

图书基本信息

书名：<<高维聚类知识发现关键技术研究及应用>>

13位ISBN编号：9787121082481

10位ISBN编号：7121082489

出版时间：2009-3

出版时间：电子工业出版社

作者：陈建斌

页数：217

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>

## 前言

计算机技术的发展总是伴随着企业计算技术的应用，企业需求又进一步刺激了计算机等信息科学技术的高速发展，数据库技术在企业业务中发挥了重要作用。

信息系统产生的数据量日益增多，企业应用中的业务数据在发挥了当期效用后，作为历史资料大量积存起来。

这些历史数据中是否蕴藏着丰富的知识呢？

这些知识能否被利用来为企业决策服务呢？

这个问题在企业界和学术界受到越来越多的关注。

随着知识经济的发展，社会经济生活发生了翻天覆地的变化，企业在日常决策中感受到的压力越来越大。

这种压力来自于：（1）决策速度的压力。

面对竞争激烈的市场，企业需要快速决策，需要对市场快速作出反应；（2）决策成本的压力。

快速发展中的企业，决策不容失误，否则后果严重；（3）决策难度的压力。

面对纷繁复杂的市场信息，需要处理的信息繁多，需要考虑的因素众多，决策难度显著增加。

企业决策者面临前所未有的压力。

自从信息技术在企业实践中得到广泛应用以来，人们一直在寻找决策支持系统的良好解决方案，希望信息技术的优势能在决策领域发挥重要作用。

这种决策辅助在今天显得尤为迫切。

在决策支持领域，传统的方法是由专家或信息技术人员总结并描述知识或规则，从外部输入系统，形成知识库、模型库和方法库，结合数据库进行决策辅助的活动。

然而，由于知识太多而且具有复杂性和模糊性，难以准确描述和有效利用，模型又难以精确地表述具体的应用，导致许多决策支持系统失败。

数据库知识发现技术（Knowledge Discovery in Database, KDD）的出现扭转了这种局面。

KDD是从一个系统内部自动获取知识，从大量数据中寻找蕴涵其中但尚未被发现的知识，这种数据应用技术的出现，必然会更有力度地支持企业的战略决策。

数据库知识发现是一个很新的研究领域，由于其具有为企业创造巨大经济效益的潜力，因而备受国内外研究人员关注。

聚类分析是数据知识发现的关键技术之一，高维数据又是当前的重要应用领域之一，研究高维数据的聚类知识发现，具有重要的理论意义和现实意义。

本书正是基于这种考虑，紧紧围绕高维聚类的关键技术进行研究，提出了一整套高维聚类方法，其中包括高维数据相似性的定义、高维聚类算法（包括硬聚类和软聚类）、离群点检测和高维聚类结果表示方法等，为聚类知识发现提供了一种新的思路。

全书共分为9章，内容包括：知识发现与KDD、聚类知识发现及其关键技术、高维数据相似性的定义、基于粗图模型的聚类算法研究、高维二元数据的映射聚类算法、基于蚂蚁行为的聚类算法、高维数据空间的离群点检测方法、高维数据聚类结果的表示和聚类知识发现数据建模及应用等。

本书适用于数据库知识发现领域的研究和应用人员，也可作为相关领域博士生、硕士生和高年级本科生的参考书。

在本书写作过程中，获得了北京联合大学商务学院“供应链知识管理与创新”科研基地建设项目资助，在此表示感谢。

陈云飞对本书第6章有重要贡献，在此表示感谢。

本书在写作中参阅了大量文献，在此向这些文献的作者们表示感谢！

## 内容概要

知识发现是从数据集中抽取和精化新的模式的过程，基于数据库的知识发现（Knowledge Discovery in Database, KDD）是知识发现研究的主体和热点，而聚类知识发现又是知识发现的重要方面。

如何有效处理巨量、高维的数据，是当前聚类分析的关键技术。

本书围绕高维数据的聚类问题展开研究，在讨论高维数据相似性尤其是高维二元数据相似性定义的基础上，提出了基于粗图模型的硬聚类和软聚类算法、高维二元数据的映射聚类算法、基于蚂蚁行为的聚类算法等，并进一步提出基于映射聚类的离群点检测方法；还特别讨论了高维聚类结果的表示方法问题，提出了应用粗糙集高效表达聚类结果的方法；最后探讨了聚类知识发现数据建模的基本步骤，给出了聚类知识发现的典型应用案例。

