Feature-Reduction Fuzzy Co-Clustering Algorithm for Hyperspectral Image Segmentation

Van Nha Pham
Department of Information Systems,
Le Quy Don Technical University,
236 Hoang Quoc Viet, Hanoi, Vietnam
famvannha@gmail.com

Long Thanh Ngo
Department of Information Systems,
Le Quy Don Technical University,
236 Hoang Quoc Viet, Hanoi, Vietnam
ngotlong@mta.edu.vn

Thao Duc Nguyen
MIST Institute of Science and Technology,
Hoang Sam, Hanoi, Vietnam
thaodr@gmail.com

Abstract— The fuzzy co-clustering algorithms are considered as effective technique for clustering the complex data, such as high-dimensional and large size. In general, features of data objects are considered the same importance. However, in reality, the features have different roles in data analyses; even some of them are considered redundancy in the individual case for data sets. Removing these features is a way for the dimensionality reduction, which needs to improve the performance of data processing algorithms. In this paper, we proposed an improved fuzzy co-clustering algorithm called feature-reduction fuzzy co-clustering (FRFCoC), which can automatically calculate the weight of features and put them out of the data processing. We considered the objective function of the FCoC algorithm with feature-weighted entropy and build a learning procedure for components of the objective function, then reducing the dimension of data by eliminating irrelevant features with small weights. Experiments were conducted on synthetic data sets and hyperspectral image using the robust assessment indexes. Experimental results demonstrated the proposed algorithm outperformed the previous algorithms.

Keywords— Cluster analysis; fuzzy co-clustering; dimensionality reduction; hyperspectral image proccessing

I. Introduction

The co-clustering is a useful tool for data analysis, where data objects are grouped into a number of clusters according to their similarity in multi-dimensional spaces. These techniques can simultaneously perform on both pattern space and feature space, which are suitable with problems of complex data that has the multi-dimensional data [1]. Some co-clustering techniques have been studied to address problems such as, clustering documents and keywords [2-4], color segmentation [5], categorical multivariate data [6] and high-dimensional data [7]. Another co-clustering algorithm was proposed by the usage of interval valued fuzzy sets to obtain the good clustering quality, but its computational complexity is high [8]. In general, these clustering techniques treat all features of data with equal importance. However, in practice, the data usually include some unimportant features, called redundant features, which could fatally affect to the clustering quality. The removing these feature will increase the performance of clustering algorithms [9]. Several clustering algorithm like K-Means, Fuzzy C-Means (FCM) and Fuzzy Co-Clustering (FCoC) are known as the well-known techniques. Some variants of the feature-weighted clustering algorithms had been proposed, such as weighted K-Means [10,11], weighted FCM uses feature-weight learning (WFCM) [13-16]. Although these algorithms may improve the performance of clustering, they were not to consider a feature-reduction schema.

Related to multivariate data with high-dimensional and large size, hyperspectral images have a very complex structure. The fuzzy co-clustering algorithms [7, 8] can be seen as effective tool to process kinds of multivariate data in general and hyperspectral image in particular. However, in fact, the hyperspectral image has many redundant data feature components; these data greatly affect the quality and speed of the hyperspectral image processing. Therefore, the dimensionality reduction is important task to hyperspectral image processing. Some methods of dimensionality reduction have studied such as band selection [19, 20], discriminant analysis [21, 22], principal component analysis [23] and some other studies [24, 25]. Recently, MS. Yang and Y. Nataliani have proposed an algorithm called a feature-reduction FCM (FRFCM) [9] that can automatically compute different feature weights by considering the FCM objective function with feature-weighted entropy, which has a feature-reduction schema to eliminate these irrelevant features with small weights such that the computational time can be decreased with better clustering performance.

In this paper, we proposed an improved fuzzy co-clustering algorithm called FRFCoC, which use feature-reduction schema in [9] that can automatically calculate the weight of each feature and remove these features out of the data processing, which achieve the goal to increase efficiency and reduce the processing time of clustering algorithm FCoC. We have conducted some experiments on the synthetic data sets and the hyperspectral image using the assessment indexes to demonstrate the performance of the proposed algorithm.

The rest of the paper is organized as follows. Section 2, briefly review some related works of feature-weighted FCM clustering algorithms; Section 3, presents the proposed FRFCoC algorithm; Section 4 shows experimental results with comparisons; Section 5 includes the conclusion and several future works.

978-1-5090-6034-4/17/$31.00 ©2017 IEEE

II. BACKGROUND

A. Fuzzy Co-Clustering algorithm

There are three fuzzy co-clustering algorithms were proposed theoretically. That is, the algorithm FCCM [2], FCCI [5] and Fuzzy CoDoK[6]. These algorithms share a general conception in which co-clustering is considered to be a combination between the data object partition and the feature classification.

