ARTICLE IN PRESS

Pattern Recognition ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Relative entropy collaborative fuzzy clustering method

M. Zarinbal [a,1], M.H. Fazel Zarandi [a,c,*,1], I.B. Turksen [b,c]

[a] Department of Industrial Engineering, Amirkabir University of Technology, 424 Hafez Ave, P.O. Box 15875-4413, Tehran, Iran
[b] TOBB Economics and Technology University, Ankara, Turkey
[c] Knowledge Intelligent Systems Laboratory, University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form
14 September 2014
Accepted 17 September 2014

Keywords:
Fuzzy set theory
Horizontal and vertical collaborative fuzzy clustering
Relative entropy
Relative entropy collaborative fuzzy clustering

ABSTRACT

The main task of clustering methods, especially fuzzy methods, is to find whether natural grouping exists in data and to impose identity on them. In some situations, data are stored in several data sites and to discover the global structures, clustering methods have to be aware of dependencies in all data sites. Collaborative fuzzy clustering methods have been proposed and widely studied to answer such need. In this paper, a novel collaborative fuzzy clustering method is proposed. In this method, relative entropy concept is used as the communication method, a new approach is applied to calculate the interaction coefficient between data sites, and horizontal and vertical modes of the proposed method are discussed. Performance of the proposed method is evaluated using several experiments and the results show that it has the highest quality of collaboration and could classify data more efficiently.

1. Introduction

Imposing identity on data and finding regularities are the main tasks of pattern recognition techniques, which can be done in supervised or unsupervised manner. In unsupervised or clustering methods, the main goal is to find the natural groupings exist in data [1] and it is assumed that data can only belong to one cluster. In reality, however, data can belong to more than just one cluster with some degree of belonging. This is effectively modeled by fuzzy logic. Many fuzzy clustering methods have been developed ([2–8]) and they have been applied in various areas ([9–13]).

In some situations, such as banking institutions, data have been stored in different data sites with same or different patterns and features. Clearly, having knowledge about the dependencies in all data sites is essential for discovering the global structures [14]. Collaborative fuzzy clustering (CFC) method was proposed to answer such need. CFC generally performs in two phases: fuzzy c-means (FCM) method runs independently at each data site in the first phase and in the second phase, data sites communicate their findings and each data site proceeds with its optimization by focusing on local data [15].

While the concepts of CFC have been widely studied ([16–23]), this paper proposes a number of novel concepts; as the ways of communication affect the results, in this paper a new communication method is proposed using relative entropy (RE) concept. RE enables the method to cluster data and handle noisy datasets more efficiently. That is, as discussed in [1], FCM divides data into given cluster numbers regardless of being noisy or not, whereas it is more natural for noise objects to have very low membership degrees in all clusters. The interaction coefficient between data sites is another important issue that, in most cases, has to be estimated beforehand. In this study, a new approach is applied to calculate this value. Data sites could also have same or different number of data patterns and features. Thus, in this paper, two modes of the proposed method, horizontal and vertical, are discussed and their performances are evaluated using several experiments. The obtained results are then compared with CF [14], CFC [15], CFC-c^*[21], CFC with fixed β, CFC-β_f-c^*, [21], and CFC with dynamic β, CFC-β_d, [21].

The rest of this paper is organized as follows: Section 2 discusses the CFC methods presented in literature. The proposed clustering method and its two modes are discussed in Section 3. Performance of the proposed method and some more discussions are addressed in Section 4. Conclusion is stated in Section 5. Finally, Appendix provides the needed proofs.

2. Related works

The CFC method, first proposed by [16], is a fuzzy clustering method concerns with the extension of fuzzy clustering to several data sites. This is done in two phases: in the first phase, FCM runs

* Corresponding author at: Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran.
E-mail addresses: mzarinbal@aut.ac.ir (M. Zarinbal), zarandi@aut.ac.ir (M.H. Fazel Zarandi), bturksen@etu.edu.tr (I.B. Turksen).
[1] Tel.: +98 2164541.

http://dx.doi.org/10.1016/j.patcog.2014.09.018
0031-3203/

independently at each data site and in the second phase, the findings are communicated and each data site proceeds with optimization by focusing on its local data. CFC has two important modes, vertical and horizontal. In horizontal CFC (HCFC), each data site has the same data patterns in different feature spaces, whereas in vertical CFC (VCFC), data sites consist of different data in the same feature space. Thus, data sites communicate the membership degrees and prototypes in HCFC and VCFC, respectively [14].

That is, in HCFC, for clustering N data pattern stored in Pdata sites, optimization of the second phase proceeds with [16]

$$\min J_{HCFC_1}[ii] = \sum_{k=1}^{N}\sum_{i=1}^{c} u_{ik}^2[ii] d_{ik}^2[ii] + \sum_{\substack{kk=1\\ kk\neq ii}}^{P} \alpha[ii,jj] \sum_{k=1}^{N}\sum_{i=1}^{c} (u_{ik}[ii]-u_{ik}[jj])^2 d_{ik}^2[ii]$$

$$S.T.\begin{cases} \sum_{i=1}^{c} u_{ik}[ii]=1 & \forall k, ii=1,\dots,P \\ 0<\sum_{k=1}^{N} u_{ik}[ii]<N & \forall i, ii=1,\dots,P \\ u_{ik}[ii]\in[0,1] & \forall i,k,ii=1,\dots,P \end{cases} \tag{1}$$

where, for iith data site, $u_{ik}[ii]$ is the membership degree of Kth data, $x_k[ii]$, in ith cluster, $d_{ik}[ii]$ is the distance between $x_k[ii]$ and ith cluster's prototype,$v_i[ii]$, and $c$is the number of clusters. $\alpha[ii,jj]$ is the non-negative entry of collaborative matrix and describes the intensity of interaction between iith and jjth data sites [16].

