Journal of Basic and Applied Engineering Research
Print ISSN: 2350-0077; Online ISSN: 2350-0255; Volume 1, Number 7; October, 2014 pp. 38-42
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/jbaer.html

Soft Subspace Fuzzy C-Means with Spatial Information for Clustering of Hyperspectral Images

Prem Shankar Singh Aydav[1], Sonajharia Minz[2]

[1,2]*School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi-110067, India*

Abstract: **Fuzzy c-means clustering technique is very popular for remote sensing data classification. However traditional fuzzy c-means clustering algorithm does not produce good results due to multi dimensionality and spatial relationship existence in hyperspectral remotely sensed images. Due to existence of several spectral bands, it falls in category of high dimensional data which affects the result of traditional fuzzy c-means clustering algorithm. We have used a soft subspace fuzzy c-means approach for remotely sensed data which take care of multi-dimensionality. The neighboring information is used to exploit spatial relationship existence which improves clustering results. The experimental results show that the soft subspace fuzzy c-means technique with spatial information gives better performance than traditional fuzzy c-means technique.**

Keywords: **C-means, Fuzzy c-means, Hyperspectral images, Remote Sensing, Soft subspace clustering**

1. INTRODUCTION

"Remote sensing is the science and art of obtaining information about an object or area through a device that is not in contact with object or area under investigation" [1]. Remote sensing data have wide applications in various fields like agriculture, glacier monitoring, land cover classification, land cover change detection, urban growth monitoring, identification of burned area, assessment of vegetation changes and environmental study. Due to advancement of technology, remote sensors are providing very huge amount of data with very high spectral, spatial, temporal and radiometric resolution. Remotely sensed data has many challenges like large data size, high dimensionality, spatial relationship, overlapping nature of classes, spatio-temporal relationship, noisy data etc. These challenges add multiple complications to the analysis of remote sensing images through manual analysis. Clustering techniques may apply to extract useful information from these images. Hyperspectral images contain more than hundreds bands, that's why clustering of hyper spectral images are very challenging task. The pixels in the image are very highly correlated. It means that probability of belonging into same class is very high in the neighboring pixels. Therefore, the integration of spatial information with spectral information could improve the clustering results. Traditional clustering algorithms do not take considerations of high dimensions and spatial information during clustering steps.

In literature many technique have been proposed for image data clustering. Keh-Shih Chung, Hong -Long Tzeng and Shareon Chan and Jay Wu in have described a fuzzy c means algorithm that incorporates spatial information into membership function for clustering [2]. The spatial function is summation of the membership function in the neighborhood of each pixel consideration. Yang HongLei, Peng JunHuan Xia BaiRu2 and Zhang DingXuan1 have proposed a technique for remotely sensed images using fuzzy c-means clustering with spatial constraints based on markov random field [3]. Luis O. Jiménez, Jorge L. Rivera-Medina, Eladio Rodríguez-Díaz, Emmanuel Arzuaga-Cruz, and Mabel Ramírez-Vélez have described an unsupervised method to integrate spatial information with spectral information of pixels in a local window [4]. The main characteristic of this method is that it makes simple to discover spatial structure of objects. Gökhan Bilgin, Sarp Ertürk and Tülay Yıldırım have proposed a technique for hyper spectral image clustering based which utilize spatial information not only for fuzz c-menas clustering but also in Gustafson–Kessel clustering [5].

In recent years in literature many soft subspace clustering algorithm have been proposed for clustering of high dimensional data [6, 7, 8]. Soft sub space clustering algorithms assign a weight value to each dimension for a cluster during clustering steps. We have integrated spatial information in soft subspace clustering algorithm for the classification of hyper spectral images. The proposed framework for remotely sensed data aim to

1. use fuzzy c-means to manage fuzziness in remote sensing data
2. use soft subspace clustering techniques for high dimensionality of remote sensing data
3. use spatial features to improve clustering results

In section 2 fuzzy c-means and soft subspace clustering technique have been described. Section 3 descries spatial information integration. In section 4 experimental results have been shown and discussed. Section 4 presents the conclusion of this study.

2. CLUSTERING

The process of grouping a set of objects into classes of similar objects is called clustering"[9]. A cluster is a collection of data objects that are similar to one another and are dissimilar to the objects in other cluster. Image data clustering is the process of assigning a label to each location in image on the basis of features set similarity.

