

《数据挖掘十大算法》

图书基本信息

书名：《数据挖掘十大算法》

13位ISBN编号：9787302310617

10位ISBN编号：7302310610

出版时间：2013-5

出版社：清华大学出版社

作者：（美）吴信东（Xindong Wu），（美），库玛尔，（Vipin Kumar）

页数：154

译者：李文波,吴素研

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

《数据挖掘十大算法》

内容概要

《世界著名计算机教材精选：数据挖掘十大算法》详细介绍了在实际中用途最广、影响最大的十种数据挖掘算法，这十种算法是数据挖掘领域的顶级专家进行投票筛选的，覆盖了分类、聚类、统计学习、关联分析和链接分析等重要数据挖掘研究和发展主题。《世界著名计算机教材精选：数据挖掘十大算法》对每一种算法都进行了多个角度的深入剖析，包括算法历史、算法过程、算法特性、软件实现、前沿发展等，此外，在每章最后还给出了丰富的习题和精挑细选的参考文献，对于读者掌握算法基本知识和进一步研究都非常有价值，对数据挖掘、机器学习和人工智能等学科的课程的设计有指导意义。

作者简介

吴信东 (Xindong Wu)，教授英国爱丁堡大学人工智能学博士，任美国佛蒙特大学计算机科学系主任。吴教授在数据挖掘、知识系统和Web信息开发等研究领域内颇有建树，在IEEE TKDE、TPAMI、ACMTOIS、DMKD、KAIS、IJCAI、AAAI、ICMI、KDD、ICDM和WWW等学术会议和期刊上发表了170余篇学术论文，另外，还出版了18部学术专著和会议文集。他还获得了IEEE ICTAI-2005的最佳论文奖和IEEE ICDM-2007的最佳理论 / 算法论文奖亚军。

吴博士是IEEE Transactions on Knowledge and Data Engineering (TKDE，由IEEE Computer Society主办)的主编，IEEE International Conference on Data Mining (ICDM)的创始人和指导委员会主席，Knowledge and Information Systems (KAIS，由Springer发行)的创办人和荣誉主编，IEEE Computer Society Technical Committee on Intelligent Informatics (TCII)的创始主席(2002-2006)，Springer Advanced Information and Knowledge Processing (AI&KP)系列著作的编辑。他还是ICDM'03 (the 2003 IEEE International Conference on Data Mining)程序委员会主席和KDD-07 (the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining)程序委员会联合主席。他获得了2004 ACM SIGKDD服务奖、2006 IEEE ICDM杰出服务奖，是2005年合肥科技大学“长江学者奖励计划”讲座教授。他还是很多学术会议的特邀专家 / 专题报告人，如NSF-NGDM'07、PAKDD-07、IEEE EDOC'06、IEEE ICTAI'04、IEEE/WICACM WI'04/IAT'04、SEKE 2002和PADD-97等。

Vipin Kumar，教授，明尼苏达大学计算机科学与工程系William Norris讲席教授、系主任。他于1977年获得印度鲁尔基理工学院(正式名称是鲁尔基大学)的电子和通信工程学士学位，1979年获得荷兰埃因霍温飞利浦国际学院的电子工程硕士学位，1982年获得马里兰大学帕克分校的计算机科学博士学位。Kumar教授的研究兴趣主要集中在数据挖掘、生物信息学和高性能计算领域。他提出了评估并行算法可扩展性的恒等效率度量指标，并研发了多款稀疏矩阵分解(PSPASES)和图剖分(METIS, ParMctis, hMetis)的高效并行算法及软件。他发表了200多篇研究论文，合编合著了9本学术专著，包括被广泛使用的教科书Introduction to Parallel Computing和Introduction to Data Mining，前者由Addison-Wesley出版。Kumar是众多数据挖掘和多并行计算领域的学术会议、专题研讨会的主席或共同主席，如IEEE International Conference on Data Mining (2002)、International Parallel and Distributed Processing Symposium (2001)和SIAM International Conference on Data Mining (2001)。Kumar是SIAM International Conference on Data Mining指导委员会共同主席，IEEE International Conference on Data Mining和IEEE International Conference on Bioinformatics and Biomedicine指导委员会委员。Kumar是Journal of Statistical Analysis and Data Mining的创始主编之一，IEEE Intelligent Informatics Bulletin主编和Data Mining and Knowledge Discovery系列图书(由CRC Press / Chapman Hall出版)的编辑。Kumar还担任很多其他学术刊物的编辑，如Data Mining and Knowledge Discovery、Knowledge and Information Systems、IEEE Computational Intelligence Bulletin、Annual Review of Intelligent Informatics、Parallel Computing、Journal of Parallel and Distributed Computing、IEEE Transactions of Data and Knowledge Engineering (1993-1997)、IEEE Concurrency (1997-2000)和IEEE Parallel and Distributed Technology (1995-1997)等。他是ACM会士、IEEE会士、AAAS会士和SIAM会员。Kumar由于在并行算法设计、图剖分和数据挖掘领域的杰出贡献，获得了2005 IEEE Computer Society的技术成就奖。

