IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG)
e-ISSN: 2321–0990, p-ISSN: 2321–0982.Volume 6, Issue 6 Ver. II (Nov. – Dec. 2018), PP 51-56
www.iosrjournals.org

Study on spatiotemporal dynamic changes of Yanji City based on SVM

Ke Xu[1], Qingyu Wang, Minshu Su, Xuelong Li, Hengdong Feng[*]
Department Of Geography, Yanbian University, Jilin, 133002, China[1]
Corresponding Author: Ke Xu

Abstract: *In this study, support vector machine (SVM) as a more accurate method was used to classify remote sensing images and analyze the dynamics and spatiotemporal process of regions. The study area is Yanji City. Taking five years as a time span, remote sensing images from 2007 to 2017 are downloaded. Geometric registration, clipping and other processing are performed in ArcGIS for the raw data. The color feature and texture feature of the processed remote sensing image data are extracted and the training set is established. According to the concrete analysis, kernel function is needed to establish the training model based on the training set data. There are four kinds of kernel functions in common use, and the radial basis kernel function is determined by comparing the classification accuracy. Taking radial basis kernel function as kernel function, a linear learning machine is established in high dimensional characteristic space. After the accuracy test of the training model, the image is predicted and segmented according to the corresponding features. Finally, the classification results are presented in the form of images. According to the classified images, it can be seen that the urban area of Yanji is relatively dense in the south and tends to expand to the southwest, which is in line with the characteristics of orthodox urbanization.*
***Key words**: SVM; Yanji City; Kernel function; Urbanization*

Date of Submission: 10-12-2018 Date of acceptance: 24-12-2018

I. Introduction

3S(GPS, GIS and RS), as an important technology used to better observe human living environment in modern society, is getting more and more advanced and has a good development prospect. In these technologies, remote Sensing plays a very important role in continuous acquisition of surface object information over a large area. In order to make use of remote sensing images for various topographic survey and spatial analysis, scientists have summarized various remote sensing related technologies. Among them, the support vector machine (SVM) method for remote sensing image classification was first proposed by Cortes and Vapnik in 1995, with good performance and more and more extensive application fields. Compared with the traditional Maximum Likelihood method, Minimum Distance method, IsoData, K-mean and other methods, the SVM method can control its learning ability and generalization ability by its model parameters [1], which greatly improves the accuracy of remote sensing image classification. Based on remote sensing images of different periods, this study extracted the built-up area information of Yanji City in each period based on SVM algorithm, analyzed the dynamic development status of Yanji City on the basis of urbanization theories, and provided theoretical basis for the urban development of Yanji City.

1 Study area overview and data preprocessing

1.1 Introduction of the study area

Yanji is the capital of Yanbian Korean autonomous prefecture, Jilin province, and the political, economic and cultural center of the prefecture. From the perspective of natural conditions, Yanji City is located in the small basin at the eastern foot of changbai mountain, which belongs to the mid-temperate semi-humid climate zone. The terrain is basically strip, surrounded by mountains on three sides, and the middle terrain is relatively flat [2]. Its climate is obviously continental climate where springs and autumns are dry, summers rainy and winters long. The specific location of Yanji City is shown in figure 1.

[1] Ke Xu Email：843820914@qq.com
[*] Hengdong Feng, corresponding author, Email:278252662@qq.com