The objective function J_{FCoC} (U, V, P) is represented in the form

$$J_{FCoC}(U,V,P)=\sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{K}u_{ci}v_{cj}d_{cij}+T_U\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ci}\log u_{ci}+T_V\sum_{c=1}^{C}\sum_{j=1}^{K}v_{cj}\log v_{cj} \quad (1)$$

where, N, C, K, respectively, the number of data objects, the number of clusters and the number of features; $X=(x_1, x_2, \ldots, x_N)$, $x_i \in X$ (i=1, 2, …, N), be the input data set in *K*-dimensional feature space. x_{ij} denotes feature *j* of data object *i*; $P=(p_1, p_2, \ldots, p_C)$, $p_c \in P$ (c=1, 2, …, C), set of feature-based centroids in *K*-dimensional feature space. u_{ci} and v_{cj} are the object membership function and the feature membership function; T_u and T_v are weights that indicate fuzzy level; d_{cij} be distance between x_{ij} and p_{cj}.

The components of objective function FCoC are defined by formulas (2), (3) and (4).

$$u_{ci}=\frac{e^{\left(-\sum_{j=1}^{K}\frac{v_{cj}d_{cij}}{T_u}\right)}}{\sum_{c=1}^{C}e^{\left(-\sum_{j=1}^{K}\frac{v_{cj}d_{cij}}{T_u}\right)}} \quad (2)$$

$$v_{cj}=\frac{e^{\left(-\sum_{i=1}^{N}\frac{u_{ci}d_{cij}}{T_v}\right)}}{\sum_{j=1}^{K}e^{\left(-\sum_{i=1}^{N}\frac{u_{ci}d_{cij}}{T_u}\right)}} \quad (3)$$

$$p_{cj}=\frac{\sum_{i=1}^{N}u_{ci}x_{ij}}{\sum_{i=1}^{N}u_{ci}} \quad (4)$$

B. Feature-Reduction Fuzzy Clustering Algorithm FRFCM

MS. Yang et al [9] proposed algorithm FRFCM to improve the performance of algorithm FCM. The objective function of FRFCM algorithm is indicated in the form

$$J(U,V,\mathrm{W})=\sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{D}u_{ci}^{m}\delta_j \mathrm{w}_j(\mathrm{x}_{ij}-p_{cj})^2+T_{\mathrm{w}}\sum_{j=1}^{D}(w_j log\delta_j w_j) \quad (5)$$

Where, $W=[w_j]_{1xD}$ with w_j being a feature weight of the j^{th} feature, $\delta=[\delta_j]_{1xD}$ with δ_j is used to control feature weight w_j, δ_j is calculated as follows

$$\delta_j=\left(\frac{mean(x)}{var(x)}\right)_j$$

The components of objective function FRFCM are defined by formulas (6) and (7).

$$u_{ci}=\frac{\left(\sum_{j}^{D}\delta_j w_j(x_{ij}-p_{cj})^2\right)^{-1/(m-1)}}{\sum_{k=1}^{C}\left(\sum_{j}^{D}\delta_j w_j(x_{ij}-p_{kj})^2\right)^{-1/(m-1)}} \quad (6)$$

$$w_j=\frac{\frac{1}{\delta_j}e^{\left(\frac{-C\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ci}^{m}\delta_j(\mathrm{x}_{ij}-\mathrm{p}_{cj})^2}{N}\right)}}{\sum_{p=1}^{D}\frac{1}{\delta_j}e^{\left(\frac{-C\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ci}^{m}\delta_p(\mathrm{x}_{ip}-\mathrm{p}_{cp})^2}{N}\right)}} \quad (7)$$

and p_{cj} is calculated by formula (4).

III. THE FEATURE-REDUCTION FUZZY CO-CLUSTERING ALGORITHM FRFCOC

In this section, we present a new method using feature-weighted entropy to improve the performance of the algorithm FCoC in which this technique is considered as dimensionality reduction problem, called FRFCoC. In this method, each feature has an individual weight updated through each iteration. In this process, features with small weight will be removed from datasets. Considering the D-dimensional dataset $X=\{x_1, x_2, \ldots, x_N\}$, the weights of features are denoted as a row matrix $W=\{w_1, w_2, \ldots, w_D\}$. The objective function of algorithm FRFCoC is to be minimized in the form,

$$J_{FRFCoC}(U,V,P,\mathrm{W},\delta)=\sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{D}u_{ci}v_{cj}\delta_j \mathrm{w}_j d_{cij}+T_u\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ci}\log u_{ci}+ \quad (8)$$

$$T_v\sum_{c=1}^{C}\sum_{j=1}^{D}v_{cj}\log v_{cj}+T_{\mathrm{w}}\sum_{j=1}^{D}\mathrm{w}_j\log\delta_j \mathrm{w}_j$$