2.1 C-Means

C-means is partition based clustering algorithm. The C-means algorithm obtain the partition of the data set into c cluster so that the resulting intra cluster similarity is high but the inter cluster similarity is low. The c-means algorithm allocates each object to one of the c clusters to minimize sum of intra-cluster distances:

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

$$J(V) = \sum_{i=1}^{c} \sum_{k \epsilon A_i} \|X_k - V_i\|^2 \quad (1)$$

$$V_i = \frac{\sum_{k=1}^{N_i} x_k}{N_i}, X_k \varepsilon A_i \quad (2)$$

Where V_i is the ith cluster center, N_i is the number objects in ith cluster, A_i is the set of objects in the ith cluster and V_i is the center of the ithcluster.

2.2 Fuzzy C-Means

Fuzzy c-means attempts to find fuzzy partitioning of a given data set by minimizing the objective function [10]. The minimizing of objective function denotes a nonlinear optimization problem.

$$J(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m \|X_k - V_i\|^2 \quad (3)$$

s.t. $0 \le \mu_{ij} \le 1, \sum_{j=1}^{c} \mu_{ij} = 1$

The update equation is given by:

$$V_i = \frac{\sum_{j=1}^{j=n} (\mu_{ij})^m x_j}{\sum_{j=1}^{j=n} (\mu_{ij})^m} \quad (4) \qquad \mu_{ij} = \frac{1}{\sum_{k=1}^{k=c} \left(\frac{d_{ji}}{d_{jk}}\right)^{\frac{2}{m-1}}} \quad (5)$$

$$d_{jk} = \sqrt{\|X_k - V_i\|^2} \quad (6)$$

where c is the number of clusters, n the number of samples in dataset, m the fuzzifier value, U the fuzzy partition matrix and V the cluster centers matrix. Let the μ_{ik} denote the degree of membership value of the kth object in the ith cluster, V_i denotes the center of the ith cluster and d_{jk} denote the distance of the kth object from the ith cluster.

2.3 Fuzzy weighting subspace clustering

Fuzzy c-mean clustering algorithm utilize Euclidian distance measure in which all features are weighted equally but some dataset contain features in which all are not equally important.

Soft subspace clustering techniques assign a feature weight vector to each cluster. Fuzzy weighting try to find different clusters from weighting subspace. In fuzzy weighting clustering techniques a weight is assigned to each feature of clusters. The objective function is given as [11]:

$$J(W,U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^m \sum_{k=1}^{d} w_{ik}{}^{l} (x_{jk} - v_{ik})^2 \quad (7)$$

s.t. $0 \le \mu_{ij} \le 1, \sum_{j=1}^{c} \mu_{ij} = 1$
$0 \le w_{ik} \le 1, \sum_{k=1}^{d} w_{ik} = 1$

By using langrage multipliers the update equation is given by:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{k=c} \left(\frac{d_{ji}}{d_{jk}}\right)^{\frac{2}{m-1}}} \quad (8)$$

$$V_i = \frac{\sum_{j=1}^{j=n} (\mu_{ij})^m x_j}{\sum_{j=1}^{j=n} (\mu_{ij})^m} \quad (9)$$

$$d_{ij} = \sum_{k=1}^{d} w_{ik}{}^{l} (x_{jk} - v_{ik})^2 \quad (10)$$

$$w_{ik} = \frac{(q_{ik})^{-1/l-1}}{\sum_{p=1}^{d} (q_{ip})^{-1/l-1}} \quad (11)$$

$$q_{ik} = \sum_{j=1}^{n} \mu_{ij} (x_{jk} - v_{ik})^2 \quad (12)$$

Where μ_{ik} denotes the membership value of kth object to ith cluster, V_i the center of ith cluster, d_{jk} distance between kth object to ith cluster, w_{ik} the weight of ith feature with kth cluster, d_{ij} the distance between kth object to ith cluster. The value of m and l denotes the fuzzifier value.

