Optimization Model to Estimate Mount Tai Forest Biomass Based on Remote Sensing

Yanfang Diao[1], Chengming Zhang[2,3,*], Jiping Liu[2], Yong Liang[3], Xuelian Hou[4], Xiaomin Gong[1]

[1]College of Water Conservancy and Civil Engineering, Shandong Agricultural University, Taian, China
[2]Chinese Academy of Surveying and Mapping, Beijing, China
[3]School of Information Science and Engineering, Shandong Agricultural University, Taian, China
[4]Information Center of Shandong Electric Power School, Taian, China

Abstract. The development of low-carbon economy and the promotion of energy conservation are becoming a basic consensus of all countries. Therefore, global carbon cycle becomes a widespread concern research topic in scientific community. About 77% of the vegetation carbon stores in forest biomass in terrestrial ecosystems. So forest biomass is the most important parameter in terrestrial ecosystem carbon cycle. In this paper, for estimating the forest biomass of Mount Tai, a support vector machine (SVM) optimization model based on remote sensing is proposed. The meteorological data, terrain data, remote sensing data are taken into account in this model. In comparison the results of SVM with that of regressive analysis method, both the training accuracy and testing accuracy of regressive analysis method are lower than those of SVM, so SVM could obtain higher accuracy.

Key word: forecast biomass; remote sensing; support vector machine (SVM)

1 Introduction

Because forest ecosystem is the most important part of terrestrial ecosystems, its research occupies a pivotal position in the global. Forest biomass is the energy-based of the entire forest ecosystem running and the sources of nutrients, which is the basis of studying biological productivity, net primary productivity (NPP), carbon cycle and global change research. Forest biomass includes tree layer biomass, living ground cover layer (shrub layer, herb layer and moss, lichen layer) biomass and animal and microbial biomass. There are three basic traditional research approaches to estimate biomass: by measuring photosynthesis; by measuring the carbon dioxide produced by respiration change; and by measuring the stock of organism. But above mentioned methods are difficult to estimate the biomass and NPP of large scale of forecast.

Spectral information of remote sensing images with good comprehensive and current trend has a correlation with forest biomass [1]. In our country, in spite of a large number of survey of biomass sample plots are carried out [2, 3], an abundant of data are accumulated, and the local site biomass data are gotten, but these data are difficult to estimate the biomass of large scale of forecast. How to combine these biomass plots survey data, topographical and meteorological data with remote sensing data, explore the establishment of remote sensing to study the biomass estimation model, is a worthy of study. In this paper, Mount Tai, for example, combining forest biomass with meteorological data, topographical data and remote sensing data, a support vector machine optimization model based on remote sensing information to estimate forest biomass in Mount tai is established.

* corresponding author. Email: chming@sdau.edu.cn.

2 The source and process of Data

2.1 Description of Mount Tai

Mount Tai is located in the central of Shandong province, 36°5'N~36°15'N, 117°5'E~117°24'E, and the highest peak is Yuhuangding, which altitude is 1545m. Mount Tai is a warm-humid and semi-humid monsoon climate and the vegetation type is the warm temperate zone deciduous broad-leaved forest. At present, the forest coverage rate reached 81.57%, and the vegetation coverage rate more than 90%. Vegetation types could be roughly classified into 19 types, including 11 types of forest, 2 types of shrub meadow, 3 types of meadow and 3 types of others. In the main forest types, Pinus tabulaeformis, Platycladus orientalis, Robinia pseudoacacia, Quercus variabilis account for about 92.9% [4].

2.2 The source of data

Air temperature, precipitation, humidity, sunshine, wind direction, wind speed and evaporation etc. are important ecological factors, which are also important factor that affect the forest ecosystem biomass. Meanwhile, the elevation and the types of vegetation are the very important factors. In order to determine above-mentioned factors' influence on the Mount Tai biomass, meteorological data, terrain data, remote sensing data are used to estimate forest biomass.