To get optimal clustering results, J_{FRFCoC} is minimized subject to the following constraints:

$$\sum_{c=1}^{C}u_{ci}=1, u_{ci}\in[0,1], \forall i=\overline{1,N}$$
$$\sum_{j=1}^{D}v_{cj}=1, v_{cj}\in[0,1], \forall c=\overline{1,\mathrm{C}} \quad (9)$$
$$\sum_{j}^{D}\mathrm{w}_j=1, \mathrm{w}_j\in[0,1], \forall \mathrm{j}=\overline{1,D}$$

Where, T_u, T_v and T_w are weights that indicate fuzzy level of U, V and W. δ_j is used to adjust feature weight w_j, the learning procedure for δ_j is analyzed below.

To minimize the objective function J_{FRFCoC} with constraints are given by (9), we construct an energy function with Lagrange coefficients λ_1 for constraint $\sum_{c=1}^{C}u_{ci}=1$, λ_2 for $\sum_{j=1}^{D}v_{cj}=1$ and λ_3 for $\sum_{j=1}^{D}\mathrm{w}_j=1$, according to Lagrange multiplier we obtain:

$$J_{FRFCoC}(U,V,P,\mathrm{W},\delta)=\sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{D}u_{ci}v_{cj}\delta_j \mathrm{w}_j d_{cij}+T_u\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ci}\log u_{ci}+T_v\sum_{c=1}^{C}\sum_{j=1}^{D}v_{cj}\log v_{cj}+T_{\mathrm{w}}\sum_{j=1}^{D}\mathrm{w}_j\log\delta_j \mathrm{w}_j+\lambda_1(\sum_{c=1}^{C}u_{ci}-1)+\lambda_2(\sum_{j=1}^{D}v_{cj}-1)+\lambda_3(\sum_{j=1}^{D}\mathrm{w}_j-1) \quad (10)$$

First, we calculate the membership function U by fixed V, P and W, then minimizing the objective function (10) according to the U, and taking derivatives of energy function given in(10) with respect to the fuzzy object memberships and setting them to zero, we obtain,

$$\frac{\partial J_{FRFCoC}}{\partial u_{ci}} = \sum_{j=1}^{K} v_{cj}\delta_j \mathrm{w}_j d_{cij} + T_u(\log u_{ci}+1)+\lambda_1 = 0 \qquad (11)$$

By some algebraic simplifications in (11), we obtain,

$$u_{ci} = \frac{e^{-\sum_{j=1}^{D}\frac{v_{cj}\delta_j \mathrm{w}_j d_{cij}}{T_u}}}{e^{\frac{\lambda_1}{T_u}}} \qquad (12)$$

Because of the constraint $\sum_{c=1}^{C} u_{ci} = 1$, the Lagrange multiplier λ_1 are eliminated as

$$\sum_{c=1}^{C} u_{ci} = \sum_{c=1}^{C}\frac{e^{-\sum_{j=1}^{D}\frac{v_{cj}\delta_j \mathrm{w}_j d_{cij}}{T_u}}}{e^{\frac{\lambda_1}{T_u}}} = \frac{\sum_{c=1}^{C} e^{-\sum_{j=1}^{D}\frac{v_{cj}\delta_j \mathrm{w}_j d_{cij}}{T_u}}}{e^{\frac{\lambda_1}{T_u}}} = 1 \qquad (13)$$

$$\Rightarrow e^{\frac{\lambda_1}{T_u}} = \sum_{c=1}^{C} e^{-\sum_{j=1}^{D}\frac{v_{cj}\delta_j \mathrm{w}_j d_{cij}}{T_u}} \qquad (14)$$

By plugging (14) into (12), the closed-form solution for the optimal object membership function is obtained as

$$u_{ci} = \frac{e^{-\sum_{j=1}^{D}\frac{v_{cj}\delta_j \mathrm{w}_j d_{cij}}{T_u}}}{\sum_{c=1}^{C} e^{-\sum_{j=1}^{D}\frac{v_{cj}\delta_j \mathrm{w}_j d_{cij}}{T_u}}} \qquad (15)$$