In VCFC, the data sites consists of different data pattern, $N[1],\dots,N[P]$, in the same feature space, so, collaboration phase proceeds with [14]

$$\min J_{VCFC_1}[ii] = \sum_{k=1}^{N[ii]}\sum_{i=1}^{c} u_{ik}^2[ii] d_{ik}^2[ii] + \sum_{\substack{jj=1\\ jj\neq ii}}^{P} \beta[ii,jj] \sum_{k=1}^{N[ii]}\sum_{i=1}^{c} u_{ik}^2[ii]||v_i[ii]-v_i[jj]||^2$$

$$S.T.\begin{cases} \sum_{i=1}^{c} u_{ik}[ii]=1 & \forall k, ii=1,\dots,P \\ 0<\sum_{k=1}^{N[ii]} u_{ik}[ii]<N[ii] & \forall i, ii=1,\dots,P \\ u_{ik}[ii]\in[0,1] & \forall i,k,ii=1,\dots,P \end{cases} \tag{2}$$

where, $\beta[ii,jj]$ is the collaboration coefficient of ith and jjth data sites.

In both modes, $\alpha[ii,jj]$ and $\beta[ii,jj]$ have to be determined in advance. To solve this problem in HCFC, Falcon et al. [17,19] proposed an approach to seek the most suitable values of $\alpha[ii,jj]$. In their approach, initial values of $\alpha[ii,jj]$ are determined using rough threshold method and a particle swarm optimization (PSO) driven tuning process is used to optimize two predefined fitness functions.

To solve this problem in VCFC, a new CFC method is proposed by [15], in which the collaboration phase proceeds with:

$$\min J_{VCFC_2}[ii] = \sum_{k=1}^{N[ii]}\sum_{i=1}^{c} u_{ik}^2[ii] d_{ik}^2[ii] + \beta \sum_{\substack{jj=1\\ jj\neq ii}}^{P}\sum_{k=1}^{N[ii]}\sum_{i=1}^{c} (u_{ik}[ii]-\tilde{u}_{ik}[ii|jj])^2 d_{ik}^2[ii]$$

$$S.T.\begin{cases} \sum_{i=1}^{c} u_{ik}[ii]=1 & \forall k, ii=1,\dots,P \\ 0<\sum_{k=1}^{N[ii]} u_{ik}[ii]<N[ii] & \forall i, ii=1,\dots,P \\ u_{ik}[ii]\in[0,1] & \forall i,k,ii=1,\dots,P \end{cases} \tag{3}$$

where, β is a non-negative collaboration number and $\tilde{u}_{ik}[ii|jj]$ is the induced membership degree computed using [15]

$$\tilde{u}_{ik}[ii|jj] = \frac{1}{\sum_{l=1}^{c}\left(\frac{||x_k[ii]-v_i[jj]||}{||x_k[ii]-v_l[jj]||}\right)^2} \tag{4}$$

Using CFC formulation in Eq. (2), four data-driven approaches is proposed by [21] to improve CFC. For vertical collaboration, CFC-β_f and CFC-β_d are proposed for known c and CFC -β_f-c^* for unknown c. CFC-c^* is also proposed for horizontal collaboration with unknown c. In CFC-β_f and CFC β_d-c^*, $\beta[ii,jj]$ remains fixed for each pair of data sites during the collaboration phase, whereas CFC-β_d and CFC-c^* dynamically adjust $\beta[ii,jj]$ for every pair of data sites at every collaboration stage. In these methods $\beta[ii,jj]$ is estimated by [21]

$$\beta[ii,jj] = \min\left\{1, \frac{J_{CFC_2}[ii]}{\sum_{k=1}^{N[ii]}\sum_{i=1}^{c}\tilde{u}_{ik}^2[ii|jj]d_{ik}^2}\right\} \tag{5}$$

Furthermore, a PSO driven CFC method and a learning approach based on self-organizing map (SOM) are proposed by [20] and [22] for both mode of collaboration and for determining optimum sets of $\alpha[ii,jj]$ and $\beta[ii,jj]$.

Table 1 summarizes the abovementioned approaches.

The effectiveness of CFC depends on the way of communication and the communicated findings, which are membership degrees in horizontal mode and prototypes in vertical mode [15]. Hence,

Table 1
Collaborative clustering methods

CFC Mode	Authors	Proposed approach
Horizontal	Falcon et al. [17,19]	Rough threshold method is used to determine $\alpha[ii,jj]$ and a particle swarm optimization driven tuning process is used to optimize the two predefined fitness functions.
	Depaire et al. [20]	A particle swarm optimization driven CFC is proposed to determine $\alpha[ii,jj]$.
	Coletta et al. [21]	CFC-c^* method is proposed for unknown c.
	Ghassany et al. [22]	A learning approach based on self-organizing map is proposed to estimate the collaboration parameter during the collaboration phase.
Vertical	Pedrycz, and Rai [15]	$\beta[ii,jj]$ is replaced with β, a nonnegative collaboration parameter. The collaboration phase proceeds with Eq. (3).
	Depaire et al. [20]	A particle swarm optimization driven CFC is proposed to determine $\beta[ii,jj]$.
	Coletta et al. [21]	CFC-β_f and CFC-β_d methods are proposed for known c and CFC-β_f-c^* method is proposed for unknown c.
	Ghassany et al. [22]	A learning approach based on self-organizing map is proposed to estimate the collaboration parameter.

minimizing the distance between membership degrees and prototypes are general ways of communication. RE or Kullback–Leibler number is another measure of distance or dissimilarity between two distributions [1]. That is, suppose that there are two distinct probability distributions, $p(x)$ and $q(x)$, on a finite set $X, x \in X$. RE measures the inefficiency of assuming that the distribution is $q(x)$ while the true distribution is $p(x)$[1]

$$D(p||q) = \sum_x p(x) \ln \frac{p(x)}{q(x)} \tag{6}$$

Although RE is not a true distance, it has two important properties; it is a nonnegative measure, $D(p||q) \geq 0$, with equity if and only if $p(x) = q(x)$ and it is a convex function of $p(x)$ and $q(x)$ [1]. Hence, as RE is a function of probability distribution and a non-negative convex function and as membership degrees in FCM have probabilistic interpretation [1], it can be used to measure the dissimilarity or distance between membership degrees. Using this concept, a new fuzzy clustering method, relative entropy fuzzy c-means, is proposed by [1] in which RE is added as a regularization function to FCM objective function. They showed that in this case reasonable results especially in noisy datasets would obtain.