2.4 Fuzzy weighting subspace clustering with spatial information

Traditional clustering algorithms do not use spatial information during clustering process. To include spatial information in spatial data set Cai et al. [12] have proposed a robust technique which gives better performance in comparison of other technique. We have used their technique in our approach. By using equation image is transformed as:

$$\epsilon_i = \frac{\sum_{j \in N_i} s_{ij} x_j}{\sum_{j \in N_i} s_{ij}} \quad (13)$$

$$s_{ij} = \exp\left(\frac{-(\mathrm{x_i}-\mathrm{x_j})^2}{r\times\sigma_{g\_i}{}^2}\right) \quad (14)$$

$$\sigma_{g\_i} = \sqrt{\frac{\sum_{j\in Ni}(\mathrm{x_j}-\mathrm{x_i})^2}{N_R}} \quad (15)$$

Where $\in_i$ is new value which is calculated by using neighboring pixels, $\mathrm{x_i}$ is the center of local window, $\mathrm{x_j}$ is the neighboring pixels of $\mathrm{x_i}$, r is the global spread factor of the similarity, N_i is the set of all neighboring pixels, N_R is the number of pixels in the window, σ_{g_i} is the mean distance between the center and neighboring pixels. σ_{g_i} is the measure of homogeneity of local spatial window. The smaller value denotes more homogenous of neighboring pixels in a local window. It determines the local spatial relationship among pixels automatically. The main advantages of this technique that it assigns weight value to each pixel according to homogeneity in local spatial window. On transformed dataset the fuzzy weighting subspace clustering can be applied as:

$$J(W,U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{ij})^m \sum_{k=1}^{d} w_{ik}{}^l\,(\in_{jk} - v_{ik})^2 \quad (16)$$

s.t. $0 \le \mu_{ij} \le 1, \sum_{j=1}^{c}\mu_{ij} = 1$

$0 \le w_{ik} \le 1, \sum_{k=1}^{d} w_{ik} = 1$

By using langrage multipliers the update equation is given by:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{k=c}\left(\frac{d_{ji}}{d_{jk}}\right)^{\frac{2}{m-1}}} \quad (17)$$

$$V_i = \frac{\sum_{j=1}^{j=n}(\mu_{ij})^m \in_j}{\sum_{j=1}^{j=n}(\mu_{ij})^m} \quad (18)$$

$$d_{ij} = \sum_{k=1}^{d} w_{ik}{}^l\,(\in_{jk} - v_{ik})^2 \quad (19)$$

$$w_{ik} = \frac{(q_{ik})^{-1/l-1}}{\sum_{p=1}^{d}(q_{ip})^{-1/l-1}} \quad (20)$$

$$q_{ik} = \sum_{j=1}^{n}\mu_{ij}\,(\in_{jk} - v_{ik})^2 \quad (21)$$

Where μ_{ik} denotes the membership value of k^{th} object to i^{th} cluster, V_i the center of i^{th} cluster and d_{jk} the distance between k^{th} object to the i^{th} cluster, w_{ik} denotes the weight of i^{th} feature with kth cluster, d_{ij} distance between k^{th} object to i^{th} cluster and m, l the fuzzifier value.

The procedure for Procedure for soft subspace fuzzy c-means with spatial information is summarized in table 1.

3. EXPERIMENTAL RESULTS

The experiment were performed with four approach fuzzy c means (FCM), soft subspace fuzzy c means (SFCM), fuzzy c means with spatial information (FCM_s) and soft subspace fuzzy c means with spatial information (SFCM_s). The experiment has been performed with hyperspectral remote sensing image which is acquired by the 224 band AVIRIS sensor over Salinas Valley, California. It is characterized by high spatial resolution (3.7-meter pixels). The Salinas Valley image and its ground truth based image corresponding to the sixteen classes is presented in fig. 1a and 1b respectively. The data set contains sixteen classes and its description in terms of class name and the number of pixels has been shown in table 2.

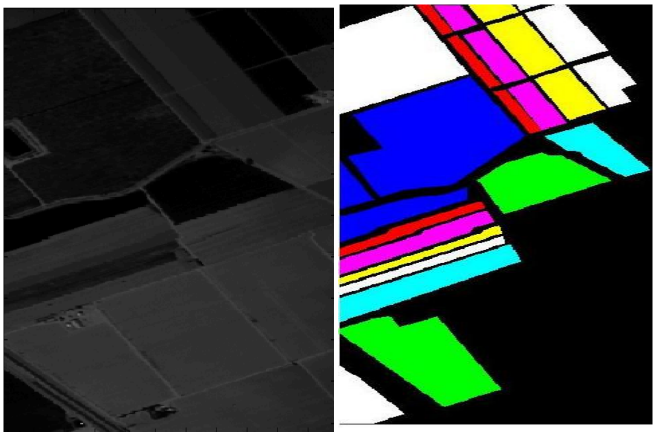
Table 1. Procedure for soft subspace fuzzy c-means with spatial information