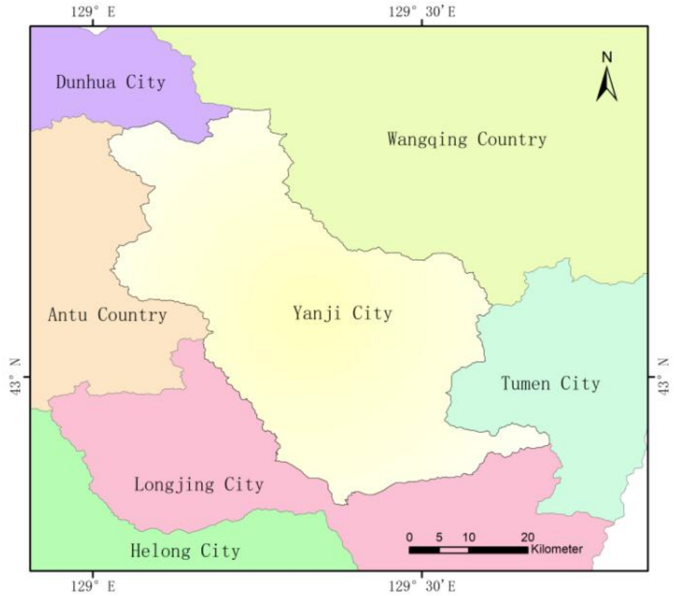


Figure 1 Location map of Yanji City

1.2 Data source and preprocessing

The raw data of this study is remote sensing data downloaded from Geospatial Data Cloud (http://www.gscloud.cn/). Landsat7 ETM+ data in November 2007 and August 2012 and Landsat8 OLI data in June 2017 is included. The data bar number is 115 and the row number is 30, covering all the study areas. For the selected data, the cloud is less and the image quality is better than others, which can be applied to this study. Due to the internal problems of the sensor, the ETM+ data in 2007 and 2012 need to process the strips first, and then the remote sensing images in three periods are geometrically registered, and the remote sensing images in Yanji City are cut out as the research data.

1.3 Extraction of regions of interest (ROIs)

In Matlab R2014b, the clipped remote sensing image of Yanji City is loaded and visually interpreted to determine the texture, color and other characteristics of the region of interest to be selected. This study mainly studies the urban changes of Yanji City. Based on the resolution of existing image data, five types of ground objects including buildings, woodlands, grasslands, arable lands and water bodies are determined to be studied. Real color image and false color image are combined to interpret the ground object when extracting the area of interest. Band 3,2,1 combination to obtain true color image; Standard false-color images are obtained by combining band 4, 3, and 2 (Landsat8 OLI data is band 5, 4, and 3).

2 City information extraction based on SVM

2.1 Theories of support vector machine (SVM)

Support Vector Machine (SVM) is a mathematical method to classify samples by establishing models based on training samples and test samples. SVM includes two main ideas . One is to analyze the case of linear separability. For the case of linear separability, the non-linear mapping algorithm is used to map it to high-dimensional space, and the linear algorithm is used for analysis. Secondly, the optimal segmentation hyperplane is constructed in the feature space to minimize the global risk. The specific calculation process is as follows.

The training set is established according to the color feature and texture feature of the image data extracted from the region of interest. Assume that the training data is:

$$(x_1, y_1),...,(x_i, y_i), x \in R^n, y \in \{+1,-1\} \tag{1}$$

Divide the training data into a hyperplane:

$$(w \cdot x)+b=0, \mathrm{w} \in \mathrm{R}^N, b \in R \tag{2}$$

$$y_i((w \cdot x_i)+b) \geq 1, i=1,...,l \tag{3}$$

In (2) and (3), w and x are both n-dimensional column vectors, x is the point on the plane, w is the normal vector on the plane, and b is the distance from the point to the hyperplane. The establishment of hyperplane is the starting point of SVM algorithm to solve linear indivisibility problem. For linearly indivisible data, SVM usually uses mapping algorithm to make it have high-dimensional space attributes and conduct linear analysis of samples in high-dimensional space. In this process, it is difficult to implement the nonlinear mapping algorithm, and the kernel function is introduced to in the subsequent steps to solve this problem skillfully.

