

实战

大数据

MATLAB数据挖掘
详解与实践

深入探究数据挖掘之有效工具
深入讲解数据挖掘各个环节技术
利用数据挖掘算法和MATLAB进行数据分析

许国根
贾瑛 著

可视化大数据分析

未知信息数据挖掘

大数据分析

未知信息数据挖掘

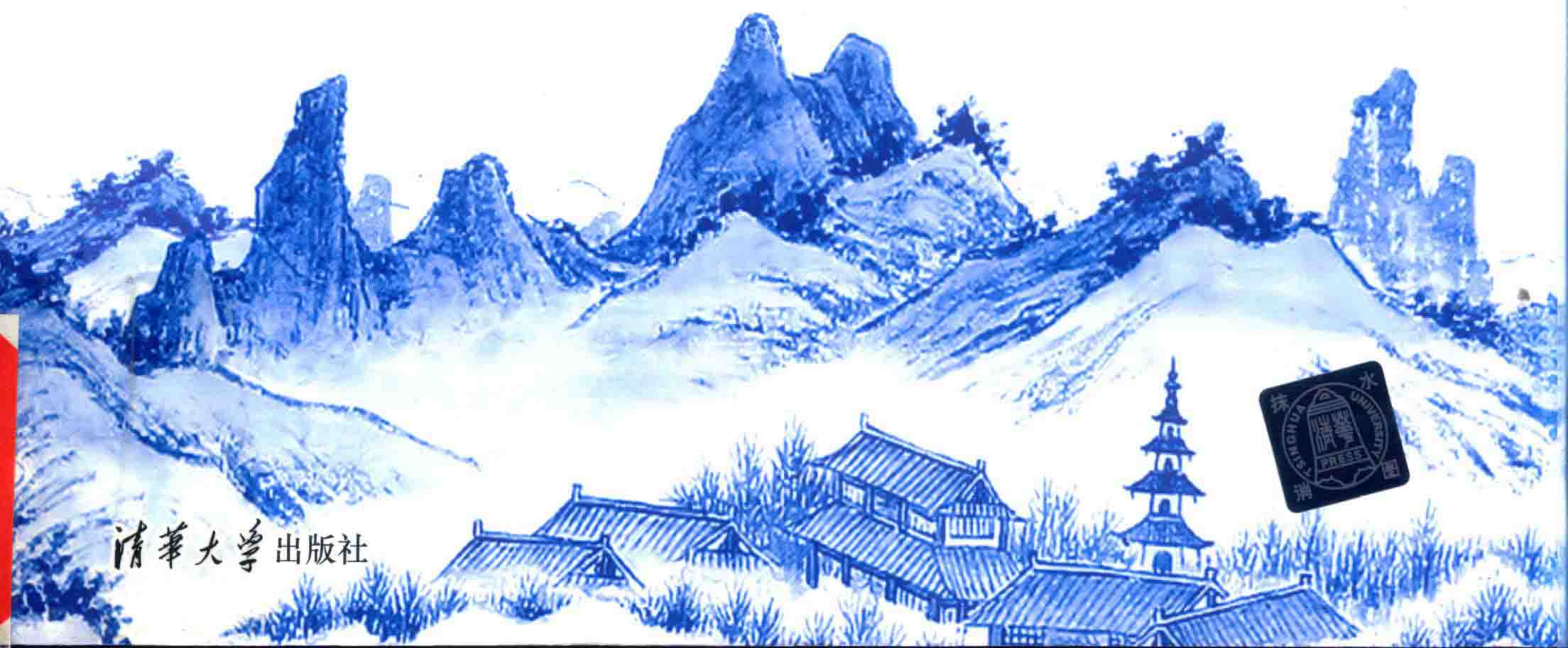
大数据分析

未知信息数据挖掘

大数据分析

未知信息数据挖掘

大数据分析



清华大学出版社





实战 大数据

MATLAB数据挖掘
详解与实践

许国根 贾瑛 著

清华大学出版社
北京

内 容 简 介

大数据时代,我们需要对各种海量数据进行筛选、清洗、挖掘,在这个过程中,获取有效数据的方式方法和模型算法成为了整个数据挖掘过程的重点, MATLAB 作为一个数据挖掘工具,如何正确和准确地使用它成为了重中之重。

