

## Fuzzy clustering with the generalized entropy of feature weights

Kai Li\* and Yan Gao

College of Computer Science and Technology, Hebei University, Baoding, China

Received: 23-July-2016; Revised: 20-September-2016; Accepted: 23-September-2016

©2016 ACCENTS

### Abstract

*Fuzzy c-means (FCM) is an important clustering algorithm. However, it does not consider the impact of different feature on clustering. In this paper, we present a fuzzy clustering algorithm with the generalized entropy of feature weights FCM (GEWFCM). By introducing feature weights and adding regularized term of their generalized entropy, a new objective function is proposed in terms of objective function of FCM. In GEWFCM, minimization of the dispersion within clusters and maximization of the generalized entropy of feature weights simultaneously obtain the optimal clustering results. Moreover, GEWFCM is viewed as a generalization of the maximum entropy-regularized weighted FCM (EWFCM). Experiments on data sets selected from University of California Irvine (UCI) machine learning repository demonstrate the effectiveness of presented method.*

### Keywords

*Fuzzy clustering, Fuzzy entropy, Generalized fuzzy entropy, Feature weights.*

### 1. Introduction

Clustering based on objective function is a commonly used method, which is attributed as an optimization problem with the constrained conditions, such as k-means clustering. As this method has some flaws, the researchers conduct some improvements on k-means clustering algorithm. It is worth mentioning that the researchers introduced the feature weights into objective function of optimization problems involved and presented a lot of different clustering algorithms. For example, Huang et al. [1] introduced feature weights into the objective function of k-means clustering and proposed WK-Means clustering algorithm. Renato et al. [2] further extended the Euclidean distance to Minkowski distance and studied the relationship between feature weights and measure of distance. In addition, an improved method for k-means clustering algorithm or its variants is proposed to the objective function of optimization problem. Since then, researchers presented fuzzy c-means (FCM) clustering algorithms and its variants [3-5]. Some researchers proposed the maximum entropy clustering algorithm in order to overcome the deficiencies of FCM [6, 7].

Meanwhile, fuzzy clustering algorithm with feature weights is obtained by taking into account the different roles of the features of data samples in the clustering process. In terms of them, researchers introduced feature weights into the fuzzy clustering algorithm [8, 9] and get the better clustering results. In fact, for problem of feature weights, it can be seen as a generalization of feature selection in pattern recognition, that is, feature selection can be seen as a special case of feature weights. Recently, Zhou et al. [10] presents a fuzzy clustering algorithm with the entropy-regularized weighted FCM (EWFCM). They mainly introduced the attribute weights into objective function and used fuzzy entropy of attribute weights as a regularized term. For obtaining the better clustering results, in this paper, we generalize this method presented by Zhou et al. [10] and obtain a new objective function by introducing feature weights and adding regularized term with the generalized entropy of feature weights in terms of objective function of FCM. By minimizing the dispersion within clusters and maximizing the generalized entropy of feature weights simultaneously, the optimal clustering results are obtained.

In this paper, we introduce feature weights and add regularized term of their generalized fuzzy entropy into objective function of FCM to obtain a new objective function. Based on this new objective

---

\*Author for correspondence

This work is supported by postgraduate innovation project of Hebei University under Grant X2016056.

function, a fuzzy clustering algorithm has been presented with the generalized entropy of feature weights FCM (GEWFCM). It is viewed as a generalization of the maximum entropy-regularized weighted FCM (EWFCM). In the experiments, we select seven data sets from UCI to demonstrate effectiveness of presented method.

## 2. Fuzzy Shannon entropy and generalized fuzzy entropy

Let  $X = \{x_1, x_2, \dots, x_n\}$  be a data set, where  $x_i \in R^s$ ,  $c$  is a positive integer greater than one and  $m \geq 1$  is fuzzy index. The  $u_{ij}$  is the degree of membership  $j$ th sample belonging to  $i$ th cluster.  $U$  is a  $c$ -by- $n$  matrix containing element  $u_{ij}$ . The generalized fuzzy entropy is defined as follows:

$$H(U, \alpha) = \sum_{j=1}^n (2^{1-\alpha} - 1)^{-1} \left( \sum_{i=1}^c u_{ij}^\alpha - 1 \right), \quad \alpha > 0, \alpha \neq 1, \quad (1)$$

Where  $\alpha$  is called as the generalized entropy index. It is seen that when  $\alpha \rightarrow 1$ ,  $H(U, \alpha)$  is fuzzy Shannon entropy, i.e.  $\lim_{\alpha \rightarrow 1} H(U, \alpha) = -\sum_{j=1}^n \sum_{i=1}^c u_{ij} \log_2 u_{ij}$ . In the following, we give a simplified proof.

Proof: When  $\alpha \rightarrow 1$ , it can be seen that  $H(U, \alpha)$  is an indeterminate form with 0/0 type. According to the L'Hopital rule,  $H(U, \alpha)$ 's limit is expressed as

$$\begin{aligned} & \lim_{\alpha \rightarrow 1} H(U, \alpha) \\ &= \lim_{\alpha \rightarrow 1} \sum_{j=1}^n (2^{1-\alpha} - 1)^{-1} \left( \sum_{i=1}^c u_{ij}^\alpha - 1 \right) \\ &= \lim_{\alpha \rightarrow 1} \sum_{j=1}^n \sum_{i=1}^c \frac{u_{ij}^\alpha \ln u_{ij}}{-2^{1-\alpha} \ln 2} \\ &= -\sum_{j=1}^n \sum_{i=1}^c u_{ij} \log_2 u_{ij} \end{aligned}$$

## 3. Maximum of the generalized fuzzy entropy with feature weights

### 3.1 The model of fuzzy clustering with the generalized fuzzy entropy of feature weights

In 2016, Zhou et al. [10] presents a fuzzy clustering with the entropy of attribute weights. They mainly introduced the attribute weights into objective function and used fuzzy entropy of attribute weights as a regularized term. To obtain the better cluster structure, in the following, we introduce the generalized fuzzy entropy into the objective function of optimization problem for FCM and add a regularized term with the generalized fuzzy entropy of feature weights. In reality, the model presented by us is viewed as a generalization of model presented

by Zhou et al.

The model of fuzzy clustering with the generalized fuzzy entropy of feature weights is as follows:

$$\begin{aligned} \min_{U, V, W} J(U, V, W) &= \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \sum_{j=1}^s w_{ij}^\alpha (x_{kj} - v_{ij})^2 \\ &+ \gamma^{-1} \sum_{j=1}^s (2^{1-\alpha} - 1)^{-1} \left( \sum_{i=1}^c w_{ij}^\alpha - 1 \right) \quad (2) \end{aligned}$$

$$s.t. \sum_{j=1}^s w_{ij} = 1, i = 1, 2, \dots, c,$$

$$\sum_{i=1}^c u_{ik} = 1, k = 1, 2, \dots, n,$$

Where  $c$  is the number of cluster,  $n$  is number of samples and  $s$  is dimension. Moreover, center of cluster is  $v_i (i = 1, 2, \dots, c)$  and  $v_i = (v_{i1}, v_{i2}, \dots, v_{is})$ ,  $w_{ij} (i = 1, 2, \dots, c; j = 1, 2, \dots, s)$  is  $j$ th feature belonging to  $i$ th cluster.

### 3.2 Fuzzy clustering algorithm with the generalized entropy of feature weights

To obtain fuzzy clustering algorithm with the generalized entropy of feature weights, we solve the optimization problem (2) with constrained condition. Thus, Lagrange function for (2) is

$$\begin{aligned} L(U, V, W; \lambda, \beta) &= \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \sum_{j=1}^s w_{ij}^\alpha (x_{kj} - v_{ij})^2 \\ &+ \gamma^{-1} \sum_{j=1}^s (2^{1-\alpha} - 1)^{-1} \left( \sum_{i=1}^c w_{ij}^\alpha - 1 \right) \\ &- \sum_{i=1}^c \lambda_i \left( \sum_{j=1}^s w_{ij} - 1 \right) - \sum_{k=1}^n \beta_k \left( \sum_{i=1}^c u_{ik} - 1 \right) \quad (3) \end{aligned}$$

By setting the gradient of  $L(U, V, W; \lambda, \beta)$  to zero with respect to  $u_{ik}$ ,  $v_{ij}$ ,  $w_{ij}$ ,  $\beta_k$  and  $\lambda_i$ , we obtain

$$\frac{\partial L}{\partial u_{ik}} = m u_{ik}^{m-1} \sum_{j=1}^s w_{ij}^\alpha (x_{kj} - v_{ij})^2 - \beta_k = 0, \quad (4)$$

$$\frac{\partial L}{\partial v_{ij}} = -2 \sum_{k=1}^n u_{ik}^m (x_{kj} - v_{ij}) w_{ij}^\alpha = 0, \quad (5)$$

$$\begin{aligned} \frac{\partial L}{\partial w_{ij}} &= \alpha w_{ij}^{\alpha-1} \sum_{k=1}^n u_{ik}^m (x_{kj} - v_{ij})^2 \\ &+ (\gamma^{-1} (2^{1-\alpha} - 1)^{-1} \alpha w_{ij}^{\alpha-1}) - \lambda_i = 0 \quad (6) \end{aligned}$$

$$\frac{\partial L}{\partial \beta_k} = \sum_{i=1}^c u_{ik} - 1 = 0, \quad (7)$$

$$\frac{\partial L}{\partial \lambda_i} = \sum_{j=1}^s w_{ij} = 1, \tag{8}$$

From (4), we obtain

$$u_{ik} = \left( \frac{\beta_k}{m \sum_{j=1}^s w_{ij}^\alpha (x_{kj} - v_{ij})^2} \right)^{1/(m-1)}, \tag{9}$$