When the hyperplane classification interval is the maximum, it is equal to:

$$\|w\|^2 \tag{4}$$

Lagrange function and KKT (karush-kuhn-tucher) conditions are applied to optimize the formula. For the formula to be optimized, it is necessary to convert it into the objective function and formulate the constraint conditions. The KKT condition is set in the constraint condition, so that the equality constraint optimization method is applied to the inequality and the solving process is simplified. After setting up the target function and constraint conditions, the relevant Lagrange function is defined and the partial derivative equation is obtained for further operation. According to (4), set up the function: $f(\vec{w})$, establish the minimum objective function, introduce the KKT condition into the constraint condition, and finally list the equation:

$$\begin{cases} \frac{\partial J}{\partial w} = 0 \\ \frac{\partial J}{\partial \beta_i} = 0 \\ \alpha_i \cdot g_i(\vec{w}) = 0, i = 1,2,3...,k \\ g_i(\vec{w}) \le 0, i = 1,2,3...,k \\ \alpha_i(\vec{w}) \le 0, i = 1,2,3...,k \end{cases} \tag{7}$$

Solve the equation according to Lagrange multiplier and KKT conditions, and transform the problem into:

$$Q(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j d_i d_j x_i^T x_j \tag{8}$$

Its decision-making surface is:

$$\sum_{i=1}^{l} \alpha_i^* d_i x_i^T x + b^* = \mathrm{O} \tag{9}$$

α_i^* is the optimal solution of the problem.

By introducing KKT condition to optimize the relaxation variables, the linear function can solve most linear separable problems. However, for some samples that are linearly indivisible, applying this formula will lead to an endless loop, so kernel function is introduced. The principle of kernel function is to transform vectors in low dimensional space into high dimensional space and find their inner product value, classify samples in high dimensional space and solve linear indivisibility problems. The kernel functions commonly used in SVM classification are shown in table 1.

Table 1 Kernel functions commonly used in SVM classification

Function name	Calculation formula
Linear kernel function	$K(x,x') = \langle x, x' \rangle$
Polynomial kernel function	$K(x,x') = (1 + x^T x')^P$
Gaussian kernel function (radial basis kernel function)	$K(x,x') = \exp(-\frac{\|x-x'\|^2}{2\sigma^2})$
Sigmoid kernel function	$K(x,x') = \tanh(\gamma \langle x, x' \rangle + c)$

The parameters of the classification include the penalty coefficient C, the gaussian radius of the kernel σ and the polynomial coefficient p [3]. Generally,the penalty coefficient C is determined by the ratio of the positive and negative sample, the gaussian radius of the gaussian radius σ is between 0.1 and 1, and the polynomial kernel function coefficients p is 2 or 3[4]. In this study, the value of σ is 0.5, and p is 2.

Through existing experience, if the number of training samples is close to the total sample number, the linear kernel function is used. If the sample number is comparatively smaller than the total sample number, the radial basis function is selected. In order to ensure the accuracy of the image classification, the training sample that should be controlled by manual extraction is much smaller than the total sample number, so the linear kernel function is not considered.

2.2 Establishment of support vector machine (SVM)

In order to select the kernel function of the SVM classification in this study, take 2007's data as an example and classify the image. Introduce the confusion matrix to verify accuracy, calculate the accuracy、precision、recall、F1 based on the confusion matrix. The computational formulas for accuracy、precision、recall、F1 are as shown in table 2.

Table 2 Model accuracy verification index

Verification index	Formula
Accuracy	$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$
Precision	$precision = \frac{TP}{TP+FP}$
Recall	$recall = TP_{rate} = \frac{TP}{TP+FP}$
F_1	$F_1 = \frac{2 \times precision \times recall}{precision + recall}$

Three kernel functions are selected and programmed in Matlab R2014b [5]. TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) of the classification results are calculated [6], and the results are obtained according to the formula as shown in table 3 and table 4.

Table 3 Comparison of classifier classification accuracy based on different kernel functions

	Polynomial kernel function	Radial basis function	Sigmoid function
Kappa Coefficient	0.9579	0.9609	0.9456
Overall Accuracy	0.9752	0.9770	0.9680

As can be seen from table 3, in general, the accuracy of SVM classification using the three kernel functions is relatively high, which is suitable for remote sensing image classification in this study. But due to the small number of total samples, the accuracy has the problem of being too high. By specific comparison of Kappa coefficient and Accuracy, it can be seen that the classification Accuracy of radial basis kernel function is higher than that of other kernel functions.

