

<<粒度计算在数据挖掘中的应用研究>>

图书基本信息

书名：<<粒度计算在数据挖掘中的应用研究>>

13位ISBN编号：9787504736468

10位ISBN编号：7504736465

出版时间：2011-7

出版时间：中国物资出版社

作者：张霞

页数：153

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>

## <<粒度计算在数据挖掘中的应用研究>>

### 内容概要

《粒度计算在数据挖掘中的应用研究》是对模糊粒度计算在文本软聚类中的应用进行了深入研究，提出了一种基于模糊粒度计算的聚类方法，并且利用该聚类方法对K-means算法进行了优化。基于模糊粒度计算的聚类是通过归一化的距离函数将聚类问题映射到距离空间，调节粒度产生对文本集合D的动态聚类划分。

动态聚类既可以作为一个单独的聚类结果，也可以作为其他算法的一个预处理步骤。

K-means算法是一种经典的聚类算法，速度快、消耗资源小，但是算法对初始聚类中心点敏感，容易陷入局部最小值。

《粒度计算在数据挖掘中的应用研究》将基于模糊粒度计算的聚类方法作为K-means算法的预处理步骤，实验结果证明，这种预处理有效地消除了K-means算法的初始值敏感问题，优化了K-means算法。

## 作者简介

张霞，女，河北省石家庄市人。

曾先后于兰州商学院经济信息管理系获得工学学士、北京科技大学信息工程学院获得工学硕士。

2004年师从尹怡欣教授，就读于北京科技大学信息工程学院控制理论与控制工程专业，并于2009年6月获工学博士学位。

自1998年至今，在河北经贸大学从事计算机教学的工作。

曾主持河北省教育厅科研计划项目、河北省科学技术研究与发展计划项目；参与河北省科技支撑计划项目、河北省教育厅教改课题以及河北省哲学社会科学规划办公室等多项课题；发表近20篇核心期刊学术论文，其中4篇为EI收录论文。

主要研究方向：数据挖掘、模式识别。

## <<粒度计算在数据挖掘中的应用研究>>

### 书籍目录

1 绪论1.1 课题背景和意义1.2 课题研究内容1.3 主要创新点1.4 本书的逻辑结构2 文献综述2.1 数据挖掘概述2.2 粒度计算在数据挖掘中的应用2.3 粒度计算数据挖掘研究中需要进一步解决的问题2.4 本书的研究内容及特点2.5 小结3 基于模糊粒度计算的聚类3.1 引言3.2 模糊聚类分析3.3 模糊粒度聚类基础3.4 基于模糊粒度计算的文本聚类3.5 基于模糊粒度计算的农业经济划分3.6 基于模糊粒度计算的K-means优化算法3.7 小结4 基于粒网络生成规则的文本分类4.1 引言4.2 基于集合论的粒度分类基础4.3 基于规则的机器学习4.4 基于粒度计算的分类4.5 基于粒网络生成规则的分类模型4.6 基于粒网络生成规则的文本情感分类4.7 小结5 基于信息粒度的不完备系统遗漏值补齐5.1 引言5.2 粗糙集理论基本概念5.3 知识发现中的不完备信息问题5.4 不完备信息系统5.5 不完备信息系统粒度模型构建5.6 基于信息粒度的遗漏值补齐5.7 决策规则的不确定性表示与度量5.8 实验结果和分析5.9 小结6 结论参考文献附录后记

章节摘录

概念归纳 ( Concept Induction ) 方法。

这种方法来源于机器学习。

我们知道, 典型的示例学习把样本分成正样本和负样本, 学习的结果就是形成覆盖所有正样本但不覆盖任何负样本的概念描述。

关于这类学习算法可以在经典的机器学习的教程中找到, 这里不再赘述。

但是, 要把这种思想应用到数据挖掘中要解决两个关键问题。

第一, 必须扩大样本集的容量和范围。

传统的机器学习希望是精练的小样本集, 而数据挖掘系统必须忠实于源数据, 是面向大容量数据库等存储数据集的。

所以, 扩大后的样本集可能难于有效地精确实现“覆盖所有正样本但不覆盖任何负样本”的概念归纳目标。

要结合概率统计方法, 在检验部分正样本或负样本情况下得到概念的描述。

因此, 最大限度地使用样本进行归纳就是必须解决的关键问题之一。

第二, 对于数据挖掘系统来说, 正样本来自源数据库, 而负样本是不可能源数据库中直接存储的, 但是缺乏对比类信息概念归纳是不可靠的。

因此, 从源数据库中形成负样本 ( 或区别性信息 ) 以及相关的评价区别的度量方法等也是要解决的另一个重要问题。

多维数据分析可以看做一种广义知识挖掘的有效方法。

数据分析的经常性工作是数据的聚集, 诸如计数、求和、平均、最大值等。

既然很多聚集函数需经常重复计算, 而且这类操作的计算量一般又特别大, 因此一种很自然的想法是, 把这些汇总的操作结果预先计算并存储起来, 以便于高级数据分析使用。

最流行的存储汇集数据类的方法是多维数据库 ( Multi-dimension Database ) 技术。

多维数据库总是提供不同抽象层次上的数据视图。

例如, 可以存放每周的数据, 也可在月底形成月数据, 月数据又能形成年数据。

关于多维数据模型的操作, 已经被很好研究, 许多文献可能和数据仓库、OLAP联系起来。

其实, 这种模型, 特别是它操作的完备性 ( 如上钻、下钻等 ), 可以成为广义知识发现的基础。

.....

## <<粒度计算在数据挖掘中的应用研究>>

### 编辑推荐

《粒度计算在数据挖掘中的应用研究》是信息处理的一种新的概念和计算范式，覆盖了所有和粒度相关的理论、方法、技术和工具的信息，主要用于描述和处理不确定的、模糊的、不完整的和海量的信息以及提供一种基于粒和粒间关系的问题求解方法。

作为一个新兴领域的研究，国内外相关研究人员提供了一些粒度计算的理论基础，并且为检验知识发现过程的有效性、解决实际问题提供了一条可行途径。

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>