

The Study of Land Use Classification Based on SPOT6 High Resolution Data

Wu Song¹, Jiang Qigang²

¹College of Earth Sciences, Jilin University, Changchun, China

²College of Geo-Exploration Science and Technology, Jilin University, Changchun, China

Abstract. A method is carried out to quick classification extract of the type of land use in agricultural areas, which is based on the spot6 high resolution remote sensing classification data and used of the good nonlinear classification ability of support vector machine. The results show that the spot6 high resolution remote sensing classification data can realize land classification efficiently, the overall classification accuracy reached 88.79% and Kappa factor is 0.8632 which means that the classification result of support vector machine is ideal and better than other traditional image classification method. So, the method which is used high-resolution satellite provide a rapid and feasible way for classification of land use types.

1 Introduction

Research of land use classification provides the basic work for technical support, such as land planning and management, land change mechanism analysis and environmental protection. Remote sensing technology has become the most effective means for the acquisition of land use information as the technology has many characteristics, such as macroscopic, dynamic and rapid. At the same time, using satellite remote sensing data for automatic classification of land use and thematic information extraction has been the forefront direction of remote sensing technology application [1], [2]. So many scholars at home and abroad carried out the research about this, and the support vector machine (SVM) technology has been widely used in the automatic classification of land use with its characteristics of small sample training, support high dimensional feature space and fast convergence.

With the rapid development of high resolution remote sensing technology, the use of high resolution remote sensing data at home and abroad for the research of automatic classification of land is increasing, and has already obtained rich success [3], [6]. SPOT satellite is wide band, high spatial resolution remote sensing satellite and has been applied in many industries due to its good application performance, but its application is less applied for the automatic classification and information extraction of land use. In this paper, the high resolution remote sensing image and image automatic classification technology has combined organically [7], [8]. Analysis of the spectral characteristics and high resolution of SPOT6 satellite based on support vector machine (SVM) classification principle which realize land information

rapid extract classification, that provide evidence for monitoring land use situation, formulate comprehensive control measures and use policy [9].

2 Introduction in the study area and SPOT6 data preprocess

2.1 The geographical situation in the study area and the selection of the test data

The study area is located in the middle of Morocco, this experiment selected SPOT6 satellite images in May 23, 2013 as the remote sensing data. The image range from 30° 29' 34" N ~ 30° 33' 56" N to 527° 08' 59" W ~ 09 W° 05' 07" and we choose a typical area in this area. The total of land use category in the area is 5, its mainly has plough, forest land, construction land, water and other classes, so it can test the classification method effectively as shown in Fig. 1.



Figure 1. SPOT RS images of study area.

2.2 SPOT6 high resolution data preprocess

Spot 6 earth observation satellite was produced by the European space technology company in September 9, 2012, and it was launched by the PSLV carrier rocket in India successfully. In 22th September, SPOT6 entered the orbit that in the same orbit plane with Pleiades 1A satellite which was 695 Km high, after January 2013, the Satellite was into the formal business operation. The satellite can obtain image data of the spatial resolution of panchromatic 1.5M and multi-spectrum 6M, and it can receive 600 square kilometers image a day. The service objects mainly distributed in ecological environmental, geology and mineral resources, agriculture, forestry, environmental protection and disaster monitoring, telecom network planning, surveying and mapping, city planning and national defense [10], [11].

Table 1. Satellite parameters of SPOT6.

| Bands | Spectral range (µm) | Resolution (m) | Imaging Swath (km) | Angle of incidence (°) |
|-------|---------------------|----------------|--------------------|------------------------|
| 1 | 0.45~0.52 | 6 | 60 | ±30 |
| 2 | 0.53~0.59 | 6 | | |
| 3 | 0.62~0.69 | 6 | | |
| 4 | 0.76~0.89 | 6 | | |
| 5 | 0.45~0.75 | 1.5 | | |

2.3 Remote sensing data preprocess

2.3.1 Band synthetic

When using remote sensing technology to extract information, it is necessary to study the spectral characteristics which will be beneficial for the combination of band that identify the target. In this test, first of all, in the experimental zone, the spectral characteristics of each band of multi-spectral data will be analyzed, then calculate the mean, standard deviation and information entropy of the image grey value respectively, add up and compare the information reflected by the band, as it is shown in Table 2.

