Electrical fuzzy C-means: A new heuristic fuzzy clustering algorithm

Esmaeil Mehdizadeh and Amir Golabzaei

Abstract: Many heuristic and meta-heuristic algorithms have been successfully applied in the literature to solve the clustering problems. The algorithms have been created for partitioning and classifying a set of data because of two main purposes: at first, for the most compact clusters, second, for the maximum separation between clusters. In this paper, we propose a new heuristic fuzzy clustering algorithm based on electrical rules. The laws of attraction and repulsion of electric charges in an electric field are conducted the same as the target of clustering. The electrical fuzzy C-means (EFCM) algorithm proposed in this article use the electrical rules in electric fields and Coulomb's law to obtain the better and the realest partitioning, having respect to the maximum separation of clusters and the maximum compactness within clusters. Computational results show that our proposed algorithm in comparison with fuzzy C-means (FCM) algorithm as a well-known fuzzy clustering algorithm have good performance.

Keywords: fuzzy clustering; fuzzy C-means; validity index; electrical rules; Coulomb's law
1. Introduction

Form a general point of view, pattern recognition is defined as the process of searching for data structures and the related classification into certain categories, in which the association among the intra-categorical and inter-categorical structures is high and low, respectively. Clustering is the most fundamental and significant issue in pattern recognition and is defined as a form of data compression, in which a large number of samples are converted into a small number of representative prototypes or clusters (Klir & Yuan, 2003). Clustering plays a key role in searching for structures in data and involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible (Mehdizadeh, Sadi-Nezhad, & Tavakkoli-Moghaddam, 2008). It is a method creating groups of objects so that objects within one cluster are similar and objects in different clusters are dissimilar. In the last few years clustering has played a critical role in different domain of science and engineering applications such as image processing (Xia, Feng, Wang, Zhao, & Zhang, 2007; Yang, Wu, Wang, & Jiao, 2010), anomaly detection (Friedman, Last, Makover, & Kandel, 2007; Moshtaghi et al., 2011), medicine (Liao, Lin, & Li, 2008), construction management (Cheng & Leu, 2009), marketing (Kim & Ahn, 2008), data retrieval (Abraham, Das, & Konar, 2006; Mahdavi, Haghir Chehreghani, Abolhassani, & Forsati, 2008; Gil-García & Pons-Porrata, 2010), reliability (Taboada & Coit, 2007), portfolio optimization (Chen & Huang, 2009), cell formation problem (Mehdizadeh & Tavakkoli-Moghaddam, 2009; Mehdizadeh, 2009) selecting supplier (Che, 2012; Mehdizadeh & Tavakkoli-Moghaddam, 2007), supplier clustering (Mehdizadeh, 2009) and data envelopment analysis (Po, Guh, & Yang, 2009; Ben-Arieh & Gullipalli, 2012), support vector machines (Sabzekar & Naghibzadeh, 2013).

In real-world cases, there are very often no sharp boundaries between clusters so that fuzzy clustering will be a better choice for the data. Membership degrees between zero and one are used in fuzzy clustering instead of crisp assignments of data to clusters. In non-fuzzy (crisp environment) or hard clustering, data are divided into crisp clusters, whose data point belongs to exactly one cluster. In fuzzy clustering, these data points can belong to more than one cluster, under these circumstances, the membership grades of each of the data points represent the degree to which the point belong to each cluster (Mehdizadeh & Tavakkoli-Moghaddam, 2009).

In literature, many algorithms such as heuristic and meta-heuristic have been proposed for solving fuzzy clustering problems. One of the most applicable methods of fuzzy clustering is fuzzy C-means (FCM) algorithm. FCM is an efficient tool used for fuzzy clustering problems. This method has been successfully adapted to solve the fuzzy clustering problem. However, this problem is a combinatorial optimization problem (Zimmermann, 1996) and if the data-sets are very high dimensional or contain severe noise points, the FCM often fails to find the global optimum. In these cases, the probability of finding the global optimum can be increased by improving FCM with inspiration by natural rules. In this paper, to skip the local optimum, the FCMs algorithm is combined with the electrical rules and a new algorithm called Electrical EFCM algorithm is presented for solving fuzzy clustering problem.

In this article, the proposed fuzzy clustering algorithm uses the electrical rules and electric potential energy lies in electric fields and Coulomb’s law to obtain the realest partitioning, having respect to the maximum separation of clusters and the maximum compactness within clusters. Clustering algorithms have been created for partitioning and classify a set of data because of two main purposes: first, for the most compact clusters, second, for the maximum separation between clusters. The laws of attraction and repulsion of electric charges in an electric field are conducted exactly the same as the target of clustering. Thus the charge with a negative charge at the center of a cluster of positively charged clusters are absorbed and the positively charged cluster of data centers and other negatively charged clusters, there is gravity. Clusters that formed by clustering algorithms act like the electrical loads. Different clusters have a repulsive force or a good separation between the data and they are trying to provide better compact within clusters. The proposed algorithm (EFCM) starts in a way that randomly chose the initial centers and ends by computing a unique objective function. In each iteration of this algorithm, the existing data is displayed at the beginning of the computation center of each iteration.
The remaining of this paper is organized as follows: Section 2 presents the literature review. An overview of the FCM algorithm, Coulomb’s low, and electric potential energy is discussed in Section 3. In Section 4 we introduce our proposed EFCMs algorithm. Experimental results are summarized in Section 5. Finally, discussion and conclusions are presented in Section 6.

