TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.1, January 2014, pp. 286 ~ 291
DOI: http://dx.doi.org/10.11591/telkomnika.v12i1.3969

Sensitivity of Support Vector Machine Classification to Various Training Features

Nanhai Yang, Shuang Li*, Jingwen Liu, FulingBian
International School of Software, Wuhan University
37 Luoyu Road, Wuhan, China, 430079, Fax: +86-27-68778221
*corresponding author, e-mail: lishuang129@gmail.com

Abstract

Remote sensing image classification is one of the most important techniques in image interpretation, which can be used for environmental monitoring, evaluation and prediction. Many algorithms have been developed for image classification in the literature. Support vector machine (SVM) is a kind of supervised classification that has been widely used recently. The classification accuracy produced by SVM may show variation depending on the choice of training features. In this paper, SVM was used for land cover classification using Quickbird images. Spectral and textural features were extracted for the classification and the results were analyzed thoroughly. Results showed that the number of features employed in SVM was not the more the better. Different features are suitable for different type of land cover extraction. This study verifies the effectiveness and robustness of SVM in the classification of high spatial resolution remote sensing images.

Keywords: *Remote sensing, image classification, support vector machine, feature extraction*

1. Introduction

High spatial resolution remote sensing images have played an important role in mapping, urban planning, defense and military, land use and surveys, and many other areas [1-3]. As the improvement of spatial resolution, single land cover shows a lot of different spectral value, which increasing the probability of misclassification. The similar spectral characteristics of different land covers often lead to confusing in classification, such as shadows and water bodies, meadows and trees, are often mixed in spectral value. Thus, it is hard to obtain high classification accuracy when only the spectral information is used. Compared with the traditional classification methods, Support Vector Machine (SVM) possesses the merits of learning with small samples, high anti-noise performance, etc. Moreover, SVM also has the advantages of high learning and promotion efficiency. Therefore, SVM classification showed good performance in remote sensing image information extraction [4-6].

In this study, SVM was used for land cover classification of Wuhan district in China using Quickbird images. Various spectral and textural features were extracted for SVM classification process and classification performances were analyzed thoroughly. It should be pointed out that the selection of features has an effect on the performance of SVM. Determination of their optimum combinations is regarded as critical for the success of classification.

2. Support Vector Machine Algorithm

Support vector machine (SVM) is supervised heuristic algorithm based on statistical learning theory [7]. The aim of SVM for classification is to determine a hyper plane that optimally separates two classes. An optimum hyper plane is determined using training data sets and is verified using test data sets.

Assume 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

Received June 23, 2013; Revised August 17, 2013; Accepted September 16, 2013

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

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

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

The training data points on the hyper planes are parallel to the optimum hyper plane. The support vectorsare defined by the function $w \cdot x_i + b = \pm 1$. If a hyper plane exists and satisfies Eq. (2), the classes are linearly separable. Therefore, the margin between these planes is equal to $2/\|w\|$. Thus, the optimum hyper plane can be found by minimizing $\|w\|^2$ under the constraint Eq. (2). Determination of optimum hyper plane is equivalent to solve optimization problem given by:

$$\min\left[\frac{1}{2}\|w\|^2\right] \tag{3}$$

As nonlinearly separable data is the case in various classifications of remote sensing images, the SVM technique can be extended to allow for nonlinear decision surfaces by introducing penalty parameter C and slack variable ξ :

$$\min\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i\right] \tag{4}$$

subject to constraints,

$$\begin{gathered} y_i\left(w \cdot x_i + b\right) \geq 1 - \xi_i \\ i = 1, 2, ..., N, \ \ \xi_i \geq 0 \end{gathered} \tag{5}$$

where penalty parameter C allows striking a balance between two competing criteria of margin maximization and error minimization, whereas the slack variable ξ_i 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.

When it is not possible to define the hyper plane by linear equations, the data may be mapped into a higher dimensional space through some nonlinear mapping function ϕ. The input point x can be represented by $\phi\left(x\right)$ in high-dimensional space. The time-consuming computation of $\phi\left(x\right) \cdot \phi\left(x_i\right)$ is reduced by using a kernel function. Thus, the classification decision function is defined as:

$$f\left(x\right) = \operatorname{sgn}\left(\sum_{i=1}^{N}\alpha_i y_i K\left(x_i \cdot x\right) + b\right) \tag{6}$$

where $\operatorname{sgn}\left(\cdot\right)$ is the sign function, $K\left(\cdot\right)$ is the kernel function and the magnitude of α_i is determined by the parameter C. The widely used kernel function includes linear kernel, polynomial kernel, sigmoid kernel and Gaussian radial basis kernel. By contrast tests, the linear kernel function obtained better results in our study.

