A Method of Particle Swarm Optimized SVM Hyper-spectral Remote Sensing Image Classification

Q J Liu, L H Jing, L M Wang and Q Z Lin
Institute of Remote Sensing and Digital Earth, CAS, No.9 Dengzhuang South Road, Haidian District, Beijing, China
Email:qjliu@ceode.ac.cn

Abstract. Support Vector Machine (SVM) has been proved to be suitable for classification of remote sensing image and proposed to overcome the Hughes phenomenon. Hyper-spectral sensors are intrinsically designed to discriminate among a broad range of land cover classes which may lead to high computational time in SVM multi-class algorithms. Model selection for SVM involving kernel and the margin parameter values selection which is usually time-consuming, impacts training efficiency of SVM model and final classification accuracies of SVM hyper-spectral remote sensing image classifier greatly. Firstly, based on combinatorial optimization theory and cross-validation method, particle swarm algorithm is introduced to the optimal selection of SVM (PSSVM) kernel parameter $\sigma$ and margin parameter $C$ to improve the modelling efficiency of SVM model. Then an experiment of classifying AVIRIS in India Pine site of USA was performed for evaluating the novel PSSVM, as well as traditional SVM classifier with general Grid-Search cross-validation method (GSSVM). And then, evaluation indexes including SVM model training time, classification Overall Accuracy (OA) and Kappa index of both PSSVM and GSSVM are all analyzed quantitatively. It is demonstrated that OA of PSSVM on test samples and whole image are 85% and 82%, the differences with that of GSSVM are both within 0.08% respectively. And Kappa indexes reach 0.82 and 0.77, the differences with that of GSSVM are both within 0.001. While the modelling time of PSSVM can be only 1/10 of that of GSSVM, and the modelling. Therefore, PSSVM is an fast and accurate algorithm for hyper-spectral image classification and is superior to GSSVM.

1. Introduction
In recent years, the support vector machine (SVM) classifier has been proposed to overcome the Hughes phenomenon in hyper-spectral remote sensing image (HSRI) classification [1]. The SVM is based on statistical learning theory as proposed by Vapnik and Chervonenkis and discussed in detailed by Vapnik [2]. The aim in binary SVM is to find the optimal linear hyperplane which separates the two classes so that it creates maximum distance between the boundary training samples and the separating hyperplane. The class separation method and using kernel space make this classifier a powerful tool for classification in many applications.

Although SVM is suitable for classification of hyper-spectral images, these images usually have large amounts of data with many classes that may lead to high computational time in multiclass algorithms. Classification of such images by SVM needs a great deal of computational time. Therefore the inefficiency of multiclass SVM from the computational point of view and the efficiency of SVM from an accuracy point of view, are considerable in hyper-spectral image classification. Besides,
Model selection for SVM involving kernel and the margin parameter values selection which is usually time-consuming with traditional grid-search (GSSVM) method, impacts training efficiency of SVM model and final classification accuracies of SVM hyper-spectral remote sensing image classifier greatly. Although the relationship between these two parameters of SVM model obviously existed, their mathematical description is not clear.

In order to solve the above problems, this paper introduces a particle swarm intelligence algorithm for hyper-spectral SVM classifier model selection (PSSVM). “Swarm intelligence” is a newly developed direction in artificial intelligence. It means that individuals with simple intelligence behave complex intelligence through exchanges and cooperation [3]. Particle swarm optimization (PSO) has roots in two main component methodologies [4]. Perhaps more obvious are its ties to artificial life in general, and to bird flocking, fishing schooling, and swarming theory in particular. It is also related to evolutionary computation, and has ties to both genetic algorithms and evolutionary programming. The PSSVM has been found to be robust in solving problem of multiple optimal selections of model parameters of hyper-spectral remote sensing SVM classifier.