针对实际应用数据挖掘技术的要求,本书既介绍了数据挖掘的基础理论和技术,又较为详细地介绍了各种算法以及 MATLAB 程序。本书共分 4 篇,分别介绍了数据挖掘的基本概念、技术与算法以及应用实例。期望通过大量的实例分析帮助广大读者掌握数据挖掘技术,并应用于实际的研究中,提高对海量数据信息的处理及挖掘能力。本书针对性和实用性强,具有较高的理论和实用价值。

本书作者就职于部队高校,专攻数据挖掘,并应用于大量实际项目,本书同时得到了国内著名数据挖掘公司的技术支持,很多案例来自实际项目。

本书可作为高等院校计算机工程、信息工程、生物医学工程、化学、环境、经济、管理等学科的研究生、本科生的教材或教学参考书,亦可作为企事业单位管理者、信息分析人员、市场营销人员和研究与开发人员的参考资料。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

实战大数据——MATLAB 数据挖掘详解与实践 / 许国根, 贾瑛著. --北京:清华大学出版社, 2017

ISBN 978-7-302-45101-3

I. ①实 … II. ①许 …②贾 … III. ①MATLAB 软件 IV. ① TP317

中国版本图书馆 CIP 数据核字 (2016) 第 227229 号

责任编辑: 栾大成

装帧设计: 杨玉芳

责任校对: 徐俊伟

责任印制: 王静怡

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 三河市金元印装有限公司

经 销: 全国新华书店

开 本: 188mm×260mm 印 张: 35.25 字 数: 904 千字

版 次: 2017 年 8 月第 1 版 印 次: 2017 年 8 月第 1 次印刷

印 数: 1~4000

定 价: 89.00 元

产品编号: 058297-01

前 言

计算机技术和通信技术的迅猛发展将人类社会带入了信息时代,在最近十几年里数据库中存储的数据量急剧增大。网络经济、注意力经济等新概念的提出,以其巨大的社会效益和极富挑战与机遇的内涵,成为信息科学引人注目的研究课题。大量的信息在给人们提供方便的同时也带来了一系列的问题,由于信息量过大,超出了人们掌握、理解信息的能力,因而给正确运用信息带来了困难。诸如信息过量、难以消化;信息真假难以辨识;信息安全难以保证;信息形式不一致,难以统一处理,等等,造成了“数据丰富,知识贫乏”。

决策者很难从海量的数据中提取出有价值的知识,促使人们产生了对数据分析工具的强烈需求,如何理解已有的历史数据并用以预测未来的行为,如何从这些海量数据中发现信息,变被动的数据为主支的知识,如何快速、准确地获得有价值的网络信息和网络服务,为用户提供重要的、未知的信息或知识、指导政府决策、企业决策、获取更大的经济效益和社会效益,这些都迫使人们去寻找新的、更为有效的数据分析手段,对各种“数据矿藏”进行有效的挖掘以发挥其应用潜能,20世纪80年代后期至今,数据挖掘正是在这样的应用需求背景下产生并迅速发展起来的,它是开发信息资源的一套科学方法、算法及软件工具和环境。

利用数据分析工具所获取的信息和知识,可以广泛地用于商务管理、生产控制、市场分析、工程设计和科学研究与探索等诸多方面。它不仅是一个重要的研究领域,而且在现实世界具有重大的潜在应用价值。

数据挖掘技术涉及人工智能的机器学习、模式识别、数据库与数据仓库、统计学、可视化图形学等各个领域,吸引了大批学者从事学术研究和工具产品的开发。20世纪90年代中后期,在国外数据挖掘已经形成高潮,我国研究数据挖掘的学者数量也在迅速增长。

由于数据挖掘是一门新兴的、正在不断发展的学科,其理论仍然不成熟,没有完善的理论体系,很多人在没有经历数据挖掘成熟应用项目的情况下,投入到这方面的技术探索与实践工作中来,效果不明显,使得他们对于数据挖掘的许多方面都在打问号,认为数据挖掘是虚的,是炒作。实际上数据挖掘与很多技术一样是一项很实用的技术,它必然会给各行各业的工作带来高效益和高效率。

