Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (18): 3770-3773, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3770.3773

Land Use/cover Classification with Classification and Regression Tree Applied to MODIS Imagery

[1]Jianguang Liu, [1]Danfeng Sun, [1]Feng He, [1,2]Weiwei Zhang and [1,3]Xiaoke Guan
[1]College of Natural Resources and Environmental Science, China Agricultural University, Beijing, China
[2]Beijing Academy of Agricultural and Forestry Sciences, Beijing, China
[3]School of Political Science and Law, Zhengzhou University of Light Industry, Zhengzhou, China

Abstract: This study attempted to develop a low-cost, high-precision method to acquire land use/cover data by combining multi-temporal and multi-spectral Moderate Resolution Imaging Spectroradiometer (MODIS). The results indicate that Classification and Regression Tree (CART) algorithm clearly outperforms the Maximum Likelihood (ML) in land use/cover classification using MODIS and the first principal component (PC1) with multi-spectral MODIS image that reflected more soil information can efficiently improve the accuracy of classification based on MODIS NDVI time series.

Key words: Land use/cover, principal component analysis, classification, MODIS

INTRODUCTION

Land use/cover information is an important evaluation index of regional eco-environment and is the essential data of regional sustainable development researches. Therefore, it is indispensable to establish a low-cost, efficient and accurate method for land use/cover monitoring. Nowadays, with the rapid development of sensor technology, remote sensing images have become the mainly employed data. Particularly, MODIS sensors aboard Terra and Aqua satellites can receive high time-resolution, multi-spectral and intermediate spatial-resolution image data, which are freely available and have been applied successfully to the large and medium scale land use/cover monitoring (Wardlow and Egbert, 2008; Friedl *et al*., 2002).

Multi-temporal and multi-spectral informations are two merits of MODIS, which were used in land use/cover monitoring (Friedl *et al*., 2002). Multi-temporal information can be used to describe land cover change accurately; and multi-spectral information can reflect abundant surface features at given time (Tingting and Chuang, 2010; Reed *et al*., 1994; Fraser *et al*., 2005). How to extract and mine both informations and how to complement each other's disadvantages become a key point in land use/cover classification based on MODIS. Moreover, there are many classification algorithms of remote sensing image, such as decision trees, maximum likelihood and neural network classification (Xu *et al*., 2005). The analysis of performance of these algorithms for a given research is another key point in land use/cover classification based on MODIS.

In order to establish an accurate method to monitor land use/cover of Beijing using MODIS, sophisticated feature selection and algorithms to optimize classifier's performance is required. The contributions of this study include:

- We attempted to extract the optimized classification features that complement each other from NDVI time series and multi-spectral information
- This study compared classification effect between the classification and regression tree and the traditional maximum likelihood algorithms

Study area: Beijing is located at N39'26" to N41’03” and E115'25" to E117’30’’ and has an area of about 16,414 km^2. The city is characterized by alluvial plains in the south and east and hills and mountains in the north, northwest and west. Beijing has a climate with four distinct seasons characterized by hot, humid summers (due to the East Asian monsoon) and cold, windy, dry winters (reflecting the influence of the vast Siberian anticyclone). The average annual temperature is 12.5°C and the average precipitation is 566.1 mm. Additionally, agricultural cropping patterns are dominated by double cropping and single cropping.

MODIS data acquisition and preprocessing: In this study, we primarily employed the MOD09 product of

Corresponding Author: Danfeng Sun, College of Natural Resources and Environmental Science, China Agricultural University, Beijing, China

Table 1: MODIS data used in this study

Data	Resolution (m)	Bands	Total of time-phases	Time range
MOD09Q1	250	1, 2	46	Whole year
MOD09A1	500	1, 2, 3, 4, 5, 6, 7	3	Mar, May, Aug

MODIS of 2008, which consisted of eight days of composite surface reflectance images freely available from the MODIS Data Product website (http://modis. gsfc.nasa.gov) (Table 1). A set of 46 time-phase images can cover a full year observation period. Image preprocessings including file format conversion, image mosaic, resampling and projection conversion were implemented using the MODIS Reprojection Tool.

