High resolution remote sensing image classification based on particle swarm optimization and support vector machine

Buyi Li, Chongjing Deng, Shuang Li*

International School of Software, Wuhan University, Luoyu Road 37#, Wuhan, China, 430079

**Corresponding author e-mail: sli@whu.edu.cn*

Received 1 November 2014, www.cmnt.lv

Abstract

Many algorithms have been developed for image classification and support vector machine (SVM) is a kind of supervised classification that has been widely used recently. However, the accuracy of a SVM classifier heavily depends on the selection of a right kernel model and appropriate parameter. In this paper, a comparative analysis of the impact of four kernels (linear kernel, polynomial kernel, radial basis function kernel and sigmoid kernel) on the accuracy of SVM classifiers is conducted. Moreover, the Particle Swarm Optimization (PSO) is used to search for the optimum parameters for each kernel function in order to improve the classification accuracy of SVM classifiers. Our experiments for optimizing the kernel function parameters and assessing the robustness of SVM classifiers were carried out with classifications of QuickBird-2 images over Wuhan, China for monitoring urban land cover/land use information. The experimental results indicate that the polynomial kernel outperforms the other kernels in classifying high resolution remote sensing image. The sigmoid kernel performs worse than any other kernels. Our findings also suggest that selected parameter by PSO will improve the classification accuracy, especially for radial basis function kernel.

Keywords: high resolution remote sensing image, support vector machine classification, parameter optimization, particle swarm optimization

1 Introduction

The classification of land use and land cover (LULC) from remotely sensed imagery is a challenging topic due to the complexity of landscapes. Numerous classification algorithms have been proposed especially since more and more remote sensing images with various spatial and spectral resolutions are sent back to the earth. Among the most popular algorithms, Support Vector Machine (SVM) is a new machine learning method based on statistical learning theory, which can solve the classification problem with small sampling, non-linear and high dimensions [1].

It is well-known that the performance of SVM depends on the training features, kernel type and its corresponding parameters [2-4]. The kernel function in SVM is used to convert non-linear separating boundaries into linear ones by mapping the input data into a high-dimensional space. Thus, determine the kernel type and kernel parameters are important for image classification accuracy. There are many kinds of support vector kernels such as the linear kernel, the polynomial kernel, the radial basis function kernel, etc. For the kernel type selection, Pal (2002) suggested that the radial basis function kernel achieved higher accuracy than linear kernel, polynomial kernel and the sigmoid kernel [5]. Villa et al. (2008) concluded that polynomial kernel outperformed the Gaussian Kernel in remote sensing image classification [3]. Kavzoglu and Colkesen (2009) indicated that radial basis function kernel performed better than polynomial kernel in land cover classification [4]. However, they also indicated that further research should be conducted on the effects of kernel type and their parameters on classification accuracy.

For kernel parameter optimization techniques, the traditional way is grid search with cross validation. However, the grid search is time consuming as the model needs to be evaluated at many grid points for each parameter set. In recent years, the artificial intelligent algorithms are employed in SVM parameter optimization, i.e. genetic algorithm (GA), simulated annealing (SA), particle swarm optimization (PSO). The GA updates the population by crossover and mutation operations to generate optimal parameters. The SA technique can also be applied to ensure that the global optimum of parameter combinations. However, these methods obtain the optimal parameters from the population evolution iteratively, which require much training time in SVM classifier. Inspired by social behavior of bird flocking or fish schooling, PSO is proposed by Kennedy and Eberhart in 1995 [6]. Through the competition and collaboration among the population, each particle in the swarm can dynamically adjust its velocity according to its own and its companion's experience and finally can find the best position to land. Compared with other intelligent algorithms, PSO demonstrates its high efficiency, easy implement and powerful both global and local exploration abilities in parameter optimization in support vector machine [7-12]. However, the ACO algorithm is only used to optimize the RBF kernel. According to the above analysis, in this research, the proposed PSO-SVM model is applied for classification of remote sensing image from Quickbird-2 sensor, in which PSO is used to determine optimized parameters of support vector machine with different kernels. The remainder of this paper is organized as follows. Section 2 describes the basic idea of support vector machine and Section 3 introduces the recommended PSO and the optimization procedure for SVM kernel parameters. Section 4 testifies the performance of the proposed method and presents the analysis for the experimental results. Finally, conclusions are made in Section 5.

