2012 2nd International Conference on Industrial Technology and Management (ICITM 2012)
IPCSIT vol. 49 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V49.7

Proposed a new Fuzzy C-Means Algorithm based on Electrical Rules

Amir Golabzaei[1 +], MohsenYaghoubizadeh [2], BehruzMirkhanzadeh [1]

[1]MS Student, Industrial Engineering, Department of Industrial Engineering Science and Research Branch of Saveh

[2]MS Student, Industrial Engineering, Industrial and Mechanical Engineering's Department Qazvin Branch of Azad University

Abstract. Every day, people deal with different types of data and different types of measurements and observations. Data describe the characteristics of data and summarize the result of experiment. Clustering or classification of data is an important activity to partitioning the data point in to similarity classes. One of the problemswithFuzzy C-Means (FCM) is that this algorithm cannot produce a good partitioning where the objective function is minimizing.The proposed fuzzy clustering algorithm uses electrical rules in FCM to obtain a low degree of variation and a large separation distance between clusters with respect to minimizing the objective function of FCM algorithm.Our algorithm optimized the validity indexes of Bezdek when the number of clusters is in the best state.Experiment results show the effectiveness of our proposed clustering model.

Keywords:Fuzzy c-mean, validity index, Coulomb's law

1. Introduction

One of the most important activities for analysis of data is to classify or clustering data into a set of clusters. People usually try to identify features of data and compare these features with those of known objects based on their similarity or distance according to sum rules. Clustering algorithms separation data objects certain number of categories. Clustering has a long history [1], dating back to Aristotle. Clustering algorithm allocated object to a cluster that this is a most popular problems for crisp clustering. So fuzzy logic (Zade, 1965)[2] create extreme clustering, rather than binary or crisp ones for fuzzy clustering, this problem is solved and the object can allocate to all of the clusters with a certain degree of membership (Bezdek, 1981)[3].

FCM is a most popular algorithm of fuzzy clustering that try to find a partition for a set of objects while minimizing the objective function [3].A large number of heuristic methods have also been proposes to clustering data based on FCM. Until now, proposed optimization approaches for fuzzy partitioning, speeding up of FCM and the improvements of the drawbacks of FCM as shown as table1.

In this paper we present a method of fuzzy clustering based on coulomb's low in section2 and then we tested our method on data set from the machine learning library at UC Irvine and compare this proposed method with FCM algorithm that result of this comparing explore in section3. Finally, in section4 we explore conclusion of our research.

1+Corresponding author. Tel.: +989350518058;
E-mail address:(amirgolabzaei@yahoo.com, m.yaghoobi@qiau.ac.ir ,mrbhrz@Gmail.com).

Table1. Literature review of proposed method for improving FCM

Refrences ID	researchers	Approaches	Published year
[4]	Gath and Geva	Initialization problem by dynamically adding cluster centers	1989
[5]	Kersten	Use of fuzzy medians with a city block distance	1997
[6]	wu and Yang	New possibility algorithm for cost function	2006
[7]	Ozdemir and Akarun	Proposed a partition index maximization (PIM) algorithm	2002
[8]	Wu et al	proofing the volumes of each cluster core in PIM are equal	2005
[9]	Cheng and zhang	Entropy based subspace clustering mining numerical data	1999

2. A Fuzzy C-Mean Model Based on Coulomb's Low

The notations used in the paper are introduced follows in table 2.

Table2. Notation of our article

Row	symbols	Description
1	$X = \{X_1, \dots, X_n\}$	Data vector of j^{th} object
2	u_{ij}	Membership coefficient of the j^{th} object in the i^{th} cluster
3	$U=[u_{ij}]_{c*n}$	Fuzzy partition matrix
4	$C \in \{1, \dots, c\}$	Cluster number
5	$V=[v_1,\dots,v_c]$	Cluster center matrix
6	$m \in [1, \dots, \infty)$	Fuzzification parameter usually set to 2
7	$D_{ij}=D(x_j,m_i)$	Distance between x_j and m_i
8	D_{ij}^2	Euclidean distance function
9	J(p)	Objective function
10	q_i	Value of data point in data set X

2.1. Fuzzy C-Mean Algorithm

FCM is one of the common algorithms of fuzzy clustering methods that proposed withBezdek in 1981[10].

The objective function of FCM method in Eq.1 can be minimizingwith an iterative procedure that leads to the FCM algorithm. The memberships function and center matrix update equations all obtained through alternating optimization. The steps of the FCM method are summarizing as follows:

Step1. Select appropriate values for m, c, and a small positive number . initialize the center matrix V randomly. Set step variable t=0;

Step2. Update the membership matrix U by

$$U_{ij}^{(t+1)} = 1 \Big/ \left(\sum_{l=1}^{c} \left(D_{lj} \big/ D_{ij} \right)^{2/(1-m)} \right) \tag{2}$$

Step3. Update the center matrix V by

$$V_i = \frac{\sum_{k=1}^{n} U_{ik}^{m} X_k}{\sum_{k=1}^{n} U_{ik}^{m}}; \text{ for i=1,…,c;} \tag{3}$$

Step4. Repeat steps 2 and 3 until max$\{\|U_{t+1} - U_t\| \le \varepsilon$.