To find the optimal fuzzy feature memberships V, similar to U, we obtain,

$$v_{cj} = \frac{e^{(-\sum_{i=1}^{N}\frac{u_{ci}\delta_j \mathrm{w}_j d_{cij}}{T_v})}}{\sum_{j=1}^{D} e^{(-\sum_{i=1}^{N}\frac{u_{ci}\delta_j \mathrm{w}_j d_{cij}}{T_v})}} \qquad (16)$$

To find the cluster centroids P, we do the similar way.Note that, we use the square of Euclidean distance $\left\| x_{ij} - p_{cj} \right\|^2$ between the j^{th} feature of i^{th} pattern and the the j^{th} feature of the c^{th} cluster centroid which is computed as follows $\left\| x_{ij} - p_{cj} \right\|^2 = (x_{ij}-p_{cj})^2 = x_{ij}^2 - 2x_{ij}p_{cj} + p_{cj}^2$,
we obtain,

$$\frac{\partial J_{FRFCoC}}{\partial p_{cj}} = v_{cj}\delta_j \mathrm{w}_j \sum_{i=1}^{N} u_{ci}x_{ij} - v_{cj}\delta_j \mathrm{w}_j p_{cj}\sum_{i=1}^{N} u_{ci} = 0 \qquad (17)$$

By some algebraic simplifications in(17), we reach,

$$p_{cj} = \frac{\sum_{i=1}^{N} u_{ci}x_{ij}}{\sum_{i=1}^{N} u_{ci}} \qquad (18)$$

Next, to calculate w_j, U, V and P and the distances (x_{ij}-p_{cj}) are considered as constants, then taking derivatives of energy function (10) with respect to w_j and setting them to zero, we obtain,

$$\frac{\partial J_{FRFCoC}}{\partial \mathrm{w}_j} = \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci}v_{cj}\delta_j d_{cij} + T_{\mathrm{w}}(\log \mathrm{w}_j + 1) + \lambda_3 = 0 \qquad (19)$$

Similar to the U and V, from (19) we easily reach,

$$\mathrm{w}_j = \frac{\frac{1}{\delta_j} e^{-\sum_{c=1}^{C}\sum_{i=1}^{N}\frac{u_{ci}v_{cj}\delta_j d_{cij}}{T_{\mathrm{w}}}}}{\sum_{k=1}^{D}\frac{1}{\delta_k} e^{-\sum_{c=1}^{C}\sum_{i=1}^{N}\frac{u_{ci}v_{cj}\delta_j d_{cij}}{T_{\mathrm{w}}}}} \qquad (20)$$

Finally, we need to calculate the parameter δ_j which uses to control the w_j. Thus, the estimating the parameter is an important problem. To estimate δ_j, we review a learning procedure follows.
According to probability theory and statistics, standard deviation and variance are used to measure the dispersion of data. Another measurement as an index of dispersion is a well-known variance-to-mean ratio (VMR), defined as VMR=σ^2/μ [17]. VMR can be used to observe a dispersed or clustered data set. Smaller dispersion means the data set would be closer to the cluster centroid, while larger dispersion means the data set is far from the cluster centroid. Because we need to determine features with small dispersion and then discard features with large dispersion, we consider the inverse of VMR, i.e., mean-to-variance ratio (MVR) can generate minimum weight better than VMR [18]. Therefore, we used MVR to estimates δ_j for measuring separation between the data clusters. i.e.

$$\delta_j = \left(\frac{mean(x)}{\mathrm{var}(x)}\right)_j \qquad (21)$$

To create a feature-reduction diagram of the proposed algorithm, we need to identify small weighted features, and then remove them from the data in processing. FRFCoC algorithm is shown as follows.

Algorithm 1. Feature-Reduction Fuzzy Co-Clustering algorithm FRFCoC

Input: Data $X = \{x_i, x_i \in R^D\}$, i=1..N, the number of clusters C, fuzzy parameters T_u, T_v, T_{w}, ε_1, ε_2, the maximum number of interation τ_{max}.
1. τ=1.
2. Initialize u_{ci} according to constraint (9).
3. DO
4. Calculate δ_j using (21)

5. Update p_{cj} using (18).
6. Update v_{cj} using (16).
7. Update u_{ci} using (12).
8. Update w_j using (20).
9. Remove feature j from the data corresponding to $w_j < \varepsilon_1$.
10. $\tau = \tau + 1$.
11. WHILE $(\max(|w_j[\tau] - w_j[\tau-1]|) \le \varepsilon_2$ or $\tau = \tau_{max})$.
Output: Report clustering result.

