Optimization of KFCM Clustering of Hyperspectral Data by Particle Swarm Optimization Algorithm

Saeid Niazmardi[1], Amin Alizadeh Naeini[2], Saeid Homayouni[3], Abdolreza Safari[4], Farhad Samadzadegan[5]

Abstract

Hyperspectral sensors, by accurate sampling of object reflectance into numerous narrow spectral bands, can provide valuable information to identify different land-cover classes. Nevertheless, classification of these data has some problems. In particular, one of the most well-known of them is not having adequate training data for learning of classifiers. One possible solution to this problem is the use of unsupervised classification such as Kernel based Fuzzy C-Means (KFCM). KFCM is a kernelized version of FCM algorithm, which usually, has better performance. However, in case of hyperspectral data, accuracy of the KFCM decreases because of high dimensionality of data and its kernel parameter. In this paper, the objective is to use the KFCM clustering and optimize it based on data dimensionality and kernel parameter. To optimize this algorithm with respect to the kernel parameter and data dimensionality, particle swarm optimization method (PSO) is introduced. In other words, PSO is a powerful optimization tool inspired from bird's behavior, which can find global optimum. In this study, two new methods are defined to optimize KFCM with respect to kernel parameter and data dimensionality. The results show that the proposed methods have a better performance than the KFCM.

Keywords: Kernel-based Clustering; PSO; FCM; Hyperspectral Data.

1. PhD Student of Remote Sensing, Department of Surveying Engineering, College of Engineering, University of Tehran, Iran.
2. PhD Student of Photogrammetry, Department of Surveying Engineering, College of Engineering, University of Tehran, Iran
3. Assistant Professor, Department of Geography, University of Ottawa
4. Associated Professor, Department of Surveying Engineering, College of Engineering, University of Tehran, Iran.
5. Professor, Department of Surveying Engineering, College of Engineering, University of Tehran, Iran

1. Introduction

hyperspectral images are composed of hundreds of bands with a very high spectral resolution, from the visible to the infrared region (Villa, Benediktsson et al., 2011). These sensors provide very high spectral resolution image and make it possible to discriminate among different land cover classes that are spectrally very similar (Chi and Bruzzone, 2007). Nonetheless, classification is the most necessary and primary stage to obtain mentioned information from hyperspectral data. In the literature, two main approaches to the classification problem have been proposed: 1) the supervised algorithms which need training data and 2) the unsupervised algorithms which do not need training data (Paoli, Melgani et al., 2009). With regard to application of supervised classification algorithms in hyperspectral data, one of the main critical issues is having enough training data for learning the classification algorithms (Hsu, 2004). In the other words, these data due to having numerous spectral bands need a large number of training samples that is not available in many real worlds' applications, because collection of such number of data is both hard and time-

consuming (Villa, Benediktsson et al., 2011). Furthermore, due to, existence of redundancy in the hyperspectral data because of strong correlation between adjacent spectral bands, using all spectral bands for classification will not necessarily prepare better classifier, and even it may reduce the accuracy of classification due to the worsening in the quality of the classifier-parameter estimates (Hughes, 1968; Shen, 2007). So, by tackling these problems i.e. the number of training data and high dimensionality, hyperspectral data can be used more efficiently for preparing accurate classification maps.

One possible way for avoiding the problem of training data, is the unsupervised classification algorithms. Unsupervised algorithms, known also as clustering algorithms, perform classification just by exploiting information conveyed by the data, without requiring any training sample set. Although the supervised methods offer higher classification accuracy compared to the unsupervised ones, but in some applications, which there are not a few training data, it is better to resort to clustering techniques. Nonetheless, the performance of clustering algorithms like classification is related to subset of spectral

bands (data dimensionality) applied. With regard to this problem, it can be avoided by using an optimum subset of spectral bands from available spectral bands. In the other words, an optimum subset of spectral bands, depending on application and classes considered, can be selected in clustering process as well.

Generally, data clustering divided into fuzzy clustering and hard clustering. Fuzzy clustering algorithms aim to model fuzzy unsupervised patterns efficiently, and one widely used algorithm is the fuzzy c-means (FCM) algorithm (Bezdek, 1981). The standard FCM by using Euclidian distance as the similarity measure, attempts to determine a pre-defined number of clusters, which optimize its objective function. But, FCM usually fail in the case of hyperspectral data because not only a large number of hyperspectral bands affect its performance but also, the Euclidian distance is sensitive to noise and slight variation of spectral reflectance of objects in the scene (Bidhendi, Shirazi et al., 2007). To enhance the performance of FCM algorithm for hyperspectral data, various methods have been applied. One of them, which is kernelizing similarity function of FCM, recently gained lots of attention. So that it

has created a distinctive branch of clustering algorithms named kernel base fuzzy clustering algorithm (KFCM).