从应用的角度看,数据挖掘是一个工具,为了很好地应用数据挖掘,首先要了解数据挖掘,尽量掌握数据挖掘的技术和方法,须知道什么时候应该使用何种数据挖掘技术,了解模型内部机制,这样才可以知道如何有效地准备建模所用的数据集,以及如何使用不同参数来改进模型的输出结果。现在有关数据挖掘的书籍越来越多,但这些书籍理论性太强,满篇数学公式,即使给出算法,也只是伪代码,看不到源代码以及算法的实际效果和各种算法的对比结果,而且应用实例很少,这往往使大多数读者感到困惑,让人难以理解,不知如何下手。有时虽然可以借助现在的专业计算机软件来完成数据挖掘工作,但因涉及知识产权保护和价格等因素,不可能每个需要进行数据挖掘的读者都能拥有此类软件。对大多数读者而言,目前确实还缺少一本具有较强系统性、可比性和实用性的有关数据挖掘的参考书。基于这点考虑,作者撰写了本书,向读者介绍各种数

据挖掘技术、方法及基于 MATLAB 的具体算法。想通过系统的介绍和实例分析，让众多的读者不仅具备数据挖掘的理论，而且能掌握数据挖掘应用方法，可以在各自的学科实际研究中予以应用，也使数据挖掘技术更易于使用和开发。

本书按照理论基础、实现步骤、实例三部分内容进行阐述，避免空洞的理论说教，着重介绍应用实例，具有较强的指导性和实用性，使读者不至于面对如此丰富的理论和方法无所适从，而是通过了解各种算法的实现思路和方法，体会算法源代码的意义，这样即使所举的实例不属于读者从事的学科，也能举一反三，掌握数据挖掘技术并应用于自己从事的科学研究中。

本书分为 4 篇，每篇涵盖的内容比较广泛，既有对数据挖掘概念的讨论，也有对数据挖掘技术和原理的介绍，而且编写了大量的实例，并给出了相应的程序。第 1 篇介绍数据挖掘的相关概念以及在多个领域中的应用情况；第 2 篇介绍数据挖掘算法，包括人工神经网络、决策树、遗传算法、关联分析、统计分析方法、支持向量机及一些聚类算法；第 3 篇介绍与数据挖掘相关的一些技术，包括数据仓库、模糊集理论、粗糙集技术、目标优化技术、可视化技术、公式发现、多媒体数据挖掘技术及 Web 数据挖掘技术；第 4 篇介绍数据挖掘具体应用实例，重点介绍数据预处理技术、聚类、分类、预测、关联规则分析、时间序列分析等方法。

本书的出版得到了清华大学出版社的大力支持，编辑栾大成为本书内容等许多方面提出了宝贵的意见。书中参考了许多学者的研究成果，在此一并表示衷心的感谢！

由于数据挖掘的内容非常丰富，所涉及的学科也较多，且限于作者学识水平，书中难免存在缺点、错误以及疏漏，敬请读者批评斧正。

本书为读者提供实例中给出的 MATLAB 程序，需要者可登录清华大学出版社网站，注册之后进行下载。读者反馈：xuggsx@sina.com 或者 QQ 号 693131033，作者随时解答读者问题。

许国根

本书习题代码下载



说明：本书习题按照“篇”分类，共三个文件夹，请对应正文的习题编号进行查询。

目 录

第 1 篇 关于数据挖掘

第 1 章 绪论	3
1.1 数据挖掘概述	4
1.2 数据挖掘的分类	6
1.3 数据挖掘的过程	7
1.4 数据挖掘的任务	9
1.5 数据挖掘的对象	11
1.5.1 数据库	11
1.5.2 文本	13
1.5.3 图像与视频数据	13
1.5.4 Web 数据	14
1.6 数据挖掘建模方法	14
1.6.1 业务理解	15
1.6.2 数据理解	16
1.6.3 数据准备	16
1.6.4 建模	17
1.6.5 评估	18
1.6.6 部署	19
1.7 数据挖掘的应用	19
1.7.1 在金融领域中的应用	19
1.7.2 在零售业中的应用	20
1.7.3 在电信业中的应用	21
1.7.4 在管理中的应用	22
1.7.5 在化学研究领域中的应用	22
1.7.6 在材料研究、生产方面的应用	23
1.7.7 在机械故障诊断与监测中的应用	24
1.7.8 在医疗领域中的应用	25