From the statistical analysis of the results, it can be seen that the statistical indicators of the fourth band is greater than the other bands, which means that the

information concluded in the fourth band is the greatest, and the fourth band (near infrared) has a significant role in the vegetation type classification, in the same time it has minimum correlation with other bands, so through combining the bands together based on the fourth band which is fixed on the green channel, we get experimental comparison results, then integrate with experimental zone concluded with the class situation, select 431 band as the band combination (figure 1a), the feature difference of this combination is big and it contains rich amount of information which is beneficial for visual discrimination and the study of classification automatically.

Table 2. Basic statistical information of multispectral images

| Basictats | Min | Max | Mean | Stdev | Eigenvalue |
|-----------|-----|------|------------|------------|------------|
| Band 1 | 0 | 788 | 339.878473 | 97.053276 | 126.748302 |
| Band 2 | 0 | 1359 | 603.766984 | 198.886201 | 219.758493 |
| Band 3 | 0 | 1243 | 538.598726 | 189.697247 | 63.63284 |
| Band 4 | 0 | 1357 | 573.856963 | 211.784032 | 45.268089 |

2.3.2 Image fusion processing

Image fusion technology can make use of the different characteristics of the data in a maximize way in order to improve the visual effect of image and the ability of image feature recognition and make the image has a higher spectral and spatial resolution in the same time[12]. Make the panchromatic and multi-spectral data fusion processing can let the SPOT6 data play a role in a maximum way, there are many algorithms for remote sensing data fusion. Principal component analysis (pca), IHS transform method and wavelet transform method were selected for fusion of resolution in this text. And the information after the process of fusion was calculated (table 3). From table 3 we can see that 3 kinds of fusion method have their advantages and disadvantages respectively, but wavelet transform method is superior to other fusion methods in spectral information keeping and peak signal to noise ratio. Therefore, the method of wavelet transform was adopted to process image fusion in this text based on the aim of image automatic classification.

Table 3. Satellite parameters of SPOT6.

| | Band | Mean | Standard deviation | Snr | Entropy | meangradient |
|---------|------|---------|--------------------|--------|---------|--------------|
| PCA | R | 228.159 | 37.455 | 9.435 | 4.518 | 4.235 |
| | G | 454.573 | 85.401 | 12.163 | 5.034 | 5.435 |
| | B | 239.548 | 48.629 | 10.006 | 4.229 | 5.654 |
| IHS | R | 231.617 | 99.102 | 10.801 | 4.309 | 4.345 |
| | G | 500.855 | 196.096 | 12.436 | 5.805 | 5.257 |
| | B | 253.535 | 99.644 | 9.925 | 4.336 | 4.286 |
| WAVELET | R | 333.918 | 89.633 | 12.354 | 4.011 | 4.387 |
| | G | 698.465 | 187.232 | 13.198 | 5.903 | 5.376 |
| | B | 355.011 | 109.918 | 8.765 | 4.678 | 5.587 |

3 Image classification principle based on support vector machine

Support vector machine (SVM) classification algorithm is a kind of machine learning algorithms based on statistical learning theory, which is used structural risk minimization principle of SRM through solving quadratic programming problem under the inequality constraints [13]-[15].

Structural classification hyperplane in the training set $T = \{ (x_1, y_1), \dots, (x_n, y_n) \}$, $x_i \in \mathbb{R}^d$, $y_i \in \{-1, 1\}$, $i=1, 2, \dots, n$. Assuming that the classification of equation is $\langle X \cdot \omega \rangle + b = 0$, the equation should be satisfied with type (1):

$$Y_i (\langle X \cdot \omega \rangle + b) - 1 \geq 0 \quad (1)$$

The class interval according to analytic geometry is $D = 2 / \|\omega\|$.