Inspiring this idea, in this paper, a new relative entropy collaborative fuzzy clustering (RECFC) method is proposed, it's two modes, horizontal and vertical, are discussed and the interaction coefficients are determined.

Moreover, considering same number of clusters for all data sites is quite unrealistic [15]. This problem can be viewed from two viewpoints, whether there is prior knowledge about $c[ii]$s, and whether the data sites share the same number of clusters (Fig. 1). Note that In the case of having no prior knowledge, finding the proper $c[ii]$s has been supported by various means including clustering validity index (VI) methods [15].

3. Relative entropy collaborative fuzzy clustering method

The general form of the proposed method is as follows:

Consider $D[1], \ldots, D[P]$ data sites with $N[1], \ldots, N[P]$ data. FCM runs independently at each data site in the first phase, and in the second phase, the data sites communicate their findings and proceed with optimization. Using RE in this phase, $u_{ik}[ii]$ is calculated by

$$\min J_{RECFC}[ii] = \sum_{k=1}^{N[ii]} \sum_{i=1}^{c^*} u_{ik}^m[ii] d_{ik}^2[ii] + \gamma[ii] \sum_{k=1}^{N[ii]} \sum_{i=1}^{c^*} \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} u_{ik}[ii] \ln\left(\frac{u_{ik}[ii]}{u_{ik}[jj]}\right)$$

$$S.T. \begin{cases} \sum_{i=1}^{c^*} u_{ik}[ii] = 1 & \forall k, ii \\ 0 < \sum_{k=1}^{N[ii]} u_{ik}[ii] < N[ii] & \forall i, ii \\ u_{ik}[ii] \in [0,1] & \forall i, k, ii \end{cases} \tag{7}$$

where, m is the degree of fuzziness, c^* is the proper number of clusters for all data sites, and $\gamma[ii]$ is the nonnegative coefficient of collaboration. The first term in the objective function is the standard sum of weighted distance between clusters in iith data site and the second term minimizes the dissimilarity between $u_{ik}[ii]$ and $u_{ik}[jj]$ $jj = 1, \ldots, P, jj \neq ii$. Note that although RE is not symmetric, considering it as a measure of distance could not cause any problem, as it does not depend on data sites' processing order.

Horizontal and vertical modes are two important modes of the above method. Thus, Eq. (7) can be rewritten as Eq. (8) for horizontal mode (HRECFC) and Eq. (9) for vertical mode (VRECFC).

$$\min J_{HRECFC}[ii] = \sum_{k=1}^{N} \sum_{i=1}^{c^*} u_{ik}^m[ii] d_{ik}^2[ii] + \gamma_H[ii] \sum_{k=1}^{N} \sum_{i=1}^{c^*} \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} u_{ik}[ii] \ln\left(\frac{u_{ik}[ii]}{u_{ik}[jj]}\right)$$

$$S.T. \begin{cases} \sum_{i=1}^{c^*} u_{ik}[ii] = 1 & \forall k, ii \\ 0 < \sum_{k=1}^{N} u_{ik}[ii] < N & \forall i, ii \\ u_{ik}[ii] \in [0,1] & \forall i, k, ii \end{cases} \tag{8}$$

$$\min J_{VRECFC}[ii] = \sum_{k=1}^{N[ii]} \sum_{i=1}^{c^*} u_{ik}^m[ii] d_{ik}^2[ii] + \gamma_V[ii] \sum_{k=1}^{N[ii]} \sum_{i=1}^{c^*} \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} u_{ik}[ii] \ln\left(\frac{u_{ik}[ii]}{\tilde{u}_{ik}[jj]}\right)$$

$$S.T. \begin{cases} \sum_{i=1}^{c^*} u_{ik}[ii] = 1 & \forall k, ii \\ 0 < \sum_{k=1}^{N[ii]} u_{ik}[ii] < N[ii] & \forall i, ii \\ u_{ik}[ii] \in [0,1] & \forall i, k, ii \end{cases} \tag{9}$$

where, $\gamma_H[ii]$ and $\gamma_V[ii]$ are non-negative coefficients of collaboration for HRECFC and VRECFC, respectively.

Theorem 1. ***-HRECFC***: *$u_{ik}[ii]$ that optimizes Eq. (8) is obtained by*

$$u_{ik}[ii] = \left(\left(\frac{\frac{m(m-1)d_{ik}^2[ii]}{\gamma_H[ii]}}{W_0\left[\frac{m(m-1)d_{ik}^2[ii]}{\gamma_H[ii]} \exp\left(-(m-1)\left(\frac{\gamma_H[ii] \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} (1 - \ln(u_{ik}[jj])) - \lambda_k}{\gamma_H[ii]} \right) \right) \right]} \right)^{1/m-1} \right)^{-1} \tag{10}$$

where

$$\gamma_H[ii] = \max\left\{ 0, \frac{md_{ik}^2[ii] \exp\left(\frac{m-1}{P} \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \ln(u_{ik}[jj]) \right) - \lambda_k}{-1 + \left(1 - \frac{1}{P}\right) \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \ln(u_{ik}[jj])} \right\} \tag{11}$$

$$0 \leq \lambda_k \leq \frac{1}{c^*}\left(\gamma_H[ii] - \frac{\gamma_H[ii]}{e(m-1)} - \gamma_H[ii] \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \ln(u_{ik}[jj]) \right) \tag{12}$$

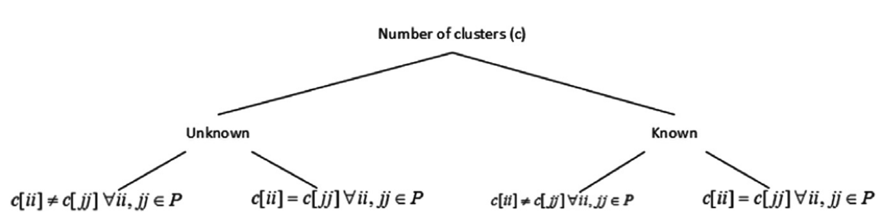

Fig. 1. Number of clusters.