2. Literature review
Clustering has a long history, dating back to Aristotle (Blacel, Hansen, & Mladenovic, 2002). Clustering algorithms allocated each object to a cluster that this is the most popular problems for crisp clustering. Fuzzy logic (Zadeh, 1965) creates approximate clustering rather than crisp clustering by using fuzzy clustering problem. This problem is solved and the object can allocate to all of the clusters with a certain degree of membership (Bezdek, 1981). In literature, many algorithms such as heuristic and meta-heuristic have been proposed for solving fuzzy clustering problems. One of the most applicable methods of fuzzy clustering is FCM algorithm. The first version of the C-means algorithms was presented by Duda and Hart (1973) that known as a hard clustering algorithm. In real word, some of the data belong to multiple clusters. In order to study this problem, Dun (1974) proposed one of the first fuzzy clustering methods based on the objective function and using Euclidean distance. This algorithm was revised several times but its final version was given by Bezdek, Ehrlich and Full (1984). For solving elliptical clustering problem, Gustafson and Kesel (1979) proposed a new fuzzy clustering algorithm using covariance matrix. They used another criterion for determining the distance instead of Euclidean distance. Required normalization of the membership degrees and the sum of membership degrees being equal to one in the fuzzy clustering algorithm lead to adverse effects in clustering on the outlying and thrown away from center data. To solve this problem, the possibility clustering algorithm was proposed by Dubois and Prade (1988) and then was corrected by Krishnapuram and Keller (1993). Many reforms and improvements have been executed on this algorithm and a general algorithm for solving the problem of distinguishing various forms have been proposed by Hathaway and Bezdek (1994). The results of FCM algorithm were greatly affected by the data scattered away from the center. To solve this, a lot of algorithms have been proposed to improve the objective function (Dave, 1991; Dave & Andsen, 1997, 1998; Frigui & Krishnapuram, 1996). Up to this time, for improving the Fuzzifier value on these algorithms according to the effectiveness of the Fuzzifier value, many algorithms were proposed based on FCM algorithm (Klawonn, 2004; Klawonn & Hoppner, 2003; Rousseeuw, Trauwaert, & Kaufman, 1995). Possibilistic fuzzy C-means (PCM) algorithms with optimizing the objective function leads to clusters that are not perceptible separation, clustering algorithm with modified objective function were proposed to avoid merging clusters which explain repulsion of clusters (Timm & Kruse, 2002; Timm, Borgelt, Do¨Ring, & Kruse, 2004). As the fact that a good clustering requires a good partitioning and a minimum objective function values, so many algorithm were proposed to improve PCM algorithms with respect to their possibility degree and membership degree (Pal, Pal, & Bezdek, 1997; Pal, Pal, Keller, & Bezdek, 2004). Before starting the FCM algorithms, it is necessary to use the number of clusters. Its performance strongly depends on the initial centroids’ vectors and may get stuck in local optima solutions (Selim & Ismail, 1984). Generally speaking, we have never seen any difficulty and in ten to twenty-five iterations, we achieve numerical convergence. Another topic in FCM is relationship between local minima of objective function and clustering of data-set. Because of the dependence of FCM objective function to initial state, the results, usually, converge to local optimum. For declining this difficulty, on each membership matrix by FCM, it is calculated by several types of cluster validity (Bezdek, 1974). For speeding up FCM and improving drawback of FCM, researchers proposed optimization approaches for fuzzy partitioning. Some of these methods have improved FCM algorithm partitioning and some of these researches improved FCM algorithm to determine the optimal number of clusters and study on validity indexes of fuzzy clustering (Dubois & Prade, 1988; Wu, Xie, & Yu, 2003).

In recent years, researchers studied on fuzzy clustering algorithm to improve relationship between clusters in objective function with different approaches (Frigui & Krishnapuram, 1996; Keller, 2000; Klawonn & Hoppner, 2003). Some of them have used heuristic and meta-heuristic methods
which inspired by social and natural rules in order to optimize the objective function of FCM. For example, Bezdek and Hathaway (1994) optimized the hard C-means (HCM) method with a genetic algorithm. Klawonn and Keller (1998) extended and applied this scheme to the FCM model. In addition, ant colony optimization (ACO) has been successfully applied to clustering problems (Handl, Knowles, & Dorigo, 2003). Similar heuristic algorithms, called ant clustering, were suggested by Kanade and Hall (2003) and (2004). Runkler (2005) introduced an ACO algorithm that explicitly minimizes the HCM and FCM cluster models. Runkler and Katz (2006) applied PSO to cluster data by considering fuzzy clustering. They introduced two new methods to minimize the two reformulated versions of the FCM objective function by PSO. A hybrid PSO and FCM clustering algorithm has been applied to clustering problem (Mehdizadeh et al., 2008). Sabzekar and Naghibzadeh (2013) used an implementation of support vector machines, namely relaxed constraints support vector machines to improve the performance of FCM algorithm. There are two hypotheses in clustering problems: (1) the most compact clusters and (2) the maximum separation between clusters. In this article, for improving the performance of FCM algorithm, we propose a new heuristic algorithm for fuzzy clustering problem based on electrical rules and called EFCMs.

3. EFCMs phenomenon

The EFCMs algorithm proposed in this article use the electrical potential energy in electric fields, Coulomb’s law, and FCM algorithm to obtain the best and the realest partitioning, having respect to the maximum separation of clusters and the maximum compactness within clusters. In this section, a brief explanation of electrical rules applied to EFCM and FCM algorithm is presented.