3. Spectral and Textural Feature Extractions

Many algorithms have been developed for image classification using SVM. Several factors (e.g. training features, kernel functions, window sizes) have significant impacts on the classification performance, which should be considered carefully by the analyst. The selection of appropriate training features depends on the knowledge of land cover types present in the image by geographers. Thus, training features selection play an important role in the classification accuracy [8].

3.1. Spectral Feature Extraction

The widely used spectral feature is mean value and the metric derived from spectral value, i.e. Normalized Difference Vegetation Index (NDVI), Ratio Index (RI), Soil Adjusted Vegetation Index (SAVI), Normalized Difference Water Index (NDWI). The metric is described in Table 1.

Table 1. The spectral features used in the study

Metric	Equation	Description
NDVI	$\frac{Band_{NIR} - Band_{Red}}{Band_{NIR} + Band_{Red}}$	It is used to extract vegetation, i.e. grassland.
RI	$\frac{Band_{Red}}{Band_{NIR}}$	It is used to extract high density vegetation, i.e. trees.
SAVI	$\frac{(Band_{NIR} - Band_{Red}) \cdot (1 + L)}{Band_{NIR} + Band_{Red} + L}$	It is used to extract soil with low vegetation cover.
NDWI	$\frac{Band_{Green} - Band_{NIR}}{Band_{Green} + Band_{NIR}}$	It is used to extract water from land covers.

3.2. Textural Features Extraction

The Gray Lever Co-occurrence Matrix (GLCM) is proposed by Haralick in 1970s, which is an important technique to analyze image texture. The GLCM is based on the second order combination of probability density function, by calculating the correlation between two points in the estimated images [9,10]. The texture features are derived from GLCM, i.e. Angular Second Moment (ASM), Contrast, Entropy and Correlation. Let $G(i,j)$ be the element in GLCM and the size of the matrix be $k*k$, the metric are described in Table 2.

Table 2. The textural features used in the study

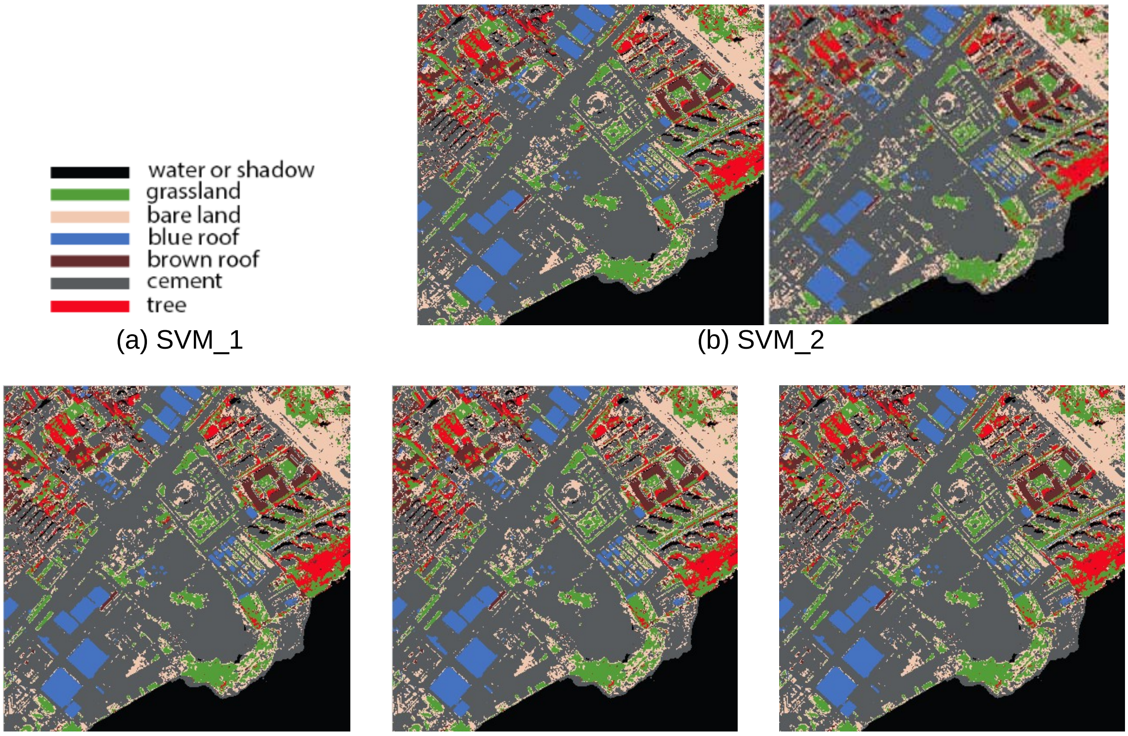
Metric	Equation	Description
ASM	$\sum_{i=1}^{k}\sum_{j=1}^{k}(G(i,j))^2$	It denotes the image gray uniformity and texture coarseness.
Contrast	$\sum_{i=0}^{k-1} n^2 \left\{\sum_{\lvert i-j \rvert = n} G(i,j)\right\}$	It reflects the texture clarity.
Entropy	$-\sum_{i=1}^{k}\sum_{j=1}^{k} G(i,j)\lg G(i,j)$	It measures the amount of information contained in the image.
Correlation	$\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{(i*j)G(i,j) - U_i U_j}{S_i S_j}$ $U_i = \sum_{i=1}^{k}\sum_{j=1}^{k} i \cdot G(i,j), U_j = \sum_{i=1}^{k}\sum_{j=1}^{k} j \cdot G(i,j)$ $S_i^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} G(i,j)(i - U_i)^2$ $S_j^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} G(i,j)(j - U_j)^2$	It describes the periodicity of texture element in a certain positional relationship.