This algorithm by using positive definite kernel functions transforms the data observed in input space in to a space of higher dimension, which, based on cover's theorem the relation between data points in this new space is simpler and more likely linear (Girolami, 2002; Zhang and Chen, 2002). In spite of its good performance, the KFCM accuracy is influenced by kernel function parameters. However, there is not any distinctive way for kernel tuning in the literature. In some literature, authors have used a try-and-error method to find the optimum value of kernel parameter (Graves and Pedrycz, 2010; Niazmardi, Homayouni et al., 2011), and in some other researches, the algorithm is run just by one value of parameter (Chen and Zhang, 2004; Kim, Lee et al., 2005; Filippone, Camastra et al., 2008; Yun-song and Yu-feng, 2010). On the other hand, the automated tuning of kernel parameter is a well-studied subject in the field of supervised and semi-supervised classification algorithms and due to its relation to these algorithms, various algorithms have been proposed for different classifications, like kernel minimum

distance (Zhang, Chen et al., 2006) and SVMs (Archibald and Fann 2007; Zhuo, Zheng et al., 2008; Wu, Tzeng et al., 2009). In the field of semi-supervised classification, lots of paper used an automated method for kernel parameter tuning (Yan and Domeniconi, 2006; Wang and Li 2008; Yan and Domeniconi, 2009; Baghshah and Shouraki, 2011). Unfortunately, kernel tuning cannot be addressed properly for clustering algorithms because it is not possible to define an appropriate dissimilarity measure or well-defined optimization problem in the absence of supervisory information (Baghshah and Shouraki, 2011).

On the one hand lack of distinctive way for selecting of kernel function and on the other hand sensitivity of KFCM to used bands, may affect good performance of KFCM. In order to improve the performance of KFCM clustering, the kernel parameter and data dimensionality have to be optimized. To solve the above-mentioned problems, in this paper PSO is introduced to KFCM to optimize it. PSO is a population based stochastic optimization technique inspired by the social behavior of bird flock (Kennedy and Eberhart, 1995). In new (proposed) method, optimized

dimensionality (an optimum subset of bands) and kernel parameter can be simultaneously achieved through clustering process.

This paper is organized as follows: Section 2 gives a short review of the KFCM, PSO and two optimized methods; in Section 3, the dataset and the performance of two proposed methods are discussed. The conclusions from these experiments are finally given in Section 4 and at last the concepts of kernel function are presented in appendix.

2. Material and Method

2.1. Kernel Based Fuzzy C-Means (KFCM)

Kernel methods transform the patterns from the input space to a new space (called Kernel space), in a way that, in this new space patterns will became more linearly separable. Kernel functions are functions that are able to calculate the dot product between the patterns in kernel space by using their values in input space (Graves and Pedrycz, 2010).

Clustering based on kernel function is firstly proposed by Girolami (2002). After that Chen and Zhang (2002) proposed a modification of FCM algorithm using kernel function and called it KFCM. In the

following years, this algorithm became more popular and used for clustering of medical imagery (Wu, Xie et al., 2003; Zhang and Chen, 2004; Zhang, Ruan et al., 2011) , the standard data mining datasets (Wu, Xie et al., 2003; Zhang and Chen, 2003) and remote sensing images (Kamel, Campilho et al., 2005; Yun-song and Yu-feng 2010; Niazmardi, Homayouni et al., 2011). Similar to FCM algorithm, KFCM clusters the data by optimizing the following objective function.

$$(1)$$

$$J(X,U,C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} \| \varphi(x_j) - \varphi(c_i) \|^2$$

$$S.T \begin{cases} \sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1,\ldots,n\} \\ \sum_{j=1}^{n} u_{ij} > 0, \forall i \in \{1,\ldots,c\} \end{cases}$$

It can be proved that, the above equation can be rewritten as follows:

$$(2)$$

$$J(X,U,C) =$$
$$\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} (K(x_j,x_j) + K(c_i,c_i) - 2K(x_j,c_i))$$

Gaussian kernel

$(k(x,x') = \exp(- \| x - x' \|^2 / \sigma^2), \sigma \in R^+)$

is used almost exclusively as kernel function in KFCM objective function, because, it makes the deviation of objective function

simpler. In addition, Chen and Zhang proved that based on Huber's robust statistics, it can give more stable results (Chen and Zhang, 2004; Zhang and Chen, 2004). By using Gaussian kernel, the KFCM's objective function can be formulated as follows:

$$(3)$$

$$J(X,U,C) = 2\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} (1 - K(x_j,c_i))$$

The priority of using this algorithm is the facts that the cluster centers reside in input space, but using kernel function, the distance is calculated in feature space (Graves and Pedrycz, 2010). By an alternative optimization theme, in each iteration, the cluster centers and fuzzy partition matrix can be calculated using the following equations:

$$u_{ij} = \frac{(1 - K(x_j,c_i))^{\frac{-1}{(m-1)}}}{\sum_{l=1}^{c} (1 - K(x_j,c_l))^{\frac{-1}{(m-1)}}} \qquad (4)$$

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} K(x_j,c_i) x_j}{\sum_{j=1}^{n} u_{ij}^{m} K(x_j,c_i)} \qquad (5)$$

From an operational point of view, KFCM is similar to FCM algorithms. In the

first step, some data points are randomly selected as cluster centers and then in an iterative way, fuzzy partition matrix and cluster centers are calculated using Equations 4 and 5 respectively. The algorithm runs for a predefined number of iterations or until some termination criteria are met.