$W_0(.)$ is the principle branch of Lambert-W function and λ_k is the Lagrangian multiplier. By considering $d_{ik}[ii]$ as Euclidean distance, the prototypes are updated by

$$v_i[ii] = \frac{\sum_{k=1}^{N} u_{ik}^m[ii] x_k[ii]}{\sum_{k=1}^{N} u_{ik}^m[ii]} \tag{13}$$

Proof. The proofs can be found in Appendix □

The main steps of HRECFC are as follows:

Algorithm 1: HRECFC
Initial Phase:
Step 1: If $c[ii]$sare known:
Step 1.1: If $c[ii]$s are the same, $c^* = c[.]$, run FCM for each data site and obtain $u_{ik}[ii]$.
Step 1.2: If $c[ii]$s are not the same, run FCM for each data site and calculate VIs, $VI[ii]$.
Step 1.2.1: Find the optimum value of VIs and the corresponding$c$,c^*.
Step 1.2.2: Re-cluster data sites using $c^*$and obtain $u_{ik}[ii]$.
Step 2: If $c[ii]$s are unknown, for each data site run FCM and obtain $c[ii]$.
Step 2.1: If $c[ii]$s are the same, $c^* = c[.]$, obtain $u_{ik}[ii]$.
Step 2.2: If $c[ii]$s are not the same:
Step 2.2.1: Find the optimum value of VIs and the corresponding$c$,c^*.
Step 2.2.2: Re-cluster data sites using $c^*$and obtain $u_{ik}[ii]$.
Collaborative phase:
Repeat until $\left(\sum_{ii=1}^{P} (\|u_{ik}^{(t-1)}[ii] - u_{ik}^{(t)}[ii]\ \|) < \varepsilon\right)$:
Step 3: Calculate $\gamma_H[ii]$ using Eq. (11).
Step 4: Determine $\lambda_k \ \ \forall k$, using Eq. (12).
Step 5: Compute $u_{ik}^{(t)}[ii] \ \ \forall i, k$, using Eq. (10).
Step 6: Update $v_i[ii] \ \ \forall i$, using Eq. (13).
Step 7: Increment t.

Theorem 2. ***-VRECFC**: $u_{ik}[ii]$ that optimizes Eq. (9) is obtained by*

$$u_{ik}[ii] = \left(\left(\frac{\frac{m(m-1)d_{ik}^2[ii]}{\gamma_V[ii]}}{W_0\left[\frac{m(m-1)d_{ik}^2[ii]}{\gamma_V[ii]}\exp\left(-(m-1)\left(\frac{\gamma_V[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}(1-\ln(\tilde{u}_{ik}[jj]))-\mu_k}{\gamma_V[ii]}\right)\right)\right]}\right)^{1/m-1}\right)^{-1} \tag{14}$$

where

$$\gamma_V[ii] = \max\left\{0, \frac{md_{ik}^2[ii]\exp\left(\frac{m-1}{P}\sum_{\substack{jj=1\\ jj\neq ii}}^{P}\ln(\tilde{u}_{ik}[jj])\right)-\mu_k}{-1+\left(1-\frac{1}{P}\right)\sum_{\substack{jj=1\\ jj\neq ii}}^{P}\ln(\tilde{u}_{ik}[jj])}\right\} \tag{15}$$

$$0 \le \mu_k \le \frac{1}{c^*}\left(\gamma_V[ii] - \frac{\gamma_V[ii]}{e(m-1)} - \gamma_V[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}\ln(\tilde{u}_{ik}[jj])\right) \tag{16}$$

and μ_k, $k = 1, \ldots, N[ii]$ is the Lagrangian multiplier. By considering $d_{ik}[ii]$ as Euclidean distance, the prototypes are updated by:

$$v_i[ii] = \frac{\sum_{k=1}^{N[ii]} u_{ik}^m[ii] x_k[ii]}{\sum_{k=1}^{N[ii]} u_{ik}^m[ii]} \tag{17}$$

Proof. The proofs can be found in Appendix. □

The main steps of VRECFC are as follows:

Algorithm 2: VRECFC
Initial phase:
Step 1: If $c[ii]$s are known:
Step 1.1: If $c[ii]$s are the same,$c^* = c[.]$, run FCM for each data site and obtain $v_i[ii]$.
Step 1.2: If $c[ii]$s are not the same, run FCM for each data site and calculate VIs, $VI[ii]$.
Step 1.2.1: Find the optimum value of VIs and the corresponding$c$,c^*.
Step 1.2.2: Re-cluster data sites using $c^*$and obtain $v_i[ii]$.
Step 2: If $c[ii]$s are unknown, for each data site run FCM and obtain $c[ii]$.
Step 2.1.: If $c[ii]$sare the same, $c^* = c[.]$, obtain $v_i[ii]$.
Step 2.2: If $c[ii]$sare not the same:
Step 2.2.1: Find the optimum value of VIs and the corresponding$c$,c^*.
Step 2.2.2: Re-cluster data sites using $c^*$and obtain $v_i[ii]$.
Collaborative phase:
Repeat until $\left(\sum_{ii=1}^{P} (\|u_{ik}^{(t-1)}[ii] - u_{ik}^{(t)}[ii]\ \|) < \varepsilon\right)$:
Step 3: Calculate$\tilde{u}_{ik}[jj] = \frac{1}{\sum_{l=1}^{c}\left(\frac{\|x_k[ii]-v_i[jj]\|}{\|x_k[ii]-v_l[jj]\|}\right)^2}$.
Step 4: Determine $\gamma_V[ii]$ using Eq. (15).
Step 5: Determine $\mu_k \ \ \forall k$, using Eq. (16).
Step 6: Compute $u_{ik}^{(t)}[ii] \ \ \forall i, k$, using Eq. (14).
Step 7: Update $v_i[ii] \ \ \forall i$, using Eq. (17).
Step 8: Increment t.