2. Methods

2.1. Traditional grid-search SVM

A simple scenario for common classification by SVM is when the classes are linearly separable. This scenario may be illustrated with the training data set comprising k cases and be represented by \( \{x_i, y_i\} \), \( i = 1, \ldots, k \), where \( x \in \mathbb{R}^N \) is an N-dimensional space and \( y \in \{-1, +1\} \) is the class label. The classification pattern is linearly separable if there exists a vector \( w \) and a scalar \( b \) such that

\[
y_i(w \cdot x_i + b) - 1 \geq 0
\]

The hypothesis space can be defined by the set of functions by

\[
f_{w,b}(x) = \text{sign}(w \cdot x + b)
\]

The separating hyperplanes which make the distance between the classes maximized can be achieved by the following constrained optimization problem:

\[
\min_{\omega, b, \xi} \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{k} \xi_i
\]

(3)

For nonlinear classes, a “slack variable” \( \xi \geq 0 \) should be introduced to relax the optimal hyperplane on the same side of the training class, to reflect the quantity proportional the number of misclassification errors in the process of maximizing the margin of hyperplane. This tradeoff between margin and misclassification error is controlled by a positive constant \( C \) such that \( \infty > C > 0 \). Thus, formula (3) can be written as

\[
\min_{\omega, b, \xi} \left[ \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{k} \xi_i \right]
\]

(4)

Besides, the feature vector \( x \) is usually mapped into a higher dimensional Euclidean Space \( F \) via a nonlinear vector \( \Phi \), with a concept of using a kernel function \( K \) in the design of nonlinear SVM. A kernel function is defined as \( K = \Phi(x_i) \cdot \Phi(x_j) \). Then formula (2) becomes

\[
f(x) = \text{sign} \left( \sum_{j=1}^{k} \lambda_j y_j K(x_j, x) + b \right)
\]

(5)

Where \( \lambda_j \) is a Lagrange multiplier. One common basic kernel function is radial basis function (RBF): \( K = \exp(-\gamma \|x-x_j\|^2) \), \( \gamma > 0 \).

SVM was originally developed to solve binary problems [5]. The implementation of SVM in multiclass classification problems could be done by combining several binary SVMs to achieve multiclass SVM. Detailed experiments and comparison of these kinds of methods and kernel functions suggest that one-against-one (OAO) and RBF are more practical for real applications [6].

For the RBF kernel used in this work, there are two parameters that need to be defined by uses, regularization parameter \( C \) and kernel width \( \gamma \). The grid-search method using cross-validation
approach is usually used to automatically determine these two parameters in SVM classification. To avoid doing an exhaustive search, a coarse grid search is performed first to identify a target region on the grid, and a finer grid search on the region is then conducted until the exact region is identified. In practice, GSSVM for hyper-spectral remote sensing classification is still time-consuming, especially when the image size is large. AVIRIS hyper-spectral remote sensing of Indian Pine USA was classified with GSSSVM. Its \((C, \gamma)\) contour values and elevation model are shown in Figure 1.

![Elevation Model of \(C\) and \(\gamma\)](image)

Figure 1. Contour value and Elevation model of \((C, \gamma)\)

2.2. Novel particle swarm optimized SVM for HSRI classification

Based on combinatorial optimization theory and cross-validation method, particle swarm algorithm is introduced to the optimal selection of SVM (PSSVM) kernel parameter \(\sigma\) and margin parameter \(C\) to improve the training efficiency of SVM model.

Particle swarm optimization (PSO) comprises a very simple concept, and paradigms are implemented in a few lines of computer code. In PSO, interactions among particles (members of the population) are governed by the so called cognitive learning part and social learning part of each particle. The cognitive learning part relates to a particle’s own experience and the social learning part relates to positive experiences of others. The idea behind PSO is simple and yet PSO has been applied to many problem domains successfully. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating each generation. At every iteration, each particle is the location of the best solution (fitness) a particle has achieved so far. This value is called pBest. Another “best” value is the location of the best solution that any neighbors, the best location is a neighborhood best and called nBest. When a particle takes all the population as its neighbors, the best location is a global best and is called gBest [7].