3.1. FCM algorithm

FCM is one of the common algorithms of fuzzy clustering methods that proposed by Bezdek (1981) and aims to find fuzzy partitioning of a given data-set by minimizing of the basic C-means objective functional as shown in Equation (1):

\[ J_p = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m d_{ik}^2 = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m \| X_k - V_i \|^2 \]  

(1)

where \( c \) is the number of clusters; \( n \) is the number of data point; the parameter \( m \) is a real number that governs the influence of membership grades. The partition becomes fuzzier with increasing \( m \); \( V_i \) is the cluster center of cluster \( i \); \( X_k \) is the vector of data point; \( \| X_k - V_i \|^2 \) represents the Euclidean distance between \( X_k \) and \( V_i \). The classification result can be expressed in terms of matrix \( U = [u_{ik}]_{c \times n} \) where \( u_{ik} \) is the membership degree of data point \( k \) to cluster \( i \) which satisfying Equations (2)–(4):

\[ 0 \leq u_{ik} \leq 1; \quad i = 1, 2, \ldots, c; \quad k = 1, 2, \ldots, n \]  

(2)

\[ \sum_{i=1}^{c} u_{ik} = 1; \quad k = 1, 2, \ldots, n \]  

(3)

\[ 0 < \sum_{k=1}^{n} u_{ik} \leq n; \quad i = 1, 2, \ldots, c \]  

(4)

Fuzzy segmentation which used an iterative procedure is achieved with the update of membership \( u_{ik} \) with Equation (5) and cluster centers \( V_i \) by Equation (6):
The steps of the FCM method can be summarized in the following algorithm:

Step 1: Initialize the membership matrix \( U \) with random values between 0 and 1 such that the constraints in Equations (2)–(4) are satisfied.

Step 2: Calculate fuzzy cluster centers \( \mathbf{V}_i \), \( i = 1, \ldots, c \) by using Equation (6).

Step 3: Compute the objective function according to Equation (1).

Step 4: Compute a new membership matrix \( U \) by using Equation (5).

Step 5: Go to Step 2.

The iterations stop when the difference between the fuzzy partitions matrices in two following iterations is lower than \( \varepsilon \).

3.2. Coulomb's law

We have been used electrical rules that described electrostatic actions between electrical points. Firstly, this rule was proposed by Charles Coulomb (Tipler, 2004). Every two charged objects will have a force on each other. Opposite charges will produce an attractive force while similar charges will produce a repulsive force. For two spherically shaped charges the formula would look like:

\[
F = \frac{kq_1q_2}{r^2}
\]

where \( q_1 \) represents the quantity of charge on object 1 (in Coulomb's), \( q_2 \) represents the quantity of charge on object 2 (in Coulomb's), and \( r \) represents the distance of separation between the two objects (in meters). The symbol \( k \) is proportionality constant and known as the Coulomb's constant. We show how these electrical points become closer the electrical energy in Figure 1.

![Figure 1. Electrostatic force between electric charges.](image-url)
4. EFCMs algorithm

The EFCMs algorithm presented in this paper is a new fuzzy clustering algorithm based on electrical rules especially Coulomb’s law. This algorithm works on the assumption that data are negative electrical charges, cluster centers are positive electrical charges, and fuzzy clusters are considered as electrical fields in an n-dimensional space. The negative electrical charges (data) in an electrical field (cluster) not only force to positive electrical charges (cluster center) in same field, but also force to positive electrical charges (cluster center) in other electrical field (cluster) as shown in Figure 2:

This algorithm is similar to FCM algorithm in minimizing the objective function. The objective function is described as follows:

The first part is derived from the incoming forces from each cluster to the cluster centers (electric field) in that it shows a high density within the cluster. In this objective function we use the membership degree of objects to show the impact of the separation between clusters which has a direct impact on the objective function. This algorithm seeks to offer the most compactness of data into clusters and to satisfy the maximum separation between clusters. The algorithm used to calculate the membership degrees of data clustering and updating the membership degree matrix of a heuristic function is based on Coulomb’s law. The proposed algorithm following by minimization of function $J(G)$ is shown in Equation (8):

$$J(G) = \left[ \sum_{k=1}^{c} \left( \frac{1}{\sum_{l=1}^{n} d^2(q_l, q_{\text{max}_l})} \right)^{m-1} \right]$$

(8)

The memberships function and center matrix update equations are as follows:

$$U^{(t+1)}_{ij} = \left[ \frac{\sqrt{n_i \cdot q_i \cdot q_j}}{D(q_i, q_j)} \right]^{m-1}$$

(9)

$$V_k = \frac{\sum_{i=1}^{n} U_{ik}^m \times q_i}{\sum_{k=1}^{n} U_{ik}^m}$$

(10)

where $n$ is data vector of objects, $c$ is number of clusters, $V$ is cluster centers matrix, $q_i$ is the center of cluster $k$, $q_{\text{max}_l}$ is the data with the highest degree of membership in a cluster $k$, $U_{ik}$ is the membership function of $i$th object in $k$th cluster, and $U$ is membership function matrix that this membership function matrix have to satisfy Equations (3)–(5).
The steps of the EFCM method can be summarized in the following algorithm:

Step 1: Initialize the membership matrix $U$ with random values between 0 and 1 such that the constraints in Equations (2)–(4) are satisfied.

Step 2: Calculate fuzzy cluster centers $V_i, i = 1, \ldots, c$ by using Equation (10).

Step 3: Compute the objective function according to Equation (8).

Step 4: Compute a new membership matrix $U$ by using Equation (9).

Step 5: Repeat Steps 2 to 4 until, $\max |U^{t+1} - U^t| \leq \varepsilon$. 

And finally:

Calculate the objective function $J(G)$ by Eq. (8)
We show the pseudo-code of our algorithm in Figure 3 and present the flowchart of the EFCM algorithm in Figure 4.

5. Experimental results

In this section, we examine performance of our proposed algorithm and compare it with FCM as a well-known fuzzy clustering algorithm on the test data-sets available in the UC Irvine Machine learning repository (Blake & Merz, 1998). We use Bezdek’s validity indexes for measuring the exactness of our proposed method. All of our experiments have been implemented using MATLAB (R2011a) running on a computer with an Intel processor (Dual-core, 1.86 GHz) and 1 GB of memory.

