

{tag} International Journal of Computer Applications
Foundation of Computer Science (FCS), NY, USA

[Volume 144](#)

-
[Number 6](#)

Year of Publication: 2016

Authors:

Nipjyoti Sarma, Arindam Saha, Adarsh Pradhan

10.5120/ijca2016910305

{bibtex}2016910305.bib{/bibtex}

Abstract

K mean clustering is a very popular clustering algorithm for clustering numerical data. . It is popular due to its simplicity of understanding and linear algorithmic complexity measure. But it has the serious limitation of clustering numerical only data. Therefore several researchers tried to improve the k mean algorithm to cluster not only numerical but also categorical dataset. In this work an effort have been made to put forward a proposed FCV mean algorithm which is a modified version of the traditional k-mean algorithm and is able to cluster objects having mixed type attributes i.e. numerical and categorical. For categorical data fuzzy set similarity is used and for numerical data differences from maximum dissimilarity is used. Experiment shows that the mixed data are highly clustered with high accuracy compared to other approach in literature.

References

1. Yiu Ming Cheung and Hong Jia in “Categorical and numerical data clustering based on a

unified similarity metric with out knowing cluster number” in Pattern Recognition 46 (2013) 2228–2238, Elsevier (2013).

2. R. A. Ahmed B. Borah D. K. Bhattacharyya and J K Kalita in HIMIC : A Hierarchical Mixed Type Data Clustering Algorithm” in”(2005) in

[http://citeseerx.isi.psu.edu/viewdoc/download?doi=10.1.1.61.6369 & rep=rep1 & type=pdf](http://citeseerx.isi.psu.edu/viewdoc/download?doi=10.1.1.61.6369&rep=rep1&type=pdf),

3. LIU Hai-tao, WEI Ru-xiang and JIANG Guo-ping in Similarity measurement for data with high-dimensional and mixed feature values through fuzzy clustering” in Proceedings of IEEE conference,2012 ,pp-617-621,

4. Z. Huang in “Clustering large data sets with mixed numeric and categorical values” in Proceedings of the First Pacific-Asia Conference on Knowledge Discovery and Data Mining, , pp. 21–34.1997,

5. Limin CHEN , Jing YANG and Jianpei ZHANG, in “An Efficient Clustering Method for Large Mixed Type Dataset” in Journal of Computational Information Systems 8: 22 (2012) 9553–9560,

6. Ming-Yi Shih, Jar-Wen Jheng and Lien-Fu Lai in “A Two Step Method for Clustering Mixed Categorical and Numerical data” in Tamkng Journal of Science and Engineering, Vol. 13, No. 1, pp. 11-19(2010)

7. Jongwoo Lim, Jongeun Jun, Seon Ho Kim and Dennis McLeod in “A Framework for Clustering Mixed Attribute Type Datasets” in Proceedings of the fourth International Conference on Emerging Databases (EBD),2102.

8. Chian Hsu and Yan-Ping Huang in “Incremental clustering of mixed data based on distance hierarchy” in Elsevier/Expert Systems with Applications 35, 1177–1185,2008,

9. Fuzzy logic and nural networks by M. Amrithavalli first, second and third chapter Fourth reprint)published by scitech publications(India)pvt. Limited, 2010.

10. UCI machine learning Repository- <http://archive.ics.uci.edu/ml/>

Index Terms

Computer Science

Fuzzy Systems

Keywords

fuzzy set, Centroid vector, dissimilarity, categorical, numerical.