We next analyze the computational complexities for the FRFCoC. This algorithm includes τ_{max} iterators to update components of the objective function (8): 1) Calculate membership functions, u_{ci}, which needs $O(NC^2D)$; 2) Calculate membership functions, v_{cj}, which needs $O(NCD^2)$; 3) Update cluster centroid, p_{cj}, which needs $O(NC)$; 4) Update the weight w_j, which needs $O(NCD^2)$. The total computational complexity for the FRFCoC algorithm is $O(NC^2D + NCD^2)$. Note that, D is larger than C, therefore, taking into account the number of iterations, the completely time complexity is $O(CD^2N\tau)$, which is the same as fuzzy co-clustering algorithms. However, in algorithm FRFCoC, D will be reduced after each iteration. Therefore, in fact, the complexity of algorithm FRFCoC is simpler than algorithm FCoC.

IV. EXPERIMENT RESULTS

In this section, we present the result of experiments on synthetic data sets and hyperspectral images using clustering algorithms FCM, FRFCM, FCoC and FRFCoC. To evaluate the quality of these clustering algorithms, we use the validity indices as Partition Coefficient index (PC) [26], Mean Squared Error index (MSE) [27], Image Quality Index (IQI) [28], Recall and Precision [12]. Note that, the larger PC, IQI, Recall and Precision and the smaller MSE index, the better the clustering quality.

TABLE 1. THE BRIEF INFORMATION OF HIGH-DIMENSIONAL DATASETS

Datasets	**No. of clusters**	**No. of features**	**No. of objects**	**No. of objects in one cluster**
Dim032	16	32	1024	64
Dim064	16	64	1024	64
Dim128	16	128	1024	64
Dim256	16	256	1024	64
Dim512	16	512	1024	64
Dim1024	16	1024	1024	64

The first experiment, we have conducted clustering on six high-dimensional synthetic data sets which is downloaded from the clustering datasets of Speech and Image Processing Unit[1]. Table 1 shows the information of these data sets.

[1] Speech and Image Processing Unit, School of Computing University of Eastern Finland, Clustering datasets [Online], <http://cs.joensuu.fi/sipu/datasets/>

TABLE 2. CLUSTERING RESULTS ON DIM032-DIM1024 USING FCM, FRFCM, FCoC AND FRFCoC.

Algorithm		**FCM**	**FRFCM**	**FCoC**	**FRFCoC**
Dim032	PC	0.85	0.90	0.90	**0.98**
	MSE	12.34	9.23	8.07	**7.63**
	IQI	0.92	0.95	0.98	**0.99**
	Prec.	0.83	0.92	0.96	**0.98**
	Rec.	0.86	0.91	**0.98**	**0.98**
	Time (s)	4.24	2.56	2.21	**2.19**
Dim064	PC	0.93	0.93	0.96	**0.98**
	MSE	13.21	7.51	5.11	**2.19**
	IQI	0.89	0.96	**0.99**	**0.99**
	Prec.	0.81	0.95	**0.97**	0.96
	Rec.	0.89	0.95	0.95	**0.97**
	Time(s)	5.67	3.38	4.04	**3.15**
Dim128	PC	0.81	0.92	0.88	**0.94**
	MSE	15.20	12.18	24.38	**3.23**
	IQI	0.83	0.96	0.96	**0.98**
	Prec.	0.91	0.96	0.97	**0.99**
	Rec.	0.89	0.95	0.97	**0.98**
	Time(s)	11.39	7.10	8.13	6.34
Dim256	PC	0.80	0.91	0.90	**0.98**
	MSE	19.10	11.42	8.22	**1.56**
	IQI	0.75	0.96	0.97	**0.98**
	Prec.	0.79	0.92	0.97	**0.98**
	Rec.	0.82	0.95	0.96	**0.99**
	Time(s)	13.53	12.65	11.71	**9.13**
Dim512	PC	0.78	0.91	0.93	**0.96**
	MSE	11.23	8.42	6.42	**2.24**
	IQI	0.86	0.93	0.98	**0.99**
	Prec.	0.91	0.96	0.93	**0.98**
	Rec.	0.93	0.92	0.95	**0.99**
	Time(s)	18.52	14.59	13.04	**12.50**
Dim1024	PC	0.75	0.93	0.95	**0.98**
	MSE	21.12	14.92	7.02	**4.28**
	IQI	0.72	0.92	0.98	**0.99**
	Prec.	0.85	0.97	0.96	**0.99**
	Rec.	0.82	0.90	0.93	**0.99**
	Time(s)	24.89	19.60	21.56	**16.8**

In this experiment, we have used set of fuzzy parameters with T_u=5, T_v=10^6, T_w=0.1 and C=16. The clustering results are shown in Table 2, the result of dimensionality reduction in Table 3. From Table 2 and Table 3, some advantages of FRFCoC can be summarized. All five indices of PC, MSE, Q, Precision and Recall producing by FRFCoC are better than the ones coming from other algorithms. Also, the consumption time of FRFCoC algorithm is lowest in comparing with FCM, FRFCM and FCoC. The dimensionality reduction results of these experiments are about 10% to 20%.

