

Fuzzy Mean Point Clustering using K-means algorithm for implementing the movecentroid function code

Inamdar S. A.

School of Computational Science,
Swami Ramanand Teerth,
Marathwada University, Nanded

Narangale S.M.

School of Media Studies, Swami
Ramanand Teerth Marathwada
University, Nanded

G. N. Shinde*

Indira Gandhi College, CIDCO,
Nanded-431603, Maharashtra,
INDIA

Abstract

The paper focus on combination of K-Means algorithm for Fuzzy Mean Point Clustering Neural Network (FMPCNN). The algorithm is implemented in JAVA program code for implementing the movecentroid function code into FMPCNN. Here we have provided movecentroid's output to Fuzzy clustering as criteria, movecentroid is the base function of K-means algorithm as in Fuzzy Mean Point Clustering Neural Network (FMPCNN) algorithm, calculation of cluster based on pre-defined criteria and scope is done. In the experiment we have used four datasets and observed results in nano seconds there is huge difference in output as time is reduced for Fuzzy Min-Max code execution of fuzzy calculations of clustering.

Keywords: Data mining, Clustering, K-means, FMPCNN, movecentroid.

1. Introduction

Web Mining is an important sub-branch of Data Mining. Data mining is extracting potential, unknown, useful

information, patterns and trends from abundant, incomplete, noise and random data. The Web includes huge amount of data, using the data mining technology on the Web, namely the Web mining technology, becomes the most important research along with the rapid development of Internet [2]. Data mining has various techniques to extract useful information in large amounts of data. Data mining is defined as a technique of finding hidden information in a database [1]. It may be called as data driven discovery, explorative data analysis, deductive learning. Data mining in general falls in to the following categories: classification patterns, clustering, association patterns [2, 3].

1.1 Clustering Analysis

Cluster is a group of patterns having some common properties. Patterns can be grouped into clusters by some predefined criterion. The clusters can be formed according to some criterion like distance, angle, curvature, symmetry, connectivity and intensity.

A main function of the clustering analysis is to give the users different classes of webs by clustering retrieval results, which enables them to make a rapid target, to obtain the automatic classification catalogues rapidly.

There are many categorizations of clustering algorithms such as hierarchical-based method, partitioning method, density-based method, grid-based method and model-based method and so forth.

In Partitioning-based method: First, data divided into K subsets, and then using the technology of iterative

location is applied with an attempt to improve the quantity of partition by transferring the object from one class to another. K-Means and K-medoids are two typical types of partitioning method [4,7,8].

1.2 K-means Algorithm

Cluster is a group of patterns having some common properties. Patterns can be grouped into clusters by some predefined criterion. The clusters can be formed according to some criterion like distance, angle, curvature, symmetry, connectivity and intensity. K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. K-Means is an iterative clustering algorithm in which items are moved among sets of clusters based on membership until the desired set is reached. The degree of membership decides whether the item is member of desired cluster or not. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroid, one for each cluster.

This algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

Where $\|x_i^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j , is an indicator of the distance of the n data points from their respective cluster centers [5].

2. Related work

P.M.Patil, U.V.Kulkarni, T.R. Sontakke given an algorithm for pattern clustering based on Fuzzy Mean point Clustering Neural Network[6]. It uses two parameters viz. centering and bunching. FMPCNN is used to bunch the patterns in the clusters, thus governs the cluster size. The mean point of patterns grouped in that cluster represents the cluster. Mean point of the cluster is calculated by considering those patterns, which gives fuzzy membership (function membership) greater than or equal to the centering parameter. Only those patterns are grouped into a cluster, which gives fuzzy membership greater than or equal to bunching factor, for that cluster. This algorithm uses a cluster fuzzy set that is defined by a mean point(centroid), a minimum distance between pattern and mean point and a fuzzy membership function. Fuzzy membership values are based on the distance between the pattern and cluster mean point. Mean points capture the natural groupings of patterns in the pattern space.