Calculate objective function J(p) as follows:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{m}\, d_{ik}^2 = \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{m} \|X_k - V_i\|^2 \tag{4}$$

2.2. Coulomb's law

We used electrical rules that described electrostatic actions between electrical points. Firstly, this rule proposed withCharls Coulomb[11]. Ration between electrical power and distance lastly discovered with Henry Cowndish[12]. Point charges that collected generate anelectric filedwith respect to characterize follows:

1. Represents the electric charge density of the electric field is create.
2. Due to the increased distance between the point charges, the electric field strength decreases.

2.3. Validity indexes

In generally clustering validity problem can be change to determination of optimum number clusters. Most studies on cluster validity design for FCM algorithm that reviewed withdubes and jane in 1979[13], 1988. The best known validity indexes were proposed for evaluate cluster validitywithBezdek as follows:

- The partition coefficient:

$$V_{\mathrm{pc}}(U) = \frac{1}{n}\left(\sum_{k=1}^{n}\sum_{i=1}^{c} U_{ik}^{2}\right) \tag{5}$$

- The partition entropy:

$$V_{\mathrm{PE}}(\mathrm{U}) = -\frac{1}{\mathrm{n}}\left(\sum_{\mathrm{k}=1}^{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{c}} \mathrm{U}_{\mathrm{ik}}\log(\mathrm{U}_{\mathrm{ik}})\right) \tag{6}$$

When VPC is the maximum value, the clusters are most compact. and when VPE is the minimum value, the clusters are the most separation.

Since many heuristic methods are proposed to improve FCM algorithm to obtain maximum compaction and separation.

3. Proposed algorithm

We present a new fuzzy clustering algorithm based on coulomb's law. This algorithm seeks to offer the most compactness of data into clusters and to satisfy the maximum separation between clusters. The algorithm is usedto calculate the membership degrees of data clustering and updating the membership degree matrix of a heuristic function based on coulomb's law.

Proposed algorithm like FCM algorithm followed by minimizing the following function:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n} U_{ik}^{m}\, d_{ik}^{2} = \sum_{i=1}^{c}\sum_{k=1}^{n} U_{ik}^{m}\, \|X_k - V_i\|^2$$

Where (n) is data vector of objects, (c) is number of clusters, V is cluster centers matrix and (U) is membership function matrix that this membership function matrix have to satisfy follows constraints:

I. $\mu_{\mathrm{ik}} \in [0,1];\ 1 \le \mathrm{i} \le \mathrm{c}, 1 \le \mathrm{k} \le \mathrm{n}$ (7)

II. $\sum_{\mathrm{i}=1}^{\mathrm{c}} \mu_{\mathrm{ik}} = 1;\ 1 \le \mathrm{k} \le \mathrm{n}$ (8)

III. $0 < \sum_{\mathrm{k}=1}^{\mathrm{n}} \mu_{\mathrm{ik}} < \mathrm{n};\ 1 \le \mathrm{i} \le \mathrm{c}$ (9)

The memberships function and center matrix update equations all obtained through alternating optimization. The steps of the FCM method are presenting as follows:

Step1. Select appropriate values for m, c, and a small positive number . initialize the center matrix V randomly. Set step variable t=0;

Step2. Update the membership matrix U by

$$U_{ij}^{(t+1)} = \frac{\left[\sqrt{q_i q_j}\Big/D^2(q_i, q_j)\right]^{m-1}}{\left(\sum_{l=1}^{c}\left[q_i q_l \big/ D^2(q_j, q_l)\right]^{m-1}\right)} \tag{10}$$

Step3. Update the center matrix V by

$$V_i = \frac{\sum_{k=1}^{n} U_{ik}^{m} q_j}{\sum_{k=1}^{n} U_{ik}^{m}} \text{ ; for i=1,…,c;} \tag{11}$$

Step4. Repeat steps 2 and 3 until max$\{\|U_{t+1} - U_t\| \leq \varepsilon\}$.

Calculate objective function J(p) as follows:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{m}\, d_{ik}^{2} = \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{m}\, \|X_k - V_i\|^2$$

4. Experimental results

4.1. Evaluating measures

There are many ways to evaluate how exactly our clustering algorithm performs.We can use many validity indexes to evaluate and compare performance between our algorithm and FCM algorithm. Therefore,we use Bezdek's validity indexes for measuring how exactly of our method.