The next experiment, we used a sample hyperspectral image size 320x660x360, i.e. the hyperspectral images consist of 360 image bands corresponding to 60 dimensions, the number of pixels (data objects) is 211200. The images were downloaded from SpecTIR's hyperspectral images gallery[2]. In this experiment, we have used set T_u=5, T_v=10^6, T_w=0.5.

[2] SpecTIR's Advanced Hyperspectral & Geospatial Solutions <http://www.spectir.com/>

Hyperspectral image is classified into 9 classes which are filled by 9 different colors. The clustering results are shown in Table 4 and Fig. 1, Fig. 2.

TABLE 3. THE DIMENSIONALITY REDUCTION RESULTS USING FRFCoC ALGORITHM

Datasets	Number of features before and after clustering		Dimensionality reduction results
Dim032	32	28	4 (12.5%)
Dim064	64	52	12 (18,7%)
Dim128	128	102	28(20.3%)
Dim256	256	203	53(20.7%)
Dim512	512	420	92(17.9%)
Dim1024	1024	912	112(10.9%)

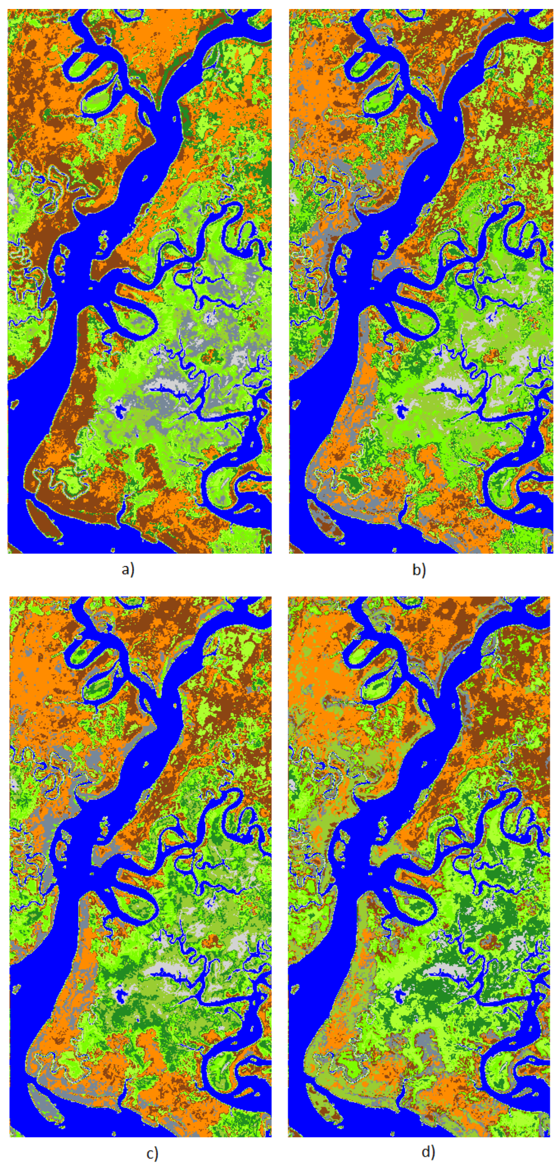

Fig. 1. Clustering results by a) FCM; b) FRFCM; c) FCoC; d) FRFCoC.