The problem can be converted into that introduce Lagrange function which is used for solving this optimization problem in order to make the function $\phi(\omega) = \|\omega\|^2 / 2$ minimum [16~17].

$$L = \|\omega\|^2 / 2 - \sum_{i=1}^n \alpha_i Y_i (\langle X_i \cdot \omega \rangle + b) + \sum_{i=1}^n \alpha_i \quad (2)$$

the $\alpha_i > 0$ is the Lagrange multiplier, the solution of the problem must be satisfied type (3) according to the KKT conditions:

$$\alpha_i \{ [\langle X_i \cdot \omega \rangle + b] Y_i - 1 \} = 0 \quad (3)$$

Therefore, the resulting discriminant function (4):

$$f(X) = \text{sgn} \left\{ \sum_{i=1}^n Y_i \alpha_i \langle X_i \cdot X \rangle + b \right\} \quad (4)$$

In general case, most of α_i are 0, the others are not 0, the samples which the α_i corresponded is SV, b calculated by whichever SV. Considering some samples that could not be classified by the hyperplane correctly, convert the optimization problem to constrain condition by introducing slack variable:

$$Y_i (\langle X_i \cdot \omega \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (5)$$

In the constrain condition above, there is $\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i$ ($C > 0$).

In the result, the most α_i are 0, referrer the sample that α_i is not 0 as support vector. The function which the support vector defined is SVM. Usually, we called the training sample which has a small amount samples as support vector that means SVM has the advantages of

sparseness, so the classification speed of SVM is better than others [18-21].

The essence of SVM: first of all, to transform the input space into a higher dimensional space through the nonlinear transformation which is defined by the appropriate product function, then to achieve the optional linear classification surface by linear regression in the higher dimensional space.

4 The process of SVM image classification

4.1 Remote sensing data preprocess

Support vector machine (SVM) classification algorithm is a kind of machine learning algorithms based on statistical learning theory, which is used structural risk minimization principle of SRM through solving quadratic programming problem under the inequality constraints [13-15].

First of all, feature extraction for high resolution data, then selecting the features as extraction algorithm, such as the mean, the standard deviation and k-l transform.

Selected the interested area of the image which the training and choice of the size of texture window is the result of test for many times. If the size is too small that will make not only the training speed reduced strongly, but also the result has no significant change; oppositely, if the size is so big that the precision will be low [29].

As it is shown in t Fig. 2.

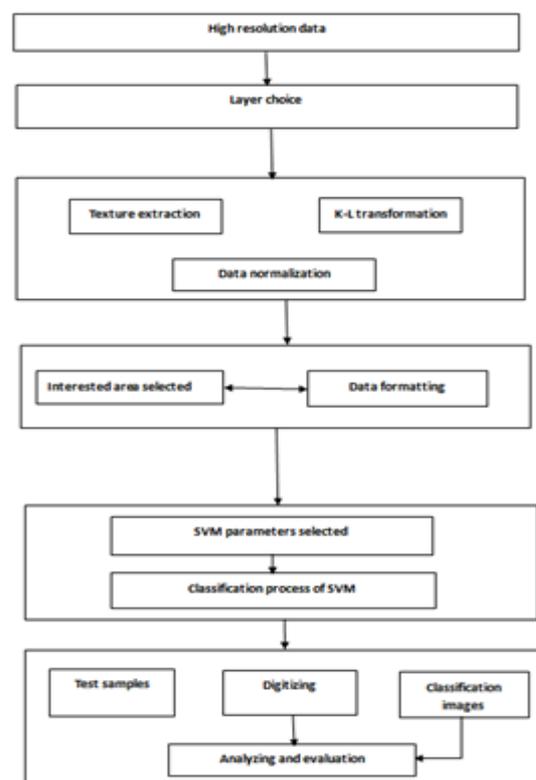


Figure 2. The flow chart of feature extraction.

4.2 Classification results and validation of precision *4.2.1 Image fusion processing*

Table 4. The confusion matrix of maximum likelihood classification.

| Category | Water | Plough | Forestland | Construtions | Others | Amounts |
|--------------|-------|--------|------------|--------------|--------|---------|
| water | 604 | 0 | 4 | 0 | 25 | 633 |
| Plough | 0 | 651 | 183 | 6 | 0 | 840 |
| Forestland | 0 | 16 | 418 | 4 | 0 | 438 |
| Construtions | 0 | 7 | 9 | 69 | 0 | 85 |
| Others | 0 | 170 | 102 | 0 | 107 | 379 |
| Amounts | 604 | 844 | 716 | 44 | 132 | 2375 |

