,邓纳姆

页数：280

译者：郭崇慧,田凤占

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>

<<数据挖掘教程>>

内容概要

本书全面系统地介绍了各种数据挖掘的基本概念、方法和算法。

全书由四部分构成：第一部分是导论，全面介绍了数据挖掘的背景信息、相关概念及其所使用的主要技术；第二部分是核心算法，系统深入地描述了用于分类、聚类和关联规则的常用算法；第三部分是高级课题，主要介绍了Web挖掘、空间数据挖掘、时序数据和序列数据挖掘；第四部分是附录，介绍了目前市场上流行的一些数据挖掘工具产品。

书中对每种算法不仅进行了详尽的解释，还给出了算例及伪码。

每章后的练习和参考文献为读者提供了进一步思考相关问题的线索。

本书适宜作为计算机专业高年级本科生、研究生教材，也可作为相关领域研究人员的参考书。

作者简介

Margaret H. Dunham 在俄亥俄州牛津市的迈阿密大学获得了数学学士和数学硕士学位，在Southern Methodist大学获得了计算机科学博士学位。

Dunham教授的研究兴趣包括主存数据库、数据挖掘、时序数据库以及移动计算。

她目前是IEEE Transactions on Knowledge and Data Engineering 杂志的副主编。

她在数据库并发控制和恢复、数据库机、主存数据库以及移动计算等研究领域发表了大量学术论文。

书籍目录

第1部分 导论第1章 概述1.1 基本数据挖掘任务1.1.1 分类1.1.2 回归1.1.3 时间序列分析1.1.4 预测1.1.5 聚类1.1.6 汇总1.1.7 关联规则1.1.8 序列发现1.2 数据挖掘与数据库中的知识发现1.2.1 数据挖掘的发展1.3 数据挖掘问题1.4 数据挖掘度量1.5 数据挖掘的社会影响1.6 从数据库观点看数据挖掘1.7 数据挖掘的未来发展1.8 练习1.9 参考文献注释第2章 相关概念2.1 数据库/OLTP系统2.2 模糊集和模糊逻辑2.3 信息检索2.4 决策支持系统2.5 维数据建模2.5.1 多维模式2.5.2 索引2.6 数据仓储2.7 OLAP2.8 Web搜索引擎2.9 统计学2.10 机器学习2.11 模式匹配2.12 小结2.13 练习2.14 参考文献注释第3章 数据挖掘技术3.1 引言3.2 数据挖掘的统计方法3.2.1 点估计3.2.2 基于汇总的模型3.2.3 贝叶斯定理3.2.4 假设检验3.2.5 回归和相关3.3 相似性度量3.4 决策树3.5 神经网络3.5.1 激励函数3.6 遗传算法3.7 练习3.8 参考文献注释第2部分 核心课题第4章 分类4.1 引言4.1.1 分类中的问题4.2 基于统计的算法4.2.1 回归4.2.2 贝叶斯分类4.3 基于距离的算法4.3.1 简单方法4.3.2 K最近邻4.4 基于决策树的算法4.4.1 ID34.4.2 C4.5 和C5.04.4.3 CART4.4.4 可伸缩的决策树技术4.5 基于神经网络的算法4.5.1 传播4.5.2 神经网络有指导学习4.5.3 径向基函数网络4.5.4 感知器4.6 基于规则的算法4.6.1 从决策树生成规则4.6.2 从神经网络生成规则4.6.3 不用决策树或神经网络生成规则4.7 组合技术4.8 小结4.9 练习4.10 参考文献注释第5章 聚类5.1 引言5.2 相似性和距离度量5.3 异常点5.4 层次算法5.4.1 凝聚算法5.4.2 分裂聚类5.5 划分算法5.5.1 最小生成树5.5.2 平方误差聚类算法5.5.3 K均值聚类5.5.4 最近邻算法5.5.5 PAM算法5.5.6 结合能量算法5.5.7 基于遗传算法的聚类5.5.8 基于神经网络的聚类5.6 大型数据库聚类5.6.1 BIRCH5.6.2 DBSCAN5.6.3 CURE算法5.7 对类别属性进行聚类5.8 比较5.9 练习5.10 参考文献注释第6章 关联规则6.1 引言6.2 大项目集6.3 基本算法6.3.1 Apriori算法6.3.2 抽样算法6.3.3 划分6.4 并行和分布式算法6.4.1 数据并行6.4.2 任务并行6.5 方法比较6.6 增量规则6.7 高级关联规则技术6.7.1 泛化关联规则6.7.2 多层关联规则6.7.3 数量关联规则6.7.4 使用多个最小支持度6.7.5 相关规则6.8 度量规则的质量6.9 练习6.10 参考文献注释第3部分 高级课题第7章 Web挖掘7.1 引言7.2 Web内容挖掘7.2.1 爬虫7.2.2 Harvest系统7.2.3 虚拟Web视图7.2.4 个性化7.3 Web结构挖掘7.3.1 PageRank7.3.2 Clever7.4 Web使用挖掘7.4.1 预处理7.4.2 数据结构7.4.3 模式发现7.4.4 模式分析7.5 练习7.6 参考文献注释第8章 空间数据挖掘8.1 引言8.2 空间数据概述8.2.1 空间查询8.2.2 空间数据结构8.2.3 主题地图8.2.4 图像数据库8.3 空间数据挖掘原语8.4 一般化和特殊化8.4.1 渐进求精8.4.2 一般化8.4.3 最近邻8.4.4 STING8.5 空间规则8.5.1 空间关联规则8.6 空间分类算法8.6.1 对ID3的扩展8.6.2 空间决策树8.7 空间聚类算法8.7.1 对CLARANS的扩展8.7.2 SD(CLARANS)8.7.3 DBCLASD8.7.4 BANG8.7.5 WaveCluster8.7.6 近似8.8 练习8.9 参考文献注释第9章 时序数据挖掘9.1 引言9.2 时序事件建模9.3 时间序列9.3.1 时间序列分析9.3.2 趋势分析9.3.3 变换9.3.4 相似性9.3.5 预测9.4 模式检测9.4.1 串匹配9.5 时序序列9.5.1 AprioriAll9.5.2 SPADE9.5.3 一般化9.5.4 特征抽取9.6 时序关联规则9.6.1 事务间关联规则9.6.2 情节规则9.6.3 趋势依赖9.6.4 序列关联规则9.6.5 日历关联规则9.7 练习9.8 参考文献注释附录A 数据挖掘产品A.1 参考文献注释附录B 参考文献词汇总表

