

《数据挖掘导论》

图书基本信息

书名：《数据挖掘导论》

13位ISBN编号：9787302381046

出版时间：2014-11-1

作者：戴红,常子冠,于宁

页数：207

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

内容概要

本书为数据挖掘入门级教材，共分8章，主要内容分为三个专题：技术、数据和评估。技术专题包括决策树技术、K-means算法、关联分析技术、神经网络技术、回归分析技术、贝叶斯分析、凝聚聚类、概念分层聚类、混合模型聚类技术的EM算法、时间序列分析和基于Web的数据挖掘等常用的机器学习方法和统计技术。数据专题包括数据库中的知识发现处理模型和数据仓库及OLAP技术。评估专题包括利用检验集分类正确率和混淆矩阵，并结合检验集置信区间评估有指导学习模型，使用无指导聚类技术评估有指导模型，利用Lift和假设检验比较两个有指导学习模型，使用MS Excel 2010和经典的假设检验模型评估属性，使用簇质量度量方法和有指导学习技术评估无指导聚类模型。

本书秉承教材风格，强调广度讲解。注重成熟模型和开源工具的使用，以提高学习者的应用能力为目标；注重结合实例和实验，加强基本概念和原理的理解和运用；注重实例的趣味性和生活性，提高学习者学习的积极性。使用章后练习、计算和实验作业巩固和检验所学内容；使用词汇表附录，解释和规范数据挖掘学科专业术语；使用适合教学的简单易用开源的Weka和通用的MS Excel软件工具实施数据挖掘验证和体验数据挖掘的精妙。

本书可作为普通高等院校计算机科学、信息科学、数学和统计学专业的入门教材，也可作为如经济学、管理学、档案学等对数据管理、数据分析与数据挖掘有教学需求的其他相关专业的教材。同时，对数据挖掘技术和方法感兴趣，致力于相关方面的研究和应用的其他读者，也可以从本书中获取基本的指导和体验。

本书配有教学幻灯片、大部分章后习题和实验的参考答案以及课程大纲。

《数据挖掘导论》

作者简介

戴红，女，副教授，研究方向为时空数据建模、数据可视化、计算机教育教学

书籍目录

第1章 认识数据挖掘	1
1.1 数据挖掘的定义	1
1.2 机器学习	2
1.2.1 概念学习	2
1.2.2 归纳学习	3
1.2.3 有指导的学习	4
1.2.4 无指导的聚类	7
1.3 数据查询	8
1.4 专家系统	8
1.5 数据挖掘的过程	9
1.5.1 准备数据	10
1.5.2 挖掘数据	10
1.5.3 解释和评估数据	10
1.5.4 模型应用	11
1.6 数据挖掘的作用	11
1.6.1 分类	11
1.6.2 估计	12
1.6.3 预测	12
1.6.4 无指导聚类	12
1.6.5 关联关系分析	13
1.7 数据挖掘技术	13
1.7.1 神经网络	14
1.7.2 回归分析	14
1.7.3 关联分析	15
1.7.4 聚类技术	16
1.8 数据挖掘的应用	16
1.8.1 应用领域	16
1.8.2 成功案例	18
1.9 Weka数据挖掘软件	19
1.9.1 Weka简介	19
1.9.2 使用Weka建立决策树模型	22
1.9.3 使用Weka进行聚类	25
1.9.4 使用Weka进行关联分析	26
本章小结	27
习题	28
第2章 基本数据挖掘技术	30
2.1 决策树	30
2.1.1 决策树算法的一般过程	31
2.1.2 决策树算法的关键技术	32
2.1.3 决策树规则	40
2.1.4 其他决策树算法	41
2.1.5 决策树小结	41
2.2 关联规则	42
2.2.1 关联规则概述	42
2.2.2 关联分析	43
2.2.3 关联规则小结	46
2.3 聚类分析技术	47