5.1. Validity indexes

Unlike crisp clustering, fuzzy clustering allows each data point that belongs to whole clusters with a special degree of membership. It can define validity indexes for fuzzy clustering in order to seek clustering schemes in which most of the data points in the data-set exhibit a high degree of membership in one cluster. The well-known validity indexes to evaluate cluster validity which proposed by Bezdek (1974) are as follows:

The partition coefficient:

\[ \text{PC}(U) = \frac{1}{n} \left( \sum_{k=1}^{c} \sum_{i=1}^{n} U_{ik}^2 \right) \]  

(11)

The partition entropy:

\[ \text{PE}(U) = -\frac{1}{n} \left( \sum_{k=1}^{c} \sum_{i=1}^{n} U_{ik} \log(U_{ik}) \right) \]  

(12)

We can see from Equation (11), the value of the PC index range is in \([1/c, 1]\). The closer the value of PC to 1/c, the fuzzier the clustering is. The lower value of PC is obtained when \(U_{ik} = 1/c\). In other words, when PC is the maximum value, the clusters are the most compact. In addition, from the definition in Equation (12), the value of the PE index range is in \([0, \log c]\). The closer the value of PE to 0, the harder the clustering is. The values of PE close to the upper bound indicate that there is no clustering structure in the data-set or the algorithm is enabling to extract it. In other words when PE is the minimum value, the clusters is the most separation.

5.2. Data-sets

In order to evaluate the performance of our proposed algorithm, seven benchmark data-sets are used. The data-sets are Soybean (small), Dermatology, Breast cancer, Iris, Wine, Ecoli, and Pima, which are available in the repository of the machine learning databases (Blake & Merz, 1998). Table 1 summarizes the main characteristics of the used data-sets. The ten data-sets are described in Table 1.

| Data-set           | Number of data objects | Number of features | Number of clusters |
|--------------------|------------------------|--------------------|--------------------|
| Soybean (small)    | 47                     | 35                 | 4                  |
| Dermatology        | 366                    | 33                 | 6                  |
| Breast cancer      | 699                    | 9                  | 2                  |
| Iris               | 150                    | 4                  | 3                  |
| Wine               | 178                    | 13                 | 3                  |
| Ecoli              | 336                    | 7                  | 8                  |
| Pima               | 768                    | 8                  | 2                  |
5.3. Results for data-sets in terms of validity indexes
In this subsection, we display the effectiveness of our proposed algorithm which obtained based on validity indexes on data-sets described in Section 5.1. Tables 2 and 3 report the value of PC and PE as validity indexes for our proposed algorithm and FCM algorithm on mentioned data-sets in Table 1. In order to compare two algorithms, mean, best, and standard deviation of each algorithm for 50 runs computed and reported. It can be seen that EFCM algorithm has better mean, best, and standard deviation in terms of PC and PE than FCM algorithm.

Tables 4 and 5 show the comparison results of the two algorithms namely FCM and EFCM algorithms for PC and PE with different cluster number values for 50 runs. The means of the indexes for two algorithms have been reported. The best value of each data-set highlighted. Results for both indexes show that EFCM algorithm has relatively better performance for different values of cluster number.

In the Appendix A, the tables show the obtained best centroids from EFCM algorithms for the Lung cancer, Soybean (small), Dermatology, Credit approval, Breast cancer, Iris, Wine, Zoo, Ecoli, and Pima data-sets. These centroids are introduced for validating the values obtained in Tables 2 and 3. Therefore, by assigning each data-set to its center in the tables, the value of that data-set would be obtained. For example, by assigning all the objects of the Lung cancer data-set to the centroids in Tables 2 and 3, the best value for Lung cancer data-set, which is reported in Tables 2 and 3, should be obtained by EFCM algorithm. This procedure can be used for checking other data-sets.

| Table 2. The comparison of EFCM algorithm with FCM algorithms in terms of PC for 50 runs |
|---------------------------------------------|--------|--------|--------|--------|--------|--------|
| Data-set                  | FCM    |        | EFCM   |        |
|                           | Mean   | Best   | Std. Dev. | Mean   | Best   | Std. Dev. |
|---------------------------|--------|--------|-----------|--------|--------|-----------|
| Soybean (small)           | 0.4756 | 0.47561| 4.00E-05  | 0.5028 | 0.50288| 2.10E-09  |
| Dermatology               | 0.4569 | 0.47123| 4.98E-06  | 0.4650 | 0.47687| 2.56E-06  |
| Breast cancer             | 0.9154 | 0.91546| 7.61E-11  | 0.9154 | 0.91550| 2.33E-08  |
| Iris                      | 0.78339| 0.78340| 2.54E-07  | 0.7842 | 0.78430| 1.42E-07  |
| Wine                      | 0.7909 | 0.79094| 1.52E-07  | 0.7990 | 0.79901| 1.22E-06  |
| Ecoli                     | 0.2832 | 0.31035| 0.00105   | 0.3191 | 0.31920| 0.00105   |
| Pima                      | 0.8242 | 0.82423| 9.72E-09  | 0.8274 | 0.82756| 6.53E-09  |

| Table 3. The comparison of EFCM algorithm with FCM algorithms in terms of PE for 50 runs |
|---------------------------------------------|--------|--------|--------|--------|--------|--------|
| Data-set                  | FCM    |        | EFCM   |        |
|                           | Mean   | Best   | Std. dev. | Mean   | Best   | Std. dev. |
|---------------------------|--------|--------|-----------|--------|--------|-----------|
| Soybean (small)           | 0.9749 | 0.97472| 4.88E-05  | 0.9277 | 0.92770| 1.70E-16  |
| Dermatology               | 1.0941 | 1.06628| 6.99E-06  | 1.0666 | 1.04317| 1.58E-08  |
| Breast cancer             | 0.1538 | 0.15381| 1.76E-10  | 0.1496 | 0.14962| 1.24E-10  |
| Iris                      | 0.3955 | 0.39540| 1.80E-07  | 0.3914 | 0.39139| 1.34E-07  |
| Wine                      | 0.38041| 0.38041| 1.73E-07  | 0.3707 | 0.37090| 1.05E-10  |
| Ecoli                     | 1.5215 | 1.53419| 0.01092   | 1.5210 | 1.52099| 0.00105   |
| Pima                      | 0.29682| 0.29683| 1.38E-08  | 0.2910 | 0.29122| 1.53E-11  |
### Table 4. Means of PC for 50 runs with respect to different cluster number on the fourteen data-sets

