

<<数据挖掘方法与应用>>

图书基本信息

书名：<<数据挖掘方法与应用>>

13位ISBN编号：9787300099705

10位ISBN编号：730009970X

出版时间：2009-1

出版时间：中国人民大学出版社

作者：吕晓玲,谢邦昌

页数：238

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>

<<数据挖掘方法与应用>>

前言

随着社会经济的不断发展、科学技术的不断进步，统计方法越来越成为人们必不可少的式具和手段。

在教学过程中，老师们也越来越感到运用统计方法解决实际问题的重耍，不少人在探索如何运用统计软件介绍和学习统计方法。

谢邦昌教授、黄登源教授在多年的教学中，积累了丰富的经验，他们热情倡议，将他们的讲稿提供出来并编写成教材，供更多的人学习和使用。

这正与我们的初衷不谋而合。

2005年开始着手这套系列教材的编写，经过不断讨论、反复的论证，形成了现在的模式。

由于有许多研究生的帮忙，又有几位年轻老师的辛劳，这套书终于问世。

在我们看来，掌握统计方法不仅要理论上弄明白，更重要的在于能够正确有效地运用这些方法，分析说明实际问题。

这套书正是试图利用实际数据，通过统计软件的实际操作，将所能够使用的统计方法加以说明，使读者不仅能够了解相应的统计方法，而且能够通过计算机操作学会运用这些方法处理分析实际数据。

希望本套书的出版能够为读者提供这样学习的工具。

由于水平有限，难免有不足之处。

恳请读者朋友们提出宝贵意见。

我们也会循着这样的思路，在教学以及和读者的交流沟通中不断积累、不断提高、不断完善，奉献给读者更多更好的成果。

感谢为这套书的编写付出汗水的研究生，感谢儿位认真用心的年轻老师，感谢中国人民大学出版社的大力支持。

谢谢读者，希望能够加强沟通和联系，为提高统计方法实际运用的能力和水平共同努力。

<<数据挖掘方法与应用>>

内容概要

随着信息技术的飞速发展，数据的产生和存储达到了空前繁荣的阶段。如何从海量的数据中提取潜在的有用信息，给传统的数据处理技术提出了严峻的考验，数据挖掘方法应运而生。

数据挖掘是一个多学科的交叉研究领域，不仅大学里的学术人员在研究它，商业公司的专家和技术人员也在密切地关注它和使用它；它不仅涉及人工智能领域以及统计学的应用，而且涉及数据库的管理和使用。

从技术上来讲，数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的，但又潜在有用的信息和知识的过程。

从商业应用来讲，数据挖掘是一种新的商业信息处理技术，其主要特点是对商业数据库中的大量业务数据进行抽取、转换、分析和其他模式化的处理，从中提取辅助商业决策的关键性数据。

本书第1章对数据挖掘进行了概述，包括数据挖掘的定义、重要性、功能、步骤和常用方法等。第2章和第3章介绍了两种数据挖掘中常用的指导的学习算法、关联规则和聚类分析，它们处理的数据的特点是没有独立的需要预测或分类的变量，而只是试图从数据中发现一些固有的模式。关联规则就是要发现两个或多个事物之间的联系；聚类分析就是要把数据中具有相似性质的放在一类，而不同类之间尽量做到有较大的不同。

第4章和第5章介绍了两种数据挖掘中常用的指导的学习算法、决策树和神经网络。它们处理的数据含有独立的需要预测或分类的变量，它们的目的就是寻找一些自变量的函数或算法对数据进行准确的预测或分类。

决策树方法在对数据处理的过程中，将数据按照树状结构分成若干分枝形成决策规则；神经网络在一定程度上模仿了人脑神经系统处理信息，存储以及检索的功能，它是一个非线性的映射系统。

第6章和第7章介绍了两种数据挖掘中常用的传统统计的方法，回归分析和时间序列。回归分析是寻找自变量和因变量之间关系的预测模型，包括线性回归和Logistic回归；时间序列分析，顾名思义，是处理以时间为序的观测数据的方法。

本书的一个特点是不仅对上述方法作了理论的阐述，还结合案例分析讲述了如何应用STATISTICA软件实现上述方法对数据的分析，是一本理论和实践相结合的理论性和应用性都很强的书。

<<数据挖掘方法与应用>>

作者简介

吕晓玲, 1977年8月生。

学历： 博士（管理科学），香港城市大学管理科学系，中国香港，2004.9–2007.8 硕士（概率论与数理统计），南开大学数学系，中国天津，1999.9–2002.7 学士（概率论与数理统计），南开大学数学系，中国天津，1995.9–1999.7 学术经历： 科研助理：香港城市大学管理科学系：数据挖掘中心，统计咨询中心，2001-2004 访问学者

：Department of Applied Statistics, Johannes Kepler University of Linz，奥地利，2007年2月至6月 科研方向： 应用数理统计：数据挖掘；电视观众收视行为统计建模（博士论文方向）；电子商务客户，网络用户行为统计分析；消费者行为数量分析，满意度研究，客户关系管理；贝叶斯预测；离散选择模型；随机系数模型等 荣誉与奖励：