4.2. Iris Data set and Wine Data set

In this section, we examine performance of proposed algorithm and compare it with FCM algorithm on the Iris and Wine database available in the UC Irvine Machine Learning Repository.The Iris database contains 150 Number of Instances, each instance has 4Number of Attributes: 4 numeric, predictive attributes and the class and the Wine Database contains 178 number of Wine recognition data kind of wines, each kind of wine has 13 Number of Attributes. Figure (1) shows the comparison results of our algorithm and FCM algorithm in Iris data set on (3,4,...,10)number of clusters for VPC value. In number clusters 3,4 our algorithm shows value of VPC better than FCM algorithm.

cluster number	VPC FCM	VPC Our algorithm
3	0.783397455	0.785937628
4	0.682650808	0.707957301
5	0.665761604	0.658932802
6	0.606769741	0.594392416
7	0.592457273	0.552972069
8	0.515994968	0.515126847
9	0.505575483	0.532902197
10	0.493301794	0.480413626

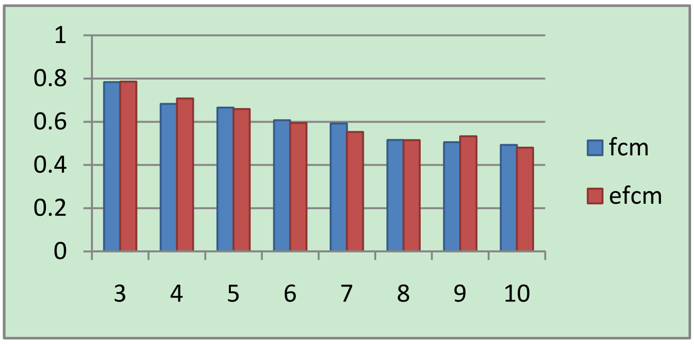

Fig 1.comparing VPC result between FCM and EFCM on Iris data set

Fig 2 shows the comparison results of our algorithm and FCM algorithm in Iris data set on (3,4,...,10) number of clusters for VPE value.In number clusters 3,4 our algorithm shows value of VPE better than FCM algorithm.

cluster number	VPE FCM	VPE Our algorithm
3	0.395491604	0.391388069
4	0.584104143	0.558889801
5	0.675143102	0.685806842
6	0.816978436	0.799686461
7	0.885942127	0.913040779
8	1.005125227	1.003042256
9	1.069546107	1.067655878
10	1.123926595	1.133000182

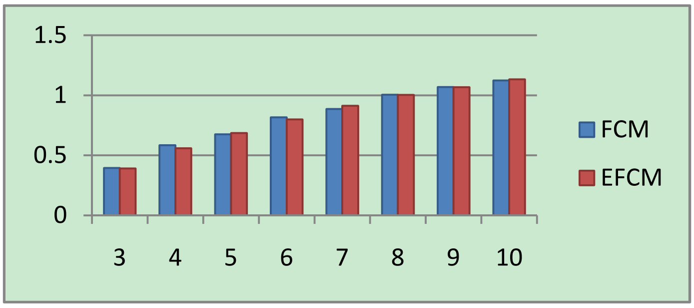

Fig 2.comparing VPE result between FCM and EFCM on Iris data set

Fig 3 shows the comparison results of our algorithm and FCM algorithm in Iris data set on (3,4,...,10)number of clusters for Objective function value.

cluster number	Objective function	
	FCM	Our algorithm
3	6050.571063	6070.610979
4	4956.572613	4172.882985
5	3273.280357	3282.031092
6	2791.493611	2480.821758
7	2421.228631	2044.637526
8	1803.964363	1809.797136
9	1552.704392	2002.155145
10	1397.982389	1371.472439

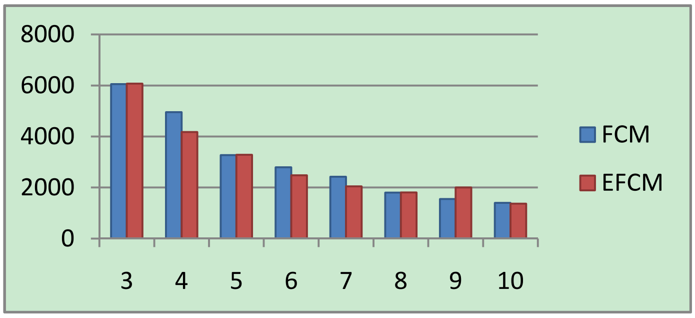

Fig 3.comparing VPE result between FCM and EFCM on Iris data set

Fig 4 shows the comparison results of our algorithm and FCM algorithm in Wine data set on (3,4,...,10) number of clusters for VPC value. In number clusters 3,4 our algorithm shows value of VPC better than FCM algorithm.