The computational complexity is another important aspect of the proposed method that has to be discussed.

Suppose that each data site has different number of data pattern, $N[ii]$, and different number of features, $F[ii]$. In the first phase, FCM runs asymptotically in $O(N[ii]c^{*2}F[ii])$time and the optimization in the second phase runs in $O(N[ii]c^{*2} F[ii]) + O(2N[ii]c^* \log(N[ii]c^*))$time. Thus, for each data site, the overall computational complexity is $O(2N[ii]c^{*2}F[ii]) + O(2N[ii]c^* \log(N[ii]c^*))$.

In fact, this value is convenient for$N[ii] \ge 100$and RECFC linearly increases the computational time. Moreover, the value of $W_0(.)$(Eq. (23)) could be easily calculated using numerical approximations ([24,25]) or using software packages, such as MATLAB and Maple.

4. Experimental results

To evaluate the performance of the proposed method, two categories of experiments, experiments on horizontal mode and experiments on vertical mode, are provided. Based on Table 1,

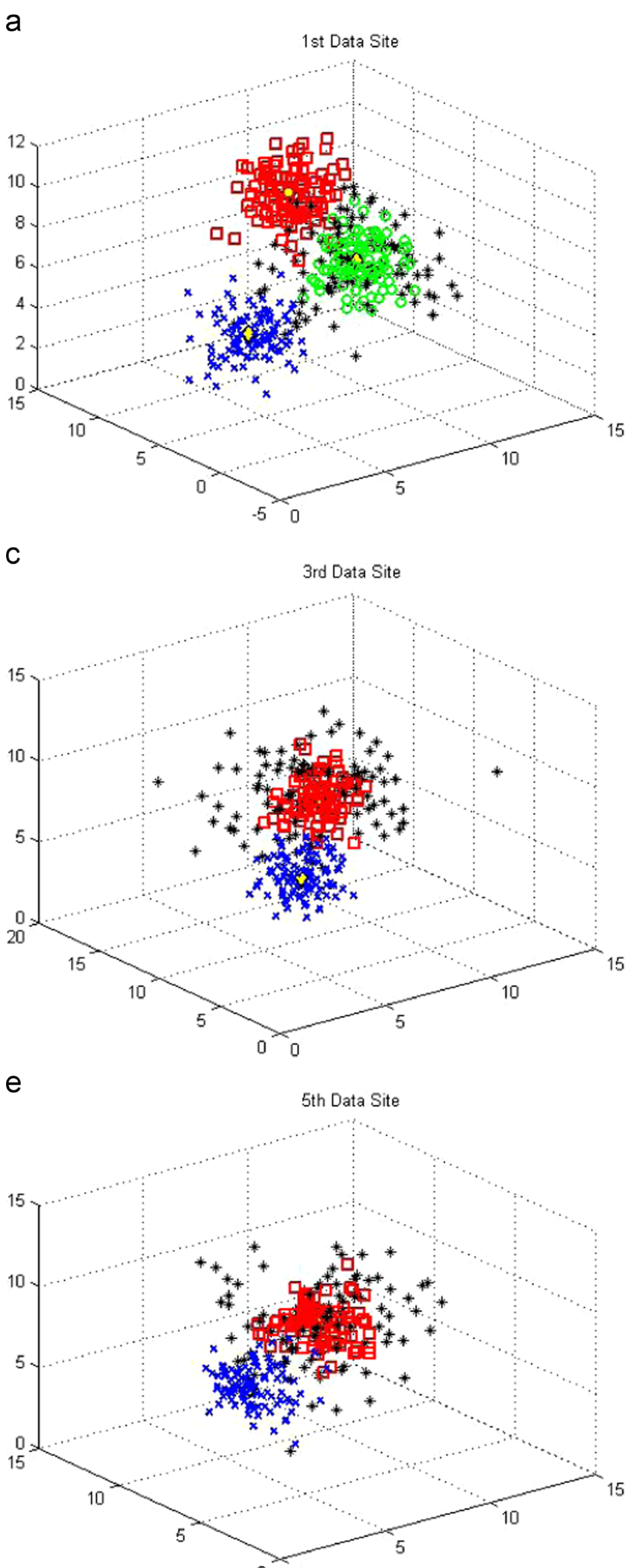

Fig. 2. The data with added noise points in (a) D[1], (b) D[2], (c) D[3], (d) D[4], and (e) D[5].

performance of HRECFC is compared with CF and CFC-c^* and performance of VRECFC is compared with CFC, CFC-$\beta_f - c^*$, and CFC-β_d.

The comparisons, in both categories of experiments, are done using global level of consistency, an indicator of quality of collaboration (Eq. (18) and Eq. (19) for horizontal and vertical modes, respectively) [15]:

$$w_H = \sum_{ii=1}^{P} \frac{2}{N^2} \sum_{jj=1}^{P} \sum_{k_1=1}^{N} \sum_{k_2>k_1}^{N} |prox(k_1,k_2)[ii] - prox(k_1,k_2)[jj]| \quad (18)$$

$$w_V = \sum_{ii=1}^{P} \frac{2}{N^2[ii]} \sum_{jj=1}^{P} \sum_{k_1=1}^{N[ii]} \sum_{k_2>k_1}^{N[ii]} |prox(k_1,k_2)[ii] - prox(k_1,k_2)[ii|jj]| \quad (19)$$

where, $prox(k_1,k_2)[ii] = \sum_{i=1}^{c^*} \min(u_{ik_1}[ii], u_{ik_2}[ii])$, $prox(k_1,k_2)[jj] = \sum_{i=1}^{c^*} \min(u_{ik_1}[jj], u_{ik_2}[jj])$ and $prox(k_1,k_2)[ii|jj] = \sum_{i=1}^{c^*} \min(u_{ik_1}[ii|jj], u_{ik_2}[ii|jj])$.