第 2 篇 数据挖掘算法

第 2 章 决策树算法	29
2.1 决策树算法概述	30

2.2	决策树基本算法	30
2.3	ID3 算法	32
2.4	C4.5 算法	34
2.5	CART 算法	35
2.6	决策树的评价标准	36
2.7	决策树的剪枝及优化	37
2.8	基于 MATLAB 的决策树分析	38
第 3 章 神经网络算法		47
3.1	神经网络概述	48
3.2	神经网络的基本模型	48
3.2.1	神经元	48
3.2.2	传递函数	49
3.2.3	神经网络的分类	50
3.3	BP 神经网络	50
3.3.1	BP 神经网络结构	50
3.3.2	BP 神经网络的学习算法	50
3.4	RBF 神经网络	51
3.4.1	RBF 网络结构	51
3.4.2	RBF 神经网络的学习算法	52
3.5	SOM 神经网络	53
3.5.1	SOM 神经网络结构	53
3.5.2	SOM 神经网络学习算法	53
3.6	反馈型神经网络 (Hopfield)	54
3.6.1	Hopfield 网络的拓扑结构	54
3.6.2	Hopfield 网络的学习算法	55
3.7	基于 MATLAB 的神经网络方法	56
3.7.1	信息表达方式	56
3.7.2	网络模型选择	56
3.7.3	网络参数选择	56
3.7.4	学习训练算法选择	56
3.7.5	系统仿真的性能对比	56
第 4 章 进化算法		65
4.1	概述	66
4.2	进化算法的基本原理	67
4.2.1	编码	67
4.2.2	适应度函数	68
4.2.3	遗传算子	69

4.2.4 基因算法的特点	71
4.3 基因算法的主要步骤	71
4.4 基本遗传算法	72
4.4.1 遗传算法的基本流程	72
4.4.2 控制参数选择	73
4.5 进化规划算法	74
4.5.1 变异算子	75
4.5.2 选择算子	75
4.6 进化策略计算	75
4.6.1 进化策略算法的基本流程	76
4.6.2 算法的构成要素	76
4.7 量子遗传算法	79
4.7.1 基本概念	79
4.7.2 量子遗传算法流程	80
4.7.3 量子算法中的控制参数	81
4.8 人工免疫算法	83
4.8.1 人工免疫算法的生物学基础	83
4.8.2 生物免疫基本原理	85
4.8.3 人工免疫算法的基本概念	86
4.8.4 免疫算子	87
4.8.5 免疫算法与免疫系统的对应	89
4.8.6 人工免疫算法与遗传算法的比较	90
4.9 基于 MATLAB 的进化算法	91
第 5 章 统计分析方法	99
5.1 假设检验	100
5.1.1 随机误差的判断	100
5.1.2 系统误差的检验	101
5.2 回归分析	103
5.2.1 一元线性回归分析	103
5.2.2 多元线性回归分析	106
5.2.3 非线性回归分析	108
5.2.4 虚拟及离散变量回归模型	110
5.2.5 异常点、高杠杆点和强影响观测值	110
5.2.6 回归假设检验	111
5.3 二项逻辑 (logistic) 回归	112
5.3.1 二项逻辑回归模型	112
5.3.2 显著性检验	114
5.3.3 回归方程的拟合优度检验	115

