ICESat-GLAS-Based Forest Type Classification Using SVM*

Licun Li and Yanqiu Xing**

Center for Forest Operations and Environment of Northeast Forestry University, Harbin 150040, China
lilicun628@126.com, yanqiuxing@nefu.edu.cn

Abstract. In order to study the forest classification effect of large footprint lidar date we used SVM(support vector machine) method to analyze the ICESAT-GLAS (Ice, Cloud and Land Elevation Satellite - Geoscience Laser Altimeter system) date in WangQing Bureau, Jilin province. In analysis we first used IDL to convert the ICESAT-GLAS original binary data into ASCII format. Then we got a waveform by using matlab software. After we were corresponded the waveform data to the field investigation data in 2006 and 2007, we could get the forest types of the waveform figure. Then waveform parameters were extracted. We applied of the SVM classification method to analyze 62 groups of training sample and established a classification model. After that we used another 62 groups of test sample to test the classification model, the result shows that the SVM classification method can better distinguish the broadleaved forest between the coniferous forest. And the classification accuracy is 82.26%.

Keywords: ICESat-GLAS, waveform parameters, SVM, forest type.

1 Introduction

Human activities make the carbon dioxide (CO2) and other gases in atmosphere continue to increase, so that the environment around us has been taken place a series of changes [1]. Forest biomass above the ground has an important impact on decrease the carbon dioxide in atmosphere. The forest biomass above the ground is related to both the height and the type of forest [2]. In recent years, lidar remote sensing technique has been proved to be a very effective technique in estimation of canopy height [3]. But there were only a few papers used lidar data to class the forest types. This article will further study on how to apply of lidar data for forest type classification. SVM (support vector machine) is a kind of universal and effective machine learning methods which has been well use in many ways [4]. Such as handwritten digit recognition [5], classification of a monkey species [6], classification of College Students' Decision about Graduation. In this paper we applied of SVM

* This research is funded by the Natural Science Foundation of China (4087119), and Project of Foundation (Gram10) by graduate school of Northeast Forestry University (NEFU).
** Corresponding author.

H. Deng et al. (Eds.): AICI 2011, Part III, LNAI 7004, pp. 630–638, 2011.

method to classify the forest type by using the parameters extraction from ICESAT-GLAS waveform data. This paper would be useful to study forest structure and the change of carbon dioxide in the atmosphere.

2 Material and Methods

The status of study area, and the access of Lidar data and field survey data are depicted as follow.

2.1 Status of the Study Area

If you have more than one surname, please make sure that the Volume Editor knows how you are to be listed in the author index. In this paper, Wangqing Forest Bureau, Jilin province of china was selected as the study area. The forest locates in the cool temperate forest ecosystem in Changbai Mountains (43 ° 05'N ~ 43 ° 40'N, 129 ° 56'E ~ 131 ° 04'E), the total area of it is 304,000 hm^2. Ground elevation is 360 ~ 1477m, slope ranges from 0 ~ 45 °.A variety of plant species range in this region and the structure of the forest is very complex. Mixed forest is the major forest type in the Mountain areas. *Pinus koraiensis*, *Picea koraiensis*, *Abies nephrolepis* are the major .coniferous trees and *Quercus mongolica*, *tilia amurensis*, *Acer mono* are the major broadleaved trees in this area.

2.2 Access of Lidar Data

ICESat-GLAS is large footprint laser radar equipment in polar orbit, altitude 600 km, using for continuous observation of the world ground. GLAS includes a laser system to measure distance. In laser system Laser pulses at 40 times per second will illuminate spots (footprints) 70 meters in diameter, spaced at 170 meters intervals along Earth's surface. ICESat-GLAS provide a total of 15 kinds of date, GLA01\GLA02\...\GLA15. GLA01 is an altimetry data product record the full waveform data which are correspond to the ground surface features in laser footprint. GLA14 record the ground footprint location and elevation data which are correspond to waveform data. The study used GLA01 and GLA14 data. I downloaded the study area data from year 2003 to 2006. GLAS data was downloaded in the U S National Snow and Ice Data Center. (http://nsidc.org/data/icesat/data.html)