Input: Image data set X, number of clusters c, set of parameters (l, m, r)
Initialization: Compute ∈ using equation (13), Initialize center matrix $\mathbf{V_0}$, initial weight matrix by $\mathbf{W_0}$, Initialize fuzzy partition matrix $\mathbf{U_0}$ by equation (17), initialize t=0
Repeat
Update each cluster center of unlabeled data by equation $\mathbf{V_t}$ (18)
Update weight matrix by equation $\mathbf{W_t}$ (11
Update fuzzy partition matrix by $\mathbf{U_t}$ (17)
t=t+1
Until $\vert \mathbf{U_t} - \mathbf{U_{t-1}}\vert \le error$ or t=t_{max}
Output: $\mathbf{V_t}$, $\mathbf{U_t}$ and **W**

Table 2. Data set descriptions

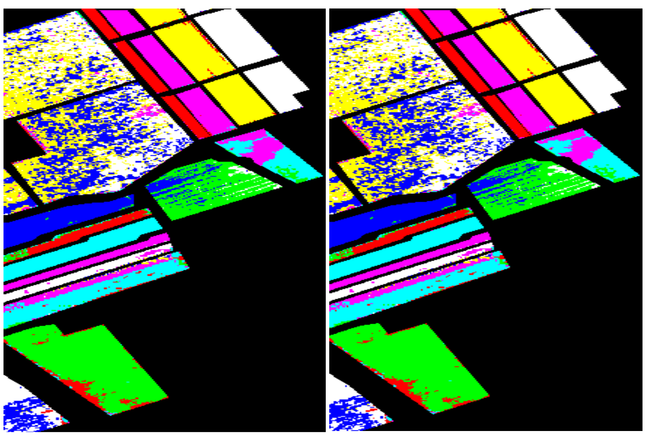
Class	Class Name	Number of samples
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807

It also includes 20 water absorption bands (108-112, 154-167, 224) which are discarded in this experiment. The experiment is performed on 16 classes of data set. The data set is publicly available for research purpose [13].



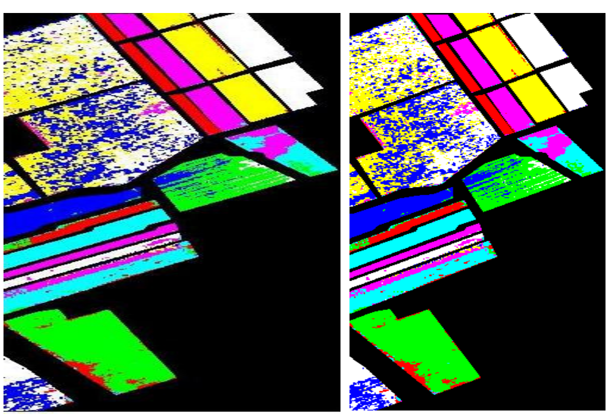
(b)

Fig. 1. (a)Salinas image (b) ground truth



(a) **(b)**

Fig. 2. **Output images (a) FCM (b) FCM_s**



(a) **(b)**

Fig. 3**. Output images (a) SFCM (b) SFCM_s**

Clustering accuracy is calculated for each technique by using following formula:

$$\text{Clustering Accuracy} = \frac{\sum_{l=1}^{c} n_l}{N}$$

where n_l denotes the number of correctly classified samples belonging to class l and N the total number of samples in dataset.

$$\text{Class Accuracy} = \frac{p_i}{N_i}$$

where p_i denotes the number of correctly classified samples and N_i the total number of samples in that class.

The accuracy of each of each class is calculated as:

First two experiments were performed with FCM and SFCM with spectral bands information. Next FCM_s and SFCM_s were performed with spatial information (equation 13) by taking 3x3 window with value of r as 6.