The particle swarm works by adjusting trajectories through manipulation of each coordinate of particle [8]. The core of PSO is the updating formulae of the particle. Equation (6) calculates a new velocity for each particle based on its previous velocity. And equation (7) updates each particle’s position in the solution hyperspace.

\[
V_{id}^{k+1} = \omega V_{id}^{k} + c_1 r_1 (P_{id}^{k} - X_{id}^{k}) + c_2 r_2 (P_{gd}^{k} - X_{id}^{k})
\]

\[
X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1}
\]

Where, \(\omega\) is inertia weight, \(c_1\) and \(c_2\) are two learning factors, \(r_1\) and \(r_2\) are random numbers between [0, 1] independently generated.

For hyper-spectral SVM classifier, an empirical coarse range of \((C, \gamma)\) is firstly given to set the original parameters of PSO, including the maximum and minimum value of particle and velocity. The fitness is supplied by 5-cross-validation approach.
3. Experimental Results
The Indian Pines data is gathered by a sensor known as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (figure 2-a). These data are obtained from an aircraft flown at 19812m altitude and operate by the National Aeronautics and Space Administration/Jet Propulsion Laboratory, with a size of $145 \times 145$ pixels and 220 spectral bands measuring approximately 20 m across on the ground [9]. From the 16 different land-cover classes available in the original ground truth, six classes, namely, Corn-notill, Grass/Pasture, Grass/Tree, Hay-windrowed, Soybeans-min and Wood, are selected for PSSVM classification experiment, and the number of samples of each class is displayed in Table 1.

![Figure 2. (a) AVIRIS of Indian Pine and (b) Distribution of ground objects](image)

Both GSSVM and PSSVM are used for AVIRIS HSRI classification experiment to compare the performance of grid-search and particle swarm optimized searching of parameters of HSRI classifier. The experiment was performed with MATLAB in Microsoft Windows 7 professional operating system. The maximum iteration generations of PSSVM is $\{5, 6, 7, 8, 9, 10, 15, 20, 25\}$, which are used one by one for the HSRI classification. The quantitative indexes including Modelling time (T), overall accuracy of test samples pixels (OAS), overall accuracy of images pixels (OAI), Kappa index of test samples pixels (KS), and Kappa index of images pixels (KI) are adopted to evaluate the performance of GSSVM and PSSVM HSRI classification, which is shown in Table 2. When iteration generations increases from 5 to 25 one by one, $T_{\text{PSSVM}}$ rises from 137s to 675s, while $T_{\text{GSSVM}}$ rises from 1413s to 1713s, and the ratio of $T_{\text{PSSVM}}/T_{\text{GSSVM}}$ decrease from 1/10 to 1/3 , So $T_{\text{PSSVM}}$ is obviously sensitive to the iteration generations. In the process, the mean values $m$ of OAS$_{\text{PSSVM}}$, OAS$_{\text{GSSVM}}$, OA$_{\text{PSSVM}}$, OA$_{\text{GSSVM}}$, KS$_{\text{PSSVM}}$, KS$_{\text{GSSVM}}$, KI$_{\text{PSSVM}}$ and KI$_{\text{GSSVM}}$ are 0.866, 0.866, 0.826, 0.826, 0.826, 0.839, 0.841 and 0.774, while the standard deviation values $\sigma$ of them are 0.012, 0.01, 0.005, 0.005, 0.013, 0.014, 0.07 and 0.005 correspondingly. The OAs and Kappa indexes of PSSVM and GSSVM make very little difference and fluctuate tiny when generations rises from 5. Therefore, the modelling time of PSSVM can be only 1/10 of that of GSSVM, and the modelling and training efficiency of SVM have been greatly improved.

4. Conclusion
In this paper, a novel particle swarm optimized SVM has been proposed for the model parameters selection of SVM in hyper-spectral remote sensing images classification. Compared to original grid-
search SVM, PSSVM shows better optimized parameters selection efficiency, and at the same time have the robust ability to keep final classification accuracies. In conclusion, the proposed PSSVM obviously improves traditional SVM hyper-spectral remote sensing classifier, and accommodates kinds of practical HSRI classification applications.

