

<<数据挖掘基础教程>>

图书基本信息

书名：<<数据挖掘基础教程>>

13位ISBN编号：9787111255437

10位ISBN编号：7111255437

出版时间：2009-1

出版时间：机械工业出版社

作者：（印度）西蒙（Soman.K.P）

页数：305

译者：范明,牛常勇

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>

## <<数据挖掘基础教程>>

### 前言

大规模数据集快速增长的今天，数据挖掘应成为一个新的学科。

我们生活在这样一个世界，即便是简单的日常任务，如打电话、使用信用卡或购买五金电器和杂货，都会留下电子印记。

诸如核物理和天体物理领域科学实验数量的增加导致每月可能产生几PB (petabytes)规模的数据。

近来，生命科学正在成为数据驱动的科学。

广泛用于商业和上述科学领域中的自动数据收集设备每小时能够产生几TB(terabytes)规模的数据，致使已有的推理方法过时。

世界上最大的数据仓库——Walmart系统包含500 TB数据。

它实在太大了，以至于无法装入任何计算机的内存。

数据挖掘技术的产生源自进行数据处理业务的企业和进行数据处理研究的科学家需要找到有效的模式来自动处理海量数据。

模式可以是简单的数据汇总、数据划分或数据内部的依赖模型。

## <<数据挖掘基础教程>>

### 内容概要

本书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。

本书讲解深入浅出，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。

本书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。

<<数据挖掘基础教程>>

作者简介

作者：(印度)西蒙 (K.P.Soman) 译者：范明 牛常勇

## &lt;&lt;数据挖掘基础教程&gt;&gt;

## 书籍目录

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) .....第5章 数据挖掘的预处理和后处理第6章 数据集第7章 关联规则挖掘第8章 用开源和商业软件进行机器学习第9章 分类和回归算法第10章 支持向量机第11章 聚类分析第12章 多维数据可视化参考文献附录A SVM公式：安全可分的线性分类器附录B 图划分的矩阵形式

章节摘录

第1章 数据挖掘1.1 引言 计算机科学家经常提到摩尔定律：计算机的处理速度大约每18个月翻一番。但是很少有人知道计算机的存储容量大约每9个月翻一番。

(Goebel和Gruenwald 1999)。

像理想气体一样，计算机的数据库迅速膨胀，占满了可用的存储空间，导致数据库中的大量数据成为未开发利用的资源。

这些数据就像一个金矿，可以从中提取信息。

然后，利用数据挖掘技术，可以将这些信息转换成有价值的知识。

很难说清楚有多少存储在全世界公司、学校、政府部门和其他机构的大型数据库中未使用的海量数据以及其当前增长率。

据估计，美国国会图书馆存储的信息量高达3PB(Lesk1997)。

Lesk估计，全世界每年大约产生160TB信息。

而且，他估计已售出的磁盘空间将超过十万TB。

很快，计算机的数据存储容量将超过人们使用该数据存储和使用其中数据的能力。

将海量数据转换为知识的过程将变得价值无限。

为此，在过去的10~15年中，一种称作数据库中知识发现(KDD)的过程逐步发展完善。

数据挖掘算法就包含在KDD过程中。

典型的数据库用户使用一种界面通过诸如SQL这样的标准技术从数据库中检索数据。

数据挖掘系统将这一过程向前推进一步，支持用户从数据中发现新的知识(Adriaans和Zantinge 1996)。

按照计算机科学家的观点，数据挖掘是一个多学科交叉领域。

诸如神经网络、遗传算法、回归、统计分析、机器学习和聚类分析等数据处理技术经常出现在数据挖掘文献中。

许多研究者认为数据挖掘还不是一个完善的学科，数据可扩展性、与数据库系统的兼容性，以及可用性和准确性都有待改进。

## <<数据挖掘基础教程>>

### 编辑推荐

《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。

#### 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>