本书学术性、知识性并重，可供从事数据仓库与数据挖掘教学、研究的师生、学者阅读，也可以为从事数据挖掘、知识发现系统等软件工程技术人員提供参考。

## 作者简介

陈建斌，男，山西长治人。

2005年获计算机应用专业博士学位。

国家职业技能鉴定专家委员会企业信息管理专家组组长，中国机械工业企业管理协会专家委员会委员，北京联合大学商务学院副教授、电子商务研究所所长，高级企业信息管理师。

主要研究方向为企业信息化管理与运作、电子商务与电子政务和商务智能。

作者多年来秉承系统工程思想，开展企业信息化管理的研究与实践，主编了《企业信息管理师培训教程》、《ERP工程师培训教程》、《电子商务与电子政务》、《电子商务与现代物流》和《信息经济学》等多部著作，参与国家自然科学基金项目多项，主持省部级科研项目多项，在国内核心期刊及国际学术会议上发表论文三十余篇。

作为高级顾问和专家组组长，主讲和辅导中央企业信息管理师职业资格培训数十期；参与多项企业信息化战略规划，数据仓库规划与建设和企业诊断等项目。

## 书籍目录

第1章 知识发现与KDD 1.1 知识与知识发现 1.1.1 知识 1.1.2 知识发现和KDD 1.1.3 知识发现的过程 1.2 数据库知识发现——KDD 1.2.1 KDD的产生与发展 1.2.2 KDD的一般机理和理论基础 1.2.3 KDD系统的基本框架 1.2.4 KDD的主要任务第2章 聚类知识发现及其关键技术 2.1 聚类问题的主要方法 2.2 聚类问题的关键技术 2.2.1 数据仓库技术 2.2.2 高维聚类技术 2.3 高维聚类关键技术研究 2.3.1 高维聚类的主要算法 2.3.2 高维聚类算法的关键技术第3章 高维数据相似性的定义 3.1 数据相似关系 3.1.1 基于距离的相似性定义 3.1.2 基于密度的相似性定义 3.1.3 基于连接的相似性定义 3.2 高维数据相似关系的定义 3.3 二元数据相似性的定义 3.3.1 属性分布特征向量 3.3.2 对象间属性分布相似性 3.4 小结第4章 基于粗图模型的聚类算法 4.1 图论基础概念 4.2 基于图论的聚类算法 4.2.1 聚集型图论聚类 4.2.2 多层粗图法 4.2.3 基于二部图的方法 4.3 图划分的关键技术 4.3.1 图的多层二分划 (Multilevel Graph Bisection) 4.3.2 增强谱分割算法 4.3.3 图的非平衡划分技术 4.4 多层粗图聚类算法的改进 4.4.1 聚类算法 4.4.2 图分割的精化算法 4.4.3 聚类质量评价 4.4.4 实验结果 4.4.5 算法评价 4.5 基于粗图模型的软聚类方法 4.5.1 引言 4.5.2 软聚类算法 4.5.3 基于图划分法的软聚类GPSC算法 4.5.4 实验分析 4.5.5 软聚类方法的评价 4.6 小结 第5章 高维二元数据的映射聚类算法第6章 基于蚂蚁行为的聚类方法第7章 高维数据空间的离群点检测方法第8章 高维数据聚类结果的表示第9章 聚类知识发现数据建模及应用参考文献

章节摘录

1.2.2KDD的一般机理和理论基础 1.一般机理 推理、联想和学习是人类智能活动的三大主要功能，推理和联想的功能必须通过学习才能不断完善、充实，因而学习是一切智能活动的基础。使计算机系统具有某些程度的学习能力，能够模拟人类的学习活动，一直是人工智能领域所追求的目标。

储存在数据库中的结构化数据，是对现实世界某种程度上符号化和数据化的抽象，是对现实世界事物某一程度、某一侧面的映射，所使用的抽象方式和抽象层次主要取决于具体的应用模式。考虑到数据采集过程中可能引入误差，因而要求数据库至少能够在总体上反映现实世界，否则数据库就不能使用。

数据库中的元组可以认为是一些低抽象程度的判断。

2.主要研究方法 KDD的主要实施对象是关系数据库。

这是因为关系数据库具有归一化的组织结构、一体化的查询语言、方便的用户接口和能进行集合处理的优点，而且在各行业中应用最广泛。

另外，关系数据库中各关系之间、各属性之间都是平等的，有利于知识发现过程中的并行计算。

由于KDD的研究对象比较特殊，一般都是大型数据库，其中的数据容量往往是一般人工智能系统所不能比拟的，因此，KDD的研究方法及技术策略就有其鲜明的特色。

首先，在研究上遵循认识的基本过程，即实践—认识—再实践—再认识。

KDD一改过去以演绎逻辑为主的策略，在本质上以归纳逻辑为主，采用从个别到一般，从感性到理性的知识抽象过程。

当然，在知识发现过程中，也不能完全抛弃演绎，而是归纳和演绎相结合。

编辑推荐

学术性、知识性并重，可供从事数据仓库与数据挖掘教学、研究的师生、学者阅读，也可以为从事数据挖掘、知识发现系统等软件工程技术人員提供参考。

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>