Table 2. Quantitative indexes of HSRI classification with PSSVM and GSSVM at different generations

| Gens | T_PSSVM | T_GSSVM | OAS_PSSVM | OAS_GSSVM | OAI_PSSVM | OAI_GSSVM | KS_PSSVM | KS_GSSVM | KI_PSSVM | KI_GSSVM |
|------|---------|---------|-----------|-----------|-----------|-----------|----------|----------|----------|----------|
| 5    | 137     | 1413    | 1/10      | 0.87      | 0.86      | 0.82      | 0.84      | 0.83      | 0.77      | 0.77      |
| 6    | 1517    | 1356    | 1/9       | 0.87      | 0.86      | 0.83      | 0.83      | 0.84      | 0.78      | 0.78      |
| 7    | 173     | 1416    | 1/8       | 0.87      | 0.87      | 0.83      | 0.83      | 0.85      | 0.78      | 0.78      |
| 8    | 195     | 1430    | 1/7       | 0.86      | 0.86      | 0.82      | 0.82      | 0.83      | 0.77      | 0.77      |
| 9    | 236     | 1496    | 1/6       | 0.86      | 0.86      | 0.82      | 0.82      | 0.84      | 0.77      | 0.77      |
| 10   | 238     | 1408    | 1/6       | 0.89      | 0.89      | 0.83      | 0.83      | 0.86      | 0.78      | 0.78      |
| 15   | 337     | 1327    | 1/4       | 0.85      | 0.86      | 0.82      | 0.82      | 0.82      | 0.76      | 0.77      |
| 20   | 408     | 1312    | 1/3       | 0.87      | 0.87      | 0.83      | 0.83      | 0.85      | 0.78      | 0.78      |
| 25   | 675     | 1713    | 1/3       | 0.85      | 0.86      | 0.83      | 0.83      | 0.83      | 0.78      | 0.78      |

Acknowledgments
This research was supported in part by National Natural Science Foundation of China under Grants 41001266, the National Key Technology R&D Program (Grand No.2012BAC16B01),100 Talents Program of The Chinese Academy of Sciences under Grants Y2ZZ03101B, and National Science & Technology Pillar Program during the Twelfth Five-year Plan Period of China under Grants 2012BAH33F00.

References
[1] Giorgos M, Jungho I M, Caesar O 2011 Support vector machines in remote sensing: a review, *ISSPRS Journal of Photogrammetry and Remote Sensing*, 66 247-259
[2] Hseini S A, Ghassemian H 2011 A new fast algorithm for multiclass hyperspectral image classification with SVM, *International Journal of Remote Sensing*, 32(23) 8657-8683
[3] Liu X, Li X, He J 2008 Classification of remote sensing images based on ant colony optimization”, *Journal of Remote Sensing*, 12(2) 253-262
[4] Eberhart R C, Kennedy J 1995 A new optimizer using particle swarm theory, *Proc. Sixth International Symposium on Micro Machine and Human Science, IEEE Service Center*, 39-43
[5] Shigeo A. 2010 Support vector machines for pattern classification (Advances in pattern recognition) (Second Edition), Springer
[6] He L M,Sen Z Q, Kong F S 2007 Study on multi-source remote sensing images classification with SVM, *Journal of Image and Graphics*, 12(4) 648 -654
[7] Xiaohui H, Yuhui S, Eberhart R 2004 “Recent advances in particle swarm”, *Proceeding of the IEEE Congress on Evolutionary Computation*, 90-96
[8] Kennedy J, Eberhart R C 1997 A discrete binary version of the particle swarm algorithm, *IEEE International Conference of on Computational Cybernetics and Simulation*, 5 4101-4108
[9] Landgrebe D 2003 *Signal theory method in multispectral remote sensing* (Hoboken, NJ:Wiley)