In addition, in order to examine, statistically, whether there is a meaningful difference between the results of global level of consistency, Kruskal–Wallis test, nonparametric one-way analysis of variance, is applied [26].

The following datasets are also used in this section:

- Artificial dataset: to evaluate the performance of the proposed method in presence of noise, five 3-dimensional data sites with 300 data pattern each are generated based on normal distribution with $\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and the following parameters:

  D[1]: $\mu_1 = [4, \ 4.5, \ 4.5], \mu_2 = [8.5, \ 9, \ 9], \ \mu_3 = [10, \ 6, \ 6]$
  D[2]:
  $\mu_1 = [4.5, \ 4.5, \ 5], \mu_2 = [9.5, \ 5, \ 6], \ \mu_3 = [10.5, \ 10, \ 8]$
  D[3]: $\mu_1 = [4.5, \ 6, \ 6], \mu_2 = [7, \ 9, \ 9]$
  D[4]: $\mu_1 = [6, \ 4, \ 4], \mu_2 = [6, \ 10, \ 10], \ \mu_3 = [10, \ 10, \ 10]$
  D[5]: $\mu_1 = [4, \ 7, \ 6], \mu_2 = [8, \ 8, \ 8]$

  100 noise objects are then added to each data site.
  Fig. 2(a)–(e) shows the scatterplots of the random data sites generated based on the above structure.
- Wisconsin breast cancer (diagnostic) dataset contains 569 data in 32 feature spaces with benign and malignant label. Using the horizontal partitioning procedure [22], this dataset is divided into four data sites with the same number of data pattern.
- Climate model simulation crashes dataset has 540 data in 18 feature spaces and in two classes of failure and success. Using the horizontal partitioning procedure [22], this dataset is divided into 2 data sites with the same number of data pattern.
- Stock market dataset [27] is a yearly collection of different companies' monthly market capital in London stock exchange from 1999 to 2013. In other words, it consists of 15 data sites with different number of data patterns in 12 feature spaces.
- Life expectancy at birth [28] is a key indicator of urban health reported by world health organization and consists of life expectancy for men and women of several countries in 1990, 2000, and 2012. This dataset consists of 3 data sites with different number of data patterns in 2 feature spaces.
- World Bank collection of development indicators [29] is an annual report of 9 critical development indicators for different countries from 2005 to 2013. This dataset consists of 9 data sites with different number of data patterns in 9 feature spaces.

4.1. Experiments on horizontal mode

In this section, performance of HRECFC (Algorithm 1) is compared with CF and CFC-c^*using artificial dataset, Wisconsin breast cancer, and climate model simulation crashes datasets. The results of global level of consistency for horizontal mode,w_H (Eq. (18)), are reported in Table 2.

Table 2
The results of global level of consistency for CF, CFC-c^* and HRECFC methods.

w_H	CF	CFC-c^*	HRECFC
Artificial dataset	39.078	18.650	1.132
Wisconsin breast cancer	42.358	20.012	7.783
Climate model simulation crashes	36.231	11.926	2.712

Table 3
The results of global level of consistency for CFC, CFC-$\beta_f - c^*$, CFC-$\beta_d$and VRECFC methods.

w_V	CFC	CFC-$\beta_f - c^*$	CFC-β_d	VRECFC
Stock market	64.621	19.567	18.625	9.292
Life expectancy at birth	14.574	13.375	12.427	5.004
Development indicators	13.085	11.239	11.364	3.897

The Kruskal–Wallis test is also applied to examine, statistically, whether there is a meaningful difference between the results ofw_Hs. It shows that at confidence level of 0.95, the difference is significant and the proposed method has the lowest value of w_Hs and therefore highest quality of collaboration.

4.2. Experiments on vertical mode

To evaluate the performance of the proposed method in vertical mode, VRECFC (Algorithm 2), stock market, life expectancy at birth and World Bank collection of development indicators datasets are used. Each dataset has different number of data patterns and unknown number of clusters but in the same feature space. VRECFC, CFC, CFC-$\beta_f - c^*$, CFC-β_d methods are applied and the global levels of consistency,w_V s, are reported in Table 3.

The Kruskal–Wallis test is also applied to examine the existence of meaningful difference between the results of w_V s. It shows that at confidence level of 0.90, the difference is significant and the proposed method has the lowest value of w_Vs and highest quality of collaboration.

5. Conclusion

In this paper, a novel CFC method is proposed, using RE concept, and its two modes, horizontal and vertical, are discussed. Several experiments are provided and the results are compared with the ones obtained from five well-known CFC methods. The results show that the proposed method has the highest global level of consistency and the highest quality of collaboration. The proposed RECFC method is based on Type-1 fuzzy logic. However, there are many uncertainties and vagueness in real world, which may cause uncertainties in fuzzy set's parameters and so assigning an exact membership degree to data (Type-1 fuzzy) is impossible. Thus, developing the proposed method using Type-2 fuzzy logic could be a potential future work of this study.

Conflict of interest

None declared.