- 2.3.1 K-means算法 48
- 2.3.2 K-means算法小结 51
- 2.4 数据挖掘技术的选择 51
- 本章小结 52
- 习题 53
- 第3章 数据库中的知识发现 55
 - 3.1 知识发现的基本过程 55
 - 3.1.1 KDD过程模型 55
 - 3.1.2 知识发现软件 57
 - 3.1.3 KDD过程的参与者 58
 - 3.2 KDD过程模型的应用 58
 - 3.2.1 步骤1：商业理解 58
 - 3.2.2 步骤2：数据理解 59
 - 3.2.3 步骤3：数据准备 60
 - 3.2.4 步骤4：建模 65
 - 3.2.5 评估 66
 - 3.2.6 部署和采取行动 66
 - 3.3 实验：KDD案例 66
- 本章小结 72
- 习题 73
- 第4章 数据仓库 74
 - 4.1 数据库与数据仓库 74
 - 4.1.1 数据(库)模型 75
 - 4.1.2 规范化与反向规范化 77
 - 4.2 设计数据仓库 79
 - 4.2.1 数据抽取、清洗、变换和加载 79
 - 4.2.2 数据仓库模型 82
 - 4.2.3 数据集市 85
 - 4.2.4 决策支持系统 86
 - 4.3 联机分析处理 87
 - 4.3.1 概述 87
 - 4.3.2 实验：使用OLAP辅助驾驶员行为分析 90
 - 4.4 使用Excel数据透视表和数据透视图分析数据 93
 - 4.4.1 创建简单数据透视表和透视图 93
 - 4.4.2 创建多维透视表和透视图 97
- 本章小结 100
- 习题 100
- 第5章 评估技术 102
 - 5.1 数据挖掘评估概述 102
 - 5.1.1 评估内容 102
 - 5.1.2 评估工具 103
 - 5.2 评估有指导学习模型 108
 - 5.2.1 评估分类类型输出模型 108
 - 5.2.2 评估数值型输出模型 109
 - 5.2.3 计算检验集置信区间 111
 - 5.2.4 无指导聚类技术的评估作用 112
 - 5.3 比较有指导学习模型 112
 - 5.3.1 使用Lift比较模型 112
 - 5.3.2 通过假设检验比较模型 114

5.4 属性评估	115
5.4.1 数值型属性的冗余检查	115
5.4.2 数值属性显著性的假设检验	117
5.5 评估无指导聚类模型	118
本章小结	118
习题	119
第6章 神经网络技术	120
6.1 神经网络概述	120
6.1.1 神经网络模型	120
6.1.2 神经网络的输入和输出数据格式	121
6.1.3 激励函数	123
6.2 神经网络训练	124
6.2.1 反向传播学习	124
6.2.2 自组织映射的无指导聚类	127
6.2.3 实验：应用BP算法建立前馈神经网络	130
6.3 神经网络模型的优势和缺点	138
本章小结	138
习题	139
第7章 统计技术	141
7.1 回归分析	141
7.1.1 线性回归分析	142
7.1.2 非线性回归	149
7.1.3 树回归	151
7.2 贝叶斯分析	152
7.3 聚类技术	156
7.3.1 分层聚类	156
7.3.2 基于模型的聚类	163
7.4 数据挖掘中的统计技术与机器学习技术	165
本章小结	165
习题	167
第8章 时间序列和基于Web的数据挖掘	169
8.1 时间序列分析	169
8.1.1 概述	169
8.1.2 线性回归分析解决时间序列问题	173
8.1.3 神经网络技术解决时间序列问题	175
8.2 基于Web的数据挖掘	176
8.2.1 概述	176
8.2.2 Web文本挖掘	178
8.2.3 Web使用挖掘	179
8.3 多模型分类技术	185
8.3.1 装袋技术	185
8.3.2 推进技术	185
本章小结	186
习题	187
附录A 词汇表	188
附录B 数据挖掘数据集	201
参考文献	

《数据挖掘导论》

精彩短评

1、很实用的数据挖掘入门书，使用简单的工具来体验数据挖掘的精彩。

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:www.tushu000.com