2.3 Field Survey Data

In September 2006 and 2007 in Wang Qing Changbai Mountains we used a pre-designed stratified random sampling method to investigate the GLAS laser footprint on ground. In investigation a total of 203 plots were selected. There are three kinds of forest types, namely, coniferous forest, mixed and broadleaved forest. In application of data we did not use the waveform data which had only one Gaussian curve waveform. Because only one Gaussian curve waveform data is only ground return data or that can not well distinguish the forest canopy return data with the ground return data. In investigation, we used GPS to local the position of laser footprints. Fig.1 shows the field survey point. Using Forest survey statistical theory we investigated the forest within the footprint on the plots effectively. First measure the slope of plots, we recorded it θ. Then we established a circular plot of 500 m^2 in horizontal projected. Then we record the vegetation distribution, forest types and vegetation cover in ICESat-GLAS footprint. We apply of the following methods to distinguish the plot forest type. When the coniferous forest volume is greater than or equal to 60% of the plot forest volume, we identified the plot forest type as coniferous. When the Broadleaved forest volume is greater than or equal to 60% of the plot forest volume, we identified the plot forest type as Broadleaved forest. Remain cases are mixed. In this paper I select 124 groups of data, the forest types is Broadleaved forest and coniferous forest. 62 groups of data were used as training sample to establish the classification model. The other 62 groups of data were used to test the model.

Fig. 1. Study area and the distribution of field investigation points

3 Support Vector Classification

SVM is a new method of machine learning which is developed basing on the statistical learning theory, in recent years this theory and algorithm have made rapid development. SVM classification method has applied to solve various practical problems, showing a lot better performance than the existing methods. C-SVC (support vector classification) is a classification method of SVM. C-SVC method is not only suit for the classification of linear separable problem, but also suit for the linear inseparable problem, so this paper applies of the C-SVC method.

3.1 C-SVC Algorithm

One: given a train set sample, T={$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$}, x_i is n dimensional vector, $x_i \in R^n$, $y_i \in \{1,-1\}, i=1,\ldots.,n$

Two: select a appropriate kernel function $K(x_i, x_j)$ and appropriate penalty factor c, construct and solve optimization problems

$$\min_a \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j a_i a_j \left(x_i \bullet x_j\right) - \sum_{j=1}^{n} a_j \tag{1}$$

$$s.t. \sum_{i=1}^{n} y_i a_i = 0, 0 \le a_i \le C, i = 1, ..., n \tag{2}$$

Then get the optimal solution $a^* = (a_1^*, \ldots, a_n^*)$.

Three: calculate $w^* = \sum_{i=1}^{n} y_i a_i^* x_i$ and select a positive component a^* $0 < a^* < C$, then calculate

$$b^* = y_i - \sum_{i=1}^{n} y_i a_i^* (x_i \bullet x_j) \tag{3}$$

Structure differentiation hyperplane $(w^* \bullet x) + b^* = 0$, then solve the decision function

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} a_i^* y_i K(x_i, x_j) + b^* \right) \tag{4}$$

3.2 Kernel Function

In the step of C-SVC algorithm involved that the SVM classification method need use kernel function $K(x_i, x_j)$, and we use ten fold cross validation method to choose the kernel function. 10 fold cross validation is a common method of accuracy test. In this method first set the training data into 10 equal parts, then 9 parts of which will do the training in turn and the other group do the test. The best value of 10 times is the result of algorithm accuracy. There are four kinds of kernel function [8].

Linear kernel function $K(x_i, x_j) = x_i \bullet x_j$

polynomina kernel function $K(x_i, x_j) = [r(x_i \bullet x_j) + 1]^q$

radial basis kernel function $K(x_i, x_j) = \exp\{-\frac{|x_i - x_j|^2}{r^2}\}$

sigmoid kernel function $K(x_i, x_j) = \tan[r(x_i \bullet x_j) + a]$

4 Data Processing

In this section, we processed the ICESat-GLAS binary format data than got waveform. After that we extracted parameters from waveform.