4. Experimental Results and Analyze

The test image is from Quickbird sensor, with the spectral band ranges from 450nm to 900 nm. The image size is 400*400, which covers water (W), grassland (GL), bare land (BL), blue roof (BlR), brown roof (BrR), cement surface (CS) and trees (T). The original image is shown in Figure 1.

Figure 1. The test image with the size of 400*400

Five types of features are selected for the SVM classification. The classifiers are SVM_1 (four features: Mean value for the four bands), SVM_2 (six features: four mean value, NDVI and SAVI), SVM_3 (eight features: four mean value, NDVI, RI, SAVI and NDWI), SVM_4 (twelve features: four mean value, mean value and standard deviation of ASM, Contrast, Entropy, Correlation), and SVM_5 (sixteen features: four mean value, NDVI, RI, SAVI and NDWI, mean value and standard deviation of ASM, Contrast, Entropy, Correlation). The number of training data is 60 (block) * 7 (classes). The classified images are shown in Figure 2.



(a) SVM_1 (b) SVM_2

(c) SVM_3 (d) SVM_4 (e) SVM_5

Figure 2. The classified images of the study area

From Figure 2, it can be seen that the SVM classifiers perform well on image classification. The water, blue roof and brown roof are detected accurately. As it is hard to distinguish between water and shadow, we classify them as a class. To compare the results from different SVMs, we calculate the overall accuracy (OA) and Kappa coefficient by confusion matrix, which is shown in Table 3 to Table 7. The test data sets were formed using random pixel selection strategy with proportional number of samples for each class. A total number of 350 pixels are selected for testing classified images.

Table 3. Confusion matrix for SVM_1

SVM_1	W	GL	BL	BlR	BrR	CS	T
W	33	2	0	0	1	5	0
GL	3	27	0	1	1	0	1
BL	0	0	43	0	6	3	1
BlR	0	0	3	12	0	1	0
BrR	0	0	0	0	13	1	0
CS	9	0	6	2	2	161	0
T	0	0	0	0	1	0	12
OA	86%						
Kappa	0.8						

Table 4. Confusion matrix for SVM_2

SVM_2	W	GL	BL	BlR	BrR	CS	T
W	33	2	0	0	1	5	0
GL	3	27	0	1	1	0	1
BL	0	0	42	0	7	3	1
BlR	0	0	3	12	0	1	0
BrR	0	0	0	0	13	1	0
CS	9	0	6	2	2	161	0
T	0	0	0	0	1	0	12
OA	86%						
Kappa	0.8						

