

<<高维数据挖掘技术研究>>

图书基本信息

书名：<<高维数据挖掘技术研究>>

13位ISBN编号：9787564109776

10位ISBN编号：7564109777

出版时间：2007-12

出版时间：东南大学出版社

作者：杨风召

页数：126

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>

<<高维数据挖掘技术研究>>

内容概要

本书从高维数据的特性出发，指出了高维数据给数据挖掘带来的影响以及高维数据挖掘的研究方向。对高维数据挖掘中的相似性搜索、高维数据聚类、高维数据异常检测、高维数据频繁模式发现及电子商务中的协同过滤技术进行了研究，提出了相关的解决方案和相应算法。

本书适用于从事数据挖掘和商业智能研究的高校教师、研究生、科研院所的科研人员以及从事商业智能项目开发的工程技术人员。

<<高维数据挖掘技术研究>>

书籍目录

- 1 绪论 1.1 研究背景 1.1.1 数据挖掘技术的产生和发展 1.1.2 高维数据挖掘的概念 1.2 高维数据挖掘所遇到的困难 1.2.1 高维数据的特点 1.2.2 维灾(the curse of dimensionality) 1.2.3 高维对数据挖掘的影响 1.3 高维数据挖掘的主要研究方向 1.3.1 高维空间中的距离函数或相似性度量函数 1.3.2 高效的高维数据相似性搜索算法 1.3.3 高效的高维数据挖掘算法 1.3.4 在高维空间中对失效的问题的处理 1.3.5 选维和降维 1.4 术语和符号约定 1.4.1 基本术语 1.4.2 符号约定 1.5 本书结2 高维数据的相似性查询处理 2.1 相似性查询 2.2 维归约 2.2.1 选维 2.2.2 降维 2.3 高维索引结构 2.4 相似性查询方法 2.4.1 RKV算法 2.4.2 HS算法 2.4.3 其他高维数据的相似性搜索算法 2.5 高维数据相似性搜索方法的讨论 2.5.1 维归约技术的局限 2.5.2 高维索引结构在性能上的局限 2.6 本章小结3 一种新的高维数据相似性度量函数Hsim() 3.1 最近邻查询的不稳定性 3.2 高维空间中的最近邻特性 3.3 高维空间中的Lk-范数特性的深入探讨 3.4 高维空间距离函数的重新设计 3.5 Hsim()函数的讨论 3.5.1 Hsim()函数的推广 3.5.2 数据的规范化 3.5.3 对高维数据中空值的处理 3.6 Hsim()与其他相似性度量方法的比较 3.6.1 由距离度量转换来的相似性度量 3.6.2 Cosine度量 3.6.3 PearSOEI相关系数 3.6.4 Jaccard系数 3.7 本章小结4 量化交易数据的相似性搜索 4.1 量化交易数据 4.2 量化交易数据的相似性度量 4.3 索引结构的建立 4.3.1 特征表 4.3.2 特征划分 4.4 相似性搜索算法 4.5 举例 4.6 性能分析 4.7 本章小结5 一种基于评分的协同过滤算法 5.1 相关研究工作 5.1.1 基于用户的推荐算法 5.1.2 基于项的推荐算法 5.1.3 两种推荐算法的比较 5.1.4 维归约技术 5.2 基于特征表的评分数据协同过滤算法[Yzs03] 5.2.1 相似性度量 5.2.2 基于特征表的协同过滤算法 5.3 实验评价 5.3.1 数据集 5.3.2 评价指标 5.3.3 实验结果 5.4 本章小结6 高维数据聚类算法分析 6.1 一般聚类算法概述 6.1.1 分层法 6.1.2 划分法 6.1.3 基于密度的方法 6.1.4 基于网格的方法 6.2 高维对聚类算法的影响及高维数据聚类方法 6.2.1 高维对聚类算法效率的影响 6.2.2 高维可能导致传统的聚类概念失去意义 6.2.3 高维数据聚类方法 6.3 子空间聚类 6.3.1 重叠划分子空间聚类算法 6.3.2 无重叠划分子空间聚类算法 6.3.3 最优投影聚类算法 6.3.4 子空间聚类算法的推广 6.4 优化的网格分割聚类方法 6.4.1 优化的网格分割 6.4.2 优化的网格分割算法 6.4.3 优化的网格分割算法性能的改进 6.5 高维类别数据聚类算法 6.6 基于对象相似性的高维数据聚类算法 6.6.1 基于对象相似性的聚类算法框架 6.6.2 基于SL树的图分割算法 6.6.3 HETIS算法 6.6.4 应用分析 6.7 本章小结7 高维数据异常检测 7.1 异常检测算法分析 7.1.1 基于统计的算法 7.1.2 基于深度的算法 7.1.3 基于偏差的算法 7.1.4 基于距离的算法 7.1.5 基于密度的算法 7.2 高维对异常检测算法的影响 7.2.1 高维对基于统计算法的影响 7.2.2 高维对基于深度算法的影响 7.2.3 高维对基于距离算法的影响 7.2.4 高维对基于密度算法的影响 7.2.5 高维异常检测的问题与出路 7.3 投影异常的概念及其检测算法 7.3.1 投影异常的定义 7.3.2 蛮力搜索算法 7.3.3 遗传算法 7.4 动态环境下局部异常的增量挖掘算法IncLOF 7.4.1 受影响对象 7.4.2 数据插入 7.4.3 数据删除 7.4.4 IncLOF的算法复杂度分析 7.4.5 性能分析 7.5 本章小结8 高维数据的频繁模式挖掘 8.1 频繁模式挖掘问题 8.1.1 关联规则挖掘问题的提出 8.1.2 频繁模式和频繁封闭模式挖掘 8.2 定义和术语 8.3 基于特征计数的频繁封闭模式挖掘算法 8.4 基于行计数的频繁封闭模式挖掘算法 8.4.1 自底向上深度优先搜索算法 8.4.2 自顶向下深度优先搜索算法 8.5 基于行计数和特征计数的混合计数频繁封闭模式挖掘算法 8.5.1 动态计数树 8.5.2 算法[PTCX04] 8.5.3 转换条件 8.6 本章小结参考文献