| Cluster number | Algorithm | Soybean (small) | Dermatology | Breast cancer | Iris | Wine | Ecoli | Pima |
|---------------|-----------|-----------------|--------------|---------------|------|------|-------|------|
| 2             | FCM       | 0.6754          | 0.7905       | 0.9155        | 0.8922 | 0.8761 | 0.6586 | 0.8242 |
|               | EFCM      | 0.6969          | 0.7916       | 0.9155        | 0.9043 | 0.8847 | 0.7119 | 0.8275 |
| 3             | FCM       | 0.6150          | 0.6770       | 0.7684        | 0.7836 | 0.7909 | 0.5700 | 0.7635 |
|               | EFCM      | 0.6406          | 0.6981       | 0.7663        | 0.7842 | 0.7990 | 0.6228 | 0.7635 |
| 4             | FCM       | 0.4756          | 0.5792       | 0.7852        | 0.7068 | 0.7830 | 0.4482 | 0.6639 |
|               | EFCM      | 0.5029          | 0.6271       | 0.7842        | 0.7080 | 0.7723 | 0.4960 | 0.6639 |
| 5             | FCM       | 0.3787          | 0.5261       | 0.7760        | 0.6658 | 0.7470 | 0.3966 | 0.5358 |
|               | EFCM      | 0.4270          | 0.5312       | 0.7657        | 0.6310 | 0.7467 | 0.4356 | 0.5523 |
| 6             | FCM       | 0.3626          | 0.4569       | 0.7787        | 0.5853 | 0.7374 | 0.3446 | 0.5251 |
|               | EFCM      | 0.3740          | 0.4650       | 0.7519        | 0.5818 | 0.7473 | 0.3801 | 0.5232 |
| 7             | FCM       | 0.2945          | 0.4044       | 0.7687        | 0.5923 | 0.7154 | 0.3127 | 0.4395 |
|               | EFCM      | 0.3507          | 0.4057       | 0.7788        | 0.5530 | 0.7175 | 0.3472 | 0.4757 |
| 8             | FCM       | 0.2748          | 0.3611       | 0.7446        | 0.5160 | 0.6945 | 0.2832 | 0.4217 |
|               | EFCM      | 0.3359          | 0.4023       | 0.7820        | 0.5426 | 0.7276 | 0.3191 | 0.4313 |
| 9             | FCM       | 0.2478          | 0.3247       | 0.7297        | 0.4959 | 0.7163 | 0.2640 | 0.3647 |
|               | EFCM      | 0.3251          | 0.3613       | 0.7436        | 0.5114 | 0.7158 | 0.2908 | 0.4013 |
| 10            | FCM       | 0.2734          | 0.2979       | 0.7369        | 0.4724 | 0.7092 | 0.2487 | 0.3701 |
|               | EFCM      | 0.3279          | 0.3205       | 0.7427        | 0.4804 | 0.7065 | 0.2767 | 0.3651 |

### Table 5. Means of PE for 50 runs with respect to different cluster number on the fourteen data-sets

| Cluster number | Algorithm | Soybean (small) | Dermatology | Breast cancer | Iris | Wine | Ecoli | Pima |
|---------------|-----------|-----------------|--------------|---------------|------|------|-------|------|
| 2             | FCM       | 0.5019          | 0.3451       | 0.1538        | 0.1957 | 0.2160 | 0.4062 | 0.2968 |
|               | EFCM      | 0.4762          | 0.3428       | 0.1496        | 0.1784 | 0.2123 | 0.4487 | 0.2912 |
| 3             | FCM       | 0.6993          | 0.5792       | 0.4017        | 0.3955 | 0.3804 | 0.6231 | 0.4271 |
|               | EFCM      | 0.6600          | 0.5413       | 0.4038        | 0.3914 | 0.3709 | 0.6763 | 0.4474 |
| 4             | FCM       | 0.9749          | 0.7856       | 0.4014        | 0.5611 | 0.6184 | 0.8803 | 0.6251 |
|               | EFCM      | 0.9277          | 0.7031       | 0.3992        | 0.5589 | 0.4390 | 0.9460 | 0.6468 |
| 5             | FCM       | 1.1796          | 0.9244       | 0.4457        | 0.6752 | 0.5098 | 1.0216 | 0.8617 |
|               | EFCM      | 1.1174          | 0.9060       | 0.4617        | 0.6975 | 0.5038 | 1.1071 | 0.8478 |
| 6             | FCM       | 1.3564          | 1.0941       | 0.4387        | 0.8102 | 0.5311 | 1.1772 | 0.9340 |
|               | EFCM      | 1.2916          | 1.0666       | 0.4775        | 0.8150 | 0.5214 | 1.2755 | 0.9428 |
| 7             | FCM       | 1.5135          | 1.2381       | 0.4779        | 0.8856 | 0.5980 | 1.2989 | 1.1311 |
|               | EFCM      | 1.4084          | 1.2205       | 0.4433        | 0.9130 | 0.5978 | 1.4007 | 1.0756 |
| 8             | FCM       | 1.5650          | 1.3654       | 0.5362        | 1.0051 | 0.6537 | 1.1968 | 1.0899 |
|               | EFCM      | 1.5031          | 1.2533       | 0.4455        | 0.9669 | 0.5818 | 1.5210 | 1.1949 |
| 9             | FCM       | 1.6174          | 1.4837       | 0.5672        | 1.0735 | 0.6129 | 1.4902 | 1.3685 |
|               | EFCM      | 1.5503          | 1.3775       | 0.5235        | 1.0581 | 0.6100 | 1.6154 | 1.2980 |
| 10            | FCM       | 1.7385          | 1.5837       | 0.5531        | 1.1468 | 0.6327 | 1.5703 | 1.3919 |
|               | EFCM      | 1.5871          | 1.5004       | 0.5425        | 1.1330 | 0.6385 | 1.5958 | 1.4108 |
Figures 5 and 6 show the convergence diagrams of EFCM algorithm for the best solutions for Iris and Wine data-sets. The Figures 5 and 6 show that our algorithm converges to a global optimum for Iris data-set in 28 iterations and for Wine data-set in 49 iterations.

