

<<数据挖掘原理与实践>>

图书基本信息

书名：<<数据挖掘原理与实践>>

13位ISBN编号：9787121140501

10位ISBN编号：7121140500

出版时间：2013-2

出版时间：电子工业出版社

作者：蒋盛益

页数：284

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>

<<数据挖掘原理与实践>>

内容概要

《高等学校计算机规划教材：数据挖掘原理与实践》分为数据挖掘理论和数据挖掘实践两大部分。基础理论部分的主要内容包括数据挖掘的基本概念、数据挖掘的预处理、聚类分析、分类与回归、关联规则挖掘、例外点检测。

数据挖掘实践部分讨论了数据挖掘在通信行业、文本挖掘等方面的实际应用；通过四个案例展示了在通信行业中如何利用数据挖掘进行客户细分、客户流失分析、客户社会关系挖掘、业务交叉销售；通过跨语言智能学术搜索系统和基于内容的垃圾邮件识别两个案例展示了数据挖掘在文本挖掘方面的应用。

《高等学校计算机规划教材：数据挖掘原理与实践》可作为高等院校计算机电子商务及相关专业的学生相关的教材或参考书，也可供从事数据挖掘研究、设计等工作的科研、技术人员参考。

书籍目录

上篇 理论篇第1章 绪论 1.1 数据挖掘产生的背景 1.2 数据挖掘任务及过程 1.2.1 数据挖掘定义 1.2.2 数据挖掘对象 1.2.3 数据挖掘任务 1.2.4 数据挖掘过程 1.2.5 数据挖掘常用软件简介 1.3 数据挖掘应用 1.3.1 数据挖掘在商业领域中的应用 1.3.2 数据挖掘在计算机领域中的应用 1.3.3 其他领域中的应用 1.4 数据挖掘技术的前景、研究热点 1.4.1 数据挖掘技术的价值和前景 1.4.2 数据挖掘的研究热点 1.4.3 数据挖掘的未来发展 本章小结 习题1 第2章 数据处理基础 2.1 数据 2.1.1 数据及数据类型 2.1.2 数据集的类型 2.2 数据统计特性 2.2.1 数据的中心度量 2.2.2 数据散布程度度量 2.3 数据预处理 2.3.1 数据清理 2.3.2 数据聚合 2.3.3 数据变换 2.3.4 数据归约 2.4 相似性度量 2.4.1 属性之间的相似性度量 2.4.2 对象之间的相似性度量 2.5 本章小结 习题2 第3章 分类与回归 3.1 概述 3.2 决策树分类方法 3.2.1 决策树的基本概念 3.2.2 决策树的构建 3.2.3 ID3分类算法 3.2.4 C4.5分类算法 3.2.5 CART算法 3.3 贝叶斯分类方法 3.3.1 贝叶斯定理 3.3.2 朴素贝叶斯分类算法 3.4 K-最近邻分类方法 3.4.1 K-最近邻分类算法基本概念 3.4.2 K-最近邻分类算法描述 3.4.3 k-最近邻分类算法的优缺点 3.5 神经网络分类方法 3.5.1 人工神经网络的基本概念 3.5.2 典型神经网络模型介绍 3.5.3 神经网络的优缺点 3.6 支持向量机 3.7 集成学习法 3.8 不平衡类问题 3.9 分类模型的评价 3.9.1 分类模型性能评价指标 3.9.2 分类模型的过拟合 3.9.3 评估分类模型性能的方法 3.10 回归 3.10.1 线性回归 3.10.2 非线性回归 3.10.3 逻辑回归 3.11 本章小结 习题3 第4章 聚类分析 4.1 概述 4.1.1 聚类分析研究的主要内容 4.1.2 数据挖掘对聚类算法的要求 4.1.3 典型聚类方法简介 4.2 基于划分的聚类算法 4.2.1 基本k-means聚类算法 4.2.2 二分k-means算法 4.2.3 k-means聚类算法的拓展 4.2.4 k-medoids算法 4.3 层次聚类算法 4.3.1 BIRCH算法 4.3.2 CURE算法 4.3.3 ROCK算法 4.4 基于密度的聚类算法 4.5 基于图的聚类算法 4.5.1 Chameleon聚类算法 4.5.2 基于SNN的聚类算法 4.6 一趟聚类算法 4.6.1 算法描述 4.6.2 聚类阈值的选择策略 4.7 基于模型的聚类算法 4.7.1 期望最大化方法EM 4.7.2 概念聚类 4.7.3 SOM方法 4.8 聚类算法评价 本章小结 习题4 第5章 关联分析 5.1 概述 5.2 频繁项集发现算法 5.2.1 Apriori 算法 5.2.2 Fp-growth算法 5.3 关联规则的生成 5.4 非二元属性的关联规则挖掘 5.5 关联规则的评价 5.5.1 支持度和置信度 5.5.2 相关性分析 5.5.3 辛普森悖论 5.6 序列模式 5.6.1 问题描述 5.6.2 序列模式发现算法 5.7 本章小结 习题 第6章 离群点挖掘 6.1 概述 6.2 基于统计的方法 6.3 基于距离的方法 6.4 基于相对密度的方法 6.5 基于聚类的方法 6.5.1 基于对象离群因子的方法 6.5.2 基于簇的离群因子方法 6.5.3 基于聚类的动态数据离群点检测 6.6 离群点挖掘方法的评估 6.7 本章小结 习题 下篇 实践篇第7章 数据挖掘在电信业中的应用 7.1 数据挖掘在电信业的应用概述 7.1.1 客户细分 7.1.2 客户流失预测分析 7.1.3 客户社会关系挖掘 7.1.4 业务交叉销售 7.1.5 欺诈客户识别 7.2 案例一: 客户通话模式分析 7.2.1 概述 7.2.2 数据描述 7.2.3 数据预处理 7.2.4 发现之旅 7.3 案例二: 基于通话数据的社交网络分析 7.3.1 概述 7.3.2 客户呼叫图的构建 7.3.3 客户呼叫图的一般属性及其应用 7.3.4 客户呼叫图的社区发现及应用 7.4 案例三: 客户细分与流失分析 7.4.1 概述 7.4.2 数据准备 7.4.3 数据预处理 7.4.4 客户聚类分析 7.4.5 建立分类预测模型 7.4.6 模型评估与调整优化 7.5 案例四: 移动业务关联分析 7.5.1 概述 7.5.2 数据准备 7.5.3 数据预处理 7.5.4 关联规则挖掘过程 7.5.5 规则的优化 7.5.5 模型的应用 7.6 本章小结 第8章 文本挖掘与Web数据挖掘 8.1 文本挖掘 8.1.1 分词 8.1.2 文本表示与词权重计算 8.1.3 文本特征选择 8.1.4 文本分类 8.1.5 文本聚类 8.1.6 文档自动摘要 8.2 Web数据挖掘 8.2.1 Web内容挖掘 8.2.2 Web使用挖掘 8.2.3 Web结构挖掘 8.3 案例五——跨语言智能学术搜索系统 8.3.1 混合语种文本分词 8.3.2 基于机器翻译的跨语言信息检索 8.3.3 不同语种文本的搜索结果聚类 8.3.4 基于聚类的个性化信息检索 8.3.5 基于聚类的查询扩展 8.3.6 其他检索便利工具 8.3.7 系统性能评估 8.4 案例六——基于内容的垃圾邮件识别 8.4.1 垃圾邮件识别方法简介 8.4.2 基于内容的垃圾邮件识别方法工作原理 8.4.3 一种基于聚类的垃圾邮件识别方法 8.5 本章小结 参考文献