Meteorological data include temperature data and rainfall data. The temperature data includes annual mean temperature data (TA) and greater than 0 ℃ accumulated temperature data (T0), and the rainfall data is the annual mean rainfall data (PA).

Topographic data is primarily the 1:250 000 digital topographic map data. The contour lines could be obtained by topographic maps from digital data, and the digital elevation model (DEM) and slope data (ASP) could be obtained by ARC / INFO.

Remote sensing data uses Landsat-7 image data. Landsat-7 inherits technical parameters of the previous Landsat series, and the most important feature of it is to carry an Enhanced Thematic Mapper Plus ETM+. ETM+ is in developed on the basis of the Landsat-4, 5 satellite thematic mapper TM. Its band data cover larger scope from visible to infrared region, contain an abundant of information, and each band have different characteristics. At the same time, it is obvious that surface features in different spectral band are differences, so the band data of ETM+ have high practical value [5]. As the Landsat-7 ETM+ sensor has an advantage in spatial resolution, spectral resolution, and performance price ratio and other aspects. Therefore, ETM+ data has become one of the most commonly used remote sensing data at present, which is widely used in agriculture, forest and grassland resources survey, land use and mapping, geology, hydrology and other fields [6].

A series derived data of vegetation indexes are produced from Landsat-7 ETM+ data, for example normalized differential vegetation index (NDVI), ratio vegetation index (RVI), difference vegetation index (DVI), soil adjusted vegetation index (SAVI) and so on. The computational methods of above vegetation indexes are as follows.

Difference vegetation index: DVI=ETM4-A×ETM3

Normalized differential vegetation index: NDVI=(ETM4-ETM3)/(ETM4+ETM3)

Perpendicular vegetation index: PVI=(ETM4-A×ETM3-B)/SQR(1+A^2)

Ratio vegetation index: RVI=ETM4/ETM3

Soil adjusted ratio vegetation index: SARVI=ETM4/(ETM3+B/A)

Transformative Soil adjusted ratio vegetation index: TSAVI=A×(ETM4-A×ETM3-B)/(ETM3+A×ETM4-A×B)

Transformed vegetation index: TVI= $(NDVI+0.5)^{0.5}$

where TM3 and TM4 are red light and near infrared wave bands respectively. A and B equal 0.96916 and 0.084726 respectively.

2.3 From a sample survey data of forest resources to calculate biomass data

According to the types of forest of Mount Tai, the biomass estimation models of various organs are selected (such as Pinus tabulaeformis biomass estimation models [3,7], Quercus variabilis and other broad-leaved biomass estimation models [8]). Employing the breast-height diameter of per plant to calculate biomasses of every organs; the various biomasses of every organs are summed to be the biomasses of per plant; and the biomasses of every per plant are summed to be the biomasses of forest.

3 Optimization Model to estimate Mount Tai Forest Biomass Based on Remote Sensing

3.1 Principle of Support Vector Machine (SVM)

The support vector machine (SVM) [9-11] proposed by Vapnik et al. is a novel approach for solving pattern recognition (binary) problems. SVM holds many advantages compared with other approaches. It is a quadratic programming (QP) problem, which assures that its solution is the global optimal solution. And its sparsity assures a better generalization. It implements the structural risk minimization principle, which minimizes the upper bound of the generalization error, instead of the empirical risk minimization principle. Moreover, it has a clear geometric intuition on the classification task. Due to the above merits, SVM has been successfully used in many fields and extended to regression problems [10-13].

SVM based on statistic learning theory is a novel learning machine, which balances learning ability and structural complexity of the model based on small samples, so this model can avoid effectively local optimization. And it has a small parameters set by people in advance which make it have better generalization and popularization ability [14,15]. The basic strategy of SVM is to seek decision rules with popularization ability. The analysis process of SVM is described as follows.