6. Discussion and conclusions
Clustering is a useful technique both for data mining and data analyzing. Clustering algorithms were created for partitioning and classify a set of data because of two main purposes: the most compact clusters and the maximum separation between clusters. The laws of attraction and repulsion of electric charges in an electric field are conducted the same as the target of clustering. In this article, we have proposed a new heuristic clustering algorithm for fuzzy clustering problem based on electrical rules and Coulomb’s law and called EFCMs. The proposed algorithm compared with FCM algorithm as a well-known fuzzy clustering algorithm based on two validity indexes proposed by Bezdek. The results show that the EFCM algorithm is able to achieve better solutions when compared with FCM algorithm. However, this algorithm is a heuristic algorithm and proposed to solve fuzzy
clustering problem, we can apply it for solving a wide range of combinatorial problems. The proposed algorithm is not an exact method. In future researches this algorithm can be used in hard clustering problem and compared with other hard clustering methods. Also, this algorithm can be combined with other fuzzy clustering algorithms or meta-heuristic algorithms.

Funding
The authors received no direct funding for this research.

Author details
Esmoel Mehdizadeh
E-mail: emehdi@qiau.ac.ir
ORCID ID: http://orcid.org/0000-0002-1149-905X
Amir Golabzaei
E-mail: amirgolabzaee@yahoo.com
1 Department of Industrial and Mechanical Engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran.
2 Department of Industrial Engineering, Islamic Azad University, Science and Research Branch, Saveh, Iran.

Citation information
Cite this article as: Electrical fuzzy C-means: A new heuristic fuzzy clustering algorithm, Esmoel Mehdizadeh & Amir Golabzaei, Cogent Engineering (2016), 3: 1208397.

References
Abraham, A., Das, S., & Konar, A. (2006). Document clustering using differential evolution. IEEE Congress on Evolutionary Computation CEC, 34, 1784–1791. http://dx.doi.org/10.1109/CEC.2006.1688523
Ben-Arieh, D., & Gullipalli, D. K. (2012). Data Envelopment Analysis of clinics with sparse data: Fuzzy clustering approach. Computers & Industrial Engineering, 63, 21. Bezdek, J. C., & Hathaway, R. J. (1994). Optimization of fuzzy clustering criteria using genetic algorithm (Vol. 2, pp. 589–594). Orlando, FL: Proceedings of the IEEE Conference on Evolutionary Computation.
Bezdek, J. C. (1974). Cluster validity with fuzzy sets. Journal of Cybernetics, 3, 58–73.
Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy C-means clustering algorithm. Computers & Geosciences, 10, 191–203.
Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4757-0450-1
Biscail, N., Hansen, P., & Mladenovic, N. (2002). Fuzzy J-means: A new heuristic for fuzzy clustering. Pattern Recognition, 35, 2193–2200.
Blake, C. L., & Merz, C. J. (1998). UCI repository of machine learning databases. Retrieved from http://www.ics.uci.edu/~mlearn/MLRepository.html
Che, Z. H. (2012). Clustering and selecting suppliers based on simulated annealing algorithms. Computers and Mathematics with Applications, 63, 228–238. http://dx.doi.org/10.1016/j.camwa.2011.11.014
Chen, L. H., & Huang, L. (2009). Portfolio optimization of equity mutual funds with fuzzy return rates and risks. Expert Systems with Applications, 36, 3720–3727. http://dx.doi.org/10.1016/j.eswa.2008.02.027
Cheng, Y. M., & Leu, S. S. (2009). Constraint-based clustering and its applications in construction management. Expert Systems with Applications, 36, 5761–5767. http://dx.doi.org/10.1016/j.eswa.2008.06.100
Dave, R., & Andsnes, S. (1998). Generalized noise clustering as a robust fuzzy C-M-estimators model (pp. 256–260). Pensacola Beach, FL: Proceeding of the 17th Int. Conference of the North American Fuzzy Information Processing Society: Nafips’98.
Dave, R., & Andsnes, S. (1997). On generalising the noise clustering algorithms (pp. 205–210). Prague: Proceedings of the 7th IFSA World Congress, IFSAS97.
Dave, R. (1991). Characterization and detection of noise in clustering. Pattern Recognition Letters, 12, 657–664. http://dx.doi.org/10.1016/0167-8655(91)90002-4
Dubois, D., & Prade, H. (1988). Possibility Theory: New York: Plenum Press. http://dx.doi.org/10.1007/978-1-4684-5287-7
Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis. New York, NY: Wiley.
Dunn, J. C. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters. Journal of Cybernetics, 4, 95–104. http://dx.doi.org/10.1080/01969727408546059
Friedman, M., Last, M., Makova, Y., & Kandel, A. (2007). Anomaly detection in web documents using crisp and fuzzy-based cosine clustering methodology. Information Sciences, 177, 467–475. http://dx.doi.org/10.1016/j.ins.2006.03.006
Frigui, H., & Krishnapuram, R. (1996). A robust algorithm for automatic extraction of an unknown number of clusters from noisy data. Pattern Recognition Letters, 17, 1223–1232. http://dx.doi.org/10.1016/0167-8655(96)00080-3
Gil-García, R., & Pons-Porrata, A. (2010). Dynamic hierarchical algorithms for document clustering. Pattern Recognition Letters, 31, 469–477. http://dx.doi.org/10.1016/j.patrec.2009.11.011
Gustafson, E. E., & Kessel, W. C. (1979). Fuzzy clustering with a fuzzy covariance matrix (pp. 761–766). San Diego, CA: Proceedings of the IEEE Conference on Decision and Control.
Handl, J., Knowles, J., & Dorigo, M. (2003). Strategies for the increased robustness of ant-based clustering. Engineering Self-organizing Systems, 2977, 90–104.
Hathaway, R., & Bezdek, J. (1996). Nerf C-means: Non-Euclidean relational fuzzy clustering. Pattern Recognition, 27, 429–437. http://dx.doi.org/10.1016/0031-3203(94)90119-8
Kanade, P. M., & Hall, L. O. (2004). Fuzzy ant clustering by centroids (pp. 371–376). Budapest: Proceeding of the IEEE Conference on Fuzzy Systems.
Kanade, P. M., & Hall, L. O. (2003). Fuzzy ants as a clustering concept (pp. 227–232). Chicago, IL: The 22nd International Conference of the North American Fuzzy Information Processing Society NAFIPS.
Keller, A. (2000). Fuzzy clustering with outliers (pp. 143–147). Atlanta, GA: Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society.
Kim, K. j., & Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. Expert Systems with Applications, 34, 1200–1209. http://dx.doi.org/10.1016/j.eswa.2006.12.025
Klawonn, F., & Hoppper, F. (2003). What is fuzzy about fuzzy clustering? Understanding and improving the concept of the fuzzifier Advances In Intelligent Data Analysis, 2810, 254–266.
Klawonn, F., & Keller, A. (1998). Fuzzy clustering with evolutionary algorithms. International Journal of Intelligent Systems, 13, 975–991. http://dx.doi.org/10.1002/(ISSN)1098-111X
Klawonn, F. (2004). Fuzzy clustering: Insights and a new approach. Math ware And Soft Computing, 11, 125–142.
Kim, G. J., & Yuan, B. (2003). Fuzzy sets and fuzzy logic, theory and applications. Upper Saddle River, NJ: Prentice-Hall Co.
The best centroids obtained from EFCM algorithms for the data-set.