章节摘录

版权页：插图：客户信用风险分析和欺诈行为预测对企业的财务安全非常重要，利用数据挖掘中的关联分析、离群点检测技术对企业经营管理数据进行分析，如何预测可能将发生的风险？

判定哪些因素会导致风险？

这些风险主要来自于何处？

通过准确、及时地对各种信用风险进行监视、评价、预警和管理，评价这些风险的严重性、发生的可能性及控制这些风险的成本，进而采取有效的规避和监督措施，在信用风险发生之前对其进行预警和控制，趋利避害，防范信用风险。

(3) 市场分析和数据挖掘技术可以用于市场营销，其基本假定是“消费者过去的行为是其今后消费倾向的最好说明”，通过收集、加工和处理，能够反映消费者消费行为的大量信息，来确定特定消费群体或个体的兴趣、消费习惯、消费倾向和消费需求，进而推断出相应消费群体或个体下一步的消费行为，然后以此为基础，对识别出的消费群体进行特定内容的定向营销。

这与传统的不区分消费对象特征的大规模营销手段相比，大大节省了营销成本，提高了营销效果，从而为企业带来更多的利润。

商业消费信息来自市场中的各种渠道。

例如，当客户使用信用卡消费时，商业企业就可以在信用卡结算过程中收集商业消费信息，记录下客户消费的时间、地点、感兴趣的商品或服务、愿意接收的价格水平和支付能力等数据；当客户在申办信用卡、办理驾驶执照、填写商品保修单等其他需要填写表格的场合时，客户的个人信息就存入了相应的业务数据库。

企业除了自行收集相关业务信息之外，还可以从其他公司或机构购买此类信息为己所用。

这些来自各种渠道的数据信息通过融合，商家可以挖掘出能够用于向特定消费群体或个体进行定向营销的决策信息。

在市场经济比较发达的国家和地区，许多公司都开始原有信息系统的基础上通过数据挖掘对业务信息进行深度加工，以构筑自己的竞争优势，扩大自己的市场份额。

基于数据挖掘的营销对我国当前的市场竞争具有启发意义，我们经常看到繁华商业街上一些厂商对来往行人不分对象地散发大量商品宣传广告，其结果是不需要的人随手丢弃资料，而需要的人并不一定能够轻松得到。

如果家电维修服务公司向在商店中刚刚购买家电的消费者邮寄维修服务广告，药品厂商向医院特定门诊就医的病人邮寄广告，那么其营销效果肯定会比漫无目的的营销效果要好很多。

(4) 企业危机管理危机管理是管理领域新出现的研究热点，是以市场竞争中危机的出现为研究起点，分析企业危机产生的原因和过程，研究企业预防危机、应付危机、解决危机的手段和策略，以增强企业的免疫力、应变力和竞争力，使管理者能够及时、准确地获取所需要的信息，迅速捕捉到企业可能发生危机的一切可能事件和先兆，进而采取有效的规避措施。

在危机发生之前对其进行控制，趋利避害，从而使企业能够适应迅速变化的市场环境，保持长久的竞争优势。

但是由于危机产生的原因复杂，种类繁多，许多因素难以量化，很多因素由于没有历史数据和相应的统计资料，很难进行科学的计算和评估。

数据挖掘技术在危机识别、分析和控制等方面都可以发挥作用。

<<数据挖掘原理与实践>>

编辑推荐

《数据挖掘原理与实践》是高等学校计算机规划教材之一。

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>