Table 5. Confusion matrix for SVM_3

SVM_3	W	GL	BL	BlR	BrR	CS	T
W	34	2	0	0	1	4	0
GL	3	27	0	1	1	0	1
BL	0	0	48	0	4	0	1
BlR	0	0	3	12	0	1	0
BrR	0	0	1	0	12	1	0
CS	9	0	0	2	2	167	0
T	0	0	0	0	1	0	12
OA	89%						
Kappa	0.84						

Table 6. Confusion matrix for SVM_4

SVM_4	W	GL	BL	BlR	BrR	CS	T
W	34	3	0	0	0	4	0
GL	3	27	0	1	1	0	1
BL	0	0	45	0	7	0	1
BlR	0	0	3	12	0	1	0
BrR	0	0	1	0	12	1	0
CS	9	0	3	2	2	164	0
T	0	0	0	0	1	0	12
OA	87%						
Kappa	0.82						

Table 7. Confusion matrix for SVM_5

SVM_5	W	GL	BL	BlR	BrR	CS	T
W	34	3	0	0	0	4	0
GL	3	27	0	1	1	0	1
BL	0	0	49	0	3	0	1
BlR	0	0	3	12	0	1	0
BrR	0	0	0	0	13	1	0
CS	9	0	0	2	2	167	0
T	0	0	0	0	1	0	12
OA	90%						
Kappa	0.85						

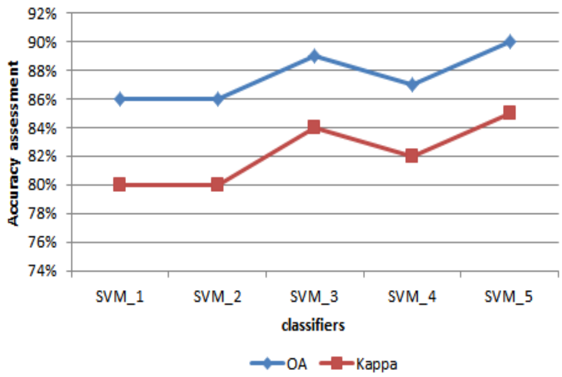


Figure 3. The accuracy trend for results images

To compare the overall accuracy and kappa coefficients for each classified images more obviously, the trend is given in Figure 3. From Figure 3, it can be seen that the accuracy of SVM_1 and SVM_2 is the same. However, the number of features used in SVM_2 is larger than SVM_1. It indicates that the NDVI and SAVI in SVM_2 do not help to increase the classification accuracy. The accuracy from SVM_3 is larger than SVM_2. It indicates that the RI and NDWI in SVM_3 increase the classification accuracy. The accuracy for SVM_4 decreases, which means the texture features are not as good as the spectral index, i.e. RI, NDWI. The accuracy for SVM_5 is the highest, which is 1% higher than the result of SVM_3. It means that the texture added in the classifiers do not significantly improve the classification accuracy.

5. Conclusion

Classification of remote sensing images is an important application for image interpretation. Support vector machine (SVM) have been recently used for many classification problems. Although it is reported that SVM produce more accurate classification results than the conventional methods, the selection of optimum training features is one of the most important issues that affect their performance. In this study, five types of features are used in the classifiers. From the experiments, several important conclusions can be drawn. Firstly, the number of features is not the more the better for the classification accuracy, i.e. SVM_3 and SVM_5. Secondly, the RI and NDWI features perform better than the texture features, including ASM, Entropy, Contrast and Correlation. This conclusions made here are based on the limited tests. More comprehensive tests will be conducted in the future.

Acknowledgement

This research was supported by a grant from 973 porject in China (Grant # 2012CB719901)

References

[1] Foody GM, Mathur A. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*. 2004; 42: 1335-1343.
[2] Han F, Li H, Wen C, Zhao W. A New Incremental Support Vector Machine Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10: 1171-1178.
[3] Yu Y, Zhou L. Acoustic Emission Signal Classification based on Support Vector Machine. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10: 1027-1032.
[4] Shao Y, Lunetta RS. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2012; 70: 78-87.
[5] Otukei J, Blaschke T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*. 2010; 12: S27-S31.
[6] Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2011; 66: 247-259.
[7] Vapnik VN. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*. 1999; 10: 988-999.
[8] Kavzoglu T, Colkesen I. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*. 2009; 11: 352-359.
[9] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*. 1973; 610-621.
[10] Dell'Acqua F, Gamba P. Texture-based characterization of urban environments on satellite SAR images. *IEEE Transactions on Geoscience and Remote Sensing*. 2003; 41: 153-159.