Substituting (9) into (7), we have

$$\sum_{i=1}^c \left( \frac{\beta_k}{m \sum_{j=1}^s w_{ij}^\alpha (x_{kj} - v_{ij})^2} \right)^{1/(m-1)} = 1, \tag{10}$$

Further, we obtain the following Lagrange multiplier  $\beta_k$  ( $k=1,2,\dots,n$ ), namely

$$\beta_k^{1/(m-1)} = \frac{1}{\sum_{i=1}^c \left( \frac{1}{m \sum_{j=1}^s w_{ij}^\alpha (x_{kj} - v_{ij})^2} \right)^{1/(m-1)}}, \tag{10}$$

Substituting (10) into (9), we have

$$u_{ik} = \frac{1}{\sum_{l=1}^c \left( \frac{\sum_{j=1}^s w_{lj}^\alpha (x_{kj} - v_{lj})^2}{\sum_{j=1}^s w_{ij}^\alpha (x_{kj} - v_{ij})^2} \right)^{1/(m-1)}}. \tag{11}$$

About  $v_{ij}$ , from (5), we directly obtain

$$v_{ij} = \frac{\sum_{k=1}^n u_{ik}^m x_{kj}}{\sum_{k=1}^n u_{ik}^m}. \tag{12}$$

Similar to method solving  $u_{ik}$ , we have

$$w_{ij} = \left( \frac{\lambda_i}{\alpha \sum_{k=1}^n u_{ik}^m (x_{kj} - v_{ij})^2 + \alpha \gamma^{-1} (2^{1-\alpha} - 1)^{-1}} \right)^{1/(\alpha-1)} \tag{13}$$

Substituting (13) into (8), we have

$$\sum_{j=1}^s \left( \frac{\lambda_i}{\alpha \sum_{k=1}^n u_{ik}^m (x_{kj} - v_{ij})^2 + \gamma^{-1} (2^{1-\alpha} - 1)^{-1} \alpha} \right)^{1/(\alpha-1)} = 1, \tag{14}$$

Further, we obtain

$$\lambda_i^{1/(\alpha-1)} = \frac{1}{\sum_{j=1}^s \left( \frac{1}{\alpha \sum_{k=1}^n u_{ik}^m (x_{kj} - v_{ij})^2 + \gamma^{-1} (2^{1-\alpha} - 1)^{-1} \alpha} \right)^{1/(\alpha-1)}} \tag{14}$$

Substituting (14) into (13), we have

$$w_{ij} = \frac{1}{\sum_{l=1}^s \left( \frac{\alpha \sum_{k=1}^n u_{lk}^m (x_{kj} - v_{lj})^2 + \gamma^{-1} (2^{1-\alpha} - 1)^{-1} \alpha}{\alpha \sum_{k=1}^n u_{ik}^m (x_{kj} - v_{ij})^2 + \gamma^{-1} (2^{1-\alpha} - 1)^{-1} \alpha} \right)^{1/(\alpha-1)}} \tag{15}$$

In the following, we give the detailed algorithm.

Input: The number of clusters  $c$  ( $2 \leq c < n$ ), fuzzy index  $m$ , the parameter  $\gamma$ , a small enough error  $\varepsilon$ , the cluster centers  $v_i$  ( $i=1,2,\dots,c$ ), all initial weights of each attribute to  $1/c$ , and the iterative index  $j=1$ .

Repeat

    Update degree of membership by (11);

    Update the cluster centers by (12);

    Update the attribute weight by (15);

$j=j+1$ ;

Until  $\|U^{(j+1)} - U^{(j)}\| < \varepsilon$ .

## 4. Experimental results and analysis

### 4.1 Data sets and evaluation criterion

For validating the effectiveness of presented method, we choose seven data sets from University of California Irvine (UCI) machine learning repository [11] for experiment.

These data sets mainly contain iris, ionosphere, haberman, heart\_scale, diabetes, sonar and glass, whose detailed features are shown in *Table 1*. In the experiment, we select the evaluation criteria with clustering or classification accuracy and normalized mutual information (MI). In addition, we also choose Jaccard coefficient (JC), partition coefficient (PC) and partition entropy (PE) as the evaluation criteria of clustering, respectively. In the following, we simply introduce two evaluation criteria PC and PE. Supposed that a fuzzy partition of data set is represented by the membership matrix  $U = [u_{ij}]$ ,

where  $u_{ij}$  represents the degree of membership of the sample  $x_i$  in the cluster  $j$ . Then partition coefficient PC is defined as  $PC = 1/n \sum_{i=1}^n \sum_{j=1}^c u_{ij}$  and  $PC \in [1/c, 1]$ ,

where  $c$  is the number of clusters and  $n$  is the number of samples. When PC is closer to 1, the partition is clearer. On the contrary, when PC is closer to  $1/c$ , the partition is fuzzier. Moreover, partition entropy is given by  $PE = -1/n \sum_{i=1}^n \sum_{j=1}^c u_{ij} \times \log_a u_{ij}$  and  $PE \in [0, \log_a c]$ .

When PE is smaller, the partition is clearer. On the contrary, when PE is closer to  $\log_a c$ , the partition is fuzzier.

**Table 1** Feature of data set

Data set	Number of samples	Number of attributes	Number of clusters
iris	150	4	3
ionosphere	351	34	2
haberman	306	3	2
heart_scale	270	13	2
diabetes	768	8	2
sonar	208	60	2
glass	214	9	7

#### 4.2 Experimental results

Firstly, we conduct experiments ten times with GEWFCM and EWFCM in terms of accuracy and MI on selected seven data sets. The experimental results are shown in *Figure 1*, where (a), (c), (e), (g), (i), (k) and (m) are accuracy of clustering result at each data set, (b), (d), (f), (h), (j), (l) and (n) are MI of clustering result at each data set. Secondly, aiming at the different generalized entropy index  $\alpha$ , we conduct experimental study on three data sets iris, ionosphere and sonar in terms of accuracy, MI and JC. The experimental results are shown in *Figure 2*. It is seen

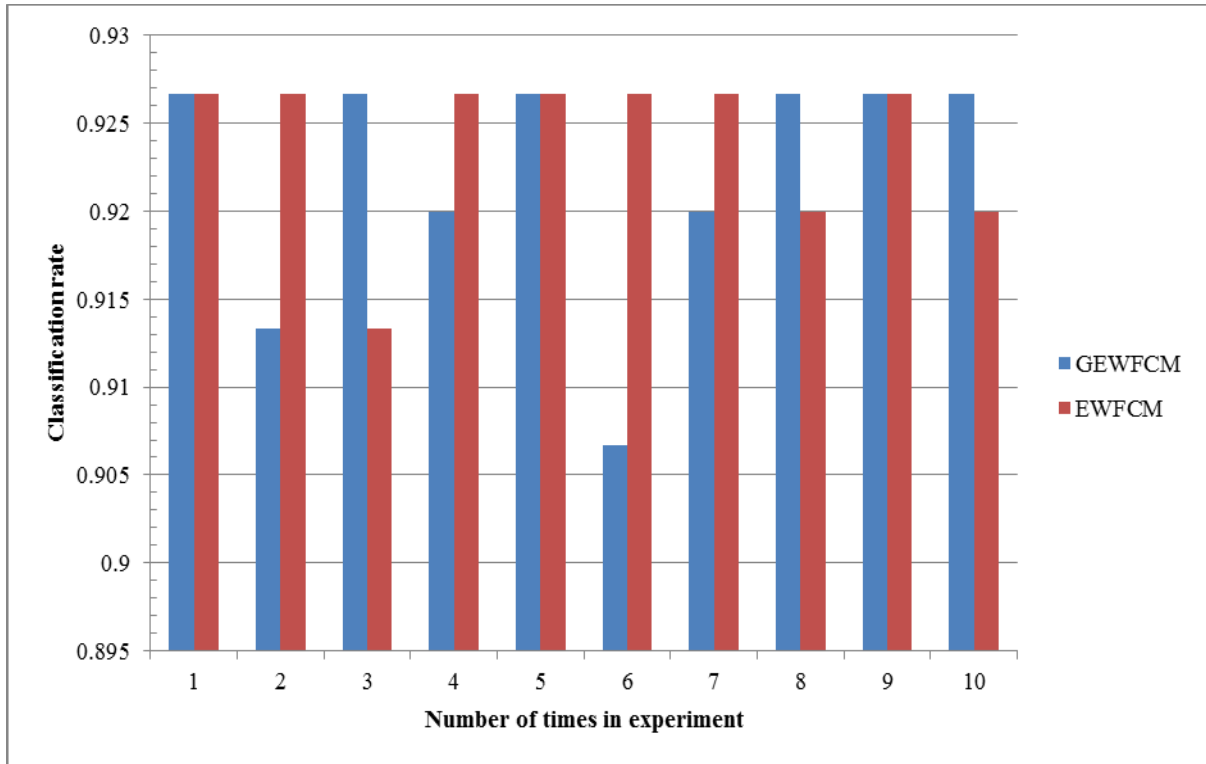
that when  $\alpha=9.5$ , GEWFCM obtains better results. Finally, aiming at the selected seven data sets, we obtain mean of performance index of GEWFCM and EWFCM about accuracy, MI, JC, PC and PE in ten times experiments. The experimental result is seen in *Table 2*, where the upper part shows the experimental results of GEWFCM and the lower part shows the experimental results of EWFCM.