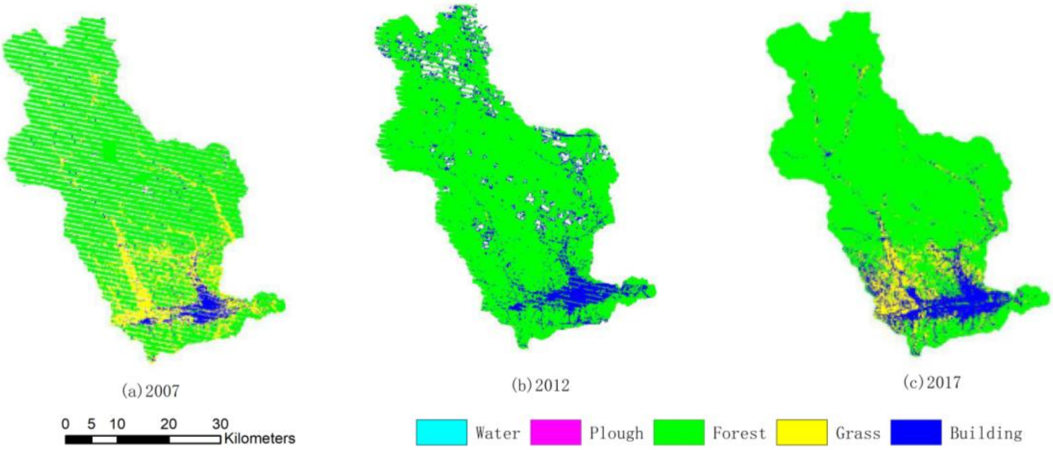
Table 4 Comparison of image classification effects of urban areas based on different kernel functions

	Polynomial kernel function	Radial basis function	Sigmoid function
precision	0.9880	0.9880	0.9880
recall	0.9572	0.9609	0.9425
F1	0.9723	0.9743	0.9647

As can be seen from table 4, for specific image classification results, the classifier precision of the three kernel functions is the same, but the recall and F1 values of radial basis functions are higher, indicating that the application of radial basis kernel functions is more sensitive and accurate than other kernel functions. In conclusion, the learning model to be established in this study should use radial basis kernel function.

2.3 Information extraction

By introducing kernel function, the formula becomes a simple linear separable problem. The established training samples are studied and the classification accuracy is detected by using the test samples. After the model obtained from the training sample passes the test, it indicates that the machine has successfully learned and established a universal model based on the extracted sample. Then, all the data to be classified can be put into the machine to get the classification results. The specific process is realized by Matlab R2014b programming. According to the classification results, buildings are drawn in blue, forest is green, grass is yellow, plough is magenta, water is blue-green, and other ground objects that cannot be classified are black. The classification results are shown in figure 2.



Due to the accuracy of the raw remote sensing image data, only three types of ground objects, namely buildings, woodlands and grasslands, are obvious in the classification results. Based on the observation of regional changes of buildings in different years, it is concluded that the urban area of Yanji is denser in the south and tends to expand to the southwest.

3 Classification results analysis

It can be seen from the classification image that Yanji city takes the southeast as the core and expands to the west continuously, and the urban expansion speed is relatively rapid. As the core of promoting political, Yanji's economic is developing rapidly, the urbanization Ⅰ is in parallel with urbanization Ⅱ, which can be called orthodox urbanization. Orthodox urbanization is defined as the coordinated development between the regional concentration of population and non-agricultural activities, the regional promotion of urban landscape and the regional diffusion of urban culture [7]. That is to say, if a city wants to develop well and urbanize in an orderly way, then it should be urbanized comprehensively from population migration and land use to spiritual construction. Since the reform and opening up, by the government's strong support, Yanji's per capita GDP continued to grow. At the same time that the city expands outwards, the urbanization inside yanji city is also proceeding in an orderly way. In the past decade, Yanji city has accelerated the construction of urban areas, vigorously promote the Korean culture, orthodox urbanization has taken shape.

