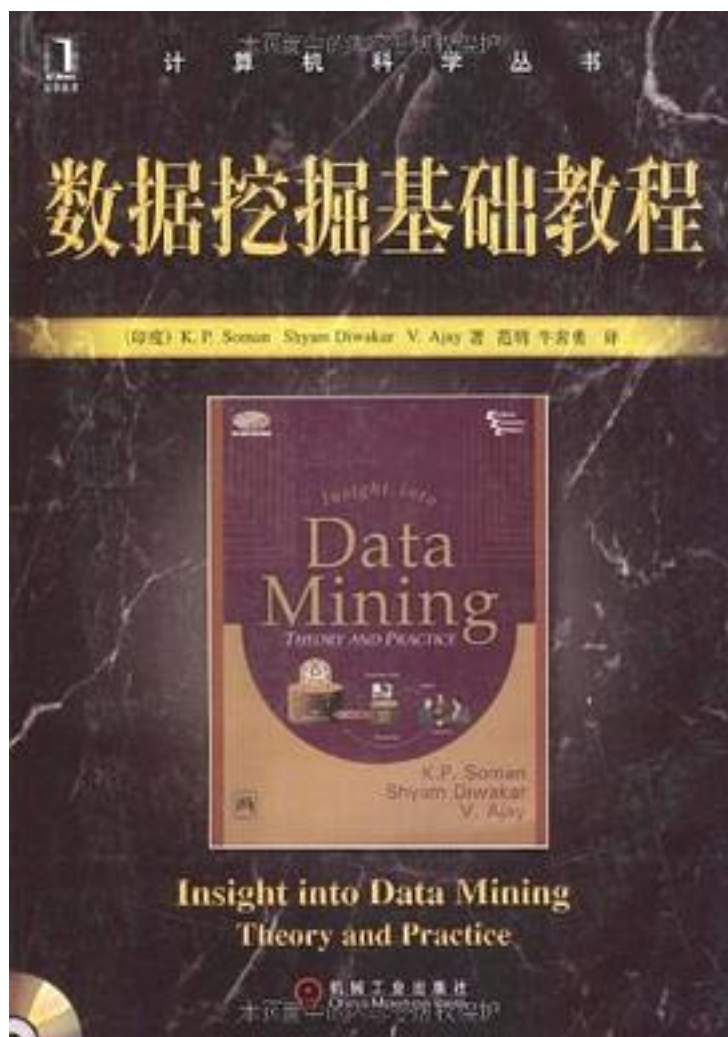


数据挖掘基础教程



[数据挖掘基础教程_下载链接1](#)

著者: (印度) 西蒙 (Soman.K.P)

出版者: 机械工业出版社

出版时间: 2009-1

装帧: 平装

isbn: 9787111255437

《数据挖掘基础教程》全面介绍数据挖掘的原理、方法和算法。主要内容包括数据挖掘

的基本概念、数据挖掘算法的数据类型、输入和输出、决策树、数据挖掘的预处理和后处理、关联规则挖掘、分类和回归算法、支持向量机、聚类分析及多维数据可视化。

《数据挖掘基础教程》讲解深入浅出，并辅以大量实例，随书光盘提供了大量数据集以及两种广泛使用的数据挖掘软件——Weka和ExcelMiner，便于读者理解数据挖掘知识。

《数据挖掘基础教程》适合作为高等院校计算机及相关专业数据挖掘课程的教材，也可供广大技术人员参考。

作者介绍:

目录: 出版者的话

译者序

前言

第1章 数据挖掘

1.1 引言

1.1.1 数据挖掘与知识发现

1.1.2 数据挖掘与数据分析

1.1.3 数据挖掘与统计学

1.1.4 数据挖掘与机器学习

1.2 数据挖掘——成功的例子

1.3 数据挖掘研究发展的主要原因

1.4 当前研究成果

1.5 图形模型和层次概率表示

1.6 新的应用

1.7 影响数据挖掘的趋势

1.8 研究挑战

1.9 实验平台和基础设施

参考文献

第2章 从商务角度看数据挖掘

2.1 引言

2.2 从数据挖掘工具到解决方案

2.3 数据挖掘系统的演变

2.4 知识发现过程

2.5 数据挖掘支撑技术概述

2.5.1 数据挖掘：验证与发现

2.5.2 决策支持系统

2.5.3 OLAP

2.5.4 桌面DSS

2.5.5 数据仓库

2.5.6 数据挖掘过程

2.6 数据挖掘技术

参考文献

第3章 数据挖掘算法的数据类型、输入和输出

3.1 引言

3.2 实例和特征

3.3 特征(数据)的不同类型

3.4 概念学习与概念描述

3.5 数据挖掘的输出——知识表示

3.5.1 分类学习算法的知识输出

3.5.2 聚类学习算法的输出

3.5.3 关联规则的输出

3.5.4 用于数值预测的树的输出

3.5.5 基于实例的学习和知识表示

参考文献

第4章 决策树——分类和回归树

4.1 引言

4.2 构造分类树

4.2.1 用于标称属性的ID3算法

4.2.2 信息论和信息熵

4.2.3 构造树

4.2.4 高分支属性

4.2.5 从ID3到C4.5

4.2.6 形象化地理解ID3和C4.5算法

4.3 CHAID

4.3.1 CHAID的数学工具

4.3.2 CHAID变量的类型

4.3.3 CHAID算法

4.3.4 CHAID算法描述

4.3.5 将CHAID用于气象数据

4.3.6 单调变量的预测子级别合并

4.4 CART(分类和回归树)

4.4.1 CART使用的不纯度度量

4.4.2 Gini指数

4.4.3 使用Gini指数——一个例子

4.4.4 双化指数

4.4.5 有序双化

4.4.6 CAR了分析的步骤

4.5 回归树

4.5.1 回归树的一个例子

4.5.2 基于树的回归

4.5.3 最小二乘方回归树

4.5.4 Ls回归树的有效生长

4.5.5 连续变量上的划分

4.5.6 离散变量上的划分

4.5.7 模型树

4.6 具有未知类值数据的类预测的一般问题

4.7 剪枝导论

4.8 模型评估

4.8.1 交叉确认：保持方法

4.8.2 模型比较

4.8.3 代价敏感的学习

习题

参考文献

第5章 数据挖掘的预处理和后处理

5.1 引言

5.2 数据预处理的步骤

5.3 离散化

5.3.1 人工方法

5.3.2 分箱

5.3.3 基于熵的离散化

5.3.4 找出分割点的其他简单方法

5.4 特征提取、选择和构造

5.4.1 特征提取

5.4.2 特征选择

5.4.3 特征构造

5.5 缺失数据及其处理方法和技术

- 5.5.1 什么是缺失数据
- 5.5.2 缺失数据的主要原因
- 5.5.3 缺失数据的机制
- 5.5.4 缺失数据的机制——一个例子
- 5.6 在决策树归纳中处理缺失数据的例子
- 5.7 后处理

参考文献

第6章 数据集

- 6.1 引言
- 6.2 隐形眼镜
- 6.3 鸢尾属植物数据库
- 6.4 乳腺癌数据库
- 6.5 工资数据库
- 6.6 信用卡数据库
- 6.7 住宅数据库
- 6.8 1985年汽车进口数据库
- 6.9 徽章问题
 - 6.9.1 问题描述
 - 6.9.2 部分数据

第7章 关联规则挖掘

- 7.1 引言
- 7.2 事务数据库中关联规则的自动发现
- 7.3 Apriori算法
- 7.4 缺点

习题

参考文献

第8章 用开源和商业软件进行机器学习

8.1 用Weka进行机器学习

- 8.1.1 开始
- 8.1.2 装入数据
- 8.1.3 选择或过滤属性
- 8.1.4 离散化
- 8.1.5 关联规则挖掘
- 8.1.6 分类
- 8.1.7 聚类

8.2 XLMINER

参考文献

第9章 分类和回归算法

- 9.1 引言
- 9.2 朴素贝叶斯
 - 9.2.1 朴素贝叶斯的零频率问题
 - 9.2.2 缺失值和数值属性
- 9.3 多元回归分析
 - 9.3.1 什么是回归分析
 - 9.3.2 简单和多元回归分析
 - 9.3.3 在市场营销中的应用
 - 9.3.4 方法
 - 9.3.5 使用Excel进行多元回归分析
 - 9.3.6 输入数据
 - 9.3.7 回归输出
- 9.4 逻辑斯谛回归
- 9.5 k-最近邻分类
 - 9.5.1 k-近邻预测
 - 9.5.2 k-NN算法的缺点