5.4	方差分析	115
5.4.1	单因素试验的方差分析	115
5.4.2	双因素试验的方差分析	116
5.5	主成分分析	118
5.5.1	主成分分析的数字模型	119
5.5.2	主成分计算步骤	119
5.5.3	主成分估计	120
5.5.4	主成分筛选	121
5.6	因子分析	121
5.6.1	因子分析的一般数学模型	121
5.6.2	因子模型中公共因子、因子载荷和变量共同度的统计意义	123
5.6.3	因子分析与主成分分析的联系与区别	123
5.6.4	Q型和R型因子分析	124
5.7	基于MATLAB的统计分析方法	124
第6章	贝叶斯网络方法	155
6.1	贝叶斯定理、先验和后验	156
6.2	贝叶斯网络	157
6.3	贝叶斯网络学习	158
6.3.1	贝叶斯网络的结构学习	158
6.3.2	贝叶斯网络的参数学习	158
6.4	主要贝叶斯网络模型	160
6.4.1	朴素贝叶斯网络	160
6.4.2	TAN贝叶斯网络	161
6.4.3	无约束贝叶斯网络	162
6.5	基于MATLAB的贝叶斯网络方法	162
第7章	支持向量机	177
7.1	支持向量机概述	178
7.2	核函数	180
7.3	基于MATLAB的支持向量机方法	182
第8章	关联分析	185
8.1	概述	186
8.1.1	关联规则的主要概念	186
8.1.2	关联规则的种类	187
8.1.3	关联规则的价值衡量的方法	187
8.2	Apriori关联规则算法	188
8.3	基于分类搜索的关联规则算法	189
8.3.1	基于分类搜索的关联规则算法特点	189

8.3.2	算法流程与实现	190
8.3.3	数据更新实现	190
8.4	时序关联规则算法	191
8.5	多值属性关联规则算法	192
8.5.1	静态离散属性关联规则	192
8.5.2	动态离散关联规则	193
8.5.3	基于距离的关联规则	193
8.6	增量关联规则算法	193
8.7	基于关联规则的分类算法	194
8.8	模糊关联分类算法	195
8.8.1	属性的模糊划分	195
8.8.2	模糊关联的定义	195
8.9	关联规则的评价	196
8.9.1	支持度—置信度框架	196
8.9.2	基于主观因素的主观度量	197
8.10	辛普森悖论	197
8.11	基于 MATLAB 的关联规则分析	198
第 9 章	其他数据挖掘方法	201
9.1	近邻法	202
9.2	K-means 聚类	203
9.3	基于 MATLAB 的近邻法及 K-means 聚类法	206

第 3 篇 数据挖掘相关技术

第 10 章	数据仓库	213
10.1	概述	214
10.1.1	数据仓库重要特性	214
10.1.2	数据仓库中几个重要概念	216
10.2	数据仓库设计	218
10.2.1	数据仓库的总体结构	218
10.2.2	数据仓库的基本功能层	219
10.2.3	数据仓库技术	220
10.2.4	数据仓库设计	221
10.2.5	数据仓库设计步骤	221
10.3	数据仓库的开发应用	222
10.3.1	数据仓库概念模型设计与开发	223
10.3.2	数据仓库的逻辑模型设计	226
10.3.3	数据仓库物理模型的设计	230

10.4	数据仓库的技术管理	232
10.5	OLAP 技术	233
10.5.1	基本概念	233
10.5.2	多维分析	234
10.5.3	维的层次关系	235
10.5.4	维的类关系	235
10.5.5	OLAP 与数据仓库的关系	235
10.6	基于 MATLAB 的数据仓库开发技术	237
10.6.1	数据库工具箱	237
10.6.2	可视查询生成器	239
10.6.3	数据的存取类型	247
10.6.4	数据输入和输出	252
第 11 章	模糊集理论	257
11.1	模糊集合	258
11.1.1	隶属度函数	258
11.1.2	模糊集运算	260
11.1.3	λ 截集	260
11.2	模糊关系	261
11.3	模糊聚类	262
11.3.1	数据标准化	263
11.3.2	相似系数和距离	263
11.3.3	模糊聚类分析	266
11.3.4	模糊 K-均值聚类	267
11.4	基于 MATLAB 的模糊集处理技术	267
第 12 章	粗糙集技术	281
12.1	粗糙集理论的基本概念	282
12.1.1	知识表达系统和决策表	282
12.1.2	等价关系	282
12.1.3	等价划分	283
12.1.4	上近似集和下近似集	283
12.1.5	粗糙集	284
12.1.6	粗糙集的非确定性的精确度 $\alpha A(Y)$ 和粗糙度 $\rho A(Y)$	284
12.2	分类规则的形成	284
12.3	知识的约简	285
12.3.1	决策表的一致性	285
12.3.2	属性约简	285
12.3.3	分辨矩阵与分辨函数	286