Combined with geographical conditions, Yanji city is surrounded by mountains on three sides, while to the west, Jilin, Changchun and other large cities often mean opportunities for economic development. Therefore, it is of political and economic significance for Yanji city to continuously develop westward. In addition, Tumen and Longjing are located on the east and west sides of Yanji city respectively. So it is a good choice to strengthen the connection with these cities in view of Yanji's development. This also proves that the development strategy of Yanlongtu Integration implemented in recent years has strong feasibility.

II. Conclusion

In this study, regions of interest are extracted from the downloaded remote sensing image data as training samples, and a classifier is established with the radial basis kernel function as the kernel function to solve the problem in high-dimensional space. After the model is verified, the remote sensing image is formally classified and the remote sensing image classification results are obtained. The classification results show that:

(1) Three kinds of SVM classification methods with different kernel functions all can be used to classify images with high classification accuracy. The remote sensing images downloaded in this study have a good description of urban changes.

(2) In the application of SVM classification method to observe the spatiotemporal dynamic change of urban, the radial basis kernel function has the highest accuracy. However, in this study, there is a problem that the basis for choosing kernel function parameters is not clear. Genetic algorithm can be considered to optimize kernel function parameters, which will be the starting point for further research.

(3) According to the classification results, it can be seen that the urban area of yanji is denser in the south and tends to expand to the southwest. At the same time, the development pattern also proves that yanji city

is moving forward steadily on the road of orthodox urbanization.

This study uses SVM remote sensing classification method to classify remote sensing images of Yanji city, which plays a supplementary role to previous research results and has important research significance. Compared with the traditional supervised classification method, SVM classification method has a great improvement in accuracy, and can be widely used in a variety of cases classified according to samples, with strong explanatory power and high credibility. However, SVM classification method also has some shortcomings. For example, it is difficult to choose the right kernel function for nonlinear problems that cannot be solved. For the remote sensing image classification in this study, SVM algorithm cannot retain relevant geographic information data, which makes it at a disadvantage compared with other supervised classification algorithms. These deficiencies will be the specific development direction of further research in the future.

References

[1]. Chunjiao Dong, Chunfu Shao, Zhihua Xiong. State discrimination of traffic flow in urban expressway network based on SVM [J]. Journal of Beijing Jiaotong University, 2011,35 (6) : 13-22.
[2]. Xuanchao Li. Analysis on the change of urban construction land in Yanji City [D] Jilin: Northeast normal university, 2009.
[3]. Xiong Tan, Xuchu Yu, Jinchun Qin, Xiangpo Wei. Classification of multi-core SVM for hyperspectral images [J]. Acta instrumentation sinica, 2014,35 (2) : 405-411.
[4]. Fei Wang, Yuquan Zhu, Geng Chen. An image labeling method based on multi-kernel function SVM [J]. Journal of kunming university of science and technology, 2012,37 (5) : 43-53.
[5]. Yubin Lei, Shankuan Zhu, Yunkai Guo. The inversion of vegetation leaf area index in road area under the assistance of extreme learning machine [J]. Bulletin of surveying and mapping, 2018(9) : 82-86.
[6]. Ikram Sumaiya Thaseen, Cherukuri Aswani Kumar. Intrusion detection model using fusion of chi-square feature selection and multi class SVM[J]. Journal of King Saud University - Computer and Information Sciences,2015.
[7]. Xueqiang Xu, Yixing Zhou, Yuemin Ning. Urban geography [M]. Beijing: Higher education press, 2009.

Ke Xu" Study on spatiotemporal dynamic changes of Yanji City based on SVM"IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG) 6.6 (2018): 51-56.