2 Support vector machines and its kernels

Consider data set

$$\{(x_1, y_1),...,(x_i, y_i),...,(x_N, y_N),\ y_i \in (1,-1)\},$$

where N is the number of samples, x_i is the training sample, y_i is the class label of x_i. Optimum hyper plane is used to maximize the margin between classes.

The hyper plane is defined as

$$w \cdot x + b = 0 \tag{1}$$

where x is a point lying on the hyper plane, w determines the orientation of the hyper plane, b is the bias that indicates the distance between hyper plane and the origin. For the linearly separable case, the hyper plane is defined as

$$y_i \left(w \cdot x_i + b \right) \geq 1 \tag{2}$$

As the margin width between both bounding hyperplanes equals to $2/\left(\|w\|^2\right)$, the constraint optimization model of soft margin based SVM is as follows:

$$\min_{w,b,\xi} \frac{1}{2} w^2 + c\sum_{i=1}^{l} \xi_i$$
$$s.t. \;\; y_i\left(\left(w\cdot x_i\right)+b\right) \geq 1-\xi_i;\ \xi_i \geq 0,\ i=1,2,...,l \tag{3}$$

where c is the penalty parameter which allows striking a balance between two competing criteria of margin maximization and error minimization, whereas ξ_i is the slack variable which indicate the distance of the incorrectly classified points from the optimal hyper plane. The larger the c value, the higher the penalty associated to misclassified samples.

To solve non-linear classification tasks, a nonlinear function $\phi(x)$ is usually employed to map the input space to a higher dimensional feature space. Thus, the input point x can be represented by $\phi(x)$ in high-dimensional space. The time-consuming computation of $\phi(x)\cdot\phi(x_i)$ is reduced by using a kernel function $K\left(x_i, x_j\right) = \phi\left(x_i\right)^T \phi\left(x_j\right)$. Thus, the classification decision function is defined as:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K\left(x_i \cdot x\right) + b\right) \tag{4}$$

where $\operatorname{sgn}(\cdot)$ is the sign function, $K(\cdot)$ is the kernel function and the magniude of α_i is Lagrange multiplier. A multiplier exits for each training data instance and data instances corresponding to non-zero α_i are support vectors.

The typical SVM kernels include linear kernel function, polynomial kernel function, radial basis kernel function and sigmoid kernel function. They are defined as follows:

Linear kernel function

$$K(x, x') = (x, x')$$

Polynomial kernel function

$$K(x, x') = \left(\gamma(x, x') + r\right)^d, \gamma > 0$$

Radial basis function

$$K(x, x') = e^{\gamma\|x-x'\|^d}, \gamma > 0$$

Sigmoid kernel function

$$K(x, x') = \tanh\left(\gamma(x, x') + r\right), \gamma > 0$$

Generally d is set to be 2 since the kernel value is related to the Euclidean distance between the two samples [13]. r is set to be 0 [14]. For the linear kernel function, only the penalty parameter c in SVM is needed for optimization. For the polynomial kernel function, radial basis function and sigmoid kernel function, the parameters (c, γ) should be set properly. c is the penalty parameter and γ is related to the kernel width.

3 Parameter optimization by particle swarm intelligent

In standard PSO algorithm, the particle swarm starts with the random initialization of a population, and each particle in the search space is characterized by two factors: its velocity and position. The velocity and position vectors of the particle $i(i=1,2,...,n)$ in d-dimensional space can be represented as $v_i = \left(v_{i1}, v_{i2}, ..., v_{id}\right)$ and $x_i = \left(x_{i1}, x_{i2}, ..., x_{id}\right)$, respectively. Then, the new velocity and position of particle i for the next generation in d-dimensional subspace is calculated as follows:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1\left(p_{best}(t) - x_i(t)\right) + c_2 r_2\left(g_{best}(t) - x_i(t)\right) \tag{5}$$

$$x_i(t+1) = x_i(t) + v_i(t) \tag{6}$$

where $v_i(t)$ represents the previous velocity and its value is limited in the range of $\left[-V_{\max}, V_{\max}\right]$. $p_{best}(t)$ is the particle's personal best position obtained so far at t *-th* generation and this part encourages the particles to move toward their own best position found so far. $g_{best}(t)$ is the global best position obtained so far by all particles and this part always pulls the particles toward the global best particle. c_1 and c_2 are constants known as acceleration coefficients which determine the relative influence of the social and cognition components. r_1 and r_2 are two independent random number uniformly distributed in the range of (0, 1). ω is the inertia weight that controls the impact of particle's previous velocity on its current generation.

