

<<数据挖掘理论与应用>>

图书基本信息

书名：<<数据挖掘理论与应用>>

13位ISBN编号：9787811232523

10位ISBN编号：7811232529

出版时间：2008-4

出版单位：清华大学出版社有限公司

作者：胡可云 编

页数：248

字数：399000

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>

<<数据挖掘理论与应用>>

内容概要

本书从数据挖掘理论与数据挖掘应用过程两方面介绍了数据挖掘的最新成果。

在理论部分，本书介绍了数据挖掘技术所涉及的基本概念、主流技术和最新成果；在应用部分，本书结合具体的实例系统论述了商业理解、数据预处理、建模、模型部署等整个数据挖掘流程。

本书既可以作为大学本科生和研究生的补充教材，也可以作为企业实施数据挖掘和商务智能的实战指导；既可以作为初次接触数据挖掘技术的入门读物，也可以作为高级研究人员的参考书。

<<数据挖掘理论与应用>>

书籍目录

第1章 导论 1.1 数据挖掘概述 1.1.1 数据挖掘的背景 1.1.2 数据挖掘的定义 1.1.3 数据挖掘的应用 1.2 数据挖掘的一般过程 1.2.1 数据挖掘中的数据集市 1.2.2 数据挖掘的任务 1.2.3 数据挖掘过程 1.3 数据挖掘的一般方法 1.3.1 分类预测型方法 1.3.2 描述型方法 1.3.3 文本/Web挖掘方法理论篇 第2章 分类方法 2.1 决策树 2.1.1 决策树基本概念 2.1.2 决策树构造过程 2.1.3 决策树的扩展 2.2 前馈神经网络 2.2.1 基本概念 2.2.2 BP训练过程 2.2.3 RBF网络 2.3 基于规则的方法 2.3.1 AQ算法 2.3.2 C4.5rules 2.3.3 RIPPER 2.4 支持向量机 2.4.1 核函数 2.4.2 线性可分模式下的最优超平面 2.4.3 线性不可分模式下的最优超平面 2.4.4 支持向量机 2.5 贝叶斯分类 2.5.1 贝叶斯理论和极大后验假设 2.5.2 贝叶斯网络和贝叶斯分类器 2.5.3 几种常见的贝叶斯分类器模型 2.5.4 贝叶斯分类器应用举例 2.6 粗糙集方法 2.6.1 粗糙集概念 2.6.2 粗糙集基本算法 2.6.3 粗糙集方法的扩展 2.7 其他分类方法 2.7.1 回归分析 2.7.2 k-最近邻分类方法 2.7.3 组合学习方法 第3章 聚类方法 3.1 聚类方法概述 3.1.1 聚类分析中的常见数据类型 3.1.2 对聚类算法的一些典型要求 3.1.3 主要的聚类方法 3.2 划分聚类 3.2.1 k-均值算法 3.2.2 二分k-均值聚类方法 3.2.3 k-中心点算法 3.3 层次聚类 3.3.1 凝聚的和分裂的层次聚类 3.3.2 BIRCH算法 3.3.3 CuRE算法 3.3.4 ROCK算法 3.3.5 Chameleon 3.4 基于密度的聚类 3.4.1 DBSCAN算法 3.4.2 OPTICS算法 3.5 Kohonen聚类 3.5.1 自组织神经网络 3.5.2 Kohonen自组织映射 3.6 孤立点分析 3.6.1 基于统计的孤立点检测 3.6.2 基于距离的孤立点检测 3.6.3 基于偏离的孤立点检测方法 3.7 概念格 3.7.1 基本概念 3.7.2 概念格的建造 3.7.3 规则提取 第4章 关联分析 4.1 基本概念与挖掘过程 4.1.1 基本概念 4.1.2 关联规则挖掘过程 4.2 频繁项集挖掘算法 4.2.1 Apriori算法 4.2.2 Apriori算法的改进 4.2.3 FP_Growth算法 4.3 关联规则生成算法 4.4 频繁闭项集挖掘 4.5 关联规则的扩展 4.5.1 多层次关联规则 4.5.2 多维关联规则 4.5.3 定量关联规则 4.5.4 加权关联规则 4.5.5 序列模式分析 第5章 文本与Web挖掘 5.1 文本挖掘 5.1.1 文本预处理 5.1.2 文本检索 5.1.3 文本分类 5.1.4 文本聚类 5.1.5 文本摘要 5.2 Web挖掘 5.2.1 概述 5.2.2 Web内容挖掘 5.2.3 Web结构挖掘 5.2.4 Web使用挖掘应用篇 第6章 业务理解 6.1 需求分析 6.1.1 需求分析的内容 6.1.2 需求分析的方法 6.1.3 需求分析的结果 6.1.4 需求分析的注意事项 6.2 实例：客户细分项目的需求分析 6.2.1 客户细分项目的内容 6.2.2 分析方法 6.2.3 分析结果 第7章 数据预处理 7.1 数据理解 7.2 数据准备 7.2.1 数据整理与合并 7.2.2 数据抽样 7.2.3 训练集和测试集的划分方法 7.2.4 类标签的确定 7.3 数据描述 7.3.1 单变量描述方法 7.3.2 多变量描述方法 7.4 数据清理 7.4.1 缺值处理 7.4.2 探测异常点与噪声清除 7.5 变量变换与合成 7.5.1 连续变量归一化 7.5.2 离散变量的数值化 7.5.3 连续变量离散化 7.5.4 变量变换 7.5.5 变量合成 7.6 变量选择 7.6.1 概述 7.6.2 包装方法 7.6.3 过滤方法 7.6.4 主成分及因子分析 7.7 一些算法对预处理的要求 7.8 实例：客户流失项目的数据预处理 7.8.1 数据理解和数据准备 7.8.2 数据描述和清理 7.8.3 数据变换与选择 第8章 建模 8.1 算法选择 8.2 模型参数调整 8.3 模型评估和性能比较 8.3.1 分类模型的评估方法 8.3.2 聚类模型的评估方法 8.4 模型导出 8.5 实例客户流失项目的建模 8.5.1 算法选择 8.5.2 参数调整 8.5.3 性能评估 8.5.4 模型导出 第9章 模型部署与维护 9.1 模型部署 9.2 模型维护 9.3 客户流失项目的模型部署与维护附录A 主要数据挖掘软件简介 A1 SAS Enterprise Miner A1.1 概述 A1.2 数据挖掘过程及模块 A2 SPSS Clementine A2.1 概述 A2.2 数据挖掘过程及模块 A3 IBM Intelligent Miner A3.1 概述 A3.2 数据挖掘过程及模块 A4 其他常见数据挖掘工具参考文献