4.1 ICESat-GLAS Date Processing

ICESat-GLAS waveform data is specifically defined in binary format, including the metadata information and data information. First, raw data format was needed to convert; we used IDL to convert the binary data into ASCII format data. In order to effectively compare the waveform data, we also standardized the waveform data, then using Gaussian filter to smooth ICESat-GLAS waveform. After the waveforms were smoothed, we made the original waveform curve into multiple Gaussian curves. Detail process in [3]. Fig. 2 is graphics obtained after deal with matlab software.

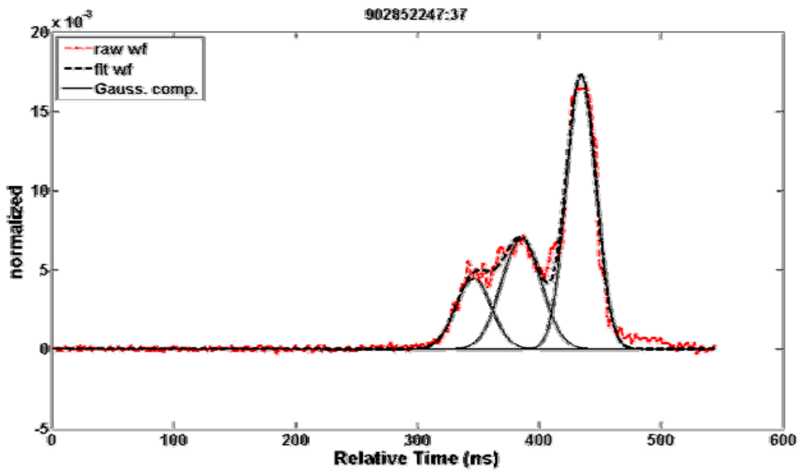

Fig. 2. Result of GLAS Waveform decomposition and Fitting

4.2 Waveform Parameters Extraction and Processing

From the physical sense, the different forest types have their own typical vertical structure and different typical LIDAR waveforms character. Moreover, from a mathematical point of view, echo waveform can be seen as a combination of several Gaussian curves, corresponding to different layers of the forest structure. Thus, by analyzing the corresponding Gaussian curves are expected to distinguish forest types. The vertical layers of coniferous tree is obvious more than the layers of Broad-leaved tree, so when lidar receive the energy from coniferous forest, the echo width is less than the Broad-leaved tree width. Thus, the slope of the Gaussian curve decomposition form waveforms of coniferous tree is larger than broad-leaved tree. Fitting the above analysis I extracted below reference points as described in this paper, extraction parameters diagram is shown as Fig. 3, t_1 is the effective echo time value of the first Gaussian curve, which is decomposition by the fitting waveform, t_2 is the time value corresponding to the peak of the first Gaussian curse, t_3 is the effective echo time value of the second Gaussian curve, t_4 is the time value corresponding to the peak of the second Gaussian curse, Q_1 is the energy value corresponding to the peak of the first Gaussian curse, Q_2 is the energy value corresponding to the peak of the last Gaussian curse. All parameters we extracted can be found in the files of the waveform generated.