**Table 1: Iris Data Centroids**

| Center | 1     | 2     | 3     | 4     |
|--------|-------|-------|-------|-------|
| 1      | 50.02836 | 34.16483 | 14.77087 | 2.509736 |
| 2      | 67.54197 | 30.44726 | 56.17733 | 20.38918 |
| 3      | 58.64082 | 27.52564 | 43.28188 | 13.79873 |

Appendix A

The best centroids obtained from EFCM algorithms for the data-set.
### Soybean (small)

| Center | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1      | 1.742123 | 0.910399 | 1.816506 | 0.362651 | 0.30256 | 2.098615 | 1.062965 | 1.697339 | 0.443504 | 1.320465 | 1     | 0.71412 |
| 2      | 0.712008 | 0.958992 | 1.853161 | 0.245584 | 0.235708 | 1.29679 | 1.06711 | 1.415419 | 0.537783 | 0.953272 | 1     | 0.592692 |
| 3      | 4.471318 | 0.024637 | 1.969837 | 0.977607 | 0.129615 | 0.34089 | 1.299135 | 0.489662 | 1.303788 | 0.989145 |       |       |
| 4      | 4.72274  | 0.010533 | 0.049455 | 1.561144 | 0.587761 | 1.678812 | 2.5078   | 1.013374 | 0.548637 | 0.956443 | 1     | 0.996382 |
| 13     | 0      | 2     | 2     | 0     | 0     | 0     | 1     | 0.146236 | 1.394214 | 1.674736 | 0.017238 | 0.557615 |
| 2      | 0      | 2     | 2     | 0     | 0     | 0     | 1     | 0.09637 | 1.273057 | 1.570919 | 0.010149 | 0.636559 |
| 3      | 0      | 2     | 2     | 0     | 0     | 0     | 1     | 0.399394 | 2.89642 | 0.690304 | 0.948119 | 0.974151 |
| 4      | 0      | 2     | 2     | 0     | 0     | 0     | 1     | 0.322978 | 0.052342 | 2.953463 | 0.010693 | 0.02025 |
| 25     | 0.184282 | 0.018656 | 0.009328 | 2.920302 | 4     | 0     | 0     | 0     | 0     | 0     | 0.677473 |       |
| 2      | 0.133457 | 0.013494 | 0.006747 | 2.949312 | 4     | 0     | 0     | 0     | 0     | 0     | 0.63957 |
| 3      | 0.017495 | 0.025586 | 0.012793 | 0.117265 | 4     | 0     | 0     | 0     | 0     | 0     | 0.018912 |
| 4      | 0.005627 | 1.948276 | 0.974138 | 0.045505 | 4     | 0     | 0     | 0     | 0     | 0     | 0.00854 |

### Breast cancer

| Center | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1      | 4.426645 | 3.136173 | 3.175885 | 2.695168 | 3.210073 | 3.313859 | 3.396564 | 2.825064 | 1.569515 |
| 2      | 4.389494 | 3.044699 | 3.240979 | 3.101898 | 3.209126 | 3.910525 | 3.481052 | 2.957879 | 1.621914 |