<<数据挖掘理论与应用>>

章节摘录

第6章 业务理解 数据挖掘过程并不是一个纯技术的过程，而是一个业务和技术相结合过程，或者说技术为业务服务的过程。

片面地追求技术指标往往使得项目失败。

数据挖掘算法相当于一个工具，如何使用这个工具非常重要。

工具先进不见得就能得出好的项目结果，就如同有一架世界顶级的钢琴摆在面前，不是谁都能弹出最动听的曲子一样。

而有些人则能在一架不起眼的钢琴上弹奏出优美的旋律。

在数据挖掘过程中，对业务的理解很大程度决定了最终结果的成败，在实际项目中，有很多这样的例子。

例如，在一个预测移动用户流失的项目中，如果只是简单地追求预测的高准确率，而不考虑项目的目的是为了挽留这些用户，那么很可能的结果就是很高的预测准确率，但是在去找这些用户的时候，发现都已经无法联系了——模型确实很准确，找到了已流失的用户，但对客户来说毫无帮助。

因此，在启动一个数据挖掘过程时，首先要做的是对业务进行理解，确定客户的需求是什么，要达到什么样的业务目标；再根据客户的业务目标制定数据挖掘的目标；然后根据对目标的理解收集相关的数据和进行数据处理。

6.1 需求分析 需求分析的目的是熟悉相关业务，明确希望通过数据挖掘达到的业务目标。只有熟悉了相关业务，才能深刻地理解数据的含义，确定项目的范围，把业务问题正确地转换为定义良好的数据挖掘过程。

6.1.1 需求分析的内容 1. 相关业务 相关业务包括组织结构、业务流程及存在问题。数据挖掘系统需要能解决实际的业务问题，才能保证它的成功。

需要确定谁是数据挖掘结果的最终用户，现有的业务流程是怎样的，存在一些什么问题，数据挖掘如何改进业务流程，要达到什么样的目标。

注意业务目标和数据挖掘目标是不同的。

业务目标是从业务角度对数据挖掘过程要获得的目标的描述。

例如，发现客户流失的模式，并利用该模式进行营销活动。

数据挖掘目标是从数据挖掘角度对数据挖掘过程要获得的目标的描述。

例如，以60%的准确率预测将要流失的客户。

在业务分析过程中，需要注意哪些因素对业务问题的影响是最重要的。

在建模过程中，一组好的变量（代表关键因素）可能对模型的最终效果起决定性作用。

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>