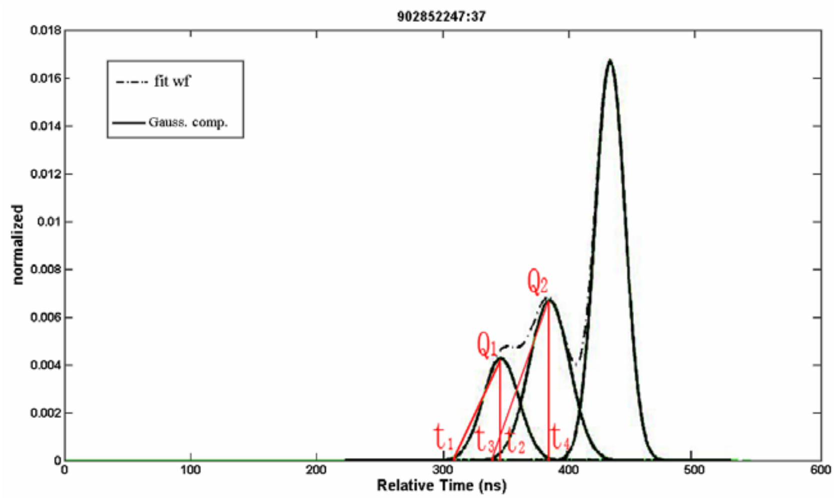

Fig. 3. Parameter extraction

When the original waveform curve was separated into many Gaussian curves. In Gaussian curve of *i*, assuming that Q_i is the standardized energy value of *i* Gaussian curve, t_{2i-1} is the effective echo time value of *i* Gaussian curve, t_{2i} is the time value corresponding to the peak of *i* Gaussian curve. We use formula 5 to calculate the slope of each Gaussian curve. The value of *n* is the number of Gaussian waveform curse besides of the last Gaussian waveform curse.

$$k_i = \frac{Q_i}{t_{2i} - t_{2i-1}} \tag{5}$$

We use formula 6 to calculate the mean slope of each Gaussian curve

$$\overline{\mathrm{K}} = \frac{\sum_{i=1}^{n} k_i}{\mathrm{n}} \tag{6}$$

We use formula 7 to calculate standard root mean square error of slope of each Gaussian curve

$$\Delta\mathrm{K} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(k_i - \bar{k}\right)^2} \tag{7}$$

In this paper we apply of the following processed parameters, the value of $\overline{\mathrm{K}}$, $\Delta\mathrm{K}$, θ, corresponding to each spot.

5 Result Analysis

We use C-SVC method to deal with the 62 groups of training samples, in C-SVC method we use ten fold cross validation to determine the kernel function. The result of ten fold cross validation is shown in Table 1. In Table 1 the best cross validation classification accuracy is 87.10%, so from Table 1 we can find that we should select Rbf as the kernel function in the classification method [9]. In Table 1 Poly is polynominal function, Rbf is radial basis function, Sig is sigmoid kernel function.

Table 1. Cross validation classification accuracy of different penalty parameter *c* and kernel function

c	0.01	0.1	1	10	100	1000
Poly	45.16	45.16	45.16	45.16	45.16	45.16
Rbf	74.19	74.19	83.87	87.10	77.52	77.52
Sig	74.19	74.19	74.19	74.19	74.19	74.19

We determine the value of penalty parameter *c* and kernel function coefficient *r* after we selected the Rbf kernel function. Table 2 shows that when *c*=10, *r*=0.01, the classification result is the best and the classification accuracy is 90.32%.

Table 2. Cross validation classification accuracy of different penalty parameter *c* and kernel function coefficient *r*

c / *r*	0.01	0.1	1	10	100	1000
0.001	74.19	74.19	85.48	82.25	83.87	79.03
0.01	74.19	74.19	87.09	90.32	82.25	82.25
0.1	74.19	74.19	85.48	87.09	79.03	80.64
1	74.19	74.19	74.19	79.03	75.80	75.80

When we select the Rbf kernel function, the value of *c*=10, *r*=0.01, we establish the classification model. The classification result when we built the model is shown in Table 3.

Table 3. The classification result of training data established model

Original Type	Divide forest type		Classification accuracy %	Total classification accuracy %	Kappa
	Broadleaved	Coniferous			
Broadleaved	44	2	95.65	90.32	0.7365
Coniferous	4	12	75.00		

Table 4. The classification result of the test data

Original Type	Divide forest type		Classification accuracy %	Total classification accuracy %	Kappa
	Broadleaved	Coniferous			
Broadleaved	42	4	91.30	82.26	0.5065
Coniferous	7	9	56.25		

Table 3 is the classification result of training data established model. Shown in Table 3, a total of training sample points is 62, 46 Broadleaved forest, 16 coniferous forest. The classification accuracy of broadleaved forest is 95.65% in which the number of correct classification is 44. The classification accuracy of coniferous forest is 75% in which the number of correct classification is 12. The total classification accuracy of training sample is 90.32%. The kappa coefficient is 0.7365. After we establish the classification model, we use the other 62 groups of test data to validation the model, the classification results shown in Table 4.