Firstly, suppose there is a sample aggregate named $\{x_i, y_i\}_{i=1}^{N}, x \in R^m$, $y_i \in \{\pm 1\}$, and then an SVM can be defined by possible mappings $\phi : R^m \to R^n$, where R^n are lower dimensional spaces aggregate and where R^n are higher dimensional spaces aggregate. The largest interval classification hyperplane can be shown as

$$A = \left[\omega \cdot \phi(x)\right] + b \tag{1}$$

where $\omega = \sum_{i=1}^{N} \alpha_i \phi(x_i)$ is obtained in the higher dimensional spaces, and where α_i and b are Lagrangian multipliers and constant, respectively.

According to the theory of Reproduce Kernel Hilbert Space (RKHS), the kernel function $K(x_i, x_j) = \left[\phi(x_i) \cdot \phi(x_j)\right]$ which satisfies the condition of Mercer can be obtained in lower dimensional spaces. Then the problem of seeking the optimal classification hyperplane using Lagrangian optimization method is conversed to the solution of its dual problem, which can be described as

$$\max Q(\alpha) = \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{2}$$

where α_i should subject to the constraint and where $K(x_i, x_j)$ is kernel function

$$\sum_{i=1}^{N} y_i \alpha_i = 0, \alpha_i \ge 0, i = 1, \cdots, N \tag{3}$$

Kernel functions usually adopt the follow three ones: Linear function, polynomial function, and the radial base function, which are shown as follows.

Linear Kernel function: $K\left(x_i, x_j\right) = x_i \cdot x_j$

Polynomial function: $K\left(x_i, x_j\right) = \left[s\left(x_i \cdot x_j\right) + c\right]^d$

Radial basis function: $K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{\sigma^2}\right)$

Only a small number of nonzero solutions α_i are obtained from Eq.(2), these vectors corresponding to nonzero solution α_i are looked on as support vectors. Then the optimization decision function can be written as

$$f\left(x\right) = \mathrm{sgn}\left\{\left[\omega \cdot \phi\left(x\right)\right] + b\right\} = \mathrm{sgn}\left\{\sum_{i=1}^{N} \alpha_i y_i K\left(x_i, x_j\right) + b\right\} \tag{4}$$

If these data can not be classified with error-free in higher dimensional spaces, a relax quantum $\xi_i\left(\xi_i \ge 0\right)$ which is used to balance the maximum classification interval and the minimum misclassification samples is introduced to support vector machine. Then the optimum equation can be described as

$$\min \phi\left(\omega\right) = \frac{1}{2}\left\|\omega\right\|^2 + C\sum_{i=1}^{N} \xi_i y_i \left(\omega' \phi\left(x_i\right) + \omega_0\right) \ge 1 - \xi_i, i = 1, \cdots, N \tag{5}$$

where C is penalty coefficient.

The optimum classification plane function can be written as

$$\max \omega\left(\alpha\right) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K\left(x_i, x_j\right) \tag{6}$$

$$C \ge \alpha_i \ge 0 \quad \sum_{i=1}^{N} y_i \alpha_i = 0$$

Then the decision function $f(x)$ can be obtained using the Karush-Kutn-Tuker rules.

3.2 Estimate Mount Tai Forest Biomass

Fifty sample plots in the same size are selected in Mount Tai, in which forty sample plots are regarded as training samples and ten are testing samples. Ten independent variables TA, T0, PA, DEM, ASP, NDVI, PVI, RVI, SARVI, TSAVT are taken as input variables in the SVM and regressive analysis method. Then the SVM and regressive analysis method are applied to train the forty samples and test the ten samples. In this paper, the forest biomass calculated by the SVM and regressive analysis method are considered as predicted values, and those calculated by the method in the section 2.3 are considered as actual values. The accuracy function can be written as

$$Accuracy = \frac{1}{n}\sum_{i=1}^{n}\left(1 - abs\left(\frac{a_i - p_i}{a_i}\right) \times 100\%\right)$$

where i is the sequence number of samples, n is the amount of samples, a is actual value and p is predicted value.

The training and testing accuracy of the SVM and regressive analysis method are shown in Table 1.