The variance (distance between two cluster centroids) within a cluster is examined. If it is too large, cluster is split. If the variance is less than a predefined threshold, they will be

call K-Means movecentroid() for defining membership criteria criteria1, criteria2;

Clustering Approach	Instances	Time to build the model (Nano Seconds)	Time Difference (Nano Seconds)	Criteria 1	Criteria 2	Relation/DataSet
weka.clusterers.SimpleKMeans	1741	357612839	351531130	0.22515795519816198	0.46065479609419874	AccidentData
FMPCNN with K Means		6081709				
weka.clusterers.SimpleKMeans	59	72510054	65690542	0.3265084745762713	0.6094745762711865	Fim
FMPCNN with K Means		6819512				
weka.clusterers.SimpleKMeans	57	144067682	136030278	0.22469487449122807	0.22469487449122807	Population
FMPCNN with K Means		8037404				
weka.clusterers.SimpleKMeans	1200	215640325	213121848	0.5	0.5	mlt1
FMPCNN with K Means		2518477				

combined.

FMPCNN algorithm uses two parameters. The centering parameter α_1 , $0 \leq \alpha_1 \leq 1$, a user-defined value used to calculate the cluster centroid. The bunching parameter β_1 , $0 \leq \beta_1 \leq 1$, is also a user-defined value that is used to bunch the patterns into a cluster. For the given set of α_1 and β_1 , the learning algorithm finds optimal mean points of pattern clusters. It uses fuzzy mean point cluster membership function m_j , of an input pattern R_h , in the j th cluster C_j , and δ i.e. the distance defining the core part of the cluster fuzzy set, is given as

$$m_j(R_h, C_j, \delta) = 1 - f(l, \delta, \gamma)$$

where $f()$ is three-parameter ramp threshold function defined as

$$f(l, \delta, \gamma) = \begin{cases} 0, & \text{if } 0 \leq l \leq \delta \\ (l - \delta)\gamma & \text{if } 0 \leq (l - \delta)\gamma \leq 1 \\ 1 & \text{if } (l - \delta) \geq (1 - \delta) \end{cases}$$

and the distance l is defined as

$$l = \left(\sum_{i=1}^n (c_{ji} - R_{hi})^2 \right)^{1/2}$$

where γ is the sensitivity parameter [6].

3. Logic of FMPCNN with K-Means

Centroid plays key role in clustering based on Fuzzy Min-Max algorithm. Criteria act as base for calculating centroid. In FMPCNN with K-Means, we propose a cluster algorithm as:

ALGORITHM FMPCNN with K-Means:

Input:

D={x1,x2,...,xn} // Set of items

k //Number of desired clusters

Output:

k //Set of clusters

FMPCNN with K-Means:

get the membership parameter for items;

assign criteria1 and criteria2 with movecentroid() function's output;

criteria1 = K-Means_Centroid1

criteria2 = K-Means_Centroid2

execute FMPCNN clustering function with criteria values, as:

cluster (criteria1, criteria2, scope);

4. Experimental Results

In the experimental setup, we are executing K-Means algorithm program code written JAVA on web usage sample data. Then, we pass K-Means's output as input to FMPCNN algorithm's as criteria1 and criteria2, as:

cluster (criteria1, criteria2, scope);

5. Tabular Comparison:

6. Conclusion

It is a novel approach of clustering with Fuzzy Min-Max and K-Means algorithm. In FMPCNN algorithm calculation of cluster based on pre-defined criteria and scope, but here we have inserted K-Means's movecentroid function for calculating mean/ mode/ median as criteria into FMPCNN algorithm. It is a JAVA based code. In the experiment we have used four datasets and observed results for building the model in nano seconds with weka clusterers. Simple K-Means and with FMPCNN with K Means for each dataset. There is huge difference in output as time is reduced for Fuzzy Min-Max code execution of fuzzy calculations of clustering which makes it effective.

7. References

- [1] Li Zhan, Liu Zhijing, ' Web Mining Based On Multi-Agents ', COMPUTER SOCIETY,IEEE(2003)
- [2] Margaret H. Dunham and Sridhar, "Data Mining, Introduction and Advanced Topics", Prentice Hall, Publication, ISBN 81-7758-785-4, 2003.