Table5. Confusion matrix of Markov distance method of classification.

| Category | Water | Plough | Forestland | Construtions | Others | Amounts |
|--------------|-------|--------|------------|--------------|--------|---------|
| water | 469 | 0 | 166 | 3 | 0 | 638 |
| Plough | 0 | 914 | 0 | 0 | 0 | 914 |
| Forestland | 0 | 0 | 504 | 10 | 0 | 514 |
| Construtions | 145 | 0 | 0 | 60 | 0 | 205 |
| Others | 171 | 221 | 118 | 6 | 132 | 648 |
| Amounts | 604 | 1135 | 788 | 79 | 132 | 2919 |

Table 6. Confusion matrix processed by SVM classification used by spectral characteristics.

| Category | Plough | Forestland | Water | Construtions | Others | Amounts |
|--------------|--------|------------|-------|--------------|--------|---------|
| Plough | 299 | 32 | 0 | 1 | 0 | 332 |
| Forestland | 50 | 259 | 0 | 8 | 0 | 317 |
| Water | 0 | 0 | 296 | 0 | 0 | 296 |
| Construtions | 0 | 0 | 0 | 61 | 0 | 61 |
| Others | 0 | 0 | 0 | 19 | 72 | 91 |
| Amounts | 349 | 291 | 296 | 89 | 72 | 1097 |

Table 7. Comparing with classification precision.

| Methods | Overall accuracy (%) | Kappa Coefficient |
|-----------------------------|----------------------|-------------------|
| Maximum likelihood method | 77.85 | 0.7013 |
| Mahalanobis distance method | 71.22 | 0.6218 |
| SVM classification method | 89.79 | 0.8632 |

In order to verify the applicability of the SVM which is used for high resolution image classification, using the markov distance method and the maximum likelihood for classification , Calculating the belonging category and making the land use classification figure of test area to the study area, as it is shown in Fig. 3.

- (a)SVM
- (b)Maximum likelihood classification
- (c)Mahalanobis distance classification

4.2.2 Classification accuracy evaluation

This experiment adopts widespread confusion matrix method to analysis classification results, selected the test sample randomly corresponding to various land use types In remote sensing image, then calculated its classification confusion matrix and its related precision index respectively based on the different results of classification above. The results as shown in table 4,5,6.

The comparison table shows(Table 7): the classification accuracy of application of SVM method is superior to maximum likelihood classification and Markov distance classification method, which verified the superiority of support vector machine (SVM) on the

nonlinear classification problem of small sample. The overall classification accuracy Kappa coefficient reached 88.79% and 0.8632.

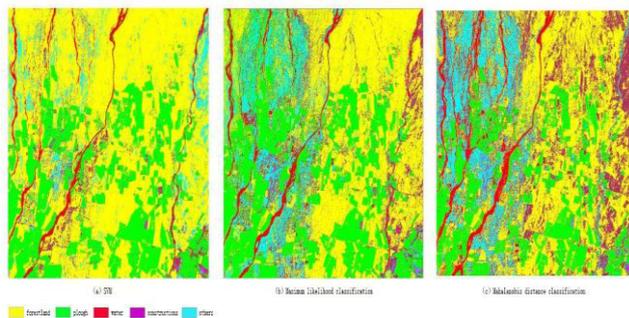


Figure 3. Comparison of classification result.

5 Conclusion

The method which realizing the rapid division of land use type based on SPOT6 high-resolution satellite data and image automatic classification technology has improved the recognition efficiency of agricultural land types.

Spot6 data was processed by training sample and predicting classification through their spectral information and the SVM classification method. The classification results shows that not only the algorithm precision of support vector machine is superior to the traditional classification algorithms, characterized by strong adaptability. The phenomenon of fault classification and miss classification is less, but also it has a high degree of stability. Therefore, in remote sensing image classification, the selection of support vector machine method for land use to classification research based on high resolution images can improve the classification precision and has a great advantage.