12.4	模糊集与粗糙集	287
12.5	基于 MATLAB 的粗糙集处理方法	287
第 13 章	目标优化技术	291
13.1	目标优化概述	292
13.2	极值问题	293
13.3	无约束非线性规划	293
13.3.1	梯度下降法	294
13.3.2	共轭梯度法	295
13.3.3	牛顿法	295
13.4	有约束非线性规划	295
13.5	大规模优化问题的分解算法	296
13.5.1	问题的描述	296
13.5.2	目标协调法	297
13.5.3	模型协调法	298
13.5.4	混合协调法	298
13.6	其他优化方法	299
13.7	基于 MATLAB 的目标优化方法	300
第 14 章	可视化技术	307
14.1	可视化技术概述	308
14.2	可视化技术分类	309
14.2.1	数据可视化	309
14.2.2	科学计算可视化	309
14.2.3	信息可视化	309
14.2.4	知识可视化	310
14.3	多维数据可视化	310
14.3.1	平行坐标表示法	311
14.3.2	雷达图	312
14.3.3	树形图	313
14.3.4	三角多项式图	314
14.3.5	散点图	315
14.3.6	星座图	316
14.3.7	基于像素的高维数据的可视化	318
14.3.8	基于非线性变换的图表示优化	318
14.3.9	高维数据降维	319
14.4	图形的特征分析	321
14.4.1	平行坐标下的聚簇分析	321
14.4.2	雷达图的图形特征方法	322

14.4.3	图形特征提取中的特征排序问题	323
14.5	基于多元图的图形分类方法	324
14.5.1	单原型图形分类器	324
14.5.2	基于平行坐标的平行筛可视化分类方法	325
14.5.3	基于平行坐标的贝叶斯可视化分类方法	325
14.6	基于色度学空间的多元图表示	326
14.7	基于 MATLAB 的数据可视化技术	327
第 15 章	公式发现	341
15.1	公式发现概述	342
15.2	公式发现系统中的知识	342
15.2.1	规则一（函数规则）	343
15.2.2	规则二（导数规则）	344
15.2.3	多维函数扩展规则	345
15.2.4	规则三	346
15.3	基于 MATLAB 的公式发现	347
第 16 章	多媒体数据挖掘技术	349
16.1	多媒体数据挖掘技术概述	350
16.1.1	数据类型	350
16.1.2	多媒体数据库管理系统（MM-DBMS）	351
16.2	文本挖掘	352
16.2.1	基于关键字的关联分析	354
16.2.2	文档分类分析	354
16.3	图像挖掘	360
16.4	视频挖掘	361
16.4.1	结构挖掘	361
16.4.2	运动挖掘	361
16.4.3	趋势挖掘	362
16.5	音频挖掘	362
16.6	复合类型数据的挖掘	363
第 17 章	Web 数据挖掘技术	365
17.1	Web 数据挖掘技术概述	366
17.2	Web 内容挖掘	366
17.2.1	爬虫	367
17.2.2	虚拟 Web 视图	367
17.2.3	个性化	368
17.3	Web 结构挖掘	369
17.3.1	PageRank	369

17.3.2	Clever	369
17.4	Web 使用挖掘	369
17.4.1	预处理	370
17.4.2	数据结构	370
17.4.3	模式发现	370
17.4.4	模式发现	371
17.4.5	基于组织协同进化的 Web 日志挖掘算法	371

