

综述评论

用于数据挖掘的聚类算法

姜园^{①②}, 张朝阳^①, 仇佩亮^①, 周东方^②

^①浙江大学信息与通信工程研究所 杭州 310027; ^②解放军信息工程大学 郑州 450002

收稿日期 2003-12-22 修回日期 2004-4-26 网络版发布日期 2008-4-14 接受日期

摘要

数据挖掘用于从超大规模数据库中提取感兴趣的信息。聚类是数据挖掘的重要工具, 根据数据间的相似性将数据库分成多个类, 每类中数据应尽可能相似。从机器学习的观点来看, 类相当于隐藏模式, 寻找类是无监督学习过程。目前已有应用于统计、模式识别、机器学习等不同领域的几十种聚类算法。该文对数据挖掘中的聚类算法进行了归纳和分类, 总结了7类算法并分析了其性能特点。

关键词 [数据挖掘](#) [聚类](#) [分层聚类](#) [分割聚类](#) [K-Means](#)

分类号 [TP391](#)

Clustering Algorithms Used in Data Mining

Jiang Yuan^{①②}, Zhang Zhao-yang^①, Qiu Pei-liang^①, Zhou Dong-fang^②

^①Institute of Information & Communication Engineering Zhejiang University Hangzhou

310027 China; ^②PLA Information Engineering University Zhengzhou 450002 China

Abstract

Data mining is used to draw interesting information from Very Large DataBases (VLDB). Clustering plays an outstanding role in data mining applications. Clustering is a division of databases into groups of similar objects based on the similarity. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learning. There are tens of clustering algorithms used in various fields such as statistics, pattern recognition and machine learning now. This paper concludes the clustering algorithms used in data mining and assorts them into 7 classes. Seven types of algorithms are summarized and their performances are analyzed here.

Key words [Data mining](#) [Clustering](#) [Hierarchical clustering](#) [Partitioning clustering](#) [K-Means](#)

DOI :

通讯作者

作者个人主页 姜园^{①②}; 张朝阳^①; 仇佩亮^①; 周东方^②

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(2423KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [参考文献\[PDF\]](#)

▶ [参考文献](#)

服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ [本刊中 包含“数据挖掘”的 相关文章](#)

▶ 本文作者相关文章

· [姜园](#)

· [张朝阳](#)

· [仇佩亮](#)

· [周东方](#)