References

1. Luo Jiancheng, Zhou Chenghu, Yang Yan. Land-cover and land-use classification based on remote sensing intelligent Geo-interpreting model[J].Journal of Natural Resources, 2001, **16(2)**:179~183.(in Chinese)
2. Li Xiubin. A review of the international researches on land use/land cover change [J].ACTA Geographica Sinica, 1996, **51(67)**:553~557.(in Chinese)
3. Mai Kaile, Zhang Wenhui. Object-oriented classification approach for remote sensing imagery information extraction in loess hilly-gully region[J].Transactions of the Chinese Society for Agricultural Machinery2011, **42(4)**:153-158. (in Chinese)
4. Zhao Chunhui, Qiao Lei. Classification of hyperspectral remote sensing image using improved LS-SVM[J]. Applied Science and Technology, 2008, **35(1)**:44~52.(in Chinese)
5. Fu Wenjie, Hong Jinyi, Lin Mingsen. A method of land use classification from remote sensing image based on support vector machines and spectral similarity scale[J]. Remote Sensing Technology and application, 2006, **21(1)**:25~30.(in Chinese)
6. Du Peijun,Liu Sicong, Zheng Hui. Land cover change detection over mining areas based on support vector machine [J]. 2012, **41(2)**:262~267. (in Chinese)
7. Sun Danfeng,Yang Yihong, Liu Shunxi. Application of high-spatial IKONS remote sensing images in land use classification and change monitoring[J]. Transactions of the CSAE, 2002, **18(2)**:160~164. (in Chinese)
8. Chen Qihao, Liu Zhimin, Liu Xiuguo, et al. Element-oriented land-use classification of mining area by high spatial resolution remote sensing image[J]. Earth Science — Journal of China University of Geosciences, 2010, **35(3)**:453~458. (in Chinese)
9. Zhou Pei, Zhou Shenli. Effect of land use on ecological benefit of farm belt in suburbs[J]. Journal of Ecology and Rural Environment, 2007, **23(4)**: 6~10. (in Chinese)
10. Zhou Yi, Wu Juan, Li Qi, et al. Test and analysis for detecting land use change by using CBERS-02C satellite image[J].Mineral Exploration.2012,**3(5)**: 688~694. (in Chinese)
11. Ma Ligang, Zhang Leping, Zheng Jinsong, et al. Land use classification using ZY1- “02C” remote sensing images [J]. Journal of Zhejiang University: Engineering Science ,2013, **47(8)**:1508~1516. (in Chinese)
12. Wang Haihun, Peng Jiexiong, Wu Wei, et al. A study of evaluation methods on performance of the multi-source remote sensing image fusion[J].Computer Engineering and Applications,2003,(25):33~37. (in Chinese)
13. hang Xuegong. Introduction to statistical learning theory and support vector machines[J]. Acta Automatica Sinica, 2000, **26(1)**:32~42. (in Chinese)
14. Vapnik V N. The Nature of Statistical Learning Theory[M].Berlin: Springer-Verlag Berlin Heidelberg,1998.
15. Hsu C, Lin C.A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks, 2002,**13(2)**:415~425.
16. Zhang Quanming, Liu Huijin. Application of LS-SVM in classification of power quality disturbances[J]. Proceedings of the CSEE,2008,**28(1)**:106~110. (in Chinese)
17. Yang Jiajia, Jiang Qigang, Chen Yongliang, et al. Lithology division for large-scale region segmentation based on LS-SVM and high resolution remote sensing images[J].Journal of China University of Petroleum(Edition of Natural Science) , 2012,**36(1)**:60~67. (in Chinese)
18. Wang Kai,Hou Zhurong,Wang Congli. Intrusion detection based on cross-validation SVM[J]. Journal of Test and Measurement Technology , 2010, **24(5)**:419~423. (in Chinese)
19. Zhang Jinshui, He Chunyang, Pan Yao, et al. The high spatial resolution RS image classification based

- on SVM method with the multi-source data, 2006,**10(1)**:49~57. (in Chinese)
20. Fauvel M, Chanussot J, Benediktsson J A. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas[J]. EURASIP Journal on Advances in Signal Processing, 2009:1~14
 21. Guo Hui, Wang Ling, Liu Heping. Integrating kernel principal component analysis with least squares support vector machines for time series forecasting problems[J]. Journal of University of Science and Technology Beijing, 2006, **28(3)**:303~306. (in Chinese)