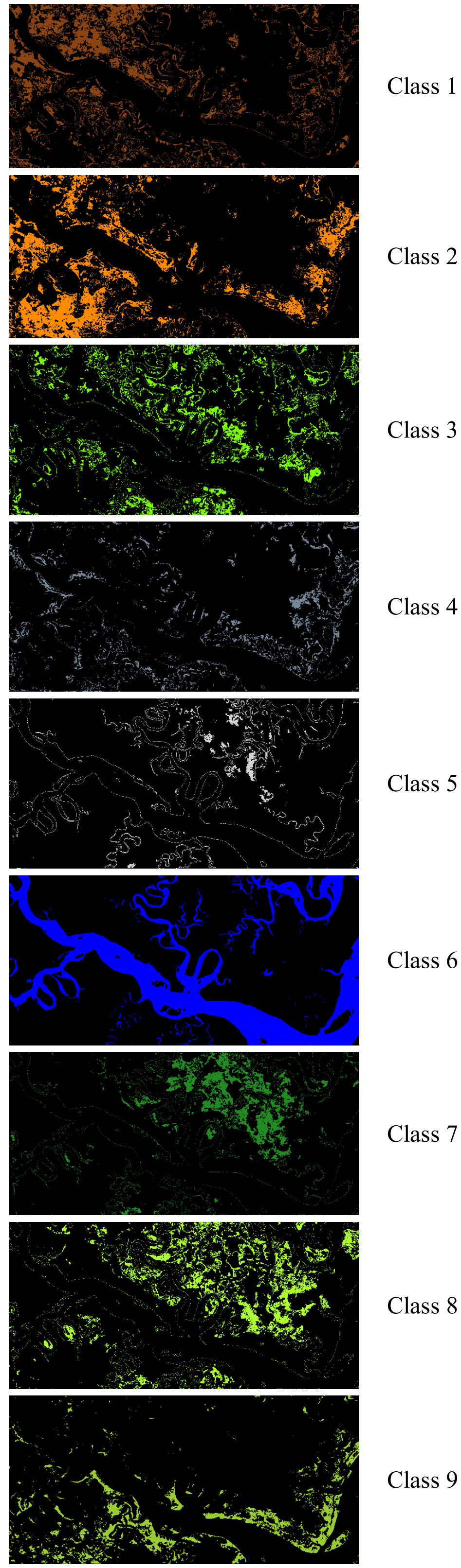

Fig. 2. Clustering result images: 9 classes producing by FRFCoC.

TABLE 4. THE CLUSTERING RESULTS ON THE HYPERSPECTRAL IMAGE DATASETS

Algorithm	**FCM**	**FRFCM**	**FCoC**	**FRFCoC**
PC	0.84	0.95	0.93	**0.98**
MSE	90.3	79. 7	80.5	**75.3**
IQI	0.89	0.96	0.95	**0.98**
Time (second)	5,239	4,126	4,537	**3,536**

In algorithm FRFCoC, validity indices PC and IQI was the largest values and index MSE was the smallest value in comparison with FCM, FRFCM and FCoC, i.e. the clustering quality coming from the proposed algorithm is better than FCM, FRFCM and FCoC through these validity indices. Also, the consumption time of FRFCoC is lowest. The dimensionality reduction result of this experiment is 11.4%.

Through these experiments, we can see algorithm FRFCoC including a dimensionality reduction technique really effective in terms of quality and speed compared to the algorithm FCM, FRFCM and FCoC. Not only that, this dimensional reduction technique is really simple and easy to implement.

V. CONCLUSION

This paper, we present a novel improved fuzzy co-clustering algorithm FRFCoC which consists of a feature-reduction schema to be able to automatically calculate the weight of each feature and then remove these features out of the data processing. The proposed algorithm is considered suitable for clustering of multivariate data types in general, hyperspectral image data in particular. Experiments were conducted on synthetic data sets and hyperspectral image using the robust validity indices. Experimental results demonstrated the proposed algorithm outperformed the algorithms FCM, FRFCM and FCoC in the quality and speed of clustering.

However, FRFCoC are sensitive to initialize and the number of clusters must be predefined. In the future, we will combine algorithm FRFCoC with other techniques to support for finding the optimal number of clusters and centroids to solve object recognition problems in the hyperspectral image.

Acknowledgement.

The result of this paper was supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED), grant number 102.05-2016.09.

REFERENCES

[1] Y. Song, S. Pan, S. Liu, F. Wei, M.X. Zhou, W Qian, “Constrained co-clustering for textual documents,” AAAI Conference on Artificial intelligence, pp. 581-586, 2010.

[2] C.-H. Oh, K. Honda, H. Ichihashi, Fuzzy clustering for categorical multivariate data, in: Joint Ninth IFSA World Congress and Twentieth NAFIPS Internat. Conf. Vol. 4, pp. 2154–2159, 2001.

[3] W. C. Tjhi, L. Chen, Possibilistic fuzzy co-clustering of large document collections, Pattern Recognition, vol. 40 (12), 3452-3466, 2007.

[4] Y. Yan, L. Chen, W. C. Tjhi, Fuzzy semi-supervised co-clustering for text documents, Fuzzy Sets and Systems 215, 74-89, 2013.

[5] M. Hanmandlua, O. P. Verma, S. Susan, V. Madasu, Color segmentation by fuzzy co-clustering of chrominance color features, Neurocomputing 120, 235-249, 2013.

[6] K. Kummamuru, A. Dhawale, R. Krishnapuram, Fuzzy Co-clustering of Documents and Keywords, IEEE International Conf. on Fuzzy Systems, Vol. 2, pp. 772-777, 2003.