The Outstanding Academic Performance Award for Research Degree Students, 香港城市大学（2006）
Research Tuition Scholarship, 香港城市大学（2005, 2006） 光华奖学金（一等），南开大学（2000）
免试保送研究生，南开大学（1999） 校三好学生，南开大学（1996, 1997, 1998, 1999） 宝洁奖学金，南开大学（1998） 校二等奖学金，南开大学（1998） “九章”基础数学奖学金，南开大学（1996, 1997） 校一等奖学金，南开大学（1996, 1997） 学术论文：
Lo, H.P., Z.N.G and Xiaoling Lu (2003)
) , Mining Loyal Customers: A Practical Use of the Repeat Buying Theory, in Wai-Ki Ching and Michael Kwok-Po Ng (2003) , Advances in Data Mining and Modeling, World Scientific, p167-181 Lu, Xiaoling and Hing-Po Lo (2007)
) , Television Audience Satisfaction: Antecedents and Consequences, Journal of Advertising Research, 47 (3)
) , 354-363 Lu, Xiaoling and Hing-Po Lo (2007)
) , Modeling Diffusion Patterns of Television Programs, Industrial Marketing Management, under review

<<数据挖掘方法与应用>>

书籍目录

第1章 数据挖掘概述1.1 数据挖掘定义1.1.1 数据挖掘的技术定义1.1.2 数据挖掘的商业定义1.2 数据挖掘的重要性及意义1.3 数据挖掘功能1.4 数据挖掘步骤和标准1.4.1 数据挖掘步骤1.4.2 数据挖掘需要的人员1.5 数据挖掘常用方法1.5.1 数据挖掘的对象1.5.2 数据挖掘的常用方法练习题第2章 关联规则2.1 关联规则介绍2.2 关联规则种类2.2.1 一般意义上的关联规则2.2.2 带有时间性的序列关联分析2.3 关联规则算法2.3.1 普通的关联规则算法2.3.2 序列关联规则算法2.4 STATISTICA中的关联规则2.5 案例分析练习题第3章 聚类分析3.1 聚类分析介绍3.2 距离定义3.2.1 点之间的距离3.2.2 类之间的距离3.3 聚类分析算法3.3.1 层次聚类3.3.2 基于划分的聚类3.3.3 EM聚类3.4 STATISTICA中的聚类分析3.5 案例分析练习题第4章 决策树建模4.1 决策树介绍4.1.1 决策树的基本知识4.1.2 决策树的应用和发展趋势4.2 树的建模过程4.2.1 数据要求4.2.2 树的生长4.2.3 有效性和风险性4.2.4 属性选择4.3 STATISTICA中的决策树4.4 案例分析练习题第5章 神经网络建模5.1 神经网络介绍5.2 神经网络的基本概念和原理5.2.1 基本组成单元5.2.2 神经网络的训练过程5.2.3 基本的神经网络模型5.3 STATISTICA中的神经网络模型5.4 案例分析练习题第6章 回归分析6.1 回归分析介绍6.2 线性回归模型6.2.1 模型的建立及未知参数的估计6.2.2 回归方程与回归参数的检验及变量的选择问题6.2.3 回归诊断和决定系数6.3 1ogistic回归模型6.3.1 1ogistic回归模型的建立6.3.2 1ogistic回归模型的参数估计6.3.3 1ogistic回归模型的检验及诊断6.3.4 1ogistic回归模型结果的解释6.3.5 1ogistic回归模型的扩展6.4 STATISTICA中的回归6.5 案例分析练习题第7章 时间序列7.1 时间序列介绍7.2 时间序列算法7.2.1 传统时间序列分析7.2.2 ARIMA模型7.3 STATISTICA中的时间序列7.4 案例分析练习题参考文献

章节摘录

第1章 数据挖掘概述 1.1 数据挖掘定义 数据挖掘是一个多学科交叉研究领域，不仅大学里的专门研究人员在使用它，商业公司的专家和技术人员也在密切地关注它；它不仅涉及人工智能领域以及统计学的应用，而且也涉及数据库的使用。

不同领域的人从不同的研究背景出发研究不同行业的数据，也就给了数据挖掘不同的内容和定义。这里我们就数据挖掘的技术定义以及商业定义展开讨论。

1.1 数据挖掘的技术定义 数据挖掘（data mining，DM）就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的，但又是潜在有用的信息和知识的过程。

与数据挖掘相近的同义词有数据融合、数据分析和决策支持等。

这个定义包括以下几层含义：（1）数据源必须是真实的、大量的、含噪声的；（2）发现的是用户感兴趣的知识；（3）发现的知识要可接受、可理解、可运用；

<<数据挖掘方法与应用>>

编辑推荐

《数据挖掘方法与应用》的一个特点是不仅对上述方法作了理论的阐述，还结合案例分析讲述了如何应用STATISTICA软件实现上述方法对数据的分析，是一本理论和实践相结合的理论性和应用性都很强的书。

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>