#### 4.3 Analysis of clustering results

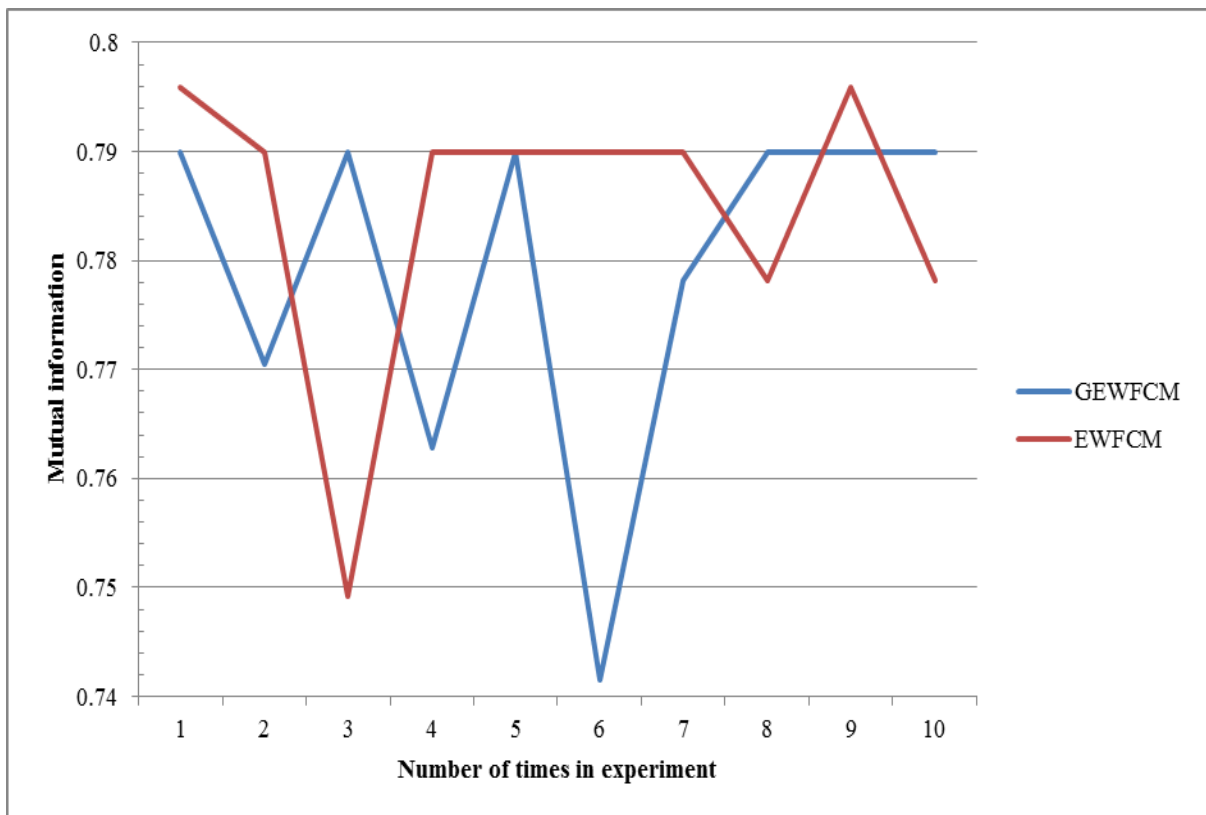
It can be seen from *Figure 1* that performance of GEWFCM presented by us is superior to EWFCM on the whole. Especially, we know that clustering results of GEWFCM are more stable than those of EWFCM except iris data set. Moreover, we also see that in experimental results of *Figure 1*, we only fix the generalized fuzzy entropy index  $\alpha$  to obtain them. In reality, to obtain better clustering results, it may find better generalized fuzzy entropy index  $\alpha$  by experiment. We conduct some experiments aiming at three data sets including iris, ionosphere and sonar in terms of accuracy, MI and JC as shown in *Figure 2*. It can find that better clustering results are obtained when  $\alpha=9.5$ . From *Table 2*, it is seen that clustering accuracies of GEWFCM are superior to those of EWFCM except iris data set whereas other three clustering indexes are basically consistent to each other. Experimental results clearly indicate that fuzzy clustering with generalized fuzzy entropy of feature weights may obtain better clustering results.

#### 5. Conclusion

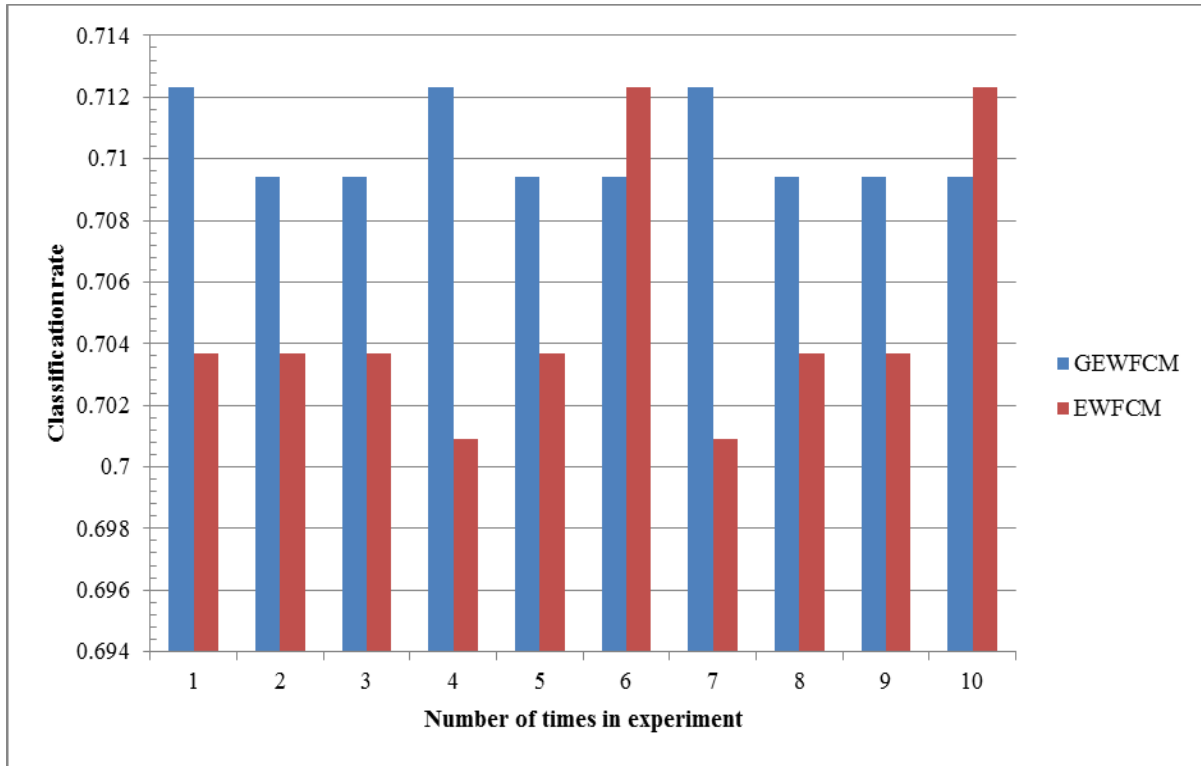
In this paper, we introduce feature weights and add regularized term of its generalized fuzzy entropy in terms of objective function of FCM to obtain a new objective function. By using Lagrange multiplier method, we propose a fuzzy clustering algorithm with the generalized entropy of feature weights, namely GEWFCM, which is viewed as a generalization of EWFCM. In GEWFCM, minimization of the dispersion within clusters and maximization of the generalized entropy of feature weights simultaneously obtain the optimal clustering results. Experiments on data sets selected from UCI repository demonstrate effectiveness of presented method. In future, we further study fuzzy clustering with the generalized feature weights based on kernel and expect to obtain better clustering results.