Appendix

Proof of Theorem 1

By including constraint term $(\sum_{i=1}^{c^*} u_{ik}[ii]=1)$ in objective function of Eq. (8) using Lagrangian multiplier, λ_k $k=1,\dots,N$, minimization of the following function is our concern:

$$J_{HRECFC}[ii]=\sum_{k=1}^{N}\sum_{i=1}^{c} u_{ik}^{m}[ii]d_{ik}^{2}[ii]+\gamma_H[ii]\sum_{k=1}^{N}\sum_{i=1}^{c^*}\sum_{\substack{jj=1\\ jj\neq ii}}^{P} u_{ik}[ii]\ln\left(\frac{u_{ik}[ii]}{u_{ik}[jj]}\right)$$

$$-\sum_{k=1}^{N}\lambda_k\left(\sum_{i=1}^{c^*}u_{ik}[ii]-1\right) \tag{20}$$

The three necessary conditions leading to the local minimum are

$$\frac{\partial J_{HRECFC}[ii]}{\partial u_{ik}[ii]}=0,\ \frac{\partial J_{HRECFC}[ii]}{\partial \gamma_H[ii]}=0,\ \frac{\partial J_{HRECFC}[ii]}{\partial \lambda_k}=0\quad \forall i,k.$$

The first necessary condition, $\partial J_{RECFC}[ii]/\partial u_{ij}[ii]=0$, results in:

$$mu_{ik}^{m-1}[ii]d_{ik}^{2}[ii]+\gamma_H[ii]\sum_{jj=1}$$

$$jj\neq ii^{P}(1+\ln(u_{ik}[ii])-\ln(u_{ik}[jj]))-\lambda_k=0 \tag{21}$$

Thus, for ii^{th} data site, $u_{ik}[ii]$ is obtained by:

$$u_{ik}[ii]=\left(\left(\frac{\frac{m(m-1)d_{ik}^2[ii]}{\gamma_H[ii]}}{W_0\left[\frac{m(m-1)d_{ik}^2[ii]}{\gamma_H[ii]}\exp\left(-(m-1)\left(\frac{\gamma_H[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}(1-\ln(u_{ik}[jj]))-\lambda_k}{\gamma_H[ii]}\right)\right)\right]}\right)^{1/m-1}\right)^{-1} \tag{22}$$

where, $W_0(.)$is defined to be the function satisfying [1]:

$$W_0(Z)\exp(W_0(Z))=Z \tag{23}$$

The second necessary condition is $\partial J_{HRECFC}[ii]/\partial\gamma[ii]=0$, so:

$$\gamma_H[ii]=\max\left\{0,\frac{md_{ik}^2[ii]\exp\left(\frac{m-1}{P}\sum_{\substack{jj=1\\ jj\neq ii}}^{P}\ln(u_{ik}[jj])\right)-\lambda_k}{-1+(1-\frac{1}{P})\sum_{\substack{jj=1\\ jj\neq ii}}^{P}\ln(u_{ik}[jj])}\right\} \tag{24}$$

For the third necessary condition, $\partial J_{RECFC}[ii]/\partial\lambda_k=0$, solving $\sum_{i=1}^{c^*}u_{ik}[ii]=1$ with respect to λ_k would not result in an exact solution, so the bounds for λ_k have to be found. This problem could be studied from two viewpoints, (1) $u_{ik}[ii]\geq 0\ \forall i,k$ and (2) $u_{ik}[ii]\leq 1\ \forall i,k$.

First, consider $u_{ik}[ii]\geq 0\ \forall i,k$:

$$u_{ik}[ii]\geq 0\rightarrow W_0\left[\frac{m(m-1)d_{ik}^2[ii]}{\gamma_H[ii]}\exp\left(-(m-1)\left(\frac{\gamma_H[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}(1-\ln(u_{ik}[jj]))-\lambda_k}{\gamma_H[ii]}\right)\right)\right]>0$$

$W_0(.)$is a monotonic increasing function with three key characteristics: $W_0(.)\geq -1$, $W_0\left(-\frac{1}{e}\right)=-1$ and $W_0(0)=0$ [1]. So,

$$\frac{m(m-1)d_{ik}^2[ii]}{\gamma_H[ii]}\exp\left(-(m-1)\left(\frac{\gamma_H[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}(1-\ln(u_{ik}[jj]))-\lambda_k}{\gamma_H[ii]}\right)\right)>0$$

which is always true. Now consider $u_{ik}[ii]\leq 1\quad \forall i,k$:

$$-\frac{1}{e}\leq -(m-1)\left(\frac{\gamma_H[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}(1-\ln(u_{ik}[jj]))-\lambda_k}{\gamma_H[ii]}\right)$$

$$\leq W_0\left[\frac{m(m-1)d_{ik}^2[ii]}{\gamma_H[ii]}\exp\left(-(m-1)\left(\frac{\gamma_H[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}(1-\ln(u_{ik}[jj]))-\lambda_k}{\gamma_H[ii]}\right)\right)\right]$$

So, the bounds for $\lambda_k$are:

$$0\leq\lambda_k\leq\frac{1}{c^*}\left(\gamma_H[ii]-\frac{\gamma_H[ii]}{e(m-1)}-\gamma_H[ii]\sum_{\substack{jj=1\\ jj\neq ii}}^{P}\ln(u_{ik}[jj])\right) \tag{25}$$

By differentiating Eq. (8) with respect to the prototypes, $v_i[ii]$ are obtained. Here, the $d_{ik}[ii]$ is considered Euclidean distance function, so the prototypes are updated by:

$$v_i[ii]=\frac{\sum_{k=1}^{N}u_{ik}^{m}[ii]x_k[ii]}{\sum_{k=1}^{N}u_{ik}^{m}[ii]} \tag{26}$$

∎End of proof.∎

Proof of Theorem 2 This theorem can be proved in the same manner as Theorem 1. ∎

References

[1] M. Zarinbal, M.H. Fazel Zarandi, I. Turksen, Relative entropy fuzzy c-means clustering, Inf. Sci. 260 (2014) 74–97.

[2] J.C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c-means clustering algorithm, Comput. Geosci. 10 (1984) 191–203.

[3] R.J. Hathaway, J.C. Bezdek, Switching regression models and fuzzy clustering, IEEE Trans. Fuzzy Syst. 1 (1993) 195–204.

[4] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst. 1 (1993) 98–110.

