

<<知识发现>>

图书基本信息

书名：<<知识发现>>

13位ISBN编号：9787302239574

10位ISBN编号：7302239576

出版时间：2011-1

出版时间：清华大学出版社

作者：史忠植

页数：491

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>

<<知识发现>>

前言

第一台电子计算机诞生于20世纪40年代。

到目前为止，计算机的发展已远远超出了其创始者的想象。

计算机的处理能力越来越强，应用面越来越广，应用领域也从单纯的科学计算渗透到社会生活的方方面面：从工业、国防、医疗、教育、娱乐直至人们的日常生活，计算机的影响可谓无处不在。

计算机之所以能取得上述地位并成为全球最具活力的产业，原因在于其高速的计算能力、庞大的存储能力以及友好、灵活的用户界面。

而这些新技术及其应用有赖于研究人员多年不懈的努力。

学术研究是应用研究的基础，也是技术发展的动力。

自1992年起，清华大学出版社与广西科学技术出版社为促进我国计算机科学与技术产业的发展，推动计算机科技著作的出版，设立了“计算机学术著作出版基金”，并将资助出版的著作列为中国计算机学会的学术著作丛书。

时至今日，本套丛书已出版学术专著近50种，产生了很好的社会影响，有的专著具有很高的学术水平，有的则奠定了一类学术研究的基础。

中国计算机学会一直将学术著作的出版作为学会的一项主要工作。

<<知识发现>>

内容概要

知识发现是从数据集中识别出有效的、新颖的、潜在有用的，以及最终可理解的模式的非平凡过程。知识发现将信息变为知识，从数据资源中发现知识宝藏，将为知识创新和知识经济的发展作出贡献。

本书全面而又系统地介绍了知识发现的方法和技术，反映了当前知识发现研究的最新成果和进展。

全书共分15章。

第1章是绪论，概述知识发现的重要概念和发展过程。

下面三章重点讨论分类问题，包括决策树、支持向量机和迁移学习。

第5章阐述聚类分析。

第6章是关联规则。

第7章讨论粗糙集和粒度计算。

第8章介绍神经网络，书中着重介绍几种实用的算法。

第9章探讨贝叶斯网络。

第10章讨论隐马尔可夫模型。

第11章探讨图挖掘。

第12章讨论进化计算和遗传算法。

第13章探讨分布式知识发现，它使海量数据挖掘成为可能。

最后两章以web知识发现、认知神经科学为例，介绍知识发现的应用。

本书内容新颖，认真总结了作者的科研成果，取材国内外最新资料，反映了当前该领域的研究水平。

论述力求概念清晰，表达准确，算法丰富，突出理论联系实际，富有启发性。

本书可以用作高等院校有关专业的研究生和高年级本科生的知识发现、数据挖掘、机器学习等课程教材，也可供从事知识发现、数据挖掘、机器学习、智能信息处理、模式识别、智能控制研究和知识管理的科技人员阅读参考。

<<知识发现>>

书籍目录

第1章 绪论	1.1 知识	1.2 知识发现的过程	1.3 知识发现的任务	1.4 知识发现的方法
	1.4.1 统计方法	1.4.2 机器学习	1.4.3 神经计算	1.4.4 可视化
	1.5 知识发现的对象	1.5.1 数据库	1.5.2 文本	1.5.3 Web信息
	1.5.4 空间数据	1.5.5 图像和视频数据	1.6 知识发现系统	第2章 决策树
	2.1 归纳学习	2.2 决策树学习	2.3 CLS学习算法	2.4 ID3学习算法
	2.4.1 信息论简介	2.4.2 信息论在决策树学习中的意义及应用	2.4.3 ID3算法	2.4.4 ID3算法应用举例
	2.4.5 C4.5算法	2.5 决策树的改进算法	2.5.1 二叉树判定算法	2.5.2 按信息比值进行估计的方法
	2.5.3 按分类信息估值	2.5.4 按划分距离估值的方法	2.6 决策树的评价	2.7 简化决策树
	2.7.1 简化决策树的动机	2.7.2 决策树过大的原因	2.7.3 控制树的大小	2.7.4 修改测试属性空间
	2.7.5 改进测试属性选择方法	2.7.6 对数据进行限制	2.7.7 改变数据结构	2.8 连续性属性离散化
	2.9 基于偏置变换的决策树学习算法	BSDT	2.9.1 偏置的形式化	2.9.2 表示偏置变换
	2.9.3 算法描述	2.9.4 过程偏置变换	2.9.5 基于偏置变换的决策树学习算法	BSDT
	2.9.6 经典案例库维护算法	TCBM	2.9.7 偏置特征抽取算法	2.9.8 改进的决策树生成算法
	GSD	2.9.9 实验结果	2.10 单变量决策树的并行处理	2.10.1 并行决策树算法
	第3章 支持向量机	第4章 迁移学习	第5章 聚类分析
	第6章 关联规则	第7章 粗糙集	第8章 神经网络	第9章 贝叶斯网络
	第10章 隐马尔可夫模型	第11章 图挖掘	第12章 进化计算	第13章 分布式知识发现
	第14章 Web知识发现	第15章 认知神经科学知识发现	参考文献	

<<知识发现>>

章节摘录

插图：不同的分类器有不同的特点。

有三种分类器评价或比较尺度：预测准确度；计算复杂度；模型描述的简洁度。

预测准确度是用得最多的一种比较尺度，特别是对于预测型分类任务，目前公认的方法是10趟分层交叉验证法。

计算复杂度依赖于具体的实现细节和硬件环境，在数据挖掘中，由于操作对象是巨量的数据库，因此空间和时间的复杂度问题将是一个非常重要的环节。

对于描述型的分类任务，模型描述越简洁越受欢迎。

例如，采用规则表示的分类器构造法就更有用，而神经网络方法产生的结果就难以理解。

另外要注意的是，分类的效果一般和数据的特点有关，有的数据噪声大，有的有缺值，有的分布稀疏，有的字段或属性间相关性强，有的属性是离散的而有的是连续值或混合式的。

目前普遍认为不存在某种方法能适合于所有不同特点的数据。

4. 聚类根据数据的不同特征，将其划分为不同的数据类。

它的目的是使得属于同一类别的个体之间的距离尽可能小，而不同类别上的个体间的距离尽可能大。

聚类方法包括统计方法、机器学习方法、神经网络方法和面向数据库的方法。

在统计方法中，聚类亦称聚类分析，它是多元数据分析的三大方法之一（其他两种是回归分析和判别分析）。

它主要研究基于几何距离的聚类，如欧氏距离、明考斯基距离等。

传统的统计聚类分析方法包括系统聚类法、分解法、加入法、动态聚类法、有序样品聚类、有重叠聚类和模糊聚类等。

这种聚类方法是一种基于全局比较的聚类，它需要考察所有的个体才能决定类的划分。

因此它要求所有的数据必须预先给定，而不能动态增加新的数据对象。

聚类分析方法不具有线性的计算复杂度，难以适用于数据库非常大的情况。

在机器学习中，聚类称为无监督或无教师归纳。

因为和分类学习相比，分类学习的例子或数据对象有类别标记，而要聚类的例子则没有标记，需要由聚类学习算法来自动确定。

在很多人工智能文献中，聚类也称概念聚类，因为这里的距离不再是统计方法中的几何距离，而是根据概念的描述来确定的。

当聚类对象可以动态增加时，概念聚类则称为概念生成。

<<知识发现>>

编辑推荐

《知识发现(第2版)》：中国计算机学会学术著作丛书

<<知识发现>>

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>