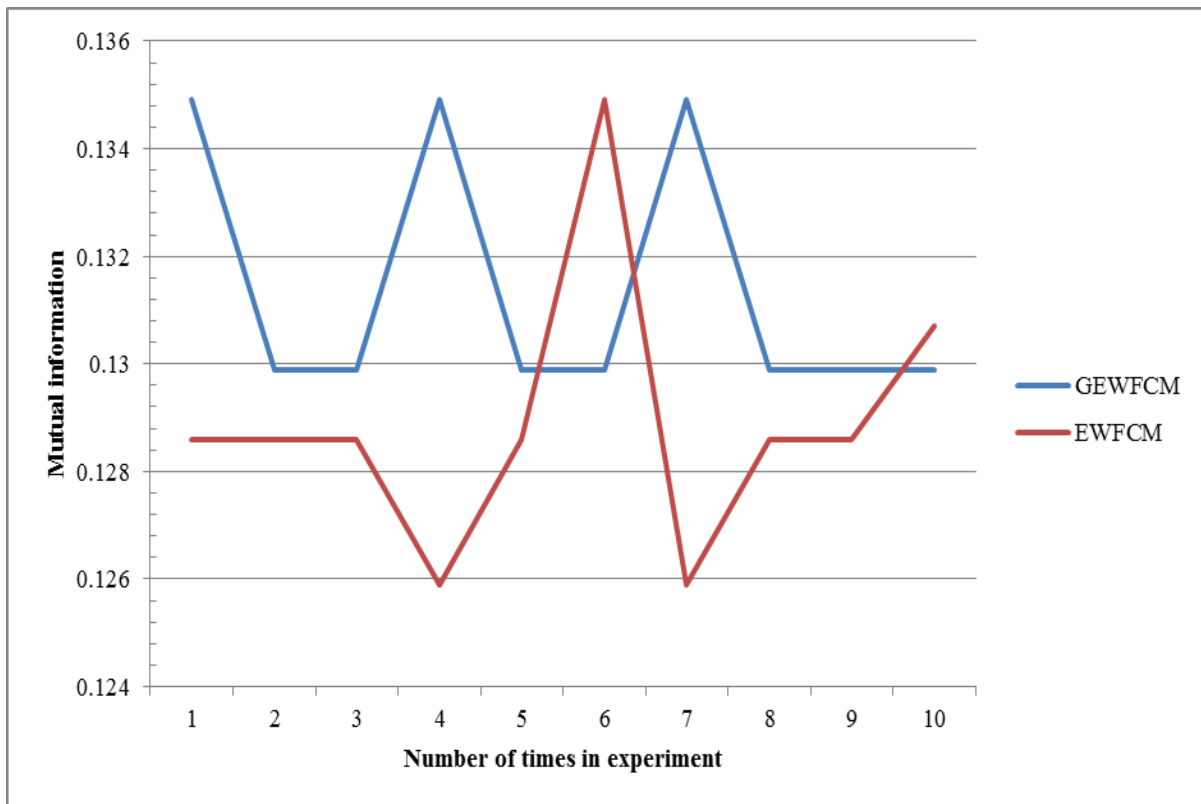
(a) Iris-accuracy



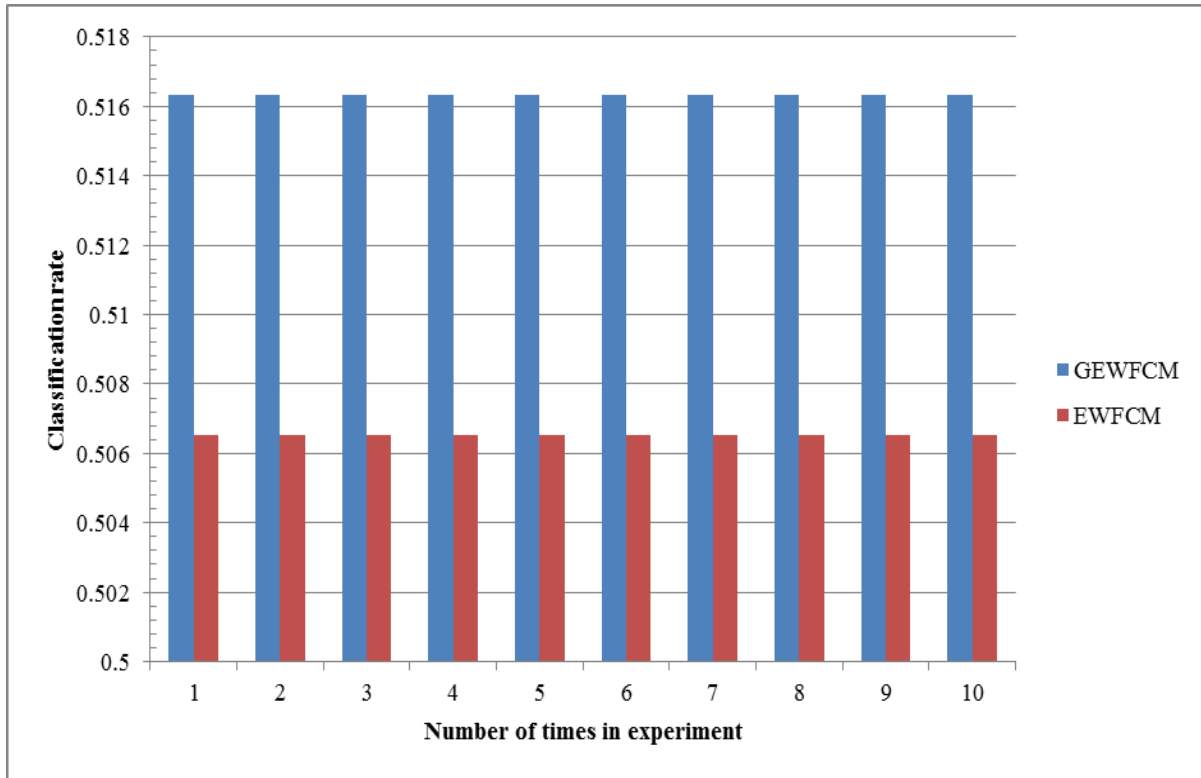
(b) Iris-MI



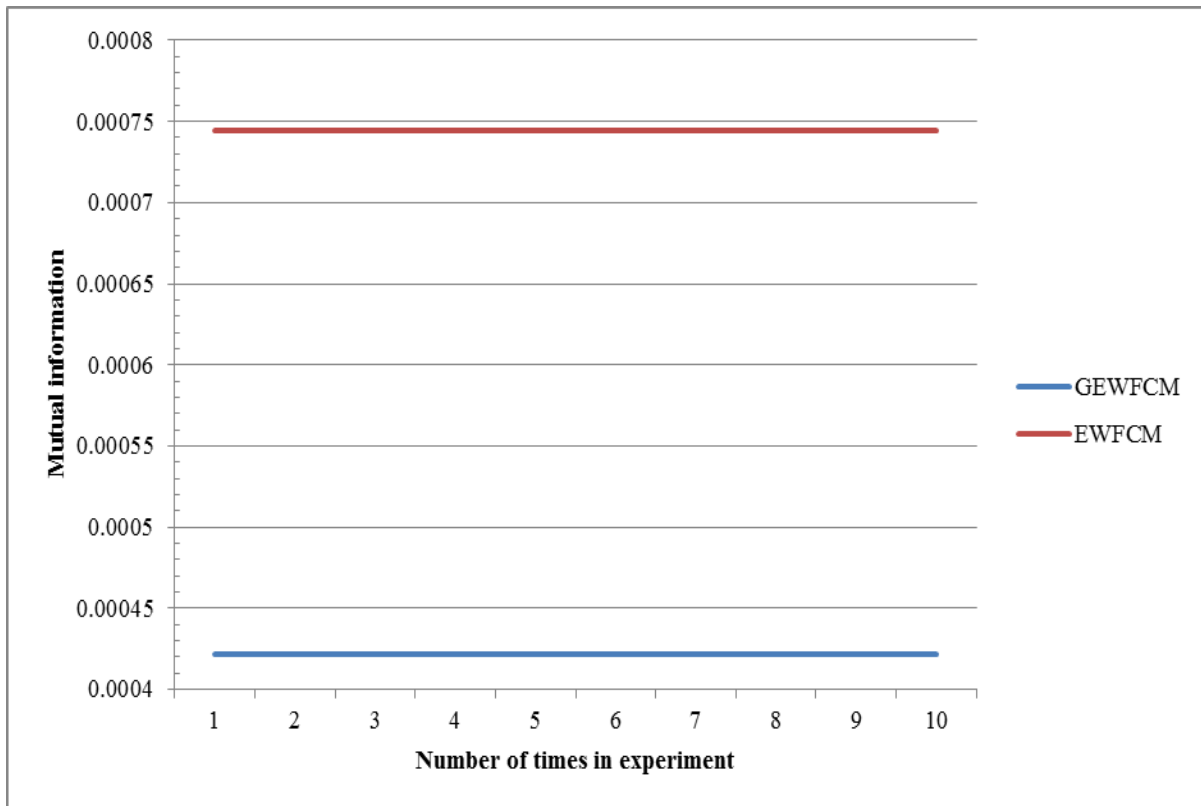
(c) Ionosphere-accuracy



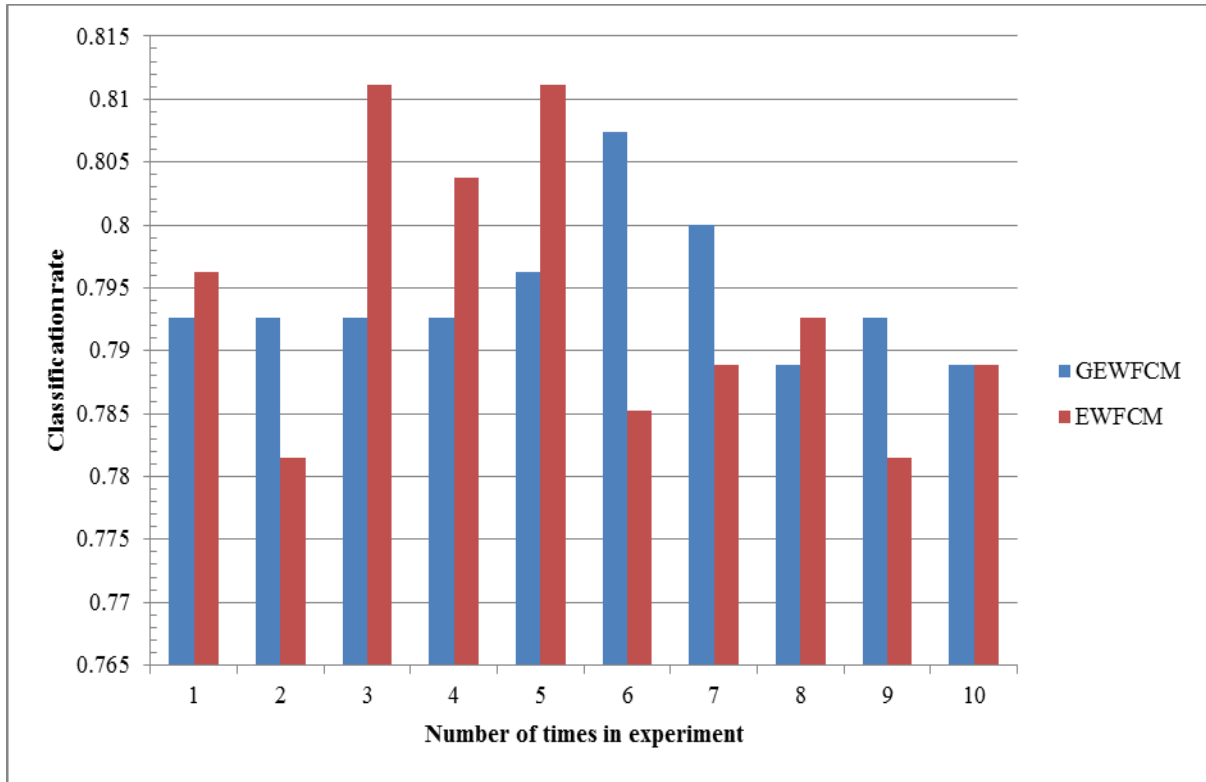
(d) Ionosphere-MI



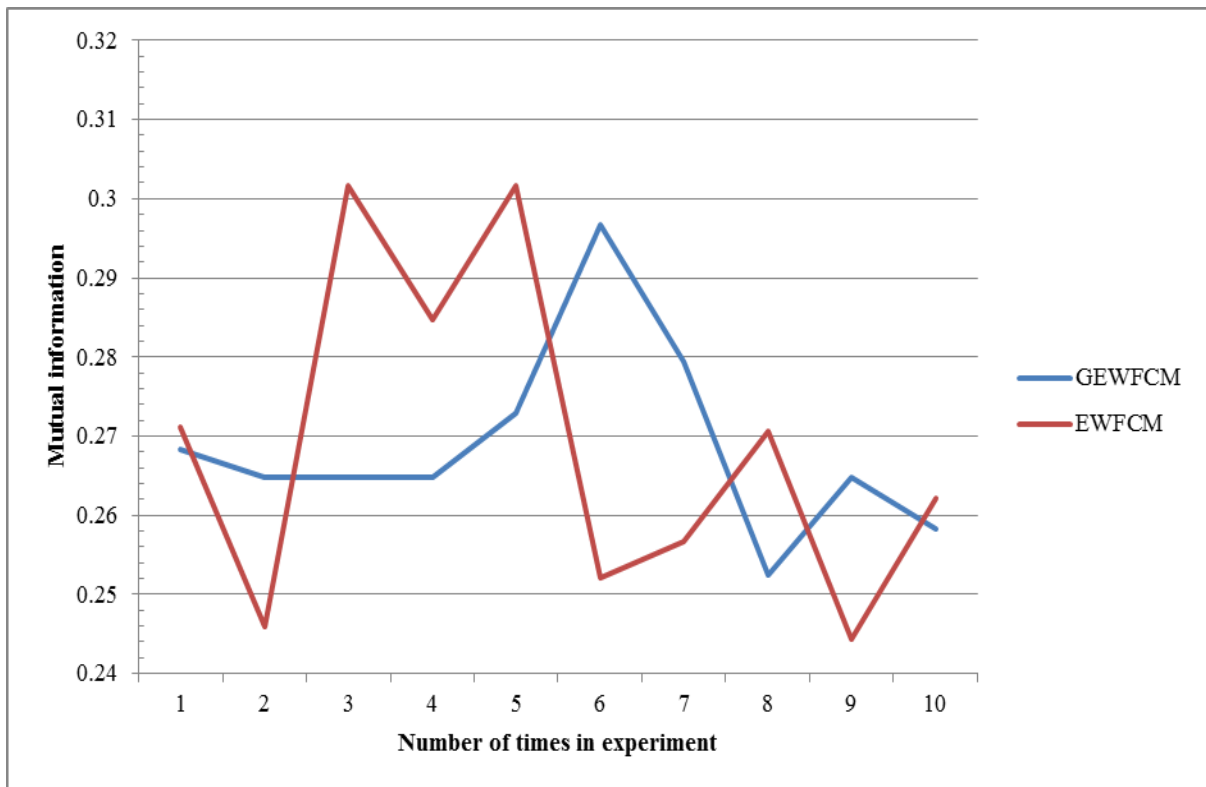
(e) Haberman-accuracy



(f) Haberman-MI

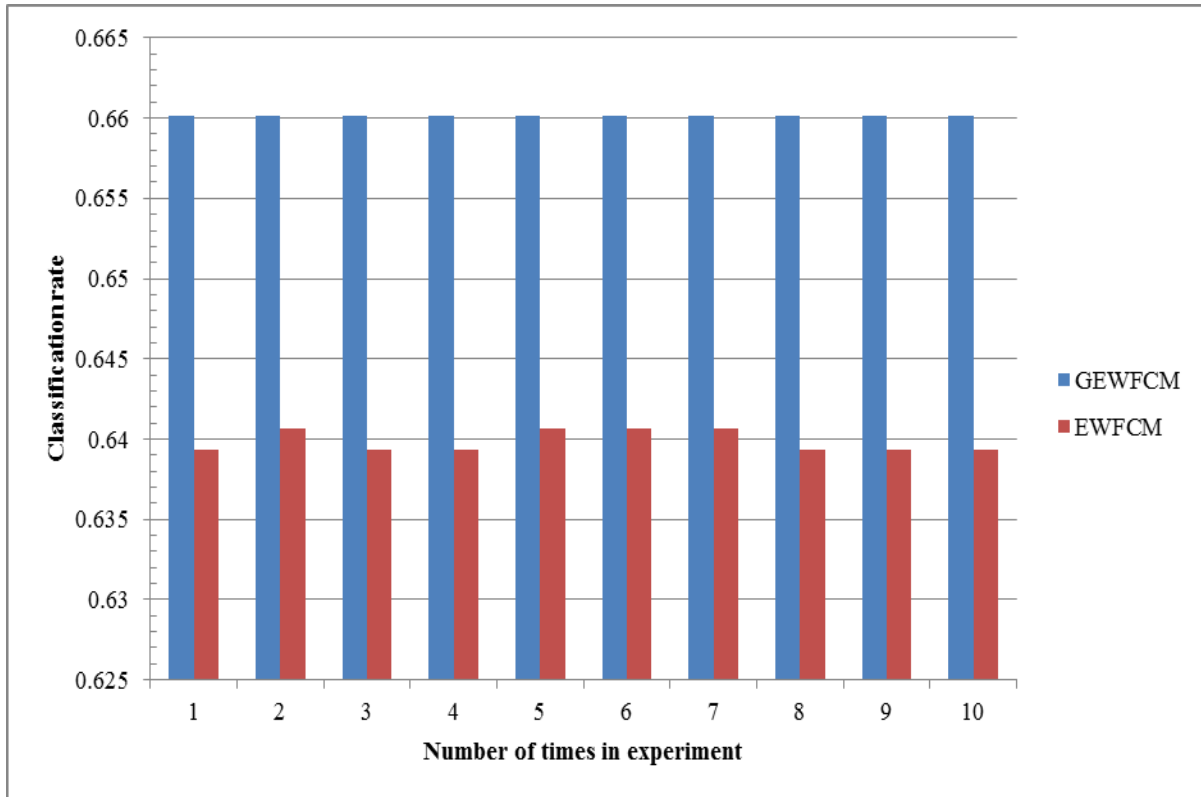


(g) Heart-scale-accuracy

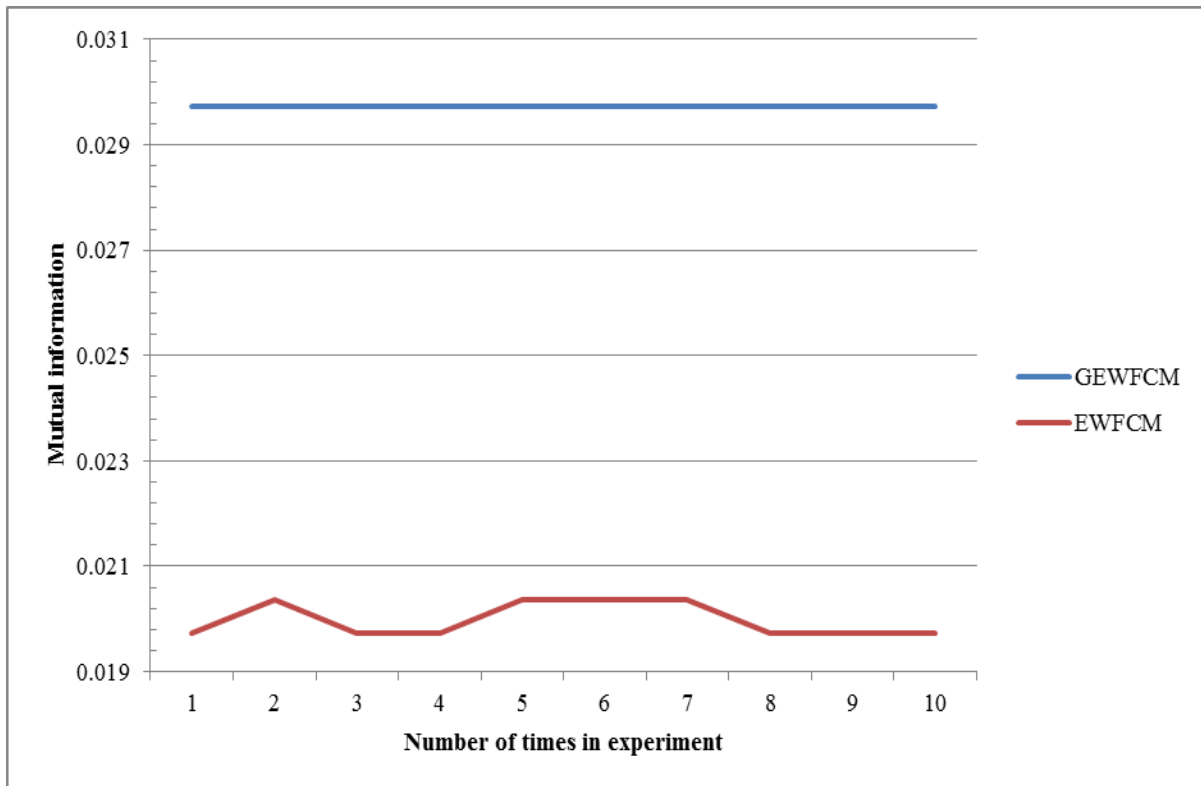


(h) Heart-scale-MI

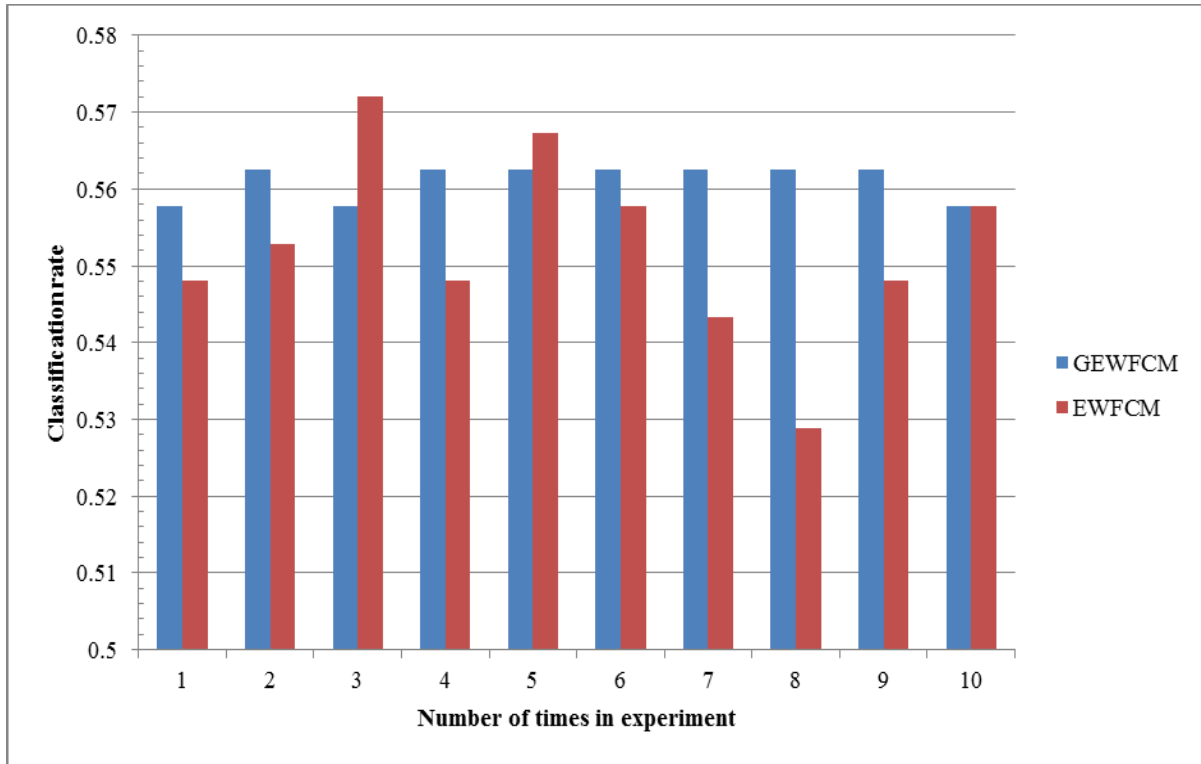




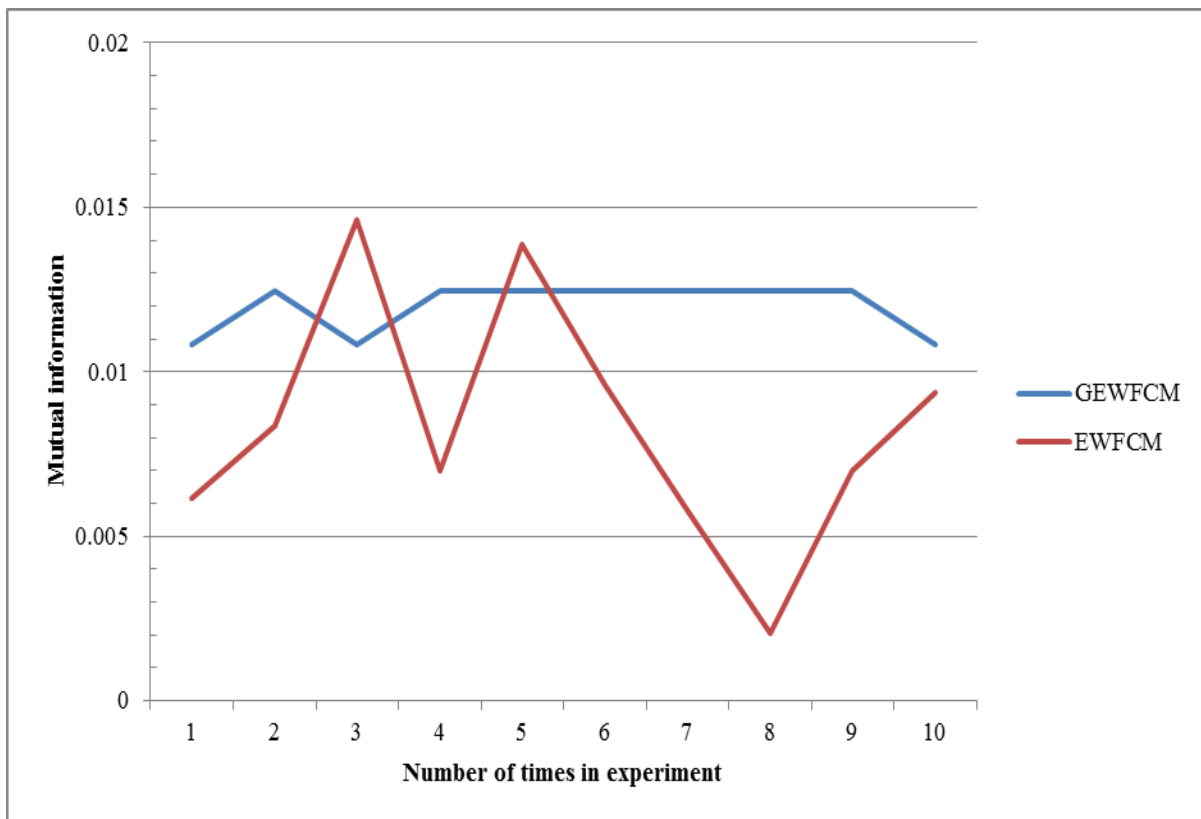
(i) Diabetes-accuracy



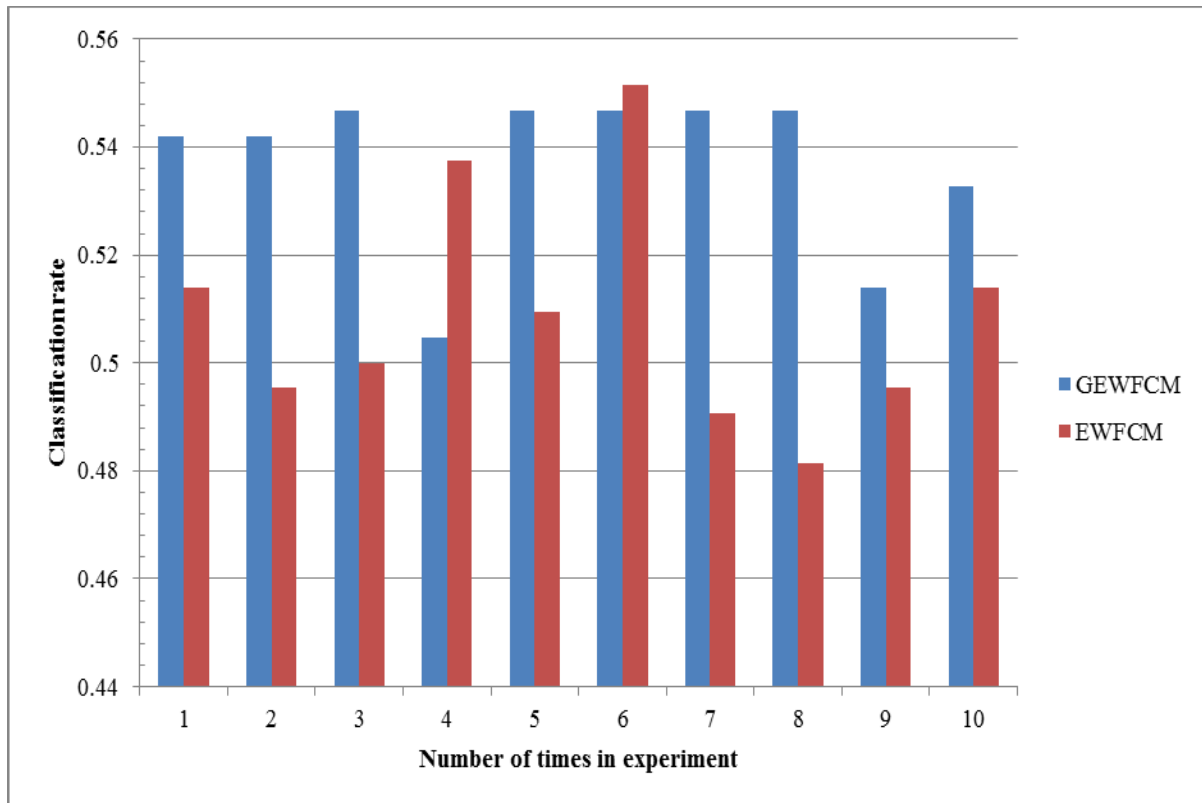
(j) Diabetes-MI



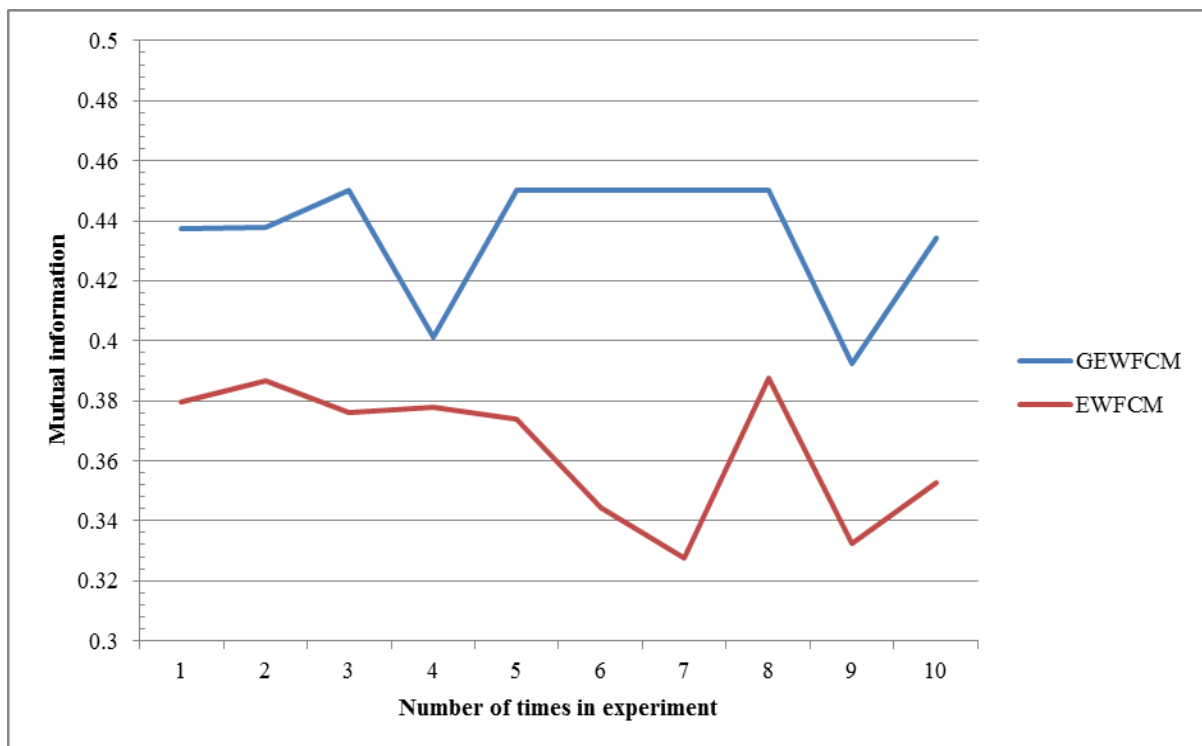
(k) Sonar-accuracy



(l) Sonar-MI

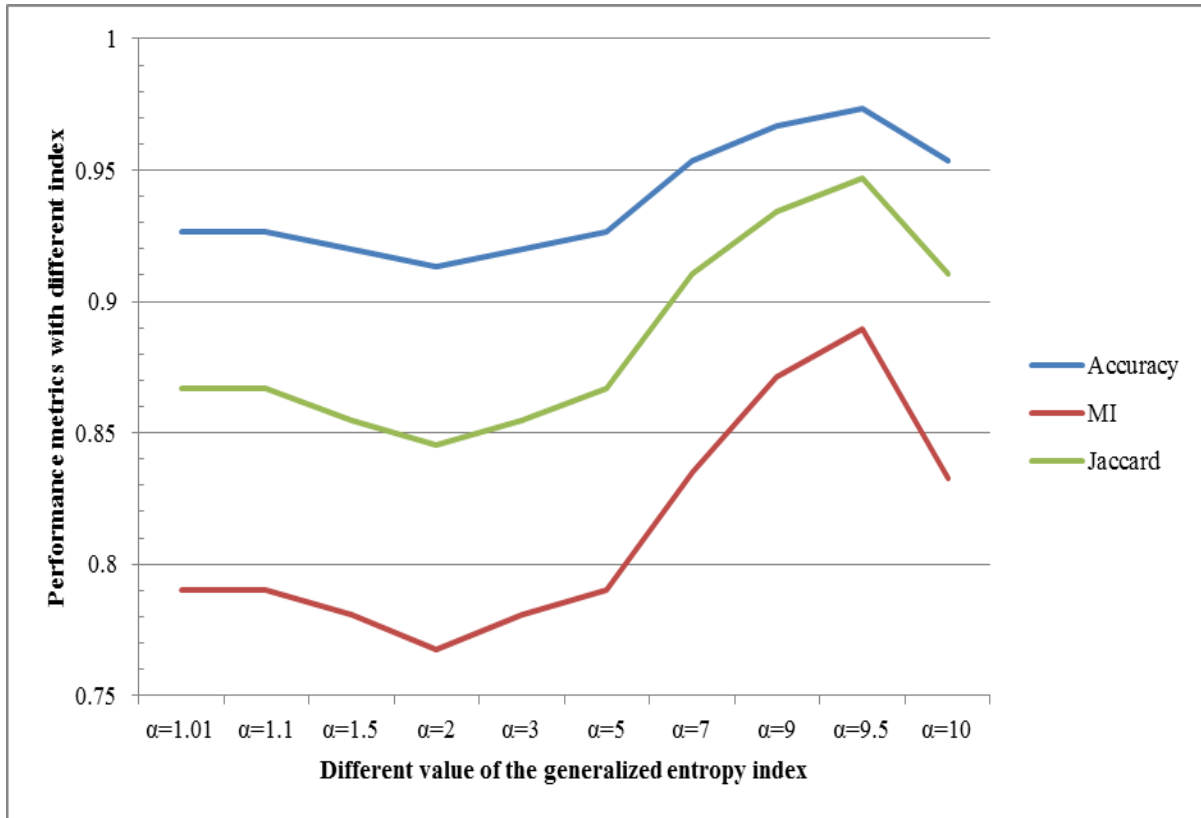


(m) Glass-accuracy

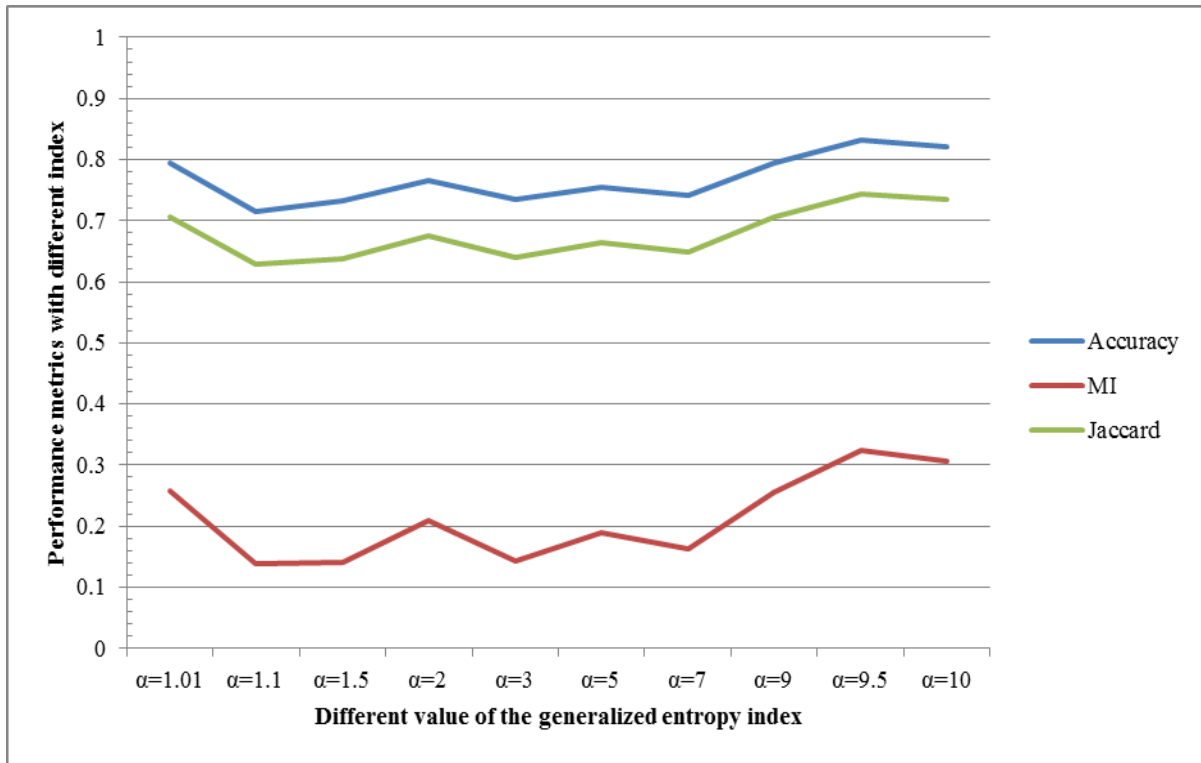


(n) Glass-MI

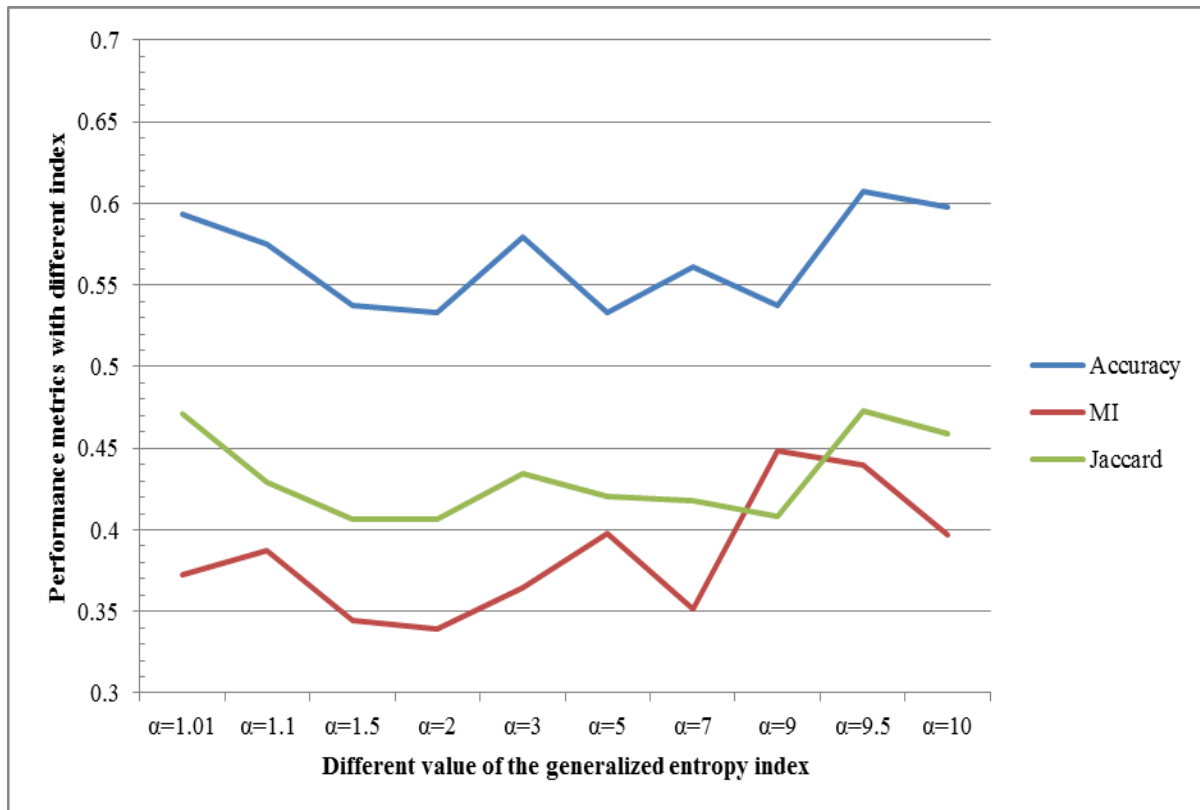
