

<<数据挖掘理论与实例>>

图书基本信息

书名：<<数据挖掘理论与实例>>

13位ISBN编号：9787514118148

10位ISBN编号：7514118144

出版时间：2012-7

出版时间：经济科学出版社

作者：王晖，王琪，何琼 著

页数：231

字数：200000

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>

<<数据挖掘理论与实例>>

内容概要

《数据挖掘理论与实例》编著者王晖、王琪何琼。

《数据挖掘理论与实例》的主要内容包括理论和实例两部分：第一、第二章介绍数据挖掘、数据仓库和数据挖掘的常用技术等基本理论；实例部分是以作者的两个研究课题为基础，第三、第四章介绍呼叫中心数据仓库的构建和数据挖掘模型与实现(MS SQL Server 2000、决策树)；第五、第六章介绍数据挖掘在QFII投资理念与持股偏好研究中的应用(Rsoftware、MySQL、多元逐步线性回归、因子分析、聚类分析等)。

<<数据挖掘理论与实例>>

书籍目录

第一章 绪论

第一节 什么是数据挖掘

第二节 基本数据挖掘任务

第三节 数据挖掘的过程

第四节 数据仓库与OLAP技术概述

第五节 数据挖掘技术的发展

第二章 数据挖掘工具

第一节 数据挖掘的统计方法

第二节 聚类分析

第三节 决策树

第四节 相关软件

第三章 呼叫中心中数据仓库的构建

第一节 数据仓库构建的实施方法及步骤

第二节 呼叫中心数据仓库模型设计

第三节 数据仓库生成

第四章 呼叫中心中的数据挖掘模型与实现

第一节 问题鉴别

第二节 解决方案

第三节 基于决策树的分类算法模型

第四节 C4.5算法构造信息需求分类和客户细分决策树实例

第五节 功能模块的实现

第六节 系统应用示例

第五章 QFII投资理念与持股偏好研究中数据收集与整理

第一节 外国机构投资者投资理念及持股偏好概述

第二节 QFII重仓股数据来源

第三节 因变量的选取

第四节 自变量的选取

第六章 QFII投资理念与持股偏好研究中数据挖掘模型与实现

第一节 重仓股家数变化趋势和行业分布

第二节 重仓股持有时间特征

第三节 描述性统计分析

第四节 相关性分析

第五节 持股偏好多元线性逐步回归分析

参考文献

附表

附表1 12家QFII基本情况及最新额度

附表2 瑞士银行的名称

附表3 12家QFII持有股票家数和行业情况汇总

附表4 各季度因变量和自变量的均值及样本个数

附表5 各季度因变量和自变量的方差及样本个数

附表6 2008年12月31日Y对上一个季度所有自变量线性相关系数

附表7 Y对上一个季度、本季度及未来一个季度的所有自变量X逐步回归模型系数情况

附表8 两种方式回归结果对比

附录

附录一 数据仓库中的数据表架构

<<数据挖掘理论与实例>>

附录二 数据仓库关系图

附录三 QFII重仓股数据的获取过程

附录四 自变量数据的获得及缺失值处理

附录五 模型建立—数据分析过程

章节摘录

版权页：插图：一、信息需求分类毫无疑问，任何信息系统都不可能满足所有客户的需求，因此，为客户提供最需要的信息成为信息系统的必然目标，而选择目标信息的基础是信息分类。

通常分类变量有地域、客户职业等。

另外，判断一个人将来需求行为的最好指标是他过去的行为，这是进行客户需求历史分析的真正意义所在。

这样我们就可以找出信息需求对哪些因素敏感，如地区差异、不同客户职业等因素。

信息需求分类挖掘所要解决的问题是，根据实际情况找出客户职业、地理区域等因素与信息需求量之间的关系并建立模式，以便决策者通过调整服务对策，给予客户更多的帮助，并对急需的信息做出预测。

二、客户细分 客户细分是指将一个大的客户群划分成为不同的细分群的动作，每个细分群的客户拥有相似的属性，属于不同细分群的客户被视为不同的客户群。

这种相似可以是地理区域的相似、收入上的相似或其他方式的相似。

细分可以让用户从比较高的层次上来看整个数据库的数据，从而使人们可以用不同的方式来对待处于不同细分群的客户，从而提高客户的满意度。

在农业呼叫中心，可以通过对特定客户背景信息的分析，预测该客户所属的客户类别。

而且分析客户的组成情况（地区差异，接人方式的差异），能更好地了解客户，从而采取相应的服务策略，这样可以有效地提高服务质量，达到最佳的效果。

这样既能够提高信息机构的服务水平，开发客户资源，避免客户流失，又能够节约资源，利用最小的投入得到较大的社会效益。

值得一提的是，利用数据挖掘技术进行信息需求分类和客户细分与OLAP的根本区别在于，前者能够以更全面的视角洞察信息需求和客户，同时发现一些隐藏在数据背后的规律；而后者则是根据人的主观意图，通过向数据库发出指令，从数据库中得到一个结果。

第二节 解决方案 处理大量、混乱又复杂的数据的一个很好方法是分类，分类可以让决策者从比较高的层次上来看整个数据库的数据。

在分类技术发展过程中，流行的几种分类技术方法包括贝叶斯分类、神经网络、遗传算法、统计学方法、粗略集合方法、最近邻居方法和决策树算法等。

<<数据挖掘理论与实例>>

编辑推荐

《数据挖掘理论与实例》是一种技术，它将传统的数据分析方法与处理大量数据的复杂算法相结合。
《数据挖掘理论与实例》的主要目标是通过不同领域的应用案例来说明数据挖掘在实际应用中的具体操作方法。
将数据库管理系统MySQL和统计软件R结合，利用数据挖掘技术帮助陷入海量数据中的组织和个人提取有用的信息。

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>