

论文与报告

## 基于数据间内在关联性的自适应模糊聚类模型

唐成龙, 王石刚

1. 上海交通大学机械及动力工程学院 上海 200240

收稿日期 2009-11-11 修回日期 2010-5-28 网络版发布日期 接受日期

摘要

提出了一种新的模糊聚类模型(Fuzzy C-means clustering model, FCM), 称为自适应模糊聚类(Adaptive FCM, AFCM). 和现有的大多数模糊聚类方法不同的是, AFCM考虑了数据集中全体数据的内在关联性, 模型中引入了自适应度向量 $W$ 和自适应指数 $p$ . 其中,  $W$ 在迭代过程中是自适应的,  $p$ 是一个给定参数.  $W$ 和 $p$ 共同作用调控聚类过程. AFCM同时输出三组参数: 模糊隶属度集 $U$ , 自适应度向量 $W$ , 以及聚类原型集 $V$ . 本文给出了两组数据实验验证AFCM的性能. 第1组实验验证AFCM的聚类性能, 以FCM为比较对象. 实验表明 AFCM可以得到更好的聚类质量, 而且通过合理选择自适应指数 $p$ , AFCM和FCM在时间复杂性上保持同一水平. 第2组实验检验了AFCM的离群点挖掘性能, 以目前常用的基于密度的LOF为比较对象. 实验表明AFCM算法具有极大的计算效率优势, 且AFCM得到的离群点是全局的, 反映的是离群点和整个数据集的关系, 离群点涵盖的信息也更丰富. 文章指出, AFCM在挖掘大数据集和实时数据中的离群点应用方面, 以及获得高质量的聚类结果的应用方面, 特别在聚类的同时需要挖掘离群点的应用方面具有独特的优势.

关键词 [模糊聚类](#) [离群点挖掘](#) [自适应聚类策略](#) [自适应度](#) [自适应指数](#)

分类号

## Adaptive Fuzzy Clustering Model Based on Internal Connectivity of All Data Points

TANG Cheng-Long, WANG Shi-Gang

1. School of Mechanical and Dynamical Engineering, Shanghai Jiao Tong University, Shanghai 200240

Abstract

This paper proposes a new kind of fuzzy C-means clustering model (FCM), which is named as adaptive fuzzy clustering (AFCM). Different from most current fuzzy clustering methods, the AFCM considers the internal connectivity of all data points. An adaptive degree vector  $W$  and an adaptive exponent  $p$  are introduced into the model to jointly influence the clustering process. The AFCM simultaneously outputs three categories of parameters: fuzzy membership degree matrix  $U$ , adaptive degree vector  $W$ , and cluster prototype matrix  $V$ . Two groups of numerical experiments, Group 1 and Group 2, were executed to evaluate the AFCM. Group 1 demonstrates the clustering performance of the AFCM, with FCM being its counterpart, and the results showed that the AFCM can obtain better clustering quality, meanwhile its time complexity can hold the same level as that of the FCM by choosing the available  $p$ . Group 2 checks the ability of the AFCM in mining the outliers, with the density-based LOF being its counterpart and the results showed that the AFCM has considerable advantages in computing efficiency, and that the outliers mined by the AFCM are global, and reflect the relationship between the outliers and the whole data set. It is pointed out that the AFCM possesses the unique advantages when mining the outliers of the large-scale or dynamic data sets, and clustering the data set for better clustering results, especially when it is necessary to simultaneously fulfill both tasks of clustering and mining outliers.

Key words [Fuzzy clustering](#) [outliers mining](#) [adaptive clustering approach](#) [adaptive degree](#) [adaptive exponent](#)

DOI: 10.3724/SP.J.1004.2010.01544

通讯作者 唐成龙 [tangchenglong@baosteel.com](mailto:tangchenglong@baosteel.com)

作者个人主页 唐成龙; 王石刚

页

扩展功能
本文信息
▶ <a href="#">Supporting info</a>
▶ <a href="#">PDF(1550KB)</a>
▶ <a href="#">[HTML全文](0KB)</a>
▶ <a href="#">参考文献[PDF]</a>
▶ <a href="#">参考文献</a>
服务与反馈
▶ <a href="#">把本文推荐给朋友</a>
▶ <a href="#">加入我的书架</a>
▶ <a href="#">加入引用管理器</a>
▶ <a href="#">复制索引</a>
▶ <a href="#">Email Alert</a>
相关信息
▶ <a href="#">本刊中 包含“模糊聚类”的 相关文章</a>
▶ 本文作者相关文章
· <a href="#">唐成龙</a>
· <a href="#">王石刚</a>