第 4 篇 数据挖掘应用实战

第 18 章	数据统计特性	377
18.1	数据关系发现	378
18.2	频率和众数	378
18.3	百分位数 (percentile)	378
18.4	中心度量	378
18.5	散布程度度量	379
18.6	数据的分布描述	380
18.7	数据的概率分布	383
第 19 章	数据预处理	385
19.1	数据预处理完毕	386
19.2	数据清理	386
19.2.1	填补缺失数据	386
19.2.2	消除噪声数据	387
19.2.3	实现数据一致性	388
19.3	数据集成与转换	388
19.3.1	数据集成	388
19.3.2	数据转换	389
19.4	数据归约与压缩	390
19.4.1	数据归约	390
19.4.2	数据压缩	395
19.4.3	数值归约	395
19.5	数值数据的概念分层与离散化	396
19.5.1	概念分层	396
19.5.2	概念分层的类型	397
19.5.3	数值数据离散化	398
19.5.4	分类数据的概念分层	399
19.6	例题	399

第 20 章 分类	411
20.1 分类概述	412
20.2 方法	412
20.3 例题	415
第 21 章 预测	421
21.1 回归分析	422
21.1.1 逐步回归	422
21.1.2 岭回归	424
21.1.3 主成分回归分析	425
21.2 时间序列预测模型	425
21.2.1 时间序列的特征量	426
21.2.2 平稳时间序列预测模型	426
21.3 马尔可夫链	429
21.4 灰色系统方法	430
21.4.1 灰色系统的基本概念	430
21.4.2 灰色序列生成算子	431
21.4.3 灰色分析	433
21.5 例题	438
第 22 章 聚类	459
22.1 聚类分析概述	460
22.2 聚类分析中的数据类型	461
22.3 相似性度量	463
22.3.1 属性间的相似性度量	464
22.3.2 对象间的相似性度量	465
22.3.3 相异度矩阵	465
22.4 聚类的特征	468
22.5 聚类准则	469
22.6 划分方法	470
22.7 层次方法	471
22.7.1 利用层次方法的平衡迭代归约及聚类	473
22.7.2 利用代表点聚类	474
22.8 基于密度的方法	474
22.9 基于网格的方法	476
22.10 基于模型的聚类方法	477
22.11 基于目标函数的方法	478
22.11.1 样本与类之间的距离	478
22.11.2 类内距离	479

22.11.3	类与类之间的距离	479
22.12	离群点检测	480
22.12.1	基于统计的离群点检测方法	481
22.12.2	基于距离的离群点检测方法	482
22.12.3	基于相对密度的离群点检测方法	483
22.12.4	基于聚类的离群点检测方法	484
22.12.5	离群点挖掘方法的评估	486
22.13	聚类有效性	487
22.13.1	内部质量评价准则	487
22.13.2	外部质量评价准则	489
22.14	例题	489
第 23 章	时序数据挖掘	505
23.1	基本定义	506
23.2	时序数据挖掘参数	507
23.3	时序关联规则	507
23.3.1	事务间关联规则	508
23.3.2	情节规则	508
23.3.3	序列关联规则	508
23.3.4	日历关联规则	509
23.4	时间序列挖掘	509
23.4.1	时间序列分析	509
23.4.2	趋势分析	509
23.4.3	相似性搜索	511
23.4.4	周期分析	512
23.5	时间序列分段线性表示	512
23.6	时间序列的预测	513
23.7	例题	513
第 24 章	关联规则挖掘	527
24.1	关联规则的类型及挖掘算法	528
24.2	基于组织进化的关联规则挖掘	528
24.2.1	组织的定义	528
24.2.2	组织适应度的计算	529
24.2.3	组织进化算子	529
24.2.4	算法步骤	529
24.3	基于组织层次进化的关联规则挖掘	530
24.3.1	聚合算子	530
24.3.2	进化种群 p_e 和最优种群 p_b	530

24.3.3 算法步骤	530
24.4 多维关联规则挖掘	531
24.4.1 染色体的编码	531
24.4.2 亲和度函数的构造	531
24.4.3 算法步骤	532
24.5 关联规则扩展	532
24.5.1 多层次关联规则	532
24.5.2 多维度关联规则	533
24.5.3 定量关联规则	533
24.5.4 基于约束的关联规则	534
24.6 例题	534
参考文献	548