9.6 GMDH

9.6.1 引言

9.6.2 数据处理群组方法的背景

9.6.3 构建决策规则

9.6.4 实验结果

9.6.5 讨论和总结

9.7 进化计算和遗传算法

9.7.1 进化理论

9.7.2 遗传算法

9.7.3 使用遗传算法进行机器学习

习题

参考文献

第10章 支持向量机

10.1 引言

10.2 线性支持向量机的基本思想

10.3 软边缘SVM：线性核

10.3.1 线性SVM的线性规划公式表示

10.3.2 有训练误差的SVM：非线性核

10.4 邻近支持向量机

10.5 生成数据集

10.5.1 螺旋数据生成器

10.5.2 棋盘格数据集

10.5.3 多元正态分布数据生成器

10.6 问题及解答

习题

参考文献

第11章 聚类分析

11.1 引言

11.1.1 相似性及其度量

11.1.2 聚类的基本类型

11.2 划分聚类

11.3 k-中心点

11.4 现代聚类方法

11.5 BIRCH

11.6 DBSCAN

11.6.1 DBSCAN算法的概念

11.6.2 DBSCAN的基本概念和算法

11.6.3 算法

11.6.4 DBSCAN算法的优点

11.7 OPTICS

11.7.1 引言

11.7.2 OPTICS算法的动机

11.7.3 OPTICS采用的概念

11.7.4 OPTICS算法

11.7.5 可达图

11.7.6 优点

11.7.7 缺点

11.8 基于图划分的聚类

11.8.1 加权图划分

11.8.2 平衡图划分——基本原则

11.8.3 k 路划分

11.9 CHAMELEON：两阶段聚类算法

11.9.1 数据建模

11.9.2 簇相似性建模

- 11.9.3 CHAMELEON的两个阶段
- 11.9.4 用例子说明CHAMELEON算法
- 11.10 COBWEB概念聚类算法
 - 11.10.1 COBWEB算法
 - 11.10.2 COBWEB：一个简单例子
- 11.11 GCLUTO：图形化聚类工具箱
 - 11.11.1 概述
 - 11.11.2 GCLUTO中的可用选项
 - 11.11.3 使用GCLUTO进行文本挖掘
- 习题
- 参考文献
- 第12章 多维数据可视化
 - 12.1 引言
 - 12.2 多维可视化的图表表示
 - 12.2.1 kiviati图
 - 12.2.2 平行坐标系
 - 12.2.3 3D散点图
 - 12.2.4 3D曲线图
 - 12.2.5 体积透视图
 - 12.2.6 房图
 - 12.2.7 Chemoff脸图
 - 12.3 可视化数据挖掘
- 参考文献
- 附录A SVM公式：完全可分的线性分类器
- 附录B 图划分的矩阵形式
 - • • • • (收起)

[数据挖掘基础教程 下载链接1](#)

标签

数据挖掘

数据分析

计算机

翻了一遍

数学与算法

思维逻辑

入门的

计算机科学

评论

讲得最好的章节是决策树那章和聚类那章

统计学 机器学习 人工智能

丰富的例子，很适合初学者，对许多理论举例说明的很清楚~~

见过。

不错， 我居然看懂了

==我会说这个是教科书么。然后学得很纠结

简单明确，彻底打开了大门。

非常适合作为入门第一本书，算法描述非常清晰！

非常基础的教程

覆盖面还是可以的，特别对于决策树，分类，聚类，最大的优点是有很多详细的例子辅助，缺点是内容稍微有些过时，有些章节没什么用处，比如weka数据集（为什么要专门花一章来罗列这玩意），weka那一章也是浪费纸张，最后一张也没讲什么内容，整体而言还是适合入门的

1、本书更适合本科生教材。

基础扫盲

翻译较差，面有点广

[数据挖掘基础教程_下载链接1](#)

书评

与传统的数据分析相比，数据挖掘更强调模拟人的学习方式，教计算机自动地识别有用的知识。

当然算法还是要人去设计出来，机器只是去执行人的算法，比如神经网络算法、聚类算法等。

目前数据挖掘已经有一些成熟的产品推出，对于个人学习者，可以通过开源的weka来练习数据挖掘...

[数据挖掘基础教程_下载链接1](#)