### Ecoli

| Center | 1     | 2     | 3     | 4     | 5     | 6     | 7     |
|--------|-------|-------|-------|-------|-------|-------|-------|
| 1      | 0.659677 | 0.724477 | 0.488393 | 0.501017 | 0.43535 | 0.456288 | 0.357954 |
| 2      | 0.450608 | 0.48213 | 0.486007 | 0.500255 | 0.55165 | 0.770883 | 0.782437 |
| 3      | 0.374177 | 0.403017 | 0.484177 | 0.500565 | 0.477581 | 0.253246 | 0.326715 |
| 4      | 0.416347 | 0.480347 | 0.489672 | 0.500935 | 0.481801 | 0.464724 | 0.504486 |
| 5      | 0.272119 | 0.38684 | 0.482788 | 0.500363 | 0.440079 | 0.315179 | 0.399156 |
| 6      | 0.713524 | 0.466542 | 0.485632 | 0.500268 | 0.567728 | 0.76409 | 0.770628 |
| 7      | 0.423254 | 0.432663 | 0.486186 | 0.500763 | 0.447253 | 0.344918 | 0.407966 |
| 8      | 0.662456 | 0.667245 | 0.496321 | 0.501496 | 0.60302 | 0.469251 | 0.349027 |

### Pima

| Center | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1      | 3.994296 | 114.3373 | 68.36761 | 15.226 | 17.63024 | 30.86682 | 0.430379 | 33.66248 |
| 2      | 3.647751 | 137.0553 | 72.02767 | 30.8569 | 22.0316 | 34.58664 | 0.563851 | 33.00666 |

### Wine

| Center | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 13    |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1      | 12.94148 | 2.582226 | 2.390735 | 19.83796 | 102.516 | 2.097608 | 1.552164 | 0.395441 | 1.489661 | 5.669304 | 0.88117 | 2.345889 | 715.5377 |
| 2      | 13.78815 | 1.881141 | 2.444485 | 17.01094 | 105.5068 | 2.85667 | 3.005874 | 0.28959 | 1.915953 | 5.747003 | 1.077468 | 3.083985 | 120.827 |
| 3      | 12.94148 | 2.376339 | 2.289644 | 20.75 | 92.1315 | 2.09355 | 1.842603 | 0.38534 | 1.47006 | 4.048857 | 0.955954 | 2.519867 | 448.022 |
### Dermatology

| Center | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | 10     | 11     | 12     |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1      | 2.109798 | 1.772281 | 1.620607 | 1.47975 | 0.777255 | 0.685174 | 0.052053 | 0.607547 | 0.318831 | 0.413534 | 0.104229 | 0.661186 |
| 2      | 2.035617 | 1.695485 | 1.169133 | 1.461357 | 0.491221 | 0.130867 | 0.22746 | 0.104543 | 0.517256 | 0.318944 | 0.10683 | 0.1279 |
| 3      | 2.043372 | 1.811306 | 1.697692 | 1.203996 | 0.538297 | 0.281876 | 0.306666 | 0.210025 | 0.828493 | 0.703574 | 0.146349 | 0.276376 |
| 4      | 2.026072 | 1.79876 | 1.506725 | 1.418711 | 0.611342 | 0.492685 | 0.086043 | 0.356553 | 0.532966 | 0.430101 | 0.065655 | 0.380445 |
| 5      | 2.065199 | 1.810505 | 1.63167 | 1.19495 | 0.624458 | 0.343721 | 0.041644 | 0.315709 | 0.483115 | 0.468728 | 0.061497 | 0.311852 |
| 6      | 1.937094 | 1.654549 | 1.063019 | 1.06125 | 0.225146 | 0.081276 | 1.197426 | 0.055124 | 1.084815 | 0.503301 | 0.429526 | 0.037079 |

| 13     | 14     | 15     | 16     | 17     | 18     | 19     | 20     | 21     | 22     | 23     | 24     |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1      | 0.203511 | 0.380751 | 0.285169 | 1.601653 | 1.965735 | 0.423262 | 1.27033 | 0.520069 | 0.764655 | 0.513583 | 0.24828 | 0.314023 |
| 2      | 0.078957 | 0.548325 | 0.508905 | 1.34075 | 1.88605 | 0.445468 | 1.108135 | 0.433044 | 1.077552 | 0.450652 | 0.31151 | 0.200749 |
| 3      | 0.077414 | 0.817261 | 0.372891 | 1.040102 | 1.926237 | 0.630214 | 1.3764 | 1.063112 | 1.501174 | 1.034152 | 0.492757 | 0.5188 |
| 4      | 0.15526 | 0.551214 | 0.294695 | 1.555232 | 1.949144 | 0.485746 | 1.278736 | 0.579244 | 0.819767 | 0.5148 | 0.221512 | 0.265035 |
| 5      | 0.183799 | 0.502028 | 0.305747 | 1.370344 | 1.953655 | 0.477762 | 1.202563 | 0.57075 | 0.887131 | 0.583101 | 0.202309 | 0.375328 |
| 6      | 0.01912 | 0.396112 | 0.431457 | 1.33872 | 1.765024 | 0.94955 | 1.324368 | 0.475046 | 0.7214 | 0.239523 | 0.147246 | 0.196156 |

| 25     | 26     | 27     | 28     | 29     | 30     | 31     | 32     | 33     |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1      | 0.666696 | 0.413168 | 0.769192 | 1.009122 | 0.736579 | 0.025807 | 0.009139 | 1.993043 | 0.903294 |
| 2      | 0.118647 | 0.303227 | 0.131905 | 1.087123 | 0.12848 | 0.112575 | 0.143052 | 1.712577 | 0.172428 |
| 3      | 0.213997 | 0.744034 | 0.308459 | 0.631355 | 0.295513 | 0.000257 | 0.000478 | 1.888739 | 0.349735 |
| 4      | 0.399166 | 0.376606 | 0.4506 | 1.068269 | 0.476109 | 0.022894 | 0.025473 | 1.927999 | 0.573176 |
| 5      | 0.292437 | 0.503741 | 0.332814 | 1.099289 | 0.357319 | 0.006732 | 0.004156 | 1.815277 | 0.402284 |
| 6      | 0.053297 | 0.05187 | 0.056266 | 1.081903 | 0.056241 | 0.988468 | 1.127724 | 1.567869 | 0.16161 |