

## Accepted Manuscript

Title: A Novel Cluster Center Fast Determination Clustering Algorithm

Authors: Chen Jinyin, Lin Xiang, Zheng Haibing, Bao Xintong



PII: S1568-4946(17)30213-2  
DOI: <http://dx.doi.org/doi:10.1016/j.asoc.2017.04.031>  
Reference: ASOC 4164

To appear in: *Applied Soft Computing*

Received date: 1-11-2016  
Revised date: 4-3-2017  
Accepted date: 18-4-2017

Please cite this article as: Chen Jinyin, Lin Xiang, Zheng Haibing, Bao Xintong, A Novel Cluster Center Fast Determination Clustering Algorithm, Applied Soft Computing Journal <http://dx.doi.org/10.1016/j.asoc.2017.04.031>

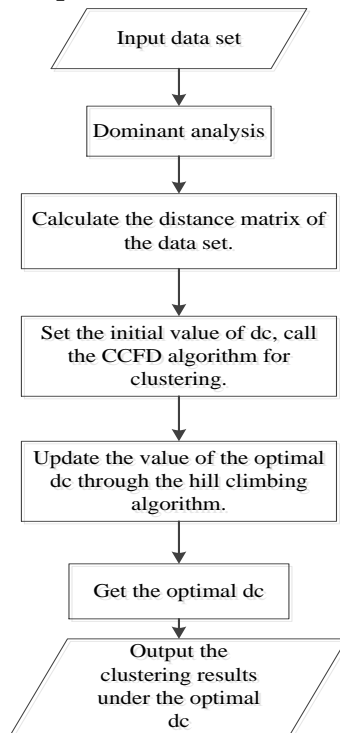
This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# A Novel Cluster Center Fast Determination Clustering Algorithm

Chen Jinyin, Lin Xiang, Zheng Haibing, Bao Xintong

College of Information Engineering, Zhejiang University of Technology, Hangzhou, China  
chenjinyin@zjut.edu.cn

## Graphical abstract



## Highlights:

- Mixed data sets are popular in real practice, a universal mixed data similarity calculation method is difficult for dealing with different kinds of mixed data attributes. We propose a universal mixed data similarity calculation mechanism in this paper. The similarity distance matrix of the data is calculated by different calculation methods for different types of data sets, so that the clustering algorithm can deal with different kinds of mixed attribute data.
- RLM algorithm made great success in applying for face recognition for its simplicity to realize. However when it is adopted for different applications, it has encounter with two challenges. The key parameter for density radius for clustering  $dc$  is dependent on manual setting and the cluster centers are selected by manual operations. In our paper, the parameter adaptation of the density radius  $dc$  is optimized based on mountain climbing algorithm. And the clustering centers are determined automatically based on normal distribution curves to fit the density distribution curve of density distance product. In this way the whole cluster process does not require human intervention.
- Abundant simulations show that the proposed algorithm has a good clustering effect for numerical data sets, classified data sets and mixed data sets, which reflects the superiority of the algorithm. Moreover, real world data set as MNIST data is testified for proving the capacity of our algorithm with satisfying performances.

**Abstract:** As one of the most important techniques in data mining, cluster analysis has attracted more and more attentions in this big data era. Most clustering algorithms have encountered with challenges including cluster centers determination difficulty, low clustering accuracy, uneven

Download English Version:

<https://daneshyari.com/en/article/4963376>

Download Persian Version:

<https://daneshyari.com/article/4963376>

[Daneshyari.com](https://daneshyari.com)