KFCM algorithm has some advantage in comparison to FCM; firstly, it is more robust to noise and less sensitive to shape of clusters (Chen and Zhang, 2004; Wang, Jin et al., 2005). Secondly, it has the same computational complexity as FCM with a more rapid convergence (Wu, Xie et al., 2003; Graves and Pedrycz, 2010). However, this algorithm can only assure finding local minimums due to using alternative optimization theme for optimizing the objective function (Yang, Sun et al., 2009).

2.2. Particle Swarm Optimization (PSO)

PSO is a population -based stochastic optimization technique, inspired from the social behavior of bird flock (and fish school, etc.), and was developed by Kennedy and Eberhart (1995). As a relatively new evolutionary paradigm, it has grown in the past decade and many studies related to PSO have been published. In PSO, each particle is an individual, and the swarm is composed of particles. The problem solution space is formulated as a search space. Each position in the search space is a correlated solution of the problem. Particles cooperate to find the best position (best solution) in the search of space (solution space). Each particle moves according to its velocity, which is computed as:

$$v_{id}(t+1) = wv_{id}(t) + c_1 r_1 \left(p_{id}(t) - x_{id}(t) \right) \\ + c_2 r_2 \left(p_{gd}(t) - x_{id}(t) \right) \tag{6}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{7}$$

In (6) and (7), x_{id} is the position of particle i at time t, v_{id} is the velocity of particle i at time t. p_{id} is the best position found by particle i itself so far, p_{gd} is the best position found by the whole swarm so far. ω is an inertia weight scaling the previous time step velocity, c_1 and c_2 are two acceleration coefficients, that scale the influence of the best personal position of the particle (p_{id}) and the best global position (p_{gd}), r_1 and r_2 are random variables between 0 and 1 (Yang, Zhang et al., 2009).

2.3.   Proposed Methods

As mentioned earlier, the performance of KFCM algorithm is completely affected by the Kernel parameter. As such, its different values result in different performance of KFCM. So, PSO is introduced to KFCM to optimize it with respect to kernel parameter. Furthermore, since accuracy of the KFCM algorithm is influenced by used dimension too, in another method, besides kernel parameter, the KFCM is optimized with respect to the dimensionality by PSO.

In an optimization problem that is formulated within a PSO framework, the solution space is explored by means of a swarm of particles whose positions point to candidate solutions.   The first task to perform consists of defining the ingredients of the PSO algorithm, namely, the particle position p and the fitness function $f(p)$.

Since our goal in this study, is to explore role of kernel parameter and used dimension in KFCM performance, cluster centers has been considered fixed. For this purpose, the K-means algorithm has been run for 50 iterations and its output has been considered as cluster centers of KFCM algorithm.

2.3.1. KFCM Optimization Based on Kernel Parameter (PSKFCMK)

As mentioned before, the kernel function calculates the dot product of mapped point in kernel space using their values in input space. In order to map the pattern correctly into the kernel space, the kernel function should be tuned. The kernel tuning can be addressed in supervised and semi-supervised classification algorithms due to presence of training data (Zhang, Chen et al., 2006; Tso and Mather, 2009; Yan and Domeniconi, 2009). However, the kernel tuning still is an open problem in unsupervised algorithms like KFCM (Baghshah and Shouraki, 2011). In this paper, it has been tried to tune kernel parameter of KFCM by means of PSO algorithm. In other words, the proposed method optimizes the KFCM accuracy with respect to the kernel parameter.

In this method, since our clustering problem consists of finding the best value of the kernel parameter, the position of each particle will simply be a cell that encodes the kernel parameter. In addition, the adaptation is carried out by optimizing the clustering accuracy, which will be

mentioned in the next section.

2.3.2. KFCM Optimization Based on both Kernel Parameter and Data Dimensionality (PSKFCMB)

Numerous bands of hyperspectral data not only influence the performance of the KFCM algorithm but also, increase its computational cost. So, besides selecting the optimum kernel parameter, an optimum subset of bands should be selected for KFCM. In this section, a new method is introduced to optimize KFCM with respect to both kernel parameter and used dimensionality.

For this purpose, it is necessary to have a search algorithm and a fitness function. The search algorithm generates possible solutions of different subsets of bands and then, the effectiveness of each solution is obtained by the fitness function (Serpico, Moser et al., 2007).

To find the optimum solution (i.e. the smallest number of bands with the highest information content for clustering purpose and the best kernel parameter for the selected subset of bands) again PSO is used. According to our objective in here, the

particle structure is composed of two different parts. The first part is the Gaussian kernel parameter, which is a real positive number and the second part is the subset of bands. Fig.1 shows the structure of each particle

Parameter	Dimensionality			
σ	*1*	*·*	*1*	*…*	*1*

1	D
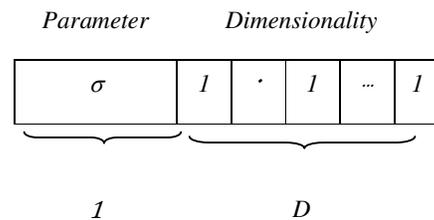
Fig1. The Structure of Particles in PS

In case of having *d* spectral bands, the particle size will be *d+1,* where the first part is continuous and the second part is a binary string in which *i*th bit is equal to one in case of selection of that band and is zero otherwise. The second important part of the PSO algorithm is its fitness function. The fitness function has to be considered such to reflect better score for particles having the fewer bands and higher clustering accuracy and vice versa. Thus, the fitness function has two different parts, the first part, is to assess of clustering and the second part is to evaluate the size of selected subset of bands. For the first part each validity index and for the second part, the number of

selected subset of bands can be used. Thus, the fitness function having to be minimized is as follows:

$$f = a(1 - I) + (1 - a)(nf / F) \quad (8)$$

Where I is clustering accuracy, nf is the number of selected bands, F is the number of all available bands and α is a constant to establish a balance between used dimensionality and clustering accuracy. Since accuracy has more importance , α is set to more than 0.5 (Zhuo, Zheng et al., 2008).

3. Results and Discussion

3.1.   Dataset

To evaluate the performance of two proposed methods on KFCM, these were run with two hyperspectral datasets. To reduce the effects of spectral bands with higher values on those having lower values,

the data are normalized. In this way, noise and outlier effects are reduced too. In the following, the characteristics of used datasets are presented.

i)   **Indiana Pine Dataset**: this data are taken over northwest Indiana's Indian Pine test site in June 1992 (1992), by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS). The image data consist of 145×145 pixels with 220 bands. Twenty water absorption's bands and fifteen noisy bands were removed, resulting in 185 spectral bands (Mojaradi, Emami et al., 2008). The original data set contains 16 land-covers classes but in this study, five classes are selected due to their suitable spatial distribution (Jia and Richards, 2007). The ground truth and a color composite image are shown in Fig. 2, and selected classes are listed in Table 1

Table 1 Class Names and Number of Pixels in Each Class

| Class		No. of Samples in Reference Data
No	Name	
1	Corn no till	1434
2	Hay Windrowed	489
3	Grass-tree	747
4	Wood	1294
5	Soybean no till	968

109

| A. Ground Truth Image | B. Color Composite of Dataset |

Fig 2 Indiana Dataset

ii) **Okavango Delta, Botswana:** This dataset is acquired by Hyperion sensor in May 2001 at 30m pixel resolution over a 7.7km strip in 242 spectral bands. These data contain 1476 ×256 pixels of study area located in the Okavango Delta, Botswana. Uncalibrated and noisy bands that cover water absorption features were removed, resulting in 145 spectral bands. The Botswana dataset consists of observations from 14 identified classes representing the land cover types in seasonal swamps, occasional swamps, and drier woodlands located in the distal portion of the delta. These classes are chosen to reflect the impact of flooding on vegetation in the study area. The class names and corresponding numbers of ground truth observations used in the experiments are listed in Table 2 (Neuenschwander, Crawford et al., 2005).

Table 2 Class Names and Number of Pixels in Each Class

Class		No. of Samples in Reference Data	Class		No. of Samples in Reference Data
No	Name		No	Name	
1	Water	270	8	Island interior	203
2	Hippo grass	101	9	Acacia woodlands	314
3	Floodplain grasses 1	251	10	Acacia shrublands	248
4	Floodplain grasses 2	215	11	Acacia grasslands	305
5	Reeds1	269	12	Short mopane	181
6	Riparian	269	13	Mixed mopane	268
7	Firescar2	259	14	Exposed soils	95

110
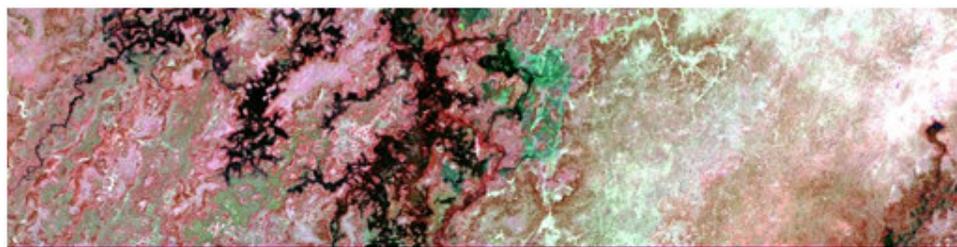
A) Color Composite of Dataset
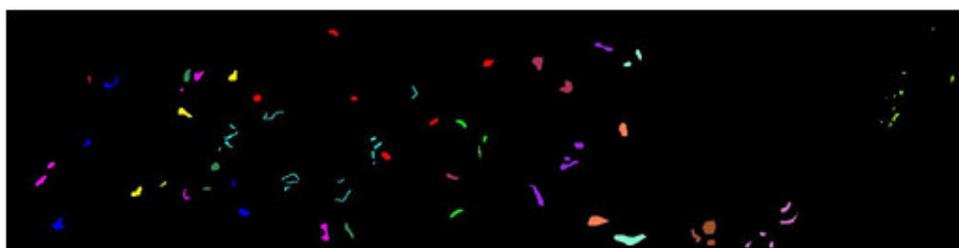
B) Ground Truth Data

Fig 3. Botswana Dataset

3.2. Evaluation Criteria

To optimize the KFCM with PSO, the clustering accuracy should be calculated. To this end, all clustering validity indices can be used. These indices categorize to three groups, named internal criteria, relative criteria and external criteria. The internal criteria use some metrics, which are based on data set and the clustering schema. The external criteria evaluate the clustering based on some user specific intuition or some reference data. The basis of the relative criteria is the comparison of the different clustering schema (Kovács, Legány et al., 2005). As in remotely sensed data clustering, the goal is finding classes, which reflects the surface model; the external criteria are better choices. Since, the clustering labels are a vector of random numbers for each class, a mapping is needed to assign each label to its correspondent class from the ground truth data, and then the external criteria can be calculated. In this study, we used kappa coefficient as external validity index. Kappa coefficient is a value between zero and one, which is calculated based on confusion matrix.

$$K = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} \times x_{+i})} \quad (9)$$

Where, K is the kappa coefficient, r is the number of columns (and rows) in a confusion matrix, x_{ii} is entry (i,i) of the confusion matrix, x_{i+} and x_{+i} are the marginal totals of row i and column j, respectively, and N is the total number of observations (Carletta, 1996; Kumar, 2004)

.

3.3. Parameter Setting

To use the PSO and KFCM algorithms, some parameters should be set. To select a subset of spectral bands, a binary random vector with length equal to number of all bands is used. The other used parameters in algorithms are listed in Table.3.

Table 3 Parameters Used in Algorithms

Algorithm	Parameters	Assigned Value
PSO	Iteration Count	150
	P_{size}	20
	W	0.72
	C_1	0.5
	C_2	0.5
KFCM	Iteration Count	50
	m	2
PSKFCM	α	0.8

3.4. KFCM Results

In order to study the performance of KFCM, this algorithm is run with different values of the kernel parameter. The selected values are between 0.1 and 1 with increment of 0.1, and between 1 and 20 with increment of 0.5. In addition, the algorithm is run with kernel parameter equal to 0.01 to survey the behavior of the KFCM in small value of parameter. For each value of kernel parameter, kappa coefficient is calculated, and the results are shown in Fig.4. The purpose of this comparison is to study the behavior the KFCM with respect to kernel parameter, and by using such a wide range of parameter we can be assured that almost the whole behavior of the KFCM algorithm is captured.

As can be observed, the behavior of KFCM is practically fixed for very large and very small values of parameter. For these values, the patterns will be transformed similar. For small values, all points are mapped near the origin in kernel space, and the algorithm cannot distinguish them as clusters. When the kernel value increases, the points are mapped more properly and the accuracy increases. Nonetheless, for large value, the accuracy

decreases from its maximum value. Although, at these values, points are similar but due to larger value of kernel function,

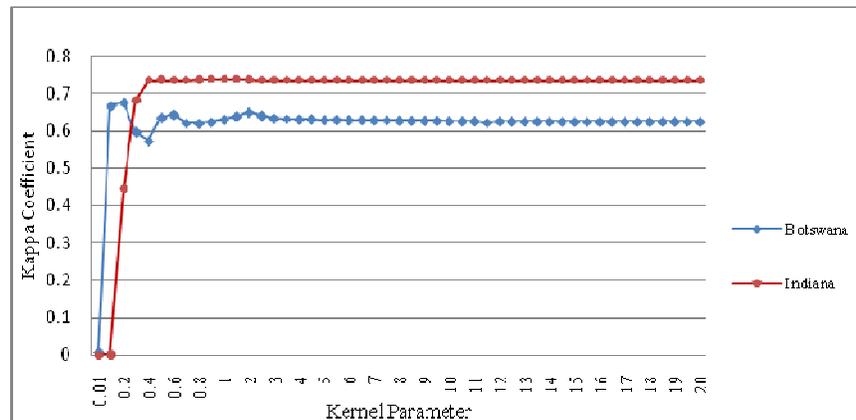the algorithm can distinguish their difference.

Fig 4. The KFCM Accuracy for Different Values of Kernel Parameter (sigma)

The other fact about these results, is that the relation of accuracy and the kernel parameter is different based on the data as it can be seen from Fig.4, the accuracy of algorithm for Botswana dataset is more sensitive to kernel parameter.

The best accuracy of KFCM for both datasets is obtained from relatively small value of kernel parameter (0.2 for Botswana dataset and 1 for Indiana dataset). The maximum value of kappa coefficient for Indiana dataset and Botswana data set is 0.7378 and 0.6764 respectively.

3.5. Results of PSKFCMK

The results of PSKFCM show that it has efficiency to tune kernel parameter automatically. Using this method, 0.7393 and 0.7396 values as kappa coefficients are obtained for Indiana and Botswana dataset respectively. In the following, the convergence of PSKFCMK is depicted.

113
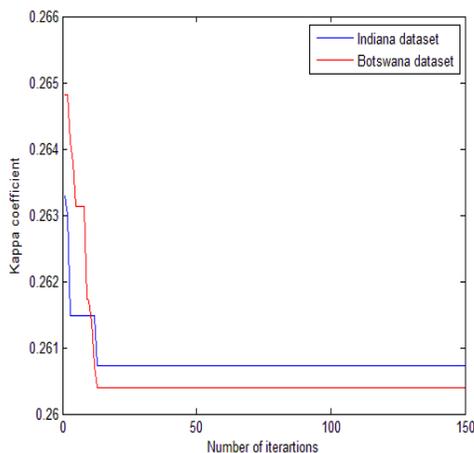
Fig 5. The Convergence of PSKFCMK Kernel Tuning

3.6. Results of PSKFCMB

To consider the impacts of numerous bands on the clustering, besides kernel parameter, KFCM is optimized with respect to data dimensionality. The results show that, selecting an optimum subset of bands; will cause dramatic increase in accuracy of clustering. This algorithm by selection of 74 bands from 185 spectral bands of Indiana dataset reaches to value of 0.8485

for kappa coefficient and by selecting 59 bands from 145 spectral bands of Botswana dataset reach to accuracy of 0.7715. As it is clear, optimization of KFCM with respect to both kernel parameter and dimensionality will cause significant increase in clustering accuracy. The convergence of PSKFCMK is depicted in Fig 6.
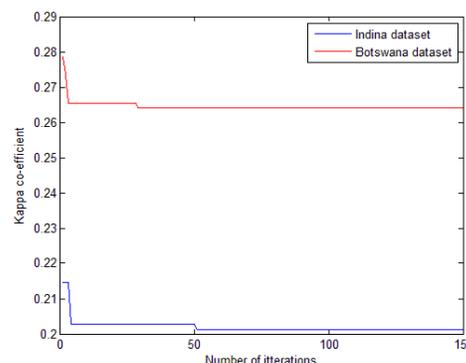
Fig 6. The Convergence of PSKFCMK

For better comparison, all the results are shown in Fig.7. In addition, the FCM algorithm is run for both datasets and the accuracy of this method is also included.
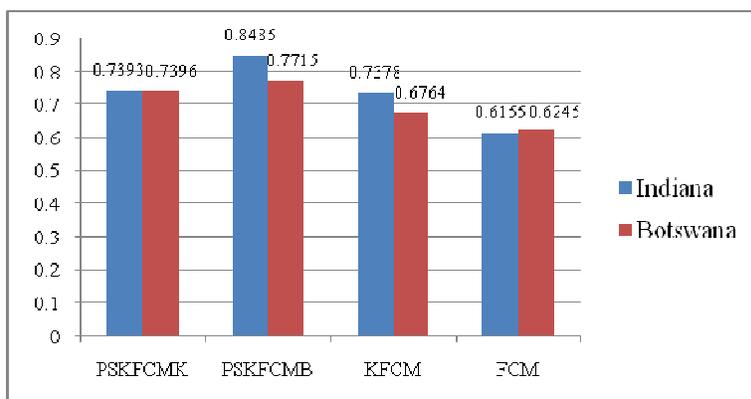
114

Fig 7 The Accuracy of Different Algorithms

As it is expected, according to Fig.7, it is clear that KFCM has a better accuracy in comparison with FCM on both datasets.

The selected bands of each dataset are listed in Table.4.

Table 4. The Band Numbers Selected by PSKFCMB

Dataset	No of Selected Bands	Selected Bands Number	Kappa
Indiana Dataset	74	1,3,6,12,13,21,23,24,25,29,32,35,36,47,50,52,53,54,58, 59,60,62,64,65,69,71,73,75,76,77,78,80,84,96,97,98,99, 101,104,108,110,111,114,115,116,117,120,121,122,124, 127,132,136,137,142,143,146,148,150,156,160,161,163, 170,171,173,174,177,178,179,180,182,184,185	0.8485
Botswana	59	3,12,14,15,19,23,24,25,26,27,28,30,31,33,34,35,36,37, 38,39,45,46,47,49,50,53,55,60,61,62,64,65,68,69,71,72, 78,79,81,83,94,104,105,106,109,112,113,116,118,121, 125,131,132,133,135,136,140,143,145	0.7715

Albeit of marginal accuracy improvement of kernel-tuning method, this method has the advantage of being fully automated. In other words, unlike the KFCM algorithm, PSKFCMK algorithm finds kernel parameter without any *a priori* knowledge

about data.

The results of PSKFCMB substantiate the claim that by using an optimum subset of hyperspectral bands, not only improve the clustering accuracy, but also decrease the computational complexity of the

115

problem. Regarding computational cost, this method decreased it 40 and 40.68 percent in Indiana and Botswana datasets respectively also, this method improved the clustering accuracy of Indiana and Botswana datasets 10.92 and 3.19 percent.

4. Conclusion

Numerous hyperspectral bands, difficulties associated with training data gathering, and parameters of classifiers are factors that mostly affect clustering accuracy. Using unsupervised classification algorithm, the problem of training data can be solved but the other two problems will still exist. In this paper, it was tried to address the remaining problems by introducing it as an optimization problem. We used KFCM as an unsupervised method. On the other hand, Particle swarm optimization is used to find the optimum parameter of kernel function in KFCM and optimum subset of bands. In the first experiment, the PSO is used for automatic tuning of kernel parameter and the results on both datasets show the ability of this method to tune kernel parameter successfully, in the second experiment, the KFCM is optimized with respect to both spectral bands and kernel parameter, which showed an incredible improvement in clustering accuracy.

In general, the results of this paper show that the KFCM is a powerful algorithm for hyperspectral data clustering if the kernel function and data dimensionality were optimized in an optimization procedure such as PSO. Accordingly, our future works will be dedicate to optimizing KFCM with respect to cluster centers as well and try other fitness functions for this purpose.

Appendix: Kernel Methods

Kernel-based learning algorithms are based on Cover's theorem (Cover, 1965). By nonlinearly transforming of a complex and nonlinearly separable patterns from input space into a higher-dimensional feature space, these patterns will became more linearly separable (Xu and Wunsch 2005). Nevertheless, transforming all data to feature space can be time-consuming, and sometimes even infeasible. To avoid working in the potentially high-dimensional space, one tries to pick a feature space in which the dot product can be evaluated directly using a nonlinear function like $K : X \times X \to \mathbb{R}$ in input space, i.e. by the mean of the kernel trick:

$$k(x, x') = < \varphi(x), \varphi(x') > \qquad (10)$$

In above equation x, x' are two patterns in input space and φ is a mapping function

from input space (X) to feature space $(\mathcal{F})$.

The most advantage of using kernel function, is that the dot product can be *implicitly* computed in feature space, without *explicitly* using or even knowing the mapping function (Muller, Mika et al. 2001). Kernel function also can be thought of as similarity measures in feature space due to the calculation of dot product between two mapped patterns (Scholkopf, 2001).

The class of kernels that can be written in the form of Eq.10 coincides with the class of positive definite or mercer kernels, which can be defined as follows.

A function like $K: X \times X \to \mathbb{R}$, is a Mercer kernel if, and only if, for each $n \in \mathbb{N}$ and $X = \{x_1, ..., x_n\}, x_i \in \mathbb{R}^d$, the $n \times n$ matrix $K_{ij} := k(x_i, x_j)$ is positive semi definite (i.e. have no negative Eigen value) (Herbrich 2002) . Usually, a kernel function is used instead of a positive, definite kernel or Mercer kernels.

One of the applications of Kennel function is calculating distance between points in feature space by their values in input space (Scholkopf 2001).

$$
\begin{aligned}
\| \varphi(x) - \varphi(x') \|^2 &=< \varphi(x), \varphi(x) > \\
&+ < \varphi(x'), \varphi(x') > -2 < \varphi(x), \varphi(x') > (11) \\
&= k(x, x) + k(x', x') - 2k(x, x')
\end{aligned}
$$

Proper definition of the kernel function and tuning of its parameter are crucial parts of all kernel-based algorithms. Since, the kernel matrix contains all the necessary information needed for learning phase of algorithm, any error in estimation of kernel parameters, could have great impacts on algorithm performance, thus, it is important to estimate the most appropriate kernel parameter based on data and algorithms (Tuia and Camps-Valls, 2011).

There are various kernel functions for different applications, but two of the most common kernel functions are Gaussian $k(x, x') = \exp(- \| x - x' \|^2 / \sigma^2), \sigma \in R^+$ and polynomial kernel $k(x, x') = (< x, x' > +1)^q, q \in \mathbf{Z}^+$.

References

[1] Archibald, R. and G. Fann (2007), "Feature selection and classification of hyperspectral images with support vector machines." *Geoscience and Remote Sensing Letters*, IEEE 4(4): 674-677.

[2] Baghshah, M. S. and S. B. Shouraki (2011), "Learning low-rank kernel matrices for

constrained clustering." *Neurocomputing* 74(12-13): 2201-2211.

[3] Bezdek, J. (1981), *Pattern recognition with fuzzy objective function algorithms*, Kluwer Academic Publishers Norwell, MA, USA.

[4] Bidhendi, S. K., A. S. Shirazi, N. Fotoohi and M. M. Ebadzadeh (2007), Material Classification of Hyperspectral Images Using Unsupervised Fuzzy Clustering Methods. Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, 2007. SITIS '07.

[5] Carletta, J. (1996), "Assessing agreement on classification tasks: the kappa statistic." *Computational Linguistics* 22(2): 249-254.

[6]Chen, S. C. and D. Q. Zhang (2004), "Robust Image Segmentation Using FCM With Spatial Constraints Based on New Kernel-Induced Distance Measure." IEEE transactions on systems, man, and cybernetics—part b: *Cybernetics* 34(4): 1907-1916.

[7] Chi, M. and L. Bruzzone (2007), "Semisupervised classification of hyperspectral images by SVMs optimized in the primal." Geoscience and Remote Sensing, *IEEE Transactions* on 45(6): 1870-1880.

[8] Cover, T. (1965), "Geometrical and statistical properties of systems of linear compute." *IEEE Trans Electron Comput* 14: 326—334.

[9] Filippone, M., F. Camastra, F. Masulli and S. Rovetta (2008), "A survey of kernel and spectral methods for clustering." *Pattern Recognition* 41(1): 176-190.

[10] Girolami, M. (2002), "Mercer kernel-based clustering in feature space." Neural Networks, IEEE Transactions on 13(3): 780-784.

[11] Graves, D. and W. Pedrycz (2010), "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study." *Fuzzy Sets and Systems* 161(4): 522-543.

[12] Herbrich, R. (2002), Kernel Classiers from a Machine Learning Perspective. Learning Kernel Classiers:theory and algorithms. T. Dietterich, G, MIT Press**:** 17-68.

[13] Hsu, P. (2004), *Feature extraction of hyperspectral images using matching pursuit*.

[14] Hughes, G. (1968), "On the mean accuracy of statistical pattern recognizers." Information Theory, IEEE Transactions on 14(1): 55-63.

[15] Jia, X. and J. Richards, A (2007), *Hyperspectral data representation. Hyperspectral Data Exploitation: Theory and Applications*. C. Chang, I, John Wiely and Sons, Inc**:** 205-225.

[16] Kamel, M., A. Campilho, X. Zhang, T. Shan, S. Wang and L. Jiao (2005), *SAR Image Segmentation Using Kernel Based Spatial FCM. Image Analysis and Recognition*, Springer Berlin / Heidelberg. 3656**:** 48-54.

[17] Kennedy, J. and R. Eberhart (1995), *Particle swarm optimization, Piscataway*, NJ: IEEE.

[18] Kim, D.-W., K. Lee, D. Lee and K. H. Lee (2005), "Evaluation of the performance of clustering algorithms in kernel-induced feature space." *Pattern Recognition* 38: 607 – 611.

[19] Kovács, F., C. Legány and A. Babos (2005), *Cluster Validity Measurement Techniques, Citeseer.*

[20] Kumar, M. (2004), Feature Selection for Classification of Hyperspectral Remotely Sensed data using NSGA-II, Citeseer.

[21] Mojaradi, B., H. Emami, M. Varshosaz and S. Jamali (2008), "A novel band selection method for hyperspectral data analysis." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.

[22] Muller, K. R., S. Mika, G. Ratsch, K. Tsuda and B. Scholkopf (2001), "An introduction to kernel-based learning algorithms." Neural Networks, IEEE Transactions on 12(2): 181-201.

[23] Neuenschwander, A., M. Crawford and S. Ringrose (2005), "Results of the EO-1 experiment-Use of Earth Observing-1 Advanced Land Imager (ALI) data to assess the vegetational response to flooding in the Okavango Delta, Botswana". *International Journal of Remote Sensing*.

[24] Niazmardi, S., S. Homayouni and A. Safari (2011), Remotely Sensed Image Clustering Based on Kernel-Based Fuzzy C-Means Algorithm. SMPR. Tehran.

بهینه‌سازی الگوریتم خوشه‌بندی KFCM برای تصاویر فراطیفی به وسیله الگوریتم بهینه‌سازی توده ذرات

سعید نیازمردی¹، امین علیزاده نایینی²، سعید همایونی³، عبدالرضا صفری⁴، فرهاد صمدزادگان⁵

سنجنده‌های فراطیفی با نمونه‌برداری دقیق از بازتاب اشیاء در قالب چندین باندطیفی باریک و به هم پیوسته، منبع اطلاعات بسیار ارزشمندی را برای تعیین کلاس‌های پوششی مختلف ایجاد می‌نمایند. با این وجود طبقه‌بندی داده‌های فراطیفی و استخراج این اطلاعات با مشکلاتی همراه است، که یکی از شناخته‌شده‌ترین آنها مربوط به در دسترس نبودن تعداد کافی از نمونه‌های آموزشی برای مرحله یادگیری الگوریتم است. یکی از راهکارهای ممکن برای رفع این مشکل استفاده از الگوریتم‌های طبقه‌بندی بدون‌نظارت (الگوریتم‌های خوشه‌بندی) مانند الگوریتم خوشه‌بندی فازی کرنل پایه (KFCM) است. الگوریتم خوشه‌بندی فازی کرنل پایه که از بازنویسی الگوریتم خوشه‌بندی فازی با استفاده از توابع کرنل حاصل می‌شود، معمولاً نتایج بهتری در مقایسه با سایر الگوریتم‌ها دارد. اما، در صورت استفاده از این الگوریتم برای خوشه‌بندی داده‌های فراطیفی، دقت الگوریتم KFCM، متاثر از ابعاد بالای داده و همچنین پارامتر تابع کرنل استفاده شده، کاهش می‌یابد. هدف از این مقاله، بهینه کردن الگوریتم KFCM نسبت به پارامتر کرنل و ابعاد داده است که برای این کار، از الگوریتم بهینه‌سازی توده ذرات (PSO) استفاده شده است. الگوریتم بهینه‌سازی توده ذرات یک ابزار قدرتمند بهینه‌سازی است که با الهام از رفتار پرندگان ایجاد شده و توانایی یافتن بهینه کلی مسائل را داراست. در این مقاله همچنین دو روش جدید برای بهینه‌سازی الگوریتم KFCM نسبت به پارامتر تابع کرنل و ابعاد داده پیشنهاد شده است. مقایسه نتایج این روش‌ها با روش‌های دیگر خوشه‌بندی نشان‌دهنده‌ی توانایی بالای آنها برای خوشه‌بندی داده‌های فراطیفی است.

واژگان کلیدی: خوشه‌بندی فازی کرنل پایه، KFCM، PSO، داده‌های فراطیفی.

١. دانشجوی دکترای سنجش از دور، دانشکده مهندسی نقشه برداری، دانشکده مهندسی، دانشگاه تهران، ایران.

٢. دانشجوی دکترای فتوگرامتری، گروه مهندسی نقشه برداری، دانشکده مهندسی، دانشگاه تهران، ایران.

٣. استادیار، گروه جغرافیا، دانشگاه اتاوا.

٤. دانشیار، گروه مهندسی نقشه برداری، دانشکده مهندسی، دانشگاه تهران، ایران.

٥. استاد گروه مهندسی نقشه برداری، دانشکده مهندسی، دانشگاه تهران، ایران.