[7] W. C. Tjhi, L. Chen, A heuristic-based fuzzy co-clustering algorithm for categorization of high-dimensional data, Fuzzy Sets and Systems 159, 371-389, 2008.

[8] V. N. Pham , L. T. Ngo, W. Pedrycz, Interval-valued fuzzy set approach to fuzzy co-clustering for data classification, Knowledge-Based Systems, Vol. 107, pp. 1–13, 2016.

[9] M. Yang, Y. Nataliani, A Feature-Reduction Fuzzy Clustering Algorithm Based on Feature-Weighted Entropy, IEEE Transactions on Fuzzy Systems, 2017, In Press.

[10] J. Z. Huang, M. K. Ng, H. Rong and Z. Li, "Automated Variable Weighting in k-Means Type Clustering," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 27(5), pp. 657-668, 2005.

[11] L. Jing, M. K. Ng and J. Z. Huang, "An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data," IEEE Transactions on Knowledge and Data Engineering, Vol. 19(8), pp. 1026-1041, 2007.

[12] I.S. Dhillon, S. Mallela, D.S. Modha, “Information-theoretic co-clustering,” ACM Int’ Conf’ on Knowledge Discovery and Data Mining, pp. 89–98, 2003.

[13] HJ. Xing, MH. Ha, "Further improvements in Feature-Weighted Fuzzy C-Means," Information Sciences, Vol. 267, pp. 1-15, 2014.

[14] Y. Unal, K. Polat, H.E. Kocer, "Pairwise FCM based feature weighting for improved classification of vertebral column disorders," Computers in Biology and Medicine, Vol. 46, pp. 61-70, 2014.

[15] K. Polat, "Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering," International Journal of Systems Science, Vol. 43(4), pp. 597-609, 2012.

[16] X.Z. Wang, Y.D. Wang, L.J. Wang, "Improving fuzzy c-means clustering based on feature-weight learning," Pattern Recognition Letters, Vol. 25(10), pp. 1123-1132, 2004.

[17] D. R. Cox, “The Statistical Analysis of Series of Events,” Springer Netherlands, 1966.

[18] Z. Bai, K. Wang, W.K. Wong, "Mean-Variance Ratio Test, A Complement to 52 Coefficient of Variation Test and Sharpe Ratio Test," Statistics & Probability Letters, Vol. 81(8), pp. 1078-1085, 2011.

[19] Y. Yuan, X. Zheng, X. Lu, "Discovering Diverse Subset for Unsupervised Hyperspectral Band Selection," IEEE Transactions on Image Processing, Vol. 26(1), pp. 51 - 64, 2017.

[20] H. Su, B. Yong, Q. Du, "Hyperspectral Band Selection Using Improved Firefly Algorithm," IEEE Geoscience and Remote Sensing Letters, Vol. 13(1), pp. 68 - 72, 2016.

[21] P. Chen, L. Jiao, F. Liu, J. Zhao, Z. Zhao, S. Liu, "Semi-supervised double sparse graphs based discriminant analysis for dimensionality reduction," Pattern Recognition, Vol. 61, pp. 361–378, 2017.

[22] W. Li, J. Liu, Q. Du, "Sparse and Low-Rank Graph for Discriminant Analysis of Hyperspectral Imagery," IEEE Transactions on Geoscience and Remote Sensing, Vol. 54(7), pp. 4094 - 4105, 2016.

[23] S. Feng, Q. Chen, C. Zuo, J. Sun, T. Tao, Y. Hu, "A carrier removal technique for Fourier transform profilometry based on principal component analysis," Optics and Lasers in Engineering 74, 80-86, 2015.

[24] Y. Qu, Z. Liu, "Dimensionality reduction and derivative spectral feature optimization for hyperspectral target recognition," Optik, Vol. 130, pp. 1349–1357, 2017.

[25] H. Binola, S. Ochilovb, M.S. Alamb, A. Bala, "Target oriented dimensionality reduction of hyperspectral data by Kernel Fukunaga–Koontz Transform," Optics and Lasers in Engineering, Vol. 89, pp. 123–130, 2017.

[26] J.C. Bezdek, “Cluster validity with fuzzy sets,” J. Cybernet., Vol. 3, pp. 58-73, 1974.

[27] Z. Wang, A.C. Bovik, “Mean squared error: love it or leave it? A new look at signal fidelity measures,” IEEE Signal Processing Magazine, 26(1), 98-117, 2009.

[28] Z. Wang, A.C. Bovik, “A universal image quality index,” IEEE Signal Processing Letters, Vol. 9(3), pp. 81-84, 2002.