**Figure 1** Performance index of two clustering algorithms GEWFCM and EWFCM



(a) Iris



(b) Ionosphere



(c) Sonar

**Figure 2** Performance index of GEWFCM on different the generalized entropy index  $\alpha$

**Table 2** Mean of performance index of GEWFCM and EWFCM

GEWFCM/EWFCM	Accuracy	MI	JC	PC	PE
Iris	0.9220	0.7793	0.7652	0.3780	0.4489
	0.9240	0.7847	0.7649	0.3760	0.4501
Ionosphere	0.7103	0.1314	0.4314	0.6510	0.2266
	0.7048	0.1289	0.4283	0.5005	0.3008
Haberman	0.5163	0.0004	0.3772	0.5460	0.2805
	0.5065	0.0007	0.3781	0.6337	0.2371
Heart_Scale	0.7944	0.2687	0.5	0.5083	0.2974
	0.7941	0.2691	0.5152	0.5000	0.3010
Diabetes	0.6602	0.0297	0.4595	0.9875	0.0090
	0.6398	0.0200	0.4336	0.9902	0.0073
Sonar	0.5611	0.0120	0.3384	0.8810	0.0869
	0.5524	0.0084	0.3359	0.5002	0.3009
Glass	0.5369	0.4353	0.3312	0.9817	0.0144
	0.5089	0.3638	0.3056	0.1455	0.8410

### Acknowledgment

None.

### Conflicts of interest

The authors have no conflicts of interest to declare.