The fitness function is used to guide the direction of search. As the classification accuracy is the object of our study, the recognition rate (RR) is used for fitness function, which is defined as follows:

$$RR = \frac{n_{correct}}{n_{total}} \times 100\% \tag{7}$$

where n_{corect} is the number of corrected classified samples, n_{total} is the total number of samples.

4 Experiments and results

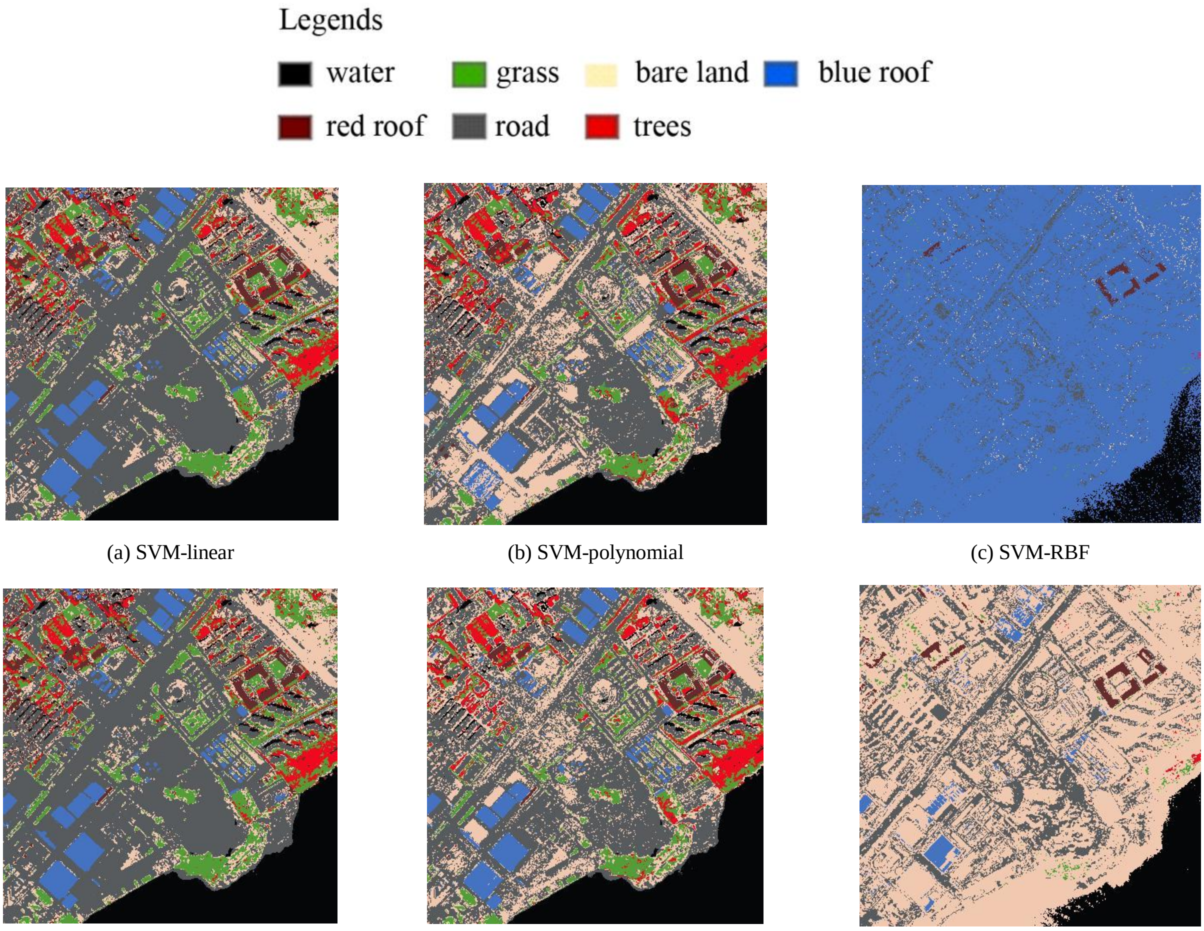
The original image is shown in Figure 1. The image size is 400*400. The image is classified into seven classes, i.e. water, grass, bare land, blue roof, red roof, road, trees.

In the standard SVM, c=2 and g=0.125, which is according to the experience. Only the spectral features are used in SVM classification, i.e. Blue, Green, Red and Near-Infrared band. The results of different kernels are shown in figure 2 (a)-(c). The result of the sigmoid kernel is not shown here as it describes only one class the grassland.

From Figure 2(a)-(c), it can be seen that polynomial kernel performs better than linear and RBF kernels. Many bare lands are misclassified into roads in the result of linear kernel SVM classifier. The result of RBF is very bad as most of study area is misclassified into blue roof.

In the PSO-SVM method, the results of different kernels are shown in figure 2 (d)-(f). From figure 2 (d)-(f), it can be seen that less bare soil is misclassified into roads for polynomial kernel. The result of RBF kernel is improved greatly by PSO. The improvement of linear kernel result is not obvious.

FIGURE 1 The original image



(a) SVM-linear (b) SVM-polynomial (c) SVM-RBF

(d) PSO-SVM-linear (e) PSO-SVM-polynomial (f) PSO-SVM-RBF

FIGURE 2 The classification results of different SVM methods

To further compare the results of different kernels, we compute the Producer's accuracy, user's accuracy, overall accuracy and kappa coefficient for the classified images. The producer's accuracy refers to the probability that a certain land-cover of an area on the ground is classified as such, which is the complementary of omission error. The user's accuracy refers to the probability that a pixel labeled as a certain land-cover class in the map is really this class, which is the complementary of commission error. The producer's accuracy and user's accuracy for any given class typically are not the same. From table1 and table 2, for SVM-Linear, an estimate for the producer's accuracy of bare land is 73%, while the user's accuracy is 79%. As a producer of classification, only 73% of all the bare land as such. As a user, roughly 79% of all the pixels classified as bare land are indeed bare land on the ground. As the producer's and user's accuracy are computed based on the diagonal of confusion matrix, the Kappa coefficient, which is calculated by all the values in the confusion matrix, is used for accuracy assessment. From table 3, it can be seen that, polynomial kernel outperforms other kernels. The PSO-SVM improves the results of SVM. For RBF kernel, the improvement of PSO-SVM is most obvious. The accuracy assessment further demonstrate the proposed PSO-SVM improve the results of classification. The improvements of different kernels by PSO-SVM are different.

TABLE 1 Producer's accuracy of classified image

	SVM			PSO-SVM		
	Linear	**Poly**	**RBF**	**Linear**	**Poly**	**RBF**
Water	90%	92%	87%	94%	95%	92%
Grassland	66%	69%	14%	73%	82%	19%
Bare land	73%	74%	9%	84%	92%	99%
Blue roof	62%	65%	97%	75%	90%	45%
Red roof	56%	59%	18%	68%	82%	79%
Road	92%	90%	19%	90%	91%	71%
Trees	68%	73%	11%	66%	92%	23%

TABLE 2 User's accuracy of classified image

	SVM			APSO-SVM		
	Linear	**Poly**	**RBF**	**Linear**	**Poly**	**RBF**
Water	94%	94%	98%	98%	100%	100%
Grassland	83%	84%	95%	96%	91%	89%
Bare land	79%	84%	49%	67%	73%	28%
Blue roof	97%	98%	19%	100%	100%	93%
Red roof	97%	97%	100%	97%	98%	100%
Road	41%	42%	58%	48%	78%	86%
Trees	77%	81%	88%	92%	94%	88%

TABLE 3 Overall accuracy (OA) and kappa coefficient (KC) of classified image

	SVM			SVM-PSO-MF		
	Linear	**Poly**	**RBF**	**Linear**	**Poly**	**RBF**
OA	72%	75%	36%	75%	89%	61%
KC	0.68	0.70	0.26	0.71	0.87	0.55

5 Conclusions

Support vector machines (SVM) are receiving increasing attention in remote sensing applications, such as image classification, land cover/land use change detection and so on. However, SVM is very sensitive to the parameters setting. In this study, a comparative analysis of the impact of four kernels (linear kernel, polynomial kernel, radial basis function kernel and sigmoid kernel) on the accuracy of SVM classifiers is conducted. Moreover, the Particle Swarm Optimization (PSO) is used to search for the optimum parameters for each kernel function in order to improve the classification accuracy of SVM classifiers. The experimental results show that the result of SVM with polynomial kernel is best while the result of sigmoid kernel is worst. The PSO improves the classification accuracy of RBF kernel most significantly, while the accuracy of linear kernel is not obvious. The PSO-SVM-Poly outperforms other methods. Of course, the experiment is limited. More experiment would be conducted in our future study, especially for feature selection in SVM classifiers.

