

A NEW MULTIPLE CLASSIFIER SYSTEM BASED ON A PSO ALGORITHM FOR THE CLASSIFICATION OF HYPERSPECTRAL IMAGES

S. H. Alizadeh Moghaddam^a, M. Mokhtarzade^{a, *}, S. A. Alizadeh Moghaddam^a

^a Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran-
h.alizadeh@email.kntu.ac.ir, m_mokhtarzade@kntu.ac.ir, h74.alizadeh@email.kntu.ac.ir

KEY WORDS: Multiple classifier systems (MCSs), Particle swarm optimization (PSO), Hyperspectral image classification

ABSTRACT:

Multiple classifier systems (MCSs) have shown great performance for the classification of hyperspectral images. The requirements for a successful MCS are 1) diversity between ensembles and 2) good classification accuracy of each ensemble. In this paper, we develop a new MCS method based on a particle swarm optimization (PSO) algorithm. Firstly, in each ensemble of the proposed method, called PSO-MCS, PSO identifies a subset of the spectral bands with a high J_2 value, which is a measure of class-separability. Then, an SVM classifier is used to classify the input image, applying the selected features in each ensemble. Finally, the classification results of the entire ensembles are integrated using a majority voting strategy. Having the benefit of the PSO algorithm, PSO-MCS selects appropriate features. In addition, due to the fact that different features are selected in different runs of PSO, diversity between the ensembles is provided. Experimental results on an AVIRIS Indian Pine image show the superiority of the proposed method over its competitor, named random feature selection method.

1. INTRODUCTION

Hyperspectral images (HSIs) have been used in various application such as agriculture and mineralogy (Moon et al., 2007; Navalgund et al., 2007; Van Ruitenbeek et al., 2006). Supervised classification methods are the most common techniques to garner information from HSIs. These techniques, however, suffer from a critical issue termed “Hughes phenomenon,” which occurs when the size of the training samples is not adequately large compared to the high dimensionality of the HSIs.

One of the major development in the supervised learning scope is the multiple classifier systems (MCSs) (Xia et al., 2016; Xu et al., 1992). The MCSs, which have been widely used for the classification of the HSIs (Xia et al., 2016; Xia et al., 2014) and multi-source data (Briem et al., 2002; Miao et al., 2012), do not refer to any specific algorithm. The MCSs in fact combine some individual classifiers, known as an ensemble, to improve the accuracy and efficiency of the classification task resulted by each individual classifier (Du et al., 2012; Xia et al., 2016).

The requirements of a powerful MCS are 1) existence of diversity between ensembles and 2) good classification accuracy of each ensemble (Du et al., 2012; Xia et al., 2016). In the literature, there are some methods proposed to provide diversity for each ensemble. These methods can be broadly divided into three categories: 1) the manipulation of training samples, such as a Bagging method (Breiman, 1996); 2) the manipulation of the input features, including random feature selection (RFS) (Barandiaran, 1998; Freund and Schapire, 1996) where each ensemble uses a random subsection of the feature space; and 3) the manipulation of both the training samples and the input features, such as random forest (Gislason et al., 2006), rotation forest (Xia et al., 2014), and rotation-based SVM (Xia et al., 2016) methods.

As mentioned earlier, in RFS, the ensembles’ diversity is achieved through the random feature selection. However, there is no guarantee that RFS will select appropriate features for the classification task. As an instance, if the noisy bands or those providing low class-separability are selected, the classification accuracy of the ensembles will decrease considerably. To address these deficiencies, we propose a new MCS based on a particle swarm optimization (PSO) algorithm in this paper.

The PSO is of the famous metaheuristic optimization algorithm (Kennedy, 2010). PSO has been widely used in the remote sensing community for different purposes (Gholinejad et al., 2018; Moghaddam et al., 2018; Naëini et al., 2018; Naëini et al., 2014; Su et al., 2014; Zhang et al., 2011). In the contexts of hyperspectral remote sensing, PSO has been applied for feature selection (FS) to identify an optimal subset of the spectral bands (Su et al., 2014). However, due to the high complexity of the FS problem, PSO seldom converges to a unique solution. In other words, different (sub)optimal solutions (here a subset of the spectral bands) will be identified in different runs of the PSO algorithm.

The diversity of the solutions resulted by PSO suggests developing a new MSC framework for the HSI classification, termed as PSO-SVM. In each ensemble of the PSO-MCS, the PSO algorithm is first applied to select the optimal features with high class-separability. Applying the identified features, PSO-MCS trains an SVM classifier to classify the input HSI. Finally, the classification results of all ensembles are integrated using a majority voting (MV), which is an effective strategy (Du et al., 2012).

We believe that PSO-MCS can be a successful MCS because 1) the classification results of the ensembles are diverse as they are trained in different feature space, and 2) the classification accuracy of the each ensemble is satisfactory enough due to the

* Corresponding author

fact that the optimal spectral bands are used for the classification task.

The main contribution of the paper is that it suggests a new MCS and, for the first time, introduces the metaheuristic optimization algorithm concept into the MCS. In this paper, we make use of the instability of the PSO algorithm in the FS task. To do so, we suggest a new MCS and combine the results of some single classifiers trained on different feature subsets, which are identified optimally thanks to PSO.

The rest of the paper is organized in the following fashion. The proposed PSO-MCS method is first introduced. In section 3, the experimental results are given. The last section will be concluded some remarks.

2. PROPOSED METHOD

2.1 PSO for feature selection

For the feature selection task, the binary version of PSO algorithm can be used to select f number features. The particle representation of the FS problem is shown in Figure 1.

$$x = \begin{array}{|c|c|c|c|c|} \hline 1 & 0 & 1 & \dots & 1 \\ \hline b_1 & b_2 & b_3 & \dots & b_N \\ \hline \end{array}$$

Figure 1. Particle representation

x is the position of a particle in an N -dimension search space, where N is the dimensionality of the input HIS. In the above representation, “1” means that the corresponding feature is selected, while “0” means the opposite.

According to the PSO rules, the position of each particle is adjusted iteratively until an optimal solution, maximizing/minimizing an objective function, is identified. In this paper, the optimization problem is identified as follows:

$$\max_x J_2(x) \quad \text{subject to} \quad \text{sum}(x) = f$$

where f is the total number of selected features, and x is the position of a particle. J_2 , considered as the objective function, is a class-separability measure calculated based on between-class and within-class scattering matrices that are achieved using training samples. We refer the reads to (Theodoridis and Koutroumbas, 1999) for details on the J_2 measure.

2.2 PSO-MCS

Figure 2 shows the flow chart of the proposed PSO-MSC method.

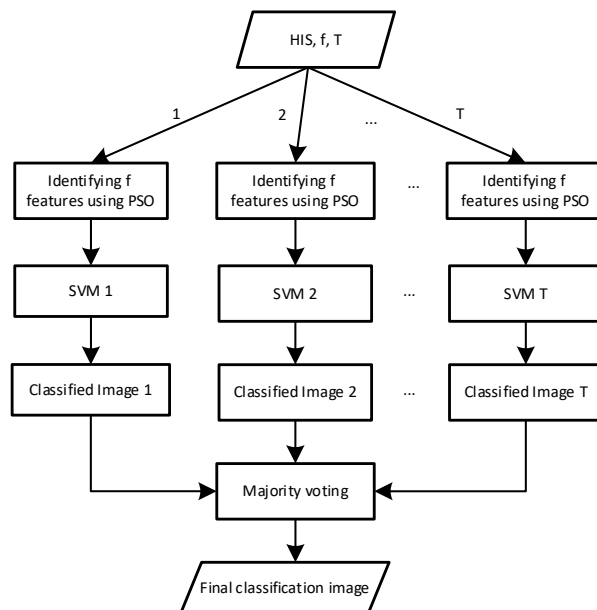


Figure 2. Flowchart of the PSO-MCS method.

The proposed method has three inputs: HIS, the total number of ensemble or ensemble size (T), and the total number of features selected by PSO in each ensemble (f). In each ensemble, firstly, PSO-MSC applies the PSO algorithm to select f number of features according to section 2.1. Applying the selected features, PSO-MCS trains an SVM classifier and classifies the HSI. T number of ensembles are generated, resulting in T number of classified images. Finally, in order to combine all T classification results, an MV strategy is applied.

3. EXPERIMENTS

3.1 Data set

We used one famous HIS known as Indian Pine in our experiment. This image was captured by the AVIRIS sensor from an agricultural field in the state of Indiana, USA. The captured image originally consist of 220 spectral bands in the spectral range of 0.4 to 2.45 micron. Twenty water absorption bands were weed out, and therefore; the experiments were conducted with the remaining 200 bands.

According to the available ground truth, this data set consists of 16 classes. In line with the literature, we excluded four of these classes with the inadequate number of sample. Figure 3 depicts the false color composite of the data set (a), and the ground truth with 12 classes of interest (b). Table 1 shows the total number of samples in each class.

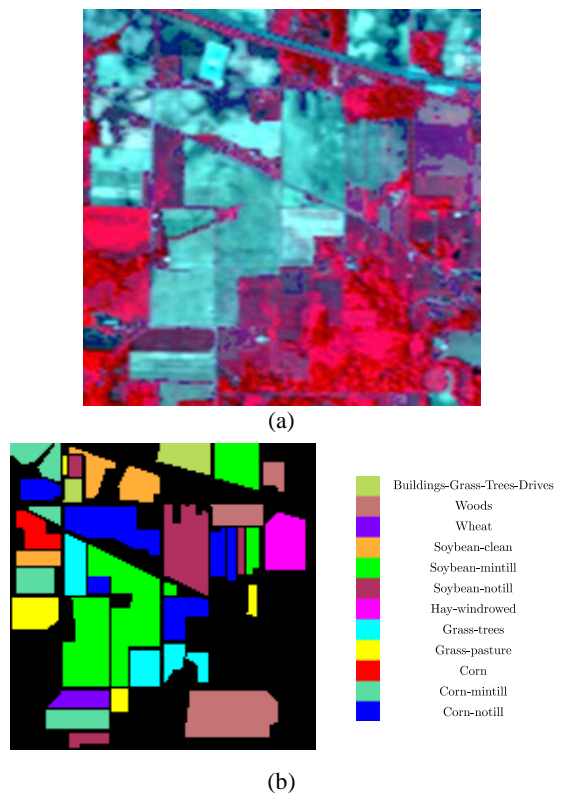


Figure 3. Indian Pine case study. (a) False color composite; (b) Ground truth map.

Class of interest	No. of Sample
Corn-notill	1428
Corn-mintill	830
Corn	237
Grass-pasture	483
Grass-trees	730
Hay-windrowed	478
Soybean-notill	972
Soybean-mintill	2455
Soybean-clean	593
Wheat	205
Woods	1265
Buildings-Grass-Trees-Drives	386

Table 1. Indian pine ground truth specification

3.2 Results and discussion

In this section, we present the results of the proposed PSO-MCS and compare them with the RFS method. The SVM was selected as the base classifier for the both methods. A free library, known as LibSVM (Chang and Lin, 2011), was applied for the implantation of the SVM. Two parameters of the SVM, namely “cost” and the RBF kernel’s parameters, were tuned applying a Grid search and five-fold cross-validation in each ensemble.

Regarding the parameters of the PSO algorithm, they are all set according to (Hassan et al., 2005). Table 2 gives the value of these parameters.

PSO parameter	Value
$c_{1\ max}$	2.5
$c_{1\ min}$	1.5
$c_{2\ max}$	2.5
$c_{2\ min}$	2
$v_{\ max}$	4
$w_{\ max}$	1.8
$w_{\ min}$	0.4
Max No. iteration	200
Number of particles	25

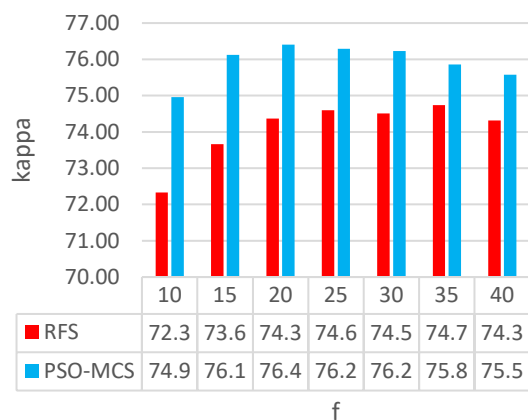
Table 2. Parameter setting of PSO.

In the above table, w_{min} , $c_{1\ min}$ and $c_{2\ min}$ are respectively the minimum values of the inertia weight and acceleration coefficients; w_{max} , $c_{1\ max}$, $c_{2\ max}$ are their corresponding upper boundary values; v_{max} is the maximum velocity value.

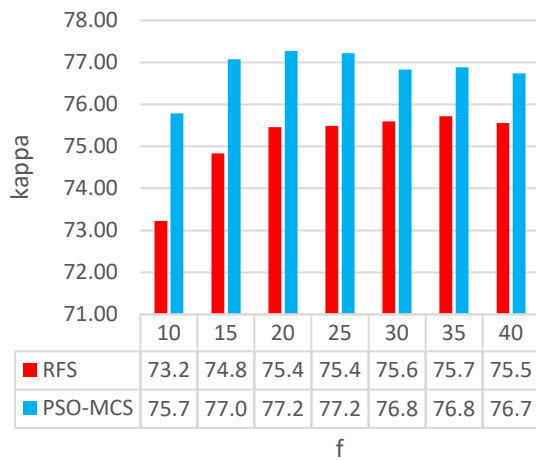
In each class, 50 samples were randomly selected from the ground truth as the training samples and the remaining samples were considered for the evaluation purpose. Each experiment was repeated ten times and the mean of kappa coefficients (multiplied by 100), as a frequently used evaluation criterion, is given in the following figures.

The PSO-MSC and RFS method have two parameters that must be set in prior: 1) the total number of features in each ensemble (f) and 2) ensemble size (T). These parameters was selected as $T = \{10,25,50\}$ and $f = \{10,15,20,25,30,35,40\}$.

Figure 4 illustrates the kappa coefficient vs. f . Note that in Figure 4(a) and 4(b), T was considered as 10 and 50, respectively.



(a)



(b)

Figure 4. Kappa coefficient vs. f . (a) $T = 10$; (b) $T = 50$.

Above figures evidently show that the proposed PSO-MSC performs better than RFS in all cases. It can be seen that there is an increasing trend in the kappa coefficient value from 10 to 20 ensemble's features. However, the accuracy declines when the number of features exceeds 20. This decreasing trend can be justified by the Hugh phenomenon, where the classification performance decreases with an increase in the number of applied features.

In order to compare the performance of MSCs against single classifier scheme, the whole image was classified using a single SVM. The result indicates that a kappa coefficient of 69.45 can be accomplished in this case. This shows the superiority of the MSCs over the single classifier.

The maximum achieved accuracy of PSO-MSC occurred in the case of 20 features; however, this situation occurred in the case of 35 features for the RFS method (Figure 4). This shows the requirement of RFS to the high number of features.

Figure 5 shows the kappa coefficient vs. ensemble size (T). Note that the number of features in each ensemble was considered as $f = 25$.

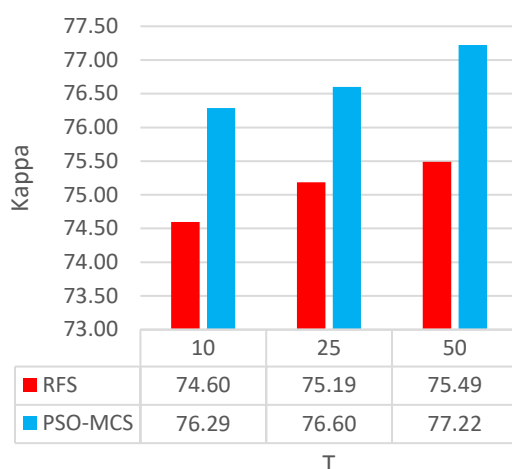


Figure 5. Kappa coefficient vs. ensemble size (T)

One can conclude from the above figure that PSO-MCS is the superior method, where it achieved higher accuracies in all cases of T .

From the diversity point of view, we measured the diversity within ensembles using a coincident failure diversity (CFD) measure (Kuncheva and Whitaker, 2003). The CFD value of the both methods, i.e., PSO-MCS and RFS were almost the same. For instance, in the case of $T = 25$, the CFD value of PSO-MCS and RFS were respectively 0.63 and 0.64 on average.

The reason behind the better performance of PSO-MCS can be sought in the applied feature space. As previously mentioned, the features in RFS were selected randomly. It means that it is highly probable that inappropriate features, e.g., noisy ones were selected in each ensemble, degrading the accuracy of the classification task. However, this is in contrast to PSO-MCS where (sub)optimal features with high class-separability were chosen thanks to J_2 measure.

4. CONCLUSION

One way to generate diverse ensembles of an MCS is to select random features, i.e., the RFS method. Although the generated ensembles in this method are diverse, there is no guarantee for the selection of appropriate features, such as those with high class-separability. In this paper, we present the PSO-MCS method that addresses the mentioned disadvantage. In fact, PSO-MCS selects the bands with high class-separability in each ensemble using the PSO algorithm. The experimental results on the AVIRIS Indian Pine data set revealed the superiority of PSO-MCS over RFS in the terms of classification accuracy.

REFERENCES

- Barandiaran, I., 1998. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence* 20.
- Breiman, L., 1996. Bagging predictors. *Machine learning* 24, 123-140.
- Briem, G.J., Benediktsson, J.A., Sveinsson, J.R., 2002. Multiple classifiers applied to multisource remote sensing data. *IEEE transactions on geoscience and remote sensing* 40, 2291-2299.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2, 27.
- Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y., Liu, S., 2012. Multiple classifier system for remote sensing image classification: A review. *Sensors* 12, 4764-4792.
- Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm, *icml*. Citeseer, pp. 148-156.
- Gholinejad, S., Naeini, A.A., Amiri-Simkooei, A., 2018. Robust Particle Swarm Optimization of RFMs for High-Resolution Satellite Images Based on K-Fold Cross-Validation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. *Pattern Recognition Letters* 27, 294-300.
- Hassan, R., Cohanim, B., De Weck, O., Venter, G., 2005. A comparison of particle swarm optimization and the genetic algorithm, 46th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, p. 1897.
- Kennedy, J., 2010. Particle swarm optimization. *Encyclopedia of machine learning*, 760-766.
- Kuncheva, L.I., Whitaker, C.J., 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning* 51, 181-207.
- Miao, X., Heaton, J.S., Zheng, S., Charlet, D.A., Liu, H., 2012. Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data. *International journal of remote sensing* 33, 1823-1849.
- Moghaddam, S.H.A., Mokhtarzade, M., Moghaddam, S.A.A., 2018. Optimization of RFM's Structure Based on PSO Algorithm and Figure Condition Analysis. *IEEE Geoscience and Remote Sensing Letters* 15, 1179-1183.
- Moon, H., Ahn, H., Kodell, R.L., Baek, S., Lin, C.-J., Chen, J.J., 2007. Ensemble methods for classification of patients for personalized medicine with high-dimensional data. *Artificial intelligence in medicine* 41, 197-207.
- Naeini, A.A., Babadi, M., Mirzadeh, S.M.J., Amini, S., 2018. Particle swarm optimization for object-based feature selection of VHRS satellite images. *IEEE Geoscience and Remote Sensing Letters* 15, 379-383.
- Naeini, A.A., Homayouni, S., Saadatesresht, M., 2014. Improving the dynamic clustering of hyperspectral data based on the integration of swarm optimization and decision analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 2161-2173.
- Navalgund, R.R., Jayaraman, V., Roy, P., 2007. Remote sensing applications: An overview. *Current Science* (00113891) 93.
- Su, H., Du, Q., Chen, G., Du, P., 2014. Optimized hyperspectral band selection using particle swarm optimization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 2659-2670.
- Theodoridis, S., Koutroumbas, K., 1999. *Pattern Recognition*. Academic Press.
- Van Ruitenbeek, F.J., Debba, P., Van Der Meer, F.D., Cudahy, T., Van Der Meijde, M., Hale, M., 2006. Mapping white micas and their absorption wavelengths using hyperspectral band ratios. *Remote Sensing of Environment* 102, 211-222.
- Xia, J., Chanussot, J., Du, P., He, X., 2016. Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples. *IEEE Transactions on Geoscience and Remote Sensing* 54, 1519-1531.
- Xia, J., Du, P., He, X., Chanussot, J., 2014. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters* 11, 239-243.
- Xu, L., Krzyzak, A., Suen, C.Y., 1992. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE transactions on systems, man, and cybernetics* 22, 418-435.
- Zhang, B., Sun, X., Gao, L., Yang, L., 2011. Endmember extraction of hyperspectral remote sensing images based on the discrete particle swarm optimization algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 49, 4173-4176.