

数据库与信息处理

基于加权快速聚类的异常数据挖掘算法

李星毅^{1,2}, 包从剑², 施化吉², 奚春海³

- 1.北京交通大学 电子信息学院, 北京 100044
- 2.江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013
- 3.亭旁中学 计算机中心, 浙江 三门 317103

收稿日期 修回日期 网络版发布日期 2007-11-29 接受日期

摘要 聚类是数据挖掘领域中最活跃的研究分支之一, 并在其他的科学领域也有广泛的应用。设计了基于加权快速聚类的异常数据挖掘算法, 以便能快速发现异常数据。首先通过对数据的每个属性赋予一定权值, 权值的大小要体现其对分类的贡献度, 并根据属性权值的特点, 选择比较优良的初始分区, 然后进行多次迭代, 得到接近最优分区, 接着运用一定规则, 发现异常数据类, 最后实践证明该技术取得很好的社会效果。

关键词 [异常数据](#) [数据挖掘](#) [学习规则](#) [K-均值聚类](#) [加权快速聚类](#)

分类号

Outlier data mining algorithms based on weighted fast clustering

LI Xing-yi^{1,2}, BAO Cong-jian², SHI Hua-ji², XI Chun-hai³

- 1.School of Electronics and Information Engineering, Beijing JiaoTong University, Beijing 100044, China
- 2.School of Computer Science and Telecommunications Engineering, JiangSu University, Zhenjiang, Jiangsu 212013, China
- 3.Center of Computer, TingPang Middle School, Sanmen, Zhejiang 317103, China

Abstract

Clustering is one of the most flourish direction of data mining, and it has been applied abroad at other scientific fields. This article promoted outlier data mining algorithms based on weighted fast clustering to inspect and deal with outlier data effectively. The processes of algorithms were described in the followings, firstly, the each property of data should be endowed with certain weight to incarnate its sort devotion degree, and choose better initialization subarea according to the weight characteristics of property, and get to the best subarea under many times iteration, and then find outlier data by the application of certain data class. Finally, the experiment demonstrated this technology obtained better social effect.

Key words [outlier data](#) [data mining](#) [learning rule](#) [K-mean clustering](#) [weighted fast clustering](#)

DOI:

通讯作者 李星毅 jianbc@126.com

扩展功能	
本文信息	
▶	Supporting info
▶	PDF(626KB)
▶	[HTML全文](0KB)
▶	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“异常数据” 的相关文章
▶	本文作者相关文章
·	李星毅
·	包从剑
·	施化吉
·	奚春海