Table3. Accuracy (%) of classes

#	FCM	FCM_s	SFCM	SFCM_s
Class1	98.41	**99.35**	91.34	70.93
Class2	77.83	84.41	99.81	**99.25**
Class3	69.03	**71.56**	52.68	55.31
Class4	**98.78**	98.28	97.27	98.28
Class5	91.06	**94.32**	83.79	84.80
Class6	93.28	97.63	99.49	**99.50**
Class7	54.63	57.39	**99.39**	98.52
Class8	41.50	**46.55**	38.86	37.78
Class9	86.75	88.76	81.49	**91.41**
Class10	58.23	61.74	60.83	**61.81**
Class11	71.50	75.66	**76.03**	68.63
Class12	50.13	52.00	99.12	**99.64**
Class13	43.20	45.52	98.03	**98.25**
Class14	89.16	91.12	91.31	**93.74**
Class15	48.21	**51.49**	39.86	37.74
Class16	30.71	28.29	63.36	**75.82**
Over all	64.31	67.74	69.34	**70.93**

The results have been presented in table 3. The outcome of FCM, FCM_s, SFCM and SFCM_s has been presented for visual interpretation in the Figure 2a, 2b, 3a, 3b respectively.

In order to compare the performance of the four clustering techniques the highest accuracy for each class has been highlighted. The experimental results show that for most of the classes (2, 6, 9, 10, 12, 13, 14, 16) the accuracy has been increased with technique SFCM_s. Experimental results show that the overall accuracy of SFCM (69.34%) is better than FCM (64.31%). It is observed that with integration of spatial information, the performances of FCM and SFCM have been increased. The overall accuracy of SFCM_s (**70.93 %**) is better than FCM_s (69.34%). The soft subspace fuzzy c means clustering technique has been observed to outperform fuzzy c-means with spatial information.

4. CONCLUSION

A frame work for hyper spectral remotely sensed data clustering was presented in this paper. Weighted feature vector for each cluster and spatial information were used in traditional FCM. The experimental results show that the soft subspace fuzzy c means and soft subspace fuzzy c means with spatial information gives better performance than fuzzy c means techniques. Therefore, the proposed methodology can be used to classification of hyper spectral remote sensing images. For future work the soft subspace clustering algorithms with more constraints can be applied.

REFERENCES

[1] Thomas M. Lillesand, Ralph W. Kiefer, and Jonathan W.Chipman, "Remote sensing and image interpretation", John Wiley & Sons (Asia), 5th edition, 2011.

[2] Keh-Shih Chung, Hong -Long Tzeng and Shareon Chan, and Jay Wu, "Fuzzy c-means clustering with spatial information for image segmentation", Computerized medical imaging and graphics, vol. 30, pp. 9-15, 2006.

[3] Yang HongLei1, Peng JunHuan1, Xia BaiRu2 and Zhang DingXuan1, "Remote sensing classification using fuzzy c-means clustering with spatial constraints based on markov random field", European Journal of Remote Sensing, vol. 46, pp. 305-316, 2013.

[4] Luis O. Jiménez, Jorge L. Rivera-Medina, Eladio Rodríguez-Díaz, Emmanuel Arzuaga-Cruz, andMabel Ramírez-Vélez, "Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyper spectral data", IEEE Transactions on Geosciences and Remote Sensing, Vol. 43, pp.844-851, 2005.

[5] Gökhan Bilgin, Sarp Ertürk and Tülay Yıldırım, "Unsupervised classification of hyperspectral-image data using fuzzy approaches that spatially exploit membership relations" IEEE Geoscience and Remote Sensing Letters, vol. 5, pp. 673-677, 2008.

[6] Chan, Y., Ching, W., Ng, M.K. and Huang, J.Z., "An optimization algorithm for clustering using weighted dissimilarity measures" Pattern Recognition, pp.943-952, 2004.

[7] Guojun Gan, Jianhong Wu, Zijiang Yang, "A fuzzy subspace algorithm for clustering high dimensional data", Proceedings of the Second international conference on Advanced Data Mining and Applications, pp.14-16, 2006.

[8] Liping Jing, Michael K. Ng, Joshua Zhexue Huang, An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data, IEEE Transactions on Knowledge and Data Engineering, v.19 n.8, pp.1026-1041, August 2007.

[9] Jiawei Han and Micheline Kamber, "Data mining concepts and techniques", Morgan Kaufmann, 2nd edition, 2006.

[10] J.C. Bezdek, "Pattern recognition with fuzzy objective function, " Plenum Press, New York, 1981.

[11] G. Gan, J. Wu, "A convergence theorem for the fuzzy subspace clustering (FSC) algorithm", Pattern Recognition vol.41, 2008.

[12] Weiling Cai, Songcan Chen, Daoqiang Zhang, "Fast and robust c-means clustering algorithms incorporating local information for image segmentation", Pattern Recognition vol.40, pp. 825 – 838, 2007.

[13] http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.