[5] P.-L. Lin, P.-W. Huang, C.H. Kuo, Y.H. Lai, A size-insensitive integrity-based fuzzy c-means method for data clustering, Pattern Recognit. 47 (2014) 2042–2056.

[6] X. Zhi, J. Fan, F. Zhao, Fuzzy Linear discriminant analysis-guided maximum entropy fuzzy clustering algorithm, Pattern Recognit. 46 (2013) 1604–1615.

[7] D. Graves, J. Noppen, W. Pedrycz, Clustering with proximity knowledge and relational knowledge, Pattern Recognit. 45 (2012) 2633–2644.

[8] S. Mitra, W. Pedrycz, B. Barman, Shadowed c-means: Integrating fuzzy and rough clustering, Pattern Recognit. 43 (2010) 1282–1291.

[9] Husseinzadeh Kashan, B. Rezaee, S. Karimiyan, An efficient approach for unsupervised fuzzy clustering based on grouping evolution strategies, Pattern Recognit. 46 (2013) 1240–1254.

[10] R. Senge, S. Bösner, K. Dembczyński, J. Haasenritter, O. Hirsch, N. Donner-Banzhoff, et al., Reliable classification: learning classifiers that distinguish aleatoric and epistemic uncertainty, Inf. Sci. 255 (2014) 16–29.

[11] M.H. Fazel Zarandi, M. Zarinbal, N. Ghanbari, I.B. Turksen, A new fuzzy functions model tuned by hybridizing imperialist competitive algorithm and simulated annealing. Application: Stock price prediction, Inf. Sci. 222 (2013) 213–228.

[12] M.H. Fazel Zarandi, M. Zarinbal, M. Izadi, Systematic image processing for diagnosing brain tumors: a Type-II fuzzy expert system approach, Appl. Soft Comput. 11 (2011) 285–294.

[13] F. Xie, A.C. Bovik, Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm, Pattern Recognit. 46 (2013) 1012–1019.

[14] W. Pedrycz, Knowledge-Based Clustering from Data to Information Granules, John Wiley and Sons, Inc., 2005.

[15] W. Pedrycz, P. Rai, Collaborative clustering with the use of Fuzzy C-Means and its quantification, Fuzzy Sets Syst. 159 (2008) 2399–2427.

[16] W. Pedrycz, Collaborative fuzzy clustering, Pattern Recognit. Lett. 23 (2002) 1675–1686.

[17] R. Falcon, G. Jeon, R. Bello, J. Jeong, Learning collaboration links in a collaborative fuzzy clustering environment, in: Proceedings of the MICAI on

Advanced Artificial Intelligence, Lecture Notes in Computer Science, 2007, pp. 483–495.
[18] F. Yu, J. Tang, R. Cai, A Necessary preprocessing in horizontal collaborative fuzzy clustering, in: Proceedings of the IEEE International Conference on Granular Computing (GRC7), 2007 p. 399.
[19] R. Falcón, B. Depaire, K. Vanhoof, A. Abraham, Towards a suitable reconciliation of the findings in collaborative fuzzy clustering, in: Proceedings of the Eighth International Conference on Intelligent System Design and Applications, 2008, pp. 652–657.
[20] B. Depaire, R. Falcon, K. Vanhoof, G. Wets, PSO driven collaborative clustering: a clustering algorithm for ubiquitous environments, Intell. Data Anal. 15 (2011) 49–68.
[21] L.F.S. Coletta, L. Vendramin, E.R. Hruschka, R.J.G.B. Campello, W. Pedrycz, Collaborative fuzzy clustering algorithms: some refinements and design guidelines, IEEE Trans. Fuzzy Syst. 20 (2012) 444–462.
[22] M. Ghassany, N. Grozavu, Y. Bennani, Collaborative clustering using prototype-based techniques, Int. J. Comput. Intell. Appl. 11 (2012).
[23] M. Ghassany, Y. Bennani, Collaborative fuzzy clustering of variational Bayesian generative topographic, 2014.
[24] D.A. Barry, J. Parlange, L. Li, H. Prommer, C.J. Cunningham, F. Stagnitti, Analytical approximations for real values of the Lambert W-function, Math. Comput. Simul. 53 (2000) 95–103.
[25] F. Chapeau-blondeau, A. Monir, Numerical evaluation of the Lambert W function and application to generation of generalized, IEEE Trans. Signal Process. 50 (2002) 2160–2165.
[26] E. Alpaydin, Introduction to Machine Learning, 2nd ed., The MIT Press, 2010.
[27] ⟨http://www.londonstockexchange.com/statistics/historic/company-files/company-files.htm⟩.
[28] ⟨http://apps.who.int/gho/data/view.main.100000⟩.
[29] ⟨http://data.worldbank.org/data-catalog/world-development-indicators⟩.

M. Zarinbal is a Ph.D. candidate at the Department of Industrial Engineering of Amirkabir University of Technology, Tehran, Iran, and a gold member of IEEE. Her main research interests are soft computing, computational intelligence, fuzzy sets and systems, image processing, and medical diagnosis.

M.H. Fazel Zarandi is a Professor at the Department of Industrial Engineering ofAmirkabir University of Technology, Tehran, Iran, and a member of the Knowledge-Information SystemsLaboratory at the University of Toronto, Ontario,Canada. His main research interests focus on intelligent information systems, soft computing, computational intelligence, fuzzy sets and systems, and multi-agent systems.

I.B. Turksen is the head of department of Industrial Engineering at TOBB Economics and Technology University Ankara, Turkey, and a member of the Knowledge-Information Systems Laboratory at University of Toronto, Ontario, Canada. His main research interests focus on System Analysis, Diagnosis and Prediction, Intelligent Manufacturing System Modeling, Knowledge/Intelligence-Data Mining and Systems Modeling for financial data analysis, consumer preference and market share analysis, Image processing and diagnosis and treatment of cancer cells, Network Simulations, and fuzzy-reinforcement learning.

Please cite this article as: M. Zarinbal, et al., Relative entropy collaborative fuzzy clustering method, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.09.018