书籍目录

第1章C4.5 1

1.1引言2

1.2算法描述3

1.3算法特性6

1.3.1决策树剪枝6

1.3.2连续型属性8

1.3.3缺失值处理8

1.3.4规则集诱导9

1.4软件实现10

1.5示例10

1.5.1 Golf数据集10

1.5.2 Soybean数据集11

1.6高级主题11

1.6.1二级存储12

1.6.2斜决策树12

1.6.3特征选择12

1.6.4集成方法12

1.6.5分类规则13

1.6.6模型重述13

1.7习题14

参考文献15

第2章k-means18

2.1引言19

2.2算法描述19

2.3可用软件22

2.4示例23

2.5高级主题27

2.6小结28

2.7习题28

参考文献29

第3章SVM：支持向量机31

3.1支持向量分类器32

3.2支持向量分类器的软间隔优化34

3.3核技巧35

3.4理论基础38

3.5支持向量回归器40

3.6软件实现41

3.7当前和未来的研究41

3.7.1计算效率41

3.7.2核的选择41

3.7.3泛化分析42

3.7.4结构化支持向量机的学习42

3.8习题43

参考文献44

第4章Apriori47

4.1引言48

4.2算法描述48

- 4.2.1挖掘频繁模式和关联规则48
- 4.2.2挖掘序列模式52
- 4.2.3讨论53
- 4.3软件实现54
- 4.4示例55
 - 4.4.1可行示例55
 - 4.4.2性能评估60
- 4.5高级主题61
 - 4.5.1改进Apriori类型的频繁模式挖掘61
 - 4.5.2无候选的频繁模式挖掘62
 - 4.5.3增量式方法63
 - 4.5.4稠密表示: 闭合模式和最大模式63
 - 4.5.5量化的关联规则64
 - 4.5.6其他的重要性/兴趣度度量方法65
 - 4.5.7类别关联规则66
 - 4.5.8使用更丰富的形式: 序列、树和图66
- 4.6小结67
- 4.7习题67
- 参考文献68
- 第5章EM72
 - 5.1引言73
 - 5.2算法描述74
 - 5.3软件实现74
 - 5.4示例75
 - 5.4.1例5.1: 多元正态混合75
 - 5.4.2例5.2: 混合因子分析78
 - 5.5高级主题80
 - 5.6习题81
 - 参考文献87
- 第6章PageRank90
 - 6.1引言91
 - 6.2算法描述92
 - 6.3一个扩展: Timed-PageRank95
 - 6.4小结96
 - 6.5习题96
 - 参考文献97
- 第7章AdaBoost98
 - 7.1引言99
 - 7.2算法描述99
 - 7.2.1符号定义99
 - 7.2.2通用推举过程100
 - 7.2.3AdaBoost算法101
 - 7.3示例103
 - 7.3.1异或问题求解103
 - 7.3.2真实数据上的性能104
 - 7.4实际应用105
 - 7.5高级主题107
 - 7.5.1理论问题107
 - 7.5.2多类别AdaBoost110

7.5.3其他高级主题	111
7.6软件实现	111
7.7习题	112
参考文献	113
第8章kNN：k-最近邻	115
8.1引言	116
8.2算法描述	116
8.2.1宏观描述	116
8.2.2若干议题	117
8.2.3软件实现	118
8.3示例	118
8.4高级主题	120
8.5习题	121
致谢	121
参考文献	122
第9章Naive Bayes	124
9.1引言	125
9.2算法描述	125
9.3独立给力	127
9.4模型扩展	128
9.5软件实现	130
9.6示例	130
9.6.1例1	130
9.6.2例2	132
9.7高级主题	133
9.8习题	133
参考文献	134
第10章CART：分类和回归树	136
10.1前身	137
10.2概述	138
10.3示例	138
10.4算法描述	140
10.5分裂准则	141
10.6先验概率和类别均衡	142
10.7缺失值的处理	144
10.8属性的重要度	145
10.9动态特征构造	146
10.10代价敏感学习	147
10.11停止准则、剪枝、树序列和树选择	147
10.12概率树	149
10.13理论基础	150
10.14 CART之后的相关研究	150
10.15可用软件	151
10.16习题	152
参考文献	153

《数据挖掘十大算法》

编辑推荐

吴信东和库玛尔编著的《数据挖掘十大算法》详细介绍了在实际中用途最广、影响最大的十种数据挖掘算法，这十种算法是数据挖掘领域的顶级专家进行投票筛选的，覆盖了分类、聚类、统计学习、关联分析和链接分析等重要数据挖掘研究和发展主题。在每章最后还给出了丰富的习题和精挑细选的参考文献，对于读者掌握算法基本知识和进一步研究都非常有价值，对数据挖掘、机器学习和人工智能等学科的课程的设计有指导意义。

《数据挖掘十大算法》

精彩短评

- 1、快速过了下，先了解下概念性的东西。
- 2、部分章节写的很难懂
- 3、当手册用，点评了数据挖掘的十大算法，给出了n多经典参考
- 4、与其说是点技能不如说是重新种了颗新树。。。
- 5、如果不懂算法直接看英文吃不消，建议看数据挖掘导论
- 6、讲解不够详细。
- 7、对于了解常用的算法，还是有帮助的。讲解比较详细，基本上都有算法运行过程的分析，对于SVM算法的说明比较透彻，学习AdaBoost算法也是通过这本书。
- 8、2016.04.06 ~ 苏州的天气呦！这书感觉更像是文献综述，适合什么人看呢 ~ 估计不适合我。看了四个算法了，搞完去搞周志华老师的《机器学习》吧
- 9、看一遍忘一遍.....

《数据挖掘十大算法》

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