Table 4 is the classification result of test data. A total of 62 test sample points are Shown in Table 4, 46 Broadleaved forests, 16 coniferous forests. The classification accuracy of broadleaved forest is 91.30% in which the number of correct classification is 42. The classification accuracy of coniferous forest is 56.25% in which the number of correct classification is 9. The total classification accuracy of test sample is 82.26%. The kappa coefficient is 0.5065. Either in modeling or classification, the classification result of broadleaved forest is better than the classification result of coniferous forest when using the C-SVC method to classify the extracted parameters. There are two reasons why the classification accuracy of coniferous forest is not very high. First, although the broadleaved forest and coniferous forest have different, in many cases, the plots of each footprint has broadleaved trees and coniferous trees and the trees within the footprint is large difference between high and low which impact the vertical structure features distinction. Second we used the parameter θ to classification the forest types. The value of θ ranges from $0 \sim 45°$. When the slope angle increases, the ground return and the forest under the slope return information will mix, then the laser radar echo data will appear superimposed waveform. This will make Broad leave forest and coniferous forest return energy change, making the classification accuracy decline.

6 Conclusion

In this paper, we used C-SVC method to analyze the waveform parameters which extracted from ICESAT-GLAS. In this method we apply ten fold cross validation method for modeling and the model accuracy reaches 90.32%. When we used the other 62 groups of test data to test the model we get the classification accuracy of 82.26%. The classification result shows that C-SVC method can well distinguish broadleaved forest between coniferous forest. In further work we will apply of other methods to study on the waveform parameters extracting from ICESAT-GLAS and compare the classification accuracy of different methods.

Acknowledgments. The Fundamental Research Funds for the Central Universities (Grant: DL09CA08). And we would like to thank Ice Data Center for providing ICESat-GLAS data.

References

1. Pang, Y., Yu, X.F., Li, Z.Y.: Waveform Length Extraction from ICEsat GLAS Data and Forest Application Analysis. Scientia Silvae Sinicae 42, 137–140 (2006)
2. Nelson, Ranson, K.J.: Estimating Siberian timber volume using MODIS and ICESat-GLAS. Remote Sensing of Environment 113, 691–701 (2009)
3. Xing, Q., Wang, L.H.: ICESat-GLAS Full Waveform based Study on Forest Canopy Height Retrieval inSloped Area —A Case Study of Forests in Changbai Mountains, Jilin. Geomatics and Information Science of Wuhan University 34, 696–700 (2009)
4. Liu, G.Z., Shi, W.Z., Li, D.R., Qin, Q.Q.: Partially Supervised Classification of remote sensed Imagery using Suppor Vector Machines. Journal of Remote Sensing 9, 363–373 (2005)
5. Suykens, J., Vandewalle, J.: Least squares support vector machine classifiers. Neural Processing Letters 9, 293–300 (1999)
6. Zhang, Y., Feng, S.Q., Pang, D.Z., Tang, Q.: The Support Vector Classification Model of the Rhinopithecus Roxellanae Subspecies. Application of Statistics and Management 27, 211–215 (2008)
7. Xu, Y., Shi, L.K., Wang, H.M.: A Study of SVM Classification of College Students'Decision about Graduation. Journal of Shijiazhuang University 10, 34–41 (2008)
8. Lin, S., Xun, P., Liu, Q.: Traffic.: classification based on support vector machine. Application Research of Computers 25, 2488–2492 (2008)
9. Yang, J.Y., Zhang, Y.Y., Zhu, Y.S.: Classification Performance of Support Vector Machine with ε -Insensitive Loss Function. Journal of Xi'an Jiaotong University 41, 1315–1321 (2007)