### References

[1] Huang JZ, Ng MK, Rong H, Li Z. Automated variable weighting in k-means type clustering. IEEE

Transactions on Pattern Analysis and Machine Intelligence.2005; 27(5):657-68.

[2] Renato C A, Boris M. Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering. Pattern Recognition. 2012; 45(3):1061-75.

[3] Dave RN, Bhaswan K. Adaptive fuzzy c-shells clustering and detection of ellipses. IEEE Transactions on Neural Networks.1992; 3(5):643-62.

- [4] Bobrowski L, Bezdek JC. C-means clustering with the  $l_1$  and  $l_\infty$  norms. IEEE Transactions on Systems, Man, and Cybernetics. 1991; 21(3):545-54.
- [5] Lin PL, Huang PW, Kuo CH, Lai YH. A size-insensitive integrity-based fuzzy c-means method for data clustering. Pattern Recognition. 2014;47(5):2042-56.
- [6] Karayiannis NB. MECA: maximum entropy clustering algorithm. Proceedings of the third IEEE conference on world congress on computational intelligence 1994 (pp. 630-5). IEEE.
- [7] Li RP, Mukaidono M. A maximum-entropy approach to fuzzy clustering. International joint conference of the fourth IEEE conference on fuzzy systems and the second international fuzzy engineering symposium 1995(pp. 2227-32). IEEE.
- [8] Fu H, Elmisery AM. A new feature weighted fuzzy c-means clustering algorithm. Algarve, Portugal. 2009.
- [9] Li J, Gao X, Jiao L. A new feature weighted fuzzy clustering algorithm. In international workshop on rough sets, fuzzy sets, data mining, and granular-soft computing 2005 (pp. 412-20). Springer Berlin Heidelberg.
- [10] Zhou J, Chen L, Chen CP, Zhang Y, Li HX. Fuzzy clustering with the entropy of attribute weights. Neurocomputing. 2016; 198: 125-34.
- [11] Lichman M. UCI machine learning repository. School Information Computer Science, University of California, Irvine, CA, USA. <http://archive.ics.uci.edu/ml>. Accessed 15 May 2016.



**Kai Li** received the B.S. and M.S. degrees in mathematics department electrical engineering department from Hebei University, Baoding, China, in 1982 and 1992, respectively. He received the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2001. He is currently a Professor in college of computer science and technology, Hebei University. His current research interests include machine learning, data mining, computational intelligence, and pattern recognition.  
Email: likai@hbu.edu.cn



**Yan Gao** received the bachelor degree in computer science and technology from Industrial and Commercial College Hebei University, Baoding, Hebei province, China, in 2011, and he is currently pursuing the M.E. degree in the computer science and technology, Hebei University, Baoding, Hebei province, China. His research interests include machine learning, data mining, and pattern recognition.