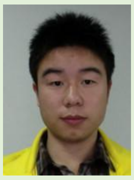
Acknowledgments

This research was supported by the Natural Science Foundation of Hubei Province (2013CFB285), the Fundamental Research Funds for the Central Universities, Wuhan University (2042014KF0054) and the State 973 project (2012CB719901) in China.

References

[1] Buciu I, Kotropoulos C, Pitas I 2006 Demonstrating the stability of support vector machines for classification *Signal Processing* **86** 2364-80
[2] Huang C, Davis L, Townshend J 2002 An assessment of support vector machines for land cover classification *International Journal of Remote Sensing* **23**(4) 725-49
[3] Villa A, Fauvel M, Chanussot J, Gamba P, Benediktsson J A, Gradient optimization for multiple kernel's parameters in support vector machines classification *Geoscience and Remote Sensing Symposium 7-11 July 2008* IGARSS 2008 IEEE International **4** 224-7
[4] Kavzoglu T, Colkesen I 2009 A kernel functions analysis for support vector machines for land cover classification *International Journal of Applied Earth Observation and Geoinformation* **11**(5) 352-9
[5] Pal M 2002 Factors influencing the accuracy of remote sensing classifications: a comparative study *University of Nottingham*
[6] Kennedy J, Eberhart R 1995 Particle swarm optimization *Proceedings of IEEE international conference on neural networks* 1942-48
[7] Wu Q, Wu S, Liu J 2010 Hybrid model based on SVM with Gaussian loss function and adaptive Gaussian PSO *Engineering Applications of Artificial Intelligence* **23**(4) 487-94
[8] Bazi Y, Melgani F 2007 Semisupervised PSO-SVM regression for biophysical parameter estimation *Geoscience and Remote Sensing, IEEE Transactions on* **45**(6) 1887-95
[9] Mao Y, Xia Z, Yin Z, Sun Y X, Wan Z 2007 Fault diagnosis based on fuzzy support vector machine with parameter tuning and feature selection *Chinese Journal of Chemical Engineering* **15**(2) 233-9
[10] Huang C L, Dun J F 2008 A distributed PSO–SVM hybrid system with feature selection and parameter optimization *Applied Soft Computing* **8**(4) 1381-91
[11] Lin S W, Ying K C, Chen S C, Lee Z J 2008 Particle swarm optimization for parameter determination and feature selection of support vector machines *Expert Systems with applications* **35**(4) 1817-24
[12] Melgani F, Bazi Y 2008 Classification of electrocardiogram signals with support vector machines and particle swarm optimization *Information Technology in Biomedicine IEEE Transactions on* **12**(5) 667-77
[13] Wu K P, Wang S D 2009 Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space *Pattern Recognition* **42**(5) 710-17
[14] Chang C C, Lin C J 2011 LIBSVM a library for support vector machines *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3) 27

Authors



Li Buyi, 1991, Xinyang, China

Current position, grades: Graduate Stundet
University studies: Wuhan University
Scientific interest: image processing and big data mining
Experience: From 2010 to 2014, studied in Wuhan University and got the Bachelor Degree in 2014, from 2014 to now, studies in Wuhan University for the Master Degree, honorable mention in Mathematical Contest In Modeling (MCM) in 2013, took part in the developing of digital library of China from 2012 to 2014.

Chongjing Deng, 1992, Changde, China

Current position, grades: Graduate Student
University studies: Wuhan University
Scientific interest: remote sensing image processing
Experience: From 2010 to 2014, studied in Wuhan University and got the Bachelor Degree in 2014, from 2014 to now, studies in Wuhan University for the Master Degree.

Li Shuang, 1982, Wuhan, China

Current position, grades: Associate professor
University works: Wuhan University
Scientific interest: image processing, remote sensings image classification and interpretation
Experience: in 2010 got Doctoral degree from State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing in Wuhan Univerity.