[3] G.K.Gupta, "Introduction to Data Mining with Case Studies", Prentice Hall of India Private Limited, New Delhi, 2006.

[4] Qi Luo, "Dynamic Fuzzy Clustering Algorithm for Web Documents Mining", International Conference on Computational Intelligence and Security, IEEE Computer Society, 2010.

[5] JinHuaXu, HongLiu, "Web User Clustering Analysis based on KMeans Algorithm", International Conference on Information, Networking and Automation (ICINA), 2010.

[6] P.M.Patil, U.V.Kulkarni, T.R.Sontakke, "FUZZY MEAN POINT CLUSTERING NEURAL NETWORK", Department of Electronics and Computer Science & Engineering, SGGS College of Engineering and Technology, Vishnupuri, Nanded. http://neuron.csie.ntust.edu.tw/homework/93/Fuzzy/%E5%9C%A8%E8%81%B7%E7%A0%94%E7%A9%B6%E6%89%80/homework_2/M9315904/main.html

[7] ODUKOYA, O.H., ADEROUNMU, G.A. AND ADAGUNODO, E.R., "An Improved Data Clustering Algorithm for Mining Web Documents", IEEE, 2010.

[8] M/s Kiran Agrawal, Mr. Ashish Mishra, "Improved K-MEAN Clustering Approach for Web Usage Mining", Second International Conference on Emerging Trends in Engineering and Technology, ICETET-09, IEEE Computer Society, 2009.

[9] Chu-Hui Lee, Yu-Hsiang Fu, "Web Usage Mining based on Clustering of Browsing Features", Eighth International Conference on Intelligent Systems Design and Applications, IEEE Computer Society, IEEE, 2008.

[10] R. Etemadi, N. Moghaddam, "An Approach in Web Content Mining for Clustering Web Pages", Digital Information Management (ICDIM), Fifth International Conference, pp.279-284, IEEE, 2010.

Short Bio-Sketch

Dr. G. N. Shinde is Principal in Indira Gandhi College, Nanded, Maharashtra, INDIA. He has received "Ideal State Teacher Award" from Government of Maharashtra, India for 2008-09 and "Best Principal Award" for 2009-2010 from S.R.T.M. University, Nanded, Maharashtra. He has received M. Sc. & Ph.D. degree from Dr. B.A. M. University, Aurangabad.



He has awarded Benjonji Jalnawala award for securing highest marks at B.Sc. He is Dean, faculty of Science of S. R. T. M. University, Nanded. Twelve research scholars were awarded Ph.D. degree under his guidance. He has published more than 75 papers in the International Journals and presented more than 50 papers in International Conferences. He was the Chairperson for F-9 session of "International Conference on Computational and Experimental Science & Engineering" (ICCES2008) which was held at Honolulu, U.S.A & Development and Application of Web 2.0 Technology for Education Purpose session of "International Conference on Multimedia & ICT in Education (m-ICT2009)", April 24, 2009, (Hall 1), Lisbon (Portugal) and "Wave Propagation and Wave Interaction with media" Progress In Electromagnetic Research Symposium (PIERS), March 25, 2010, Session 3A4 (Room D) Xian, China. In his account one book is published, which is reference book for different courses. He is also member of different academic & professional bodies such as IAENG (Hon Kong), ANAS (Jordan). He is in reviewer panel for different Journals such as IEEE (Transactions on Neural Networks), International Journal of Physical Sciences (U.S.A.), Journal of Electromagnetic Waves and Applications (JEMWA, U.S.A.). His abroad Visit includes U.S.A., Thailand, Portugal, Germany, Switzerland, Italy, Vatican City, Monaco, France, Maldives, Sri Lanka, U. K., Scotland and China. He was Chairman of Grievances Committee and member of Management Council & Senate of S.R.T.M. University, Nanded, INDIA. His research interest includes Filters, Image processing, web security and Multimedia analysis and retrieval system.