媒体关注与评论

数据挖掘技术是多学科交叉的新兴技术，它是随着数据的大量积累以及市场竞争对信息与知识的迫切需求而产生和发展起来的，并逐渐成为人们关注的热点。

人们希望通过数据挖掘技术找到蕴藏在数据中的有用信息，进而找到尚未发现的知识，为商业竞争、企业生产和管理、政府部门决策以及科学探索等提供信息与知识，这种所谓隐藏在数据中的信息与知识是人的先验知识和经验无法确定的，对于帮助人们作出适当决策是很有价值的。

数据挖掘技术是在统计学、人工智能（特别是机器学习）和数据库技术等多种技术的基础上发展起来的。

数据挖掘强调的是大数据量和算法的可伸缩性，它是一门很接近实用的学科，一出现就被许多部门所应用。

由于它的实用性和商业效益，近年来人们研究出许多数据挖掘的新方法，并开发了许多数据挖掘的新产品。

本书从数据库的角度对数据挖掘的基本方法和算法进行了系统的介绍。

全书共分三部分：第1部分包括第1~3章，介绍数据挖掘的发展和基本概念；第2部分包括第4~6章，介绍最基本的数据挖掘方法，这部分也是全书的重点；第3部分包括第7~9章，介绍了近年来出现的较新的数据挖掘方法和领域。

每章最后两节均为练习和参考文献注释。

一部分练习用于检验学生掌握书中所述概念和知识的情况，另一部分练习提出需要进一步研究和思考的问题。

每章的参考文献注释则较详细地说明了该章涉及的方法与算法的发展历程和状况，作者花费了很大精力查阅和收集这方面的资料。

本书适合作为计算机专业研究生及高年级本科生教材。

作为教科书，书中的内容有一定的深度和广度，对许多方法和算法都作了引导性的叙述。

但作为一本基础性的教科书，它不可能包括太广的内容，对近年来发展较快的一些新方法，如粗糙集、贝叶斯网络和支持向量机等，书中并未深入叙述。

要想更深入地掌握一些方法和提出改进建议，还需要查阅书中给出的参考文献和一些方法的最新进展。

本书还可作为相关领域科技人员的参考书。

<<数据挖掘教程>>

编辑推荐

《数据挖掘教程》适宜作为计算机专业高年级本科生、研究生教材，也可作为相关领域研究人员的参考书。

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>