cluster number	VPC	
	FCM	Our algorithm
3	0.790939753	0.799013319
4	0.782981911	0.783310761
5	0.767545241	0.746762379
6	0.74642678	0.747335142
7	0.757808986	0.717532853
8	0.694476051	0.727666625
9	0.716355515	0.715829333
10	0.709314668	0.706538088

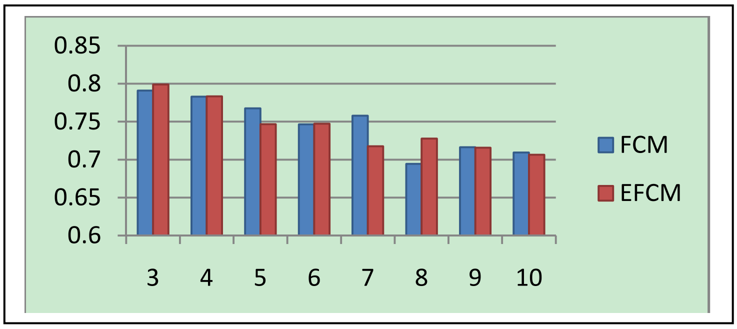

Fig 4.comparing VPC result between FCM and EFCM on Wine data set

Fig 5 shows the comparison results of our algorithm and FCM algorithm in Wine data set on (3,4,...,10) number of clusters for VPE value.In number clusters 3,4 our algorithm shows value of VPE better than FCM algorithm.

cluster number	VPE	
	FCM	Our algorithm
3	0.380407767	0.370904939
4	0.418337915	0.439005255
5	0.464952665	0.503812946
6	0.521833757	0.521349977
7	0.505360801	0.597685377
8	0.653683355	0.581739185
9	0.612783795	0.60985028
10	0.632313414	0.638376505

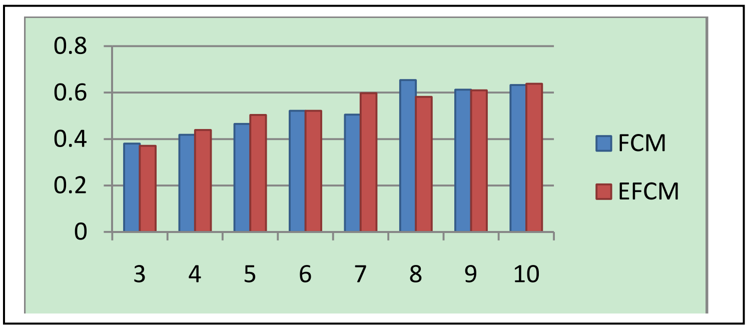

Fig 5.comparing VPE result between FCM and EFCM on Wine data set

5. Conclusion

In generally our algorithm ideas come from similarity between specifications of electrical fields and clustering algorithms and our proposed algorithm based on heuristic approach. In this paper, we present a new clustering algorithm based on coulomb's law that provide 2 validity indexes of Bezdek (VPC, VPE) compare with FCM algorithm.

We provide separation and compactness of clusters more than FCM method. Experimental results on document datasets suggest the success of our methodology.

6. References

[1] N. Blacel, P.Hansen, N.Mladenovic, fuzzy j-means: a new heuristic for fuzzy clustering, pattern recognition 35(2002)

[2] L.A.Zadeh, fuzzy sets, inf. Control 8(1965) 338-353

[3] J.C.Bezdek, “pattern recognition with fuzzy objective function algorithms”, plenum press, new york, (1981)

[4] Gath I, Geva AB, unsupervised optimal fuzzy clustering. IEEE Trans Pami (1991)

[5] P.R Kersten, L.S. Lee, T.L. Ainsworth, Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering, IEEE Trans. Geosci. Remote sensing 43(3) (2005)

[6] Wu J, Yang Z. A fuzzy subspace clustering for high dimensional categorical data. (2006)

[7] D.O¨zdemir, L.Akarun, A fuzzy algorithm for color quantization of images, PatternRecognition35 (2002)1785–1791.

[8] K.L. Wu, J. Yu, M.S. Yang, A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests, Pattern Recognition Lett.26(2005) 639–652.

[9] Cheng, C, FU, A, and Zhang, Y. Entropy-based subspace clustering for mining numerical data.ACM SIGKDD. New York (1999).

[10] Zimmermann, H.J, Hans-Jürgen, Fuzzy set theory and its applications, 2nd, Revised Edition, 1991.

[11] Roberts, Elliott, Electromagnetics History theory and applications. ISBN 978-0-7803-5384-8.

[12] James Clerk Maxwell,ed. The electrical researches ofthe honorable Cavendish, Cambridge University Press.

[13] Dubes, R.C, jain, A.k , algorithm for clustering data , Englewood cliffs,prentiee Hall,1988.