MODIS NDVI TIME SERIES ACQUISITION AND PREPROCESSING

The normalized difference vegetation index (NDVI) derived from remote sensing has been shown to be related to biophysical variables that control vegetation productivity, such as the leaf-area index and the NPP (Myneni *et al*., 1997; Rasmussen, 1998). Given that MODIS has a great advantage of high time resolution, MODSI_NDVI time series can be used to reflect vegetation activty effectively.

The NDVI was calculated by Eq. 1, where ρ_{NIR} is the near-infrared band and $\tilde{n}_R$ is the red band. The MODIS NDVI time series were then generated by layer stacking. Although, MOD09 production has removed some of noise by the Maximum Value Composite (MVC) method, there was still some noise caused by clouds and the atmosphere.

The Savitzky-Golay filter, which is a very effective method of obtaining high-quality NDVI time series data (Chen *et al*., 2004), was used in this study. The processed NDVI time series were more beneficial to trend analysis and information extraction:

$$\text{NDVI} = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R} \tag{1}$$

ESTABLISHMENT OF LAND USE/COVER CLASSIFICATION SYSTEM

In light of the biophysical characteristics of the study area, Beijing was divided into seven fundamental land use/cover categories. The classification scheme and detailed descriptions are given in Table 2.

CLASSIFICATION FEATURES FROM NDVI TIME SERIES

With the sampling statistical analysis method, the typical MODIS NDVI time series curve of each land use/cover category could be established. Analyzing differences between their characteristics, we extracted six discriminating classification features that can quantify phenophase, plants growth rate and other features by IDL and ENVI 4.4, including: (1) Cropping index; (2) Length of growing season; (3) NDVI growth rate during 18-27 time-phases; (4) Mean NDVI during 18-35 time-phases; (5) Minimum NDVI throughout the year; (6) Amplitude of NDVI throughout the year.

Table 2: Land use/cover classes used in this study

Land use/cover categories	Description
Construction	City and rural settlements.
Water	Rivers, lakes, and reservoirs
Orchards	Cash trees that are artificially planted
Forests	Coniferous forests, broadleaf forests, shrubs, etc.
Double cropping cropland	Cropland on which crops are sown twice and harvested twice a year
Single cropping cropland	Cropland on which crops are sown once and harvested once a year
Other land uses	Barren land, sparse grassland, tidal flats, etc.

CLASSIFICATION FEATURE FROM MULTI-SPECTRAL INFORMATION

Since NDVI mainly reflect vegetation information, classification features from NDVI time series are difficult to distinguish land types with sparse vegetation, such as construction, bare land and tidal flat. There are significant differences in spectral characteristics of their surface soil. Therefore, multi-spectral image of March that can show more soil information was used to improve classification accuracy. Principal component transform was used for information compression. The first principal component (PC1) with 97.7% of contribution rate was taken as one classification feature.

CLASSIFICATION AND REGRESSION TREE ALGORITHM

Decision trees have several advantages for remote sensing applications owing to their relatively simple, explicit and intuitive classification structure. Furthermore, decision tree algorithms are strictly nonparametric and therefore make no assumptions regarding the distribution of input data. Moreover, they are flexible and robust with respect to nonlinear and noisy relationships among input features and class labels (Friedl and Brodley, 1997). In this study, the Classification and Regression Tree (CART) algorithm was used and the Gini index of impurity for node splitting was selected (Breiman *et al*., 1984; Yohannes and Hoddinott, 1999). The Gini index of impurity is defined as Eq. 2, where p (j|t) is the proportion of class j at node t:

$$i(t) = 1 - p^2(j \mid t) \quad (2)$$

According to visual interpretation of MODIS images of July and August and some SPOT-5 images in 2008, 2,014 pixels were selected as training samples. