

Spectral and Wavelet-based Feature Selection with Particle Swarm Optimization for Hyperspectral Classification

Sheng Ding

College of Computer Science and Technology ,
Wuhan University of Science and Technology, Wuhan , China
Email:dingwhu@gmail.com

Abstract—Spectral band selection is a fundamental problem in hyperspectral classification. This paper addresses the problem of band selection for hyperspectral remote sensing image and SVM parameter optimization. First, we propose an evolutionary classification system based on particle swarm optimization (PSO) to improve the generalization performance of the SVM classifier. For this purpose, we have optimized the SVM classifier design by searching for the best value of the parameters that tune its discriminant function, and upstream by looking for the best subset of features that feed the classifier. Second, for making use of wavelet signal feature of pixels of hyperspectral image, we investigate the performance of the selected wavelet features based on wavelet approximate coefficients at the third level. The PSO algorithm is performed to optimize spectral feature and wavelet-based approximate coefficients to select the best discriminant features for hyperspectral remote imagery. The experiments are conducted on the basis of AVIRIS 92AV3C dataset. The obtained results clearly confirm the superiority of the SVM approach as compared to traditional classifiers, and suggest that further substantial improvements in terms of classification accuracy can be achieved by the proposed PSO-SVM classification system.

Index Terms — support vector machine (SVM) , Particle Swarm Optimization(PSO) , optimization , Feature Selection, Wavelet Decomposition

I. INTRODUCTION

Hyperspectral Classification problems have been extensively studied. Numerous factors, such as incomplete data, and the choice of values for the parameters of a given model, may affect classification results. Classification problems have previously been solved with statistical methods such as logistic regression or discriminate analysis. Technological advances have led to the development of methods for solving classification problems, including decision trees, back-propagation neural networks, rough set theory and support vector machines (SVM). SVM which is an emerging data classification technique proposed by Vapnik[1], and has been widely adopted in various fields of classification problems in recent years.

Classification problems generally involve a number of features. However, not all of these features are equally important for a specific task. Some of them may be redundant or even irrelevant. Better performance may be

achieved by discarding some features. In other circumstances, the dimensionality of input space may be decreased to save some computation effort, although this may slightly lower classification accuracy. Therefore, the classification process must be fast and accurate, using the smallest number of features. This objective can be achieved using feature selection. Feature selection strategies are often implied to explore the effect of irrelevant attributes on the performance of classifier systems.

This study attempts to increase the classification accuracy rate by employing an approach based on particle swarm optimization (PSO) in SVM. This novel approach is termed PSO-SVM. The developed PSO-SVM approach not only tunes the parameter values of SVM, but also identifies a subset of features for specific problems, maximizing the classification accuracy rate of SVM. This makes the optimal separating hyper-plane obtainable in both linear and non-linear classification problems.

In particular, they are organized so as to test the sensitivity of the SVM classifier and that of three reference classifiers used for comparison, i.e., SVM-Linear classifier, the k-nearest neighbor (K-nn) classifier and the radial basis function neural network (RBF-NN) classifier, with respect to the curse of dimensionality and the number of available training data

The rest of the paper is organized as follows. The literature review is recalled in Section II. The main concepts and principles of PSO are introduced in Section III. The proposed PSO-SVM classification system is described in Section IV. The experimental results obtained on hyperspectral remote sensing data are reported in Sections V. PSO-SVM for spectral and wavelet feature selection is illustrated in Section VI and experimental result indicates the methods is effective. Finally, conclusions are drawn in Section V.

II. LITERATURE REVIEW

Approaches for feature selection can be categorized into two models, namely a filter model and a wrapper model [2]. Statistical techniques, such as principal component analysis, factor analysis, independent component analysis and discriminate analysis can be adopted in filter-based feature selection approaches to

investigate other indirect performance measures, most of which are based on distance and information. Chen and Hsieh[3] presented latent semantic analysis and web page feature selection, which are combined with the SVM technique to extract features. Gold [4] presented a Bayesian viewpoint of SVM classifiers to tune hyper-parameter values in order to determine useful criteria for pruning irrelevant features.

The wrapper model [5] applies the classifier accuracy rate as the performance measure. Some researchers have concluded that if the purpose of the model is to minimize the classifier error rate, and the measurement cost for all the features is equal, then the classifier's predictive accuracy is the most important factor. Restated, the classifier should be constructed to achieve the highest classification accuracy. The features adopted by the classifier are then chosen as the optimal features. In the wrapper model, meta-heuristic approaches are commonly employed to help in looking for the best feature subset. Although meta-heuristic approaches are slow, they obtain the (near) best feature subset. Shon[6] employed GA to screen the features of a dataset. The selected subset of features is then fed into the SVM for classification testing. Zhang[7] developed a GA-based approach to discover a beneficial subset of features for SVM in machine condition monitoring. Samanta [8] proposed a GA approach to modify the RBF width parameter of SVM with feature selection. Nevertheless, since these approaches only consider the RBF width parameter for the SVM, they may miss the optimal parameter setting. Huang and Wang [9] presented a GA-based feature selection and parameters optimization for SVM. Moreover, Huang et al.[10] utilized the GA-based feature selection and parameter optimization for credit scoring.

Several kernel functions help the SVM obtain the optimal solution. The most frequently used such kernel functions are the polynomial, sigmoid and radial basis kernel function (RBF). The RBF is generally applied most frequently, because it can classify high-dimensional data, unlike a linear kernel function. Additionally, the RBF has fewer parameters to set than a polynomial kernel. RBF and other kernel functions have similar overall performance. Consequently, RBF is an effective option for kernel function. Therefore, this study applies an RBF kernel function in the SVM to obtain optimal solution. Two major RBF parameters applied in SVM, C and γ , must be set appropriately. Parameter C represents the cost of the penalty. The choice of value for C influences on the classification outcome. If C is too large, then the classification accuracy rate is very high in the training phase, but very low in the testing phase. Parameter γ has a much greater influence on classification outcomes than C , because its value affects the partitioning outcome in the feature space. An excessively large value for parameter γ results in over-fitting, while a disproportionately small value leads to under-fitting. Grid search [11] is the most common method to determine appropriate values for C and γ . Values for parameters C and γ that lead to the highest classification accuracy rate in this interval can be found by setting appropriate values for the upper and

lower bounds (the search interval) and the jumping interval in the search. Nevertheless, this approach is a local search method, and vulnerable to local optima. Additionally, setting the search interval is a problem. Too large a search interval wastes computational resource, while too small a search interval might render a satisfactory outcome impossible.

In addition to the commonly used grid search approach, other techniques are employed in SVM to improve the possibility of a correct choice of parameter values. Pai and Hong [12] proposed an SA-based approach to obtain parameter values for SVM, and applied it in real data; however, this approach does not address feature selection, and therefore may exclude the optimal result. As well as the two parameters C and γ , other factors, such as the quality of the feature's dataset, may influence the classification accuracy rate. For instance, the correlations between features influence the classification result. Accidental removal of important features might lower the classification accuracy rate. Additionally, some dataset features may have no influence at all, or may contain a high level of noise. Removing such features can improve the searching speed and accuracy rate.

It is worth underlining that the kernel-based implementation of SVM involves the problem of the selection of multiple parameters, including the kernel parameters (e.g., the γ and p parameters for the Gaussian and polynomial kernels, respectively) and the regularization parameters C .

Studies have also illustrated that a radial basis kernel yields the best results in remote sensing applications [13][14]. We chose to use the radial basis kernel for SVM in this study. The verification of the applicability of other specialized kernel functions for the classification of remote sensing data may be used in future studies. The equation for the radial basis kernel is

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2) \quad (1)$$

where γ represents a parameter inversely proportional to the width of the Gaussian kernel.

III. PARTICLE SWARM OPTIMIZATION AND DEVELOPED PSO-SVM APPROACH

A. PSO Concept

The particle swarm optimization (PSO), originally developed by Kennedy and Eberhart [15], is a method for optimizing hard numerical functions on metaphor of social behaviors of flocks of birds and schools of fish. It is an evolutionary computation technique based on swarm intelligence. A swarm consists of individuals, which are called particles, which change their positions over time. Each particle represents a potential solution to the problem. In a PSO system, particles fly around in a multidimensional searching space. During its flight each particle adjusts its position according to its own experience and the experience of its neighboring particles making use of the best position encountered by itself and its neighbors. The effect is that particles move towards the better solution areas, while still having the ability to

search a wide area around the better solution areas. The performance of each particle is measured according to a pre-defined fitness function, which is related to the problem being solved. The PSO has been found to be robust and fast in solving nonlinear, non-differentiable and multi-modal problems. The mathematical description and executive steps of the PSO are as follows. Let the i th particle in a D-dimensional space be represented as $x_i=(x_{i1}, \dots, x_{id}, \dots, x_{iD})$. The best previous position of the i th particle is recorded and represented as $p_i=(p_{i1}, \dots, p_{id}, \dots, p_{iD})$, which gives the best fitness value and is also called $pbest$. The index of the best $pbest$ among all the particles is represented by the symbol g . The location P_g is also called $gbest$. The velocity for the i th particle is represented as $v_i=(v_{i1}, \dots, v_{id}, \dots, v_{iD})$. The concept of the particle swarm optimization consists of changing the velocity and location of each particle towards its $pbest$ and $gbest$ locations at each time step:

$$v_{id} = wv_{id} + C_1 r_1(p_{id} - x_{id}) + C_2 r_2(p_{gd} - x_{id}), \quad (2)$$

$$x_{id} = x_{id} + v_{id}, \quad (3)$$

where w is the inertia coefficient which is a constant in the interval $[0, 1]$ and can be adjusted in the direction of linear decrease; C_1 and C_2 are learning rates which are nonnegative constants; r_1 and r_2 are generated randomly in the interval $[0, 1]$; The termination criterion for iterations is determined according to whether the maximum generation or a designated value of the fitness is reached.

In this following of the section, we describe the proposed SVM system for the classification of hyperspectral remote sensing image. As mentioned in the Introduction, the aim of this system is to optimize the SVM classifier accuracy by automatically: 1) detecting the subset of the best discriminative features (without requiring a user-defined number of desired features) and 2) solving the SVM model selection issue (i.e, estimating the best values of the regularization and kernel parameters). In order to attain this, the system is derived from an optimization framework based on PSO.

B. PSO Setup

Because of the good performances generally achieved with the nonlinear SVM classifier based on the Gaussian kernel, in the following, we shall describe the proposed classification system with this particular kernel. This study developed a PSO approach, termed PSO-SVM, for parameter determination and feature selection in the SVM. Without feature selection, two decision variables, designated C and γ are required. For the feature selection, if n features are required to decide which features are chosen, then $n+2$ decision variables must be adopted. The value of n variables ranges between 0 and 1. If the value of a variable is less than or equal to 0.5, then its corresponding feature is not chosen. Conversely, if the value of a variable is greater than 0.5, then its corresponding feature is chosen. Fig. 1 illustrates the solution representation. C is penalty cost, γ represents a

parameter inversely proportional to the width of the RBF kernel, a_n =Feature n is selected.

Spectral features(N)			C	γ
1	N	N+1	N+2
a_1	a_N	C	γ

Figure 1. Solution representation.

The following shows the while process for PSO-SVM. First, the population of particles is initialized, each particle having a random position within the D-dimensional space and a random velocity for each dimension. Second, each particle's fitness for the SVM is evaluated. The each particle's fitness in this study is the classification accuracy. If the fitness is better than the particle's best fitness, then the position vector is saved for the particle. If the particle's fitness is better than the global best fitness, then the position vector is saved for the global best. Finally the particle's velocity and position are updated until the termination condition is satisfied. Figure 2 shows the architecture of the developed PSO-based parameter determination and feature selection approach for SVM.

IV. EXPERIMENTAL DESIGN

A. Dataset Description

The hyperspectral remote sensing image used in our experiments is a section of a scene taken over northwest Indiana's Indian Pines by the AVIRIS sensor in 1992 [16]. From the 220 spectral channels acquired by the AVIRIS sensor, 20 channels are discarded because affected by atmospheric problems. From the 16 different land-cover classes available in the original ground truth, seven classes are discarded, since only few training samples were available for them (this makes the experimental analysis more significant from the statistical viewpoint). The remaining nine land-cover classes were used to generate a set of 4757 training samples (used for learning the classifiers) and a set of 4588 test samples (exploited for assessing their accuracies) (see TABLE I).

TABLE I NUMBER OF TRAINING AND TEST SAMPLES

CLASS	TRAINING	TEST
ω_1 -Corn-no till	742	692
ω_2 -Corn-min till	442	392
ω_3 -Grass/Pasture	260	237
ω_4 -Grass/Trees	389	358
ω_5 -Hay-windrowed	236	253
ω_6 -Soybean-no till	487	481
ω_7 -Soybean-min till	1245	1223
ω_8 -Soybean-clean till	305	309
ω_9 -Woods	651	643
Total	4757	4588

B. Experimental Scheme

The proposed experimental framework was articulated around the following three main experiments.

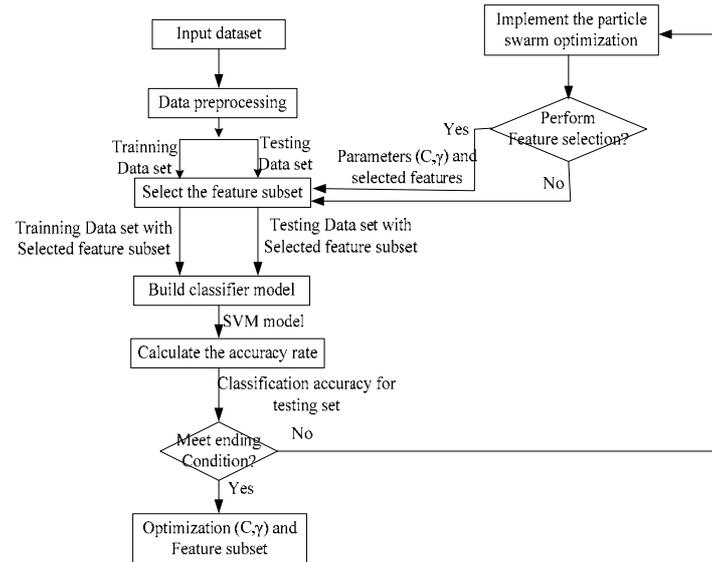


Figure 2. The architecture of the proposed PSO-SVM based parameters determination and feature selection approach for SVM.

1) The first experiment aimed at assessing the effectiveness of the SVM approach in classifying hyperspectral data directly in the whole original hyperdimensional feature space (i.e., by means of all the 200 available features). The total number of training number fixed to 4757. For comparison purpose, we implemented three other reference nonparametric classification approaches, namely, the SVM-Linear, the k-nearest neighbor (K-nn) and the radial basis function neural network(RBF-NN) classifiers .

2) In the second experiment, it was desired to explore the behavior of the SVM classifier (compared to the two reference classifiers) when integrated within a standard classification scheme based on a PCA feature reduction. In particular, the number of features was varied from 20 to 200 with a step of 20 so as to test this classifier in small as well as high-dimensional feature subspaces. And the result is compared with the original feature reduction scheme(the feature number vary from original 20 to 200 band without PCA feature transform)(see Fig.4 and Fig.5)

3) The third experimental part had for objective to assess the capability of the proposed PSO-SVM classification system to boost further the accuracy of the SVM-RBF classifier, thanks to its automatic feature detection and parameter optimization process of SVM-RBF.

C. Experiment Settings

In the experiments, we considered the nonlinear SVM based on the popular Gaussian kernel (referred to as SVM-RBF). The related parameters C and γ for this kernel were varied in the arbitrarily fixed ranges $[10^{-3}, 200]$ and $[10^{-3}, 2]$ so as to cover high and small regularization of the classification model, and fat as well as thin kernels, respectively. The experiments are implemented by LIBSVM [17], In addition, for comparison purpose, we implemented, in the first experiment, the SVM classifier with another kernel, which is the linear kernels, leading thus to another SVM

classifiers termed as SVM-Linear. The degree d of the polynomial kernel was varied in the range $[2, 5]$ in order to span polynomials with low and high flexibility. The K value and the number of hidden nodes (h) of the K-nn and the RBF-NN classifiers were tuned in the arbitrarily fixed intervals and $[10, 60]$, respectively. The other RBF parameters, which include the center and the width of each RBF (kernel), were computed by applying the K-means clustering algorithm separately to each class. Concerning the PSO algorithm, we considered the following standard parameters: swarm size $S=40$, inertia weight $w=0.4$, acceleration constants c_1 and c_2 equal to the unity, and maximum number of iterations fixed at 50.

V. EXPERIMENTAL RESULTS

A. Experiment 1: Classification in the Whole Original Hyperdimensional Feature Space

As mentioned earlier, in this experiment, we applied the SVM classifier directly on the entire original hyperdimensional feature space, which is made up of 200 features. During the training phase, the SVM parameters were selected according to am-fold cross-validation (CV) procedure [18], first by randomly splitting the 4757 training beats into 10 mutually exclusive subsets(folds) of equal size, and then, by training m times an SVM classifier modeled with predefined values: C for the linear kernel, C and γ for the Gaussian kernel. Each time we left one of the subsets out of the training, and only used it to obtain an estimate of the classification accuracy. From 10 times of training and accuracy computation, the OA yielded a prediction of the classification accuracy of the considered SVM classifier. We chose the best SVM classifier parameter values to maximize this prediction. In all experiments reported in this paper, we adopted a fivefold CV. The same procedure was adopted to find the best parameters for the K-nn and RBF-NN classifiers. We recall that this empirical parameter estimation procedure and all the classification experiments were repeated three

times, each with one of the three different training sets generated randomly.

As reported in TABLE II, the OA accuracies achieved with the SVM classifier based on the Gaussian kernel (SVM-RBF) on the test set were equal to 92.63%. These results were better than those achieved by the SVM-linear the K-nn and the RBF-NN classifiers. Indeed, the OA accuracies were equal to 88.47% for the SVM-linear classifier, 82.74% for the K-nn classifier, and 85.92% for the RBF-NN classifier. This experiment appears to confirm what was observed in other application fields, i.e., the superiority of SVM based on the Gaussian kernel as compared to traditional classifiers when dealing with feature spaces of very high dimensionality. In addition, it provides reference classification accuracies (95.25%) in order to quantify the capability of the PSO-SVM approach.

B. Experiment 2: Classification Based on Feature Reduction

In this experiment, we train the SVM classifier based on the Gaussian kernel, which proved in the previous experiments to be the most appropriate kernel for hyperspectral classification, in feature subspaces of various dimensionalities. The desired number of features varied from 20 to 200 with a step of 20, namely, from small to high-dimensional feature subspaces. Feature reduction was achieved by the traditional principal component analysis (PCA) algorithm[19], commonly used in hyperspectral classification. It is based on the idea to select the first component (i.e., the direction of maximum variance), then the second component (direction of second maximum variance), and so on, up to the desired number of components, which will compose the considered feature subspace.

Fig. 4 depicts the results obtained in terms of OA by the three considered classifiers combined with the PCA algorithm, namely, the PCA-SVM-RBF, the PCA-RBF-NN, and the PCA-K-nn classifiers. The PCA-SVM-RBF classifier maintains a clear superiority over the other three. Its best accuracy was found using a feature subspace made up of the first 160 components. The corresponding OA were 92.74%. Comparing these results with those achieved with the SVM classifier based on the Gaussian kernel in the original feature space (i.e., without feature reduction), a slight increase of 0.11% in terms of OA. As regards the PCA-K-nn and the PCA-RBF-NN classifiers, the best empirical numbers of features were found to be 80 and 60, respectively. The corresponding OA are 82.78% for the PCA-K-nn classifier, and 88.5% for the PCA-RBF-NN, respectively. Note from Fig.3 and Fig.4 that these classifiers behave much better with 20 features

by PCA feature transform than in the original hyper-dimensional feature space.

From this experiment, we can make three observations: 1) the SVM-RBF classifier shows a relatively low sensitivity to the curse of dimensionality and higher classification accuracy as compared to the SVM-Linear, K-nn and the RBF-NN classifiers[see Fig.3]; 2) the SVM-RBF classifier still preserve its superiority when integrated in a feature reduction-based classification scheme; and 3) though the SVM performs well in the whole original feature space, its accuracy can still be improved provided that a subspace of higher generalization capability can be found.

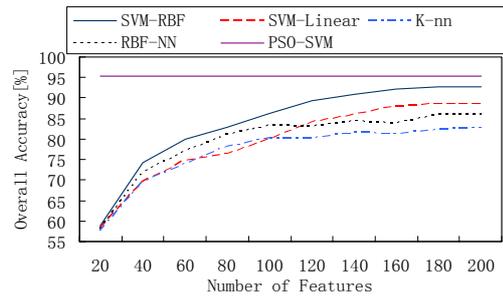


Figure 3. Overall accuracy versus the number of original features

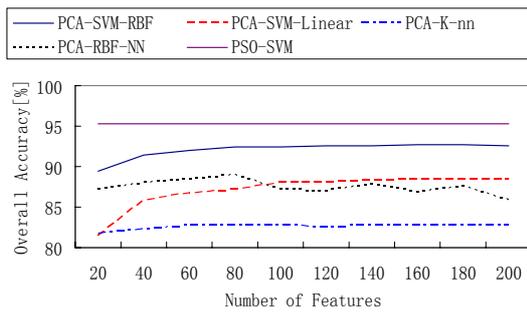


Figure 4. Overall accuracy versus the number of PCA features

C. Experiment 3: Classification With PSO-SVM

As described in Section III, the proposed PSO-SVM classification system aims at enhancing the SVM-RBF classification process from two different viewpoints [20]: 1) by automatically detecting a feature subspace of higher generalization capability in order to deal in a more effective way with the curse of dimensionality, instead of reducing the dimension of the original feature space basing on PCA and 2) by passing from an empirical tuning of the value of the two SVM parameters to their automatic optimization. This experiment is aimed at assessing the effectiveness of this methodological

TABLE II BEST OVERALL AND CLASS-BY-CLASS ACCURATES BY DIFFERENT CLASSIFIERS IN THE ORIGINAL HYPERSPECTRAL SPACE

Method	CLASSIFICATION ACCURACY[%]									OA[%]
	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	
SVM-Linear	88.29	76.53	96.62	99.16	99.6	62.37	89.37	90.94	99.22	88.47
SVM-RBF	91.91	81.07	96.62	99.44	99.6	68.19	94.94	96.12	99.53	92.63
K-nn	77.02	68.88	92.83	99.16	99.21	66.32	86.1	57.61	95.96	82.74
RBF-NN	97.44	73.51	87.67	78.13	99.21	98.04	91.98	73.72	80.06	85.92

enhancement. To this purpose, we apply the PSO-SVM classifier to the available training data. Note that each particle of the swarm was defined by position and velocity vectors of a dimension of 202. At convergence of the optimization process, we assessed the PSO-SVM-RBF classifier accuracy on the test samples. The achieved overall accuracy is 95.25% corresponding to substantial accuracy gains as compared to what is yielded either by the SVM classifier (with the Gaussian kernel) applied to all available features (+2.62%) or by the PCA-SVM classifier. Moreover, by means of this approach, the average subset feature number is 120, which is fewer than original feature number 220. The whole process is automatic and without user's interface.

In the experiments, when using the proposed intelligent optimization methods, we considered the nonlinear SVM based on the popular Gaussian kernel (referred to as SVM-RBF). The related parameters C and γ for this kernel were varied in the arbitrarily fixed ranges $[10^{-3}, 300]$ and $[10^{-3}, 3]$, so as to cover high and small regularization of the classification model, and fat as well as thin kernels, respectively. The experiments are implemented by LIBSVM [17].

LIBSVM is widely used in SVM classifier, but the value of RBF kernel parameters is always difficult to define. The default values are as follows: C is 1, and γ is the reciprocal of the dimension. In our experiment, the dimension is the band number, so the parameter value of γ is 0.005. In the same way, the default value of C of SVM parameter in ENVI is 100, and γ is the reciprocal of the dimension. In our experiment, the dimension is the band number, so the parameter value of γ is 0.005. In addition, we also select SVM parameters by grid algorithm. In grid algorithm, according to reference [17], the range of C and γ is $[2^{-5}, 2^{15}]$ and $[2^{-15}, 2^3]$, the step length is 2^2 .

Concerning the PSO algorithm, we considered the following standard parameters: swarm size $S=20$, inertia weight $w=1$, acceleration constants c_1 and c_2 equal to 2, and maximum number of iterations fixed at 300. The parameters setting is summarized in Table III.

Table III Parameters setting of different methods

PSO-SVM		Grid Algorithm	
Parameter	Value	Parameter	Value
Swarm size	20	C	$[2^{-5}, 2^{-3}, \dots, 2^{15}]$
Number of generations	300	γ	$[2^{-15}, 2^{-13}, \dots, 2^3]$
V_{max}	4		
C_1, C_2	2		

In addition, for comparison purpose, we implemented the three traditional methods and our two intelligent optimization methods for classification. The experimental comparison results are shown in Figure 5 and TABLE IV.

Table IV illustrates the classification of different classifiers. As can be seen our proposed classifier have better accuracy than traditional classifiers. The LIBSVM default setting lead to the lowest accuracy, which is 52.79%. The best percentage of classification, is 95.25%

by PSO-SVM method. The results still confirm the strong superiority of our proposed PSO-SVM over the other classifiers, with a gain in overall accuracy +12.80% and +4.27% with respect to the default SVM classifier in ENVI and the grid algorithm classifiers (see TABLE IV).

From the obtained experimental results, we conclude the proposed PSO-SVM classifier has the best classification accuracy account of its superior generalization capability as compared to traditional classification techniques.

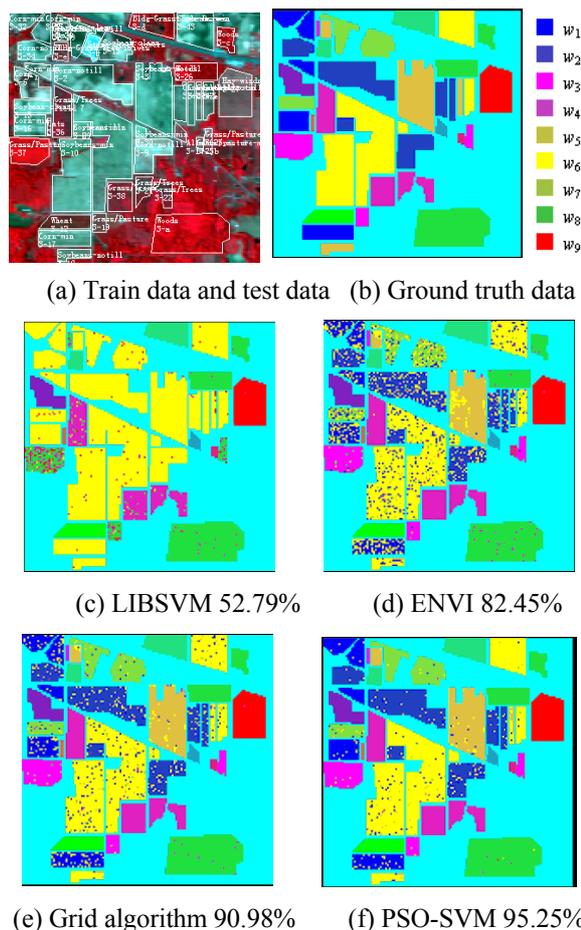


Figure 5. Classification accuracy of different classifiers

The Grid search algorithm is a common method for searching for the best C and γ . In the Grid algorithm, pairs of (C, γ) are tried and the one with the best cross-validation accuracy is chosen. After identifying a 'better' region on the grid, a finer grid search on that region can be conducted. This research conducted the experiments using the proposed PSO-SVM approaches and the Grid algorithm. The results from the proposed method were compared with that from the Grid algorithm. In all of the experiments 10-fold cross validation is used to estimate the accuracy of each learned classifier.

We apply the PSO-SVM classifier to the available training data. Note that each particle of the swarm was defined

Table IV Classification result by different parameter selection methods

Methods of selecting parameters	C	γ	Band num	Classification accuracy(%)
LIBSVM default	1	0.005	200	52.79
ENVI default	100	0.005	200	82.45
10-cross validation (grid search algorithm)	8	2	200	90.98
PSO-SVM	223.32	0.9696	120	95.25

VI. SPECTRAL AND WAVELET FEATURE SELECTION

From the perspective of signal processing ,the hyperspectral curve of each pixel can be thought as a one-dimensional (1-D) vector.By using the wavelet transform, the hyperspectral vector is transformed from the spectral space to the time-scale space. The wavelet analysis involves two compounds: approximations and details. For 1-D wavelet decomposition, starting from original hyperspectral vector s , the first step produces two sets of coefficients: approximation coefficients (scaling coefficients) a_1 and detail coefficients (wavelet coefficients) d_1 . These coefficients are computed by convolving original hyperspectral vector with the low-pass filter for approximation and with the high-pass filter for detail. The convolved coefficients are down-sampled by keeping the even indexed elements. Then, the approximation coefficients a_1 are split into two parts by using the same algorithm and are replaced by a_2 and d_2 , etc. This decomposition process is repeated until the required level is reached. The coefficient vectors are produced by down-sampling and are only half the length of the signal or the coefficient vector at the previous level. Conversely, approximations and details are constructed inverting the decomposition step by inserting zeros and convolving the approximation and detail coefficients with the reconstruction filters. Fig.6 shows a wavelet decomposition tree at level 3. Notice that the approximation preserves most of the trend of original spectrum, in our study, we select some approximate coefficient as wavelet feature. we find that when the wavelet decomposition level is 3,the approximate coefficient number is 25, and the classification is higher than wavelet decomposition level 2.

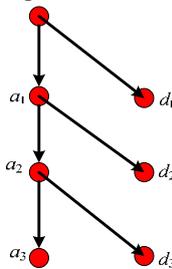


Figure 6. Wavelet decomposition tree at three levels

Figure 7 represent the spectral and wavelet feature solution. In the new feature space, when using our propsoed PSO approach to select best feature subset and parameters of SVM.First, for the feature selection, if n features are required to decide which features are chosen, then $M+N+2$ decision variables must be adopted. The value of $M+N$ variables ranges between 0 and 1. If the value of a variable is less than or equal to 0.5, then its corresponding feature is not chosen. Conversely, if the value of a variable is greater than 0.5, then its corresponding feature is chosen. Fig. 9 illustrates the solution representation. C is penalty cost, γ represents a parameter of RBF-SVM kernel function. a_n =Feature n is selected.

Spectral features(N bit)		Wavelet features(M bit)			C	γ
1	N	N+1	D+M	D+M+1
a_1	a_N	T_1	T_M	C
						γ

Figure 7. The representation of spectral 、 wavelet feature of image and parameters C and γ of SVM.

Fig.8 represent the performance process of PSO-SVM in spectral and wavelet feature space. In the flow char, $Y, W_{tr}, W_{te}, P_{tr},$ and P_{te} represent the original feature vector, spectral and wavelet feature vectors for training and test samples, and selected PSO feature vectors for training and test samples. $N, N_{tr}, N_{te}, d_{ori}, d_w,$ and d_{ps} represent the number of whole samples , training samples and test samples , dimension number of the original data vector , dimension number of spectral features and wavelet approximate coefficients at the third level,and the number of selected PSO features, respectively.The PSO and SVM parameters setting is the same as section V.From TABLE V,we can see that the classification accuray is 97.48% by PSO-SVM in spectral and wavelet feature space,the obtained result is 2.23% increase than classification accuracy (95.25%) in only spectral feature spaces.

Table V. Classification result by different feature selection methods

Feature space	C	γ	feature num	Classification accuracy(%)
spectral	223.32	0.9696	120	95.25
spectral+wavelet features	243.56	0.8945	125	97.48

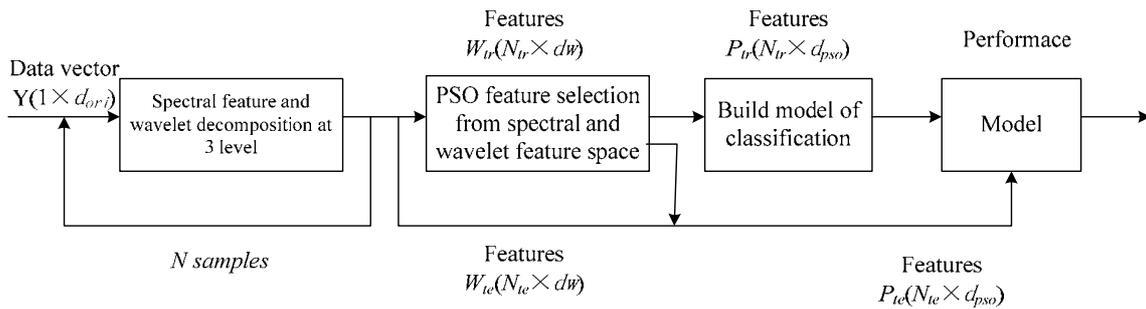


Figure 8. The process of PSO feature selection and classification

VII. CONCLUSION

From the obtained experimental results, we can strongly recommend the use of the SVM approach for classifying hyperspectral remote sensing image on account of their superior generalization capability as compared to traditional classification techniques. This capability generally provides them with higher classification accuracies and a lower sensitivity to the curse of dimensionality compared with traditional classifiers. First, we present a thorough experimental study to show the superiority of the generalization capability of the support vector machine (SVM) approach in the hyperspectral classification of remote sensing image.

Second, this study also presents a particle swarm optimization-based approach, capable of searching for the optimal parameter values for SVM, simultaneous obtaining a subset of beneficial features. This optimal subset of features is then adopted in both training and testing to obtain the optimal outcomes in classification. Comparison of the obtained results with those of other approaches demonstrates that the developed PSO-SVM approach has a better classification accuracy than others tested.

Third, traditional classifications of hyperspectral remote sensing have made use of spectral information mainly. Wavelet feature based on discrete wavelet transform are proposed to classify hyperspectral remote sensing images. Aim at the condition, the paper proposes PSO algorithms to optimize feature selection in wavelet features and spectral features. Compared with traditional spectral feature space classification, the new feature space has larger features information. Experiment indicates the classification method which combines wavelet feature and spectral feature can get better results than only using spectral feature classification.

ACKNOWLEDGMENT

This research was supported by National Natural Science Foundation of China under Grant NO. 61003127 and Hubei Provincial Natural Science Foundation of China under Grant No. 2010CDA090.

REFERENCES

[1] Vapnik Vladimir N. The Nature of Statistical Learning Theory[M]. Springer Verlag, New York, Inc., 2000.

[2] H. Liu and H. Motoda, "Feature Selection for knowledge discovery and data mining," Boston: Kluwer Academic, 1998.

[3] R. C. Chen and C. H. Hsieh, "Web page classification based on a support vector machine using a weighed vote schema," *Expert Systems with Applications*, vol.31, pp.427-435, 2006.

[4] C. Gold, A. Holub and P. Sollich, "Bayesian approach to feature selection and parameter tuning for support vector machine classifiers," *Neural Networks*, vol. 18, pp.693-701, 2005.

[5] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol.97, pp. 273-324, 1997.

[6] T. Shon, Y. Kim and J. Moon, "A machine learning framework for network anomaly detection using SVM and GA," In *Proceedings of IEEE Workshop on Information Assurance and Security 2*, pp. 176-183, 2005.

[7] L. Zhang, L. Jack and A. K. Nandi, "Fault detection using genetic programming," *Mechanical Systems and Signal Processing*, pp.271-289, 2005.

[8] B. Samanta, K. R. Al-Balushi and S. A. Al-Arjami, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol.16, pp.657-665, 2003.

[9] C. L. Huang, M. C. Chen and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications*, vol.33, no.4, pp 847-856, 2007.

[10] C. L. Huang and C. L. Wang, "A GA-based feature selection and parameters optimization for support vector machines," *Expert Systems with Applications*, vol.31, no.2, 231-240, 2006.

[11] C. W. Hsu, C. C. Chang and C. J. Lin, "A practical guide to support vector classification". Technical report, University of National Taiwan, Department of Computer Science and Information Engineering, July, pp. 1-12, 2003.

[12] P. F. Pai and W. C. Hong, "Support vector machines with Simulated annealing algorithms in electricity load forecasting," *Energy Conversion and Management*, vol.46, pp.266-2688, 2005.

[13] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote sensing*, vol.42, no.8, pp.1778-1790, 2004.

[14] G. M. Foody and A. A. Mathur, "relative evaluation of multiclass image classification by support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol.42, no. 6, pp.1335-1343, 2004.

[15] Kennedy Y J, Eberhart R. Particle swarm optimization. In: *IEEE International Conference on Neural Networks*. Perth: IEEE Neural Networks Society, 1995: 1942-1948.

[16] Aviris Indiana's IndianPines1 DataSet[Online] ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.lan(originalfiles)ftp://

- /ftp.ecn.purdue.edu/biehl/PCMultiSpeeThyFiles.zip(ground truth)
- [17] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
 - [18] Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification. Technical report, University of National Taiwan, Department of Computer Science and Information Engineering, July, pp. 1 - 12.
 - [19] Jackson, J. E., A User's Guide to Principal Components, John Wiley and Sons, Inc. 1991. pp. 1-25.
 - [20] Frohlich, H., & Chapelle, O. (2003). Feature selection for support vector machines by means of genetic algorithms. Proceedings of the 15th IEEE international conference on tools with artificial intelligence, Sacramento, CA, USA pp. 142-148.

Sheng Ding received the B.S. and M.S. degrees in computer sciences from China Center Normal University, Wuhan University of Science and Technology, Wuhan, China, in 1998 and 2005, respectively, and the Ph.D. degree in sciences and techniques of remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, in 2010.

He is currently associate professor in College of Computer Science and Technology, Wuhan University of Science and Technology. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote-sensing applications.