Table 1 The accuracy of SVM and regressive analysis method

Accuracy \ Method	SVM	Regressive analysis
Training accuracy	86.09%	84.06%
Testing accuracy	84.62%	82.93%

4 Conclusion

(1) Because the training accuracy and testing accuracy of SVM and the regressive analysis method are high in Table 1, these are a close correlation between meteorological data, terrain data, remote sensing data and forest biomass. It is also shown that the independent variables selected in this paper could reflect the information of the Mount Tai forest biomass.
(2) From the Table 1, we could see that both the training accuracy and testing accuracy of regressive analysis method are lower than those of SVM, so SVM could obtain higher accuracy and it is a suitable method to estimate the forest biomass.
(3) In the current conditions, the proposed method to estimate Mount Tai forest biomass using remote sensing could replace the field survey methods. Its advantages are to reduce the amount of the field survey work, save the survey funds, reduce the labor intensity and have great social and economic benefits.
(4) The proposed method is relatively simple which is suitable to estimate biomass in larger area.

Acknowledgements

This study is funded by: (1) Basic Research Fund of Chinese Academy of Surveying and Mapping under Grant (Contract Number: 7771109); (2) Award Fund of Outstanding Young Scientist of Shandong Province (No. BS2011DX031).

References

1. Spencer R, Green M, Biggs P. Integrating eucalypt forest inventory and GIS in western Australia. Photogrammetric engineering and remote sensing. 63(12), 1345-1351(1997)
2. Jingyun Fang, Anping Chen, Shuqing Zhao, Longjun Ci. Estimation biomass carbon of China’s Forecast: Supplementary Notes on Report Published in Science(291: 2320-2322) by Fang etal.(2001). Acta Phytoecologica Sinica. 26(2), 243-249(2002) (in Chinese)
3. Jingyun Fang, Guohua Liu, Songling Xu. Biomass and net production of forest vegetation in China. ACTA ECOLOGICA SINICA. 16(5), 497-508(1996) (in Chinese)
4. Luo Guo, Shixiao Yu. Landscape Pattern of Taishan Mountain and Its Spatio Temporal Variation. ACTA SCIENTIARUM NATURALIUM UNIVERSITATIS SUNYATSENI. 43(4), 61-65(2004) (in Chinese)
5. Lexiang Qian. Remote sensing digital image processing and extracting geographic feature. Science Press, Beijing (2004) (in Chinese)
6. Hanqiu Xu. Study on Data Fusion and Classification of Landsat 7 ETM + Imagery. Journal of Remote Sensing. 9(2), 186-194(2005) (in Chinese)

7. Gang Wu, Zongwei Feng. Study on the social characteristics and biomass of the pinus tabulaeformis forest systems in China. ACTA ECOLOGICA SINICA. 14(4), 415-422(1994) (in Chinese)

8. Zhaolu Wu, Chenglin Dang. The biomass and net primary productivity of Pinus kesiya var langbianensis stands in Chang-Nin district, Yunnan. Journal of Yunnan University. 14(2),137-145(1992) (in Chinese)

9. Christianini, V and Shawe-Taylor, J. An Introduction to Support Vector Machines. Cambridge University Press, Cambridge (2002)

10. Vapnik, V. The nature of statistical learning theory. Springer Verlag, New York(1995)

11. Vapnik, V. Statistical learning theory. Springer-Verlag, New York(1998)

12. Mukherjee, S., Osuna, E., and Girosi, F. Nonlinear prediction of chaotic time series using support vector machines. In: Proceedings of the 1997 IEEE Workshop, pp. 511-520. IEEE Press, Amelia Island(1997)

13. Jeng, J. T., Chuang, C. C., and Su, S. F. Support vector interval regression networks for interval regression analysis. Fuzzy Sets and Systems. 138(2), 283-300(2003)

14. Xuegong Zhang. Introduction to statistical learning theory and support vector machines. Acta Automatica Sinica. 26(1), 32-42(2000)

15. Burges J C C. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery. 2(2), 121-167(1998)