<<高维数据挖掘技术研究>>

章节摘录

1.1 研究背景 1.1.1 数据挖掘技术的产生和发展 在过去的三十年,随着计算机硬件技术、数据收集技术和数据存储技术的快速发展,各行各业都逐步建立起各自的数据库体系。在这些数据库中存放着大量的数据,如何能有效地利用这些信息,使之能为生产实践所利用,成为人们所关注的问题。

但相对于堆积成山的丰富的数据而言,人们缺乏强有力的分析手段和分析工具,因而造成了“数据丰富而信息缺乏”的状况。

显然,数据库的检索和查询难以满足人们的需要,虽然伴随着数据仓库出现的联机分析处理(On—Line Analytical Processing, OIAP)技术具有总结、概化和聚集的功能,可以从不同角度来观察数据,支持多维分析和决策支持,但它不能进行更深层次的分析,挖掘出大量数据背后所蕴藏的知识。在这种情况下,数据挖掘技术便应运而生。

数据挖掘指的是从大量的数据中提取人们感兴趣的知识,这些知识是隐含的、事先未知的、并且是潜在有用的信息[FPSU96]。

它是计算机技术研究中的一个很有应用价值的新领域,融合了数据库、人工智能、机器学习、统计学等多个领域的理论和技术,目前已成为国际上数据库和信息决策领域中最前沿的研究方向之一,引起了学术界和工业界的广泛关注。

一些国际上高级别的工业研究实验室,例如IBM Almaden和GTE,众多的学术单位,例如UC Berkeley,都在这个领域开展了各种各样的研究计划。

其研究的主要目标是发展有关的方法论、理论和工具,以支持从大量数据中提取有用的和让人感兴趣的知识和模式。

数据挖掘,也叫数据库中发现知识(Knowledge Discovery in Databases, KDD)。

KDD一词首次出现在1989年8月举行的第11届国际联合人工智能学术会议上。

随着KDD在学术界和工业界的影响越来越大,国际KDD组委会于1995年把专题讨论会更名为国际会议,在加拿大蒙特利尔市召开了第1届KDD国际学术会议,以后每年召开一次。

迄今为止,由美国人工智能协会主办的KDD国际研讨会已经召开了13次,规模由原来的专题讨论会发展成为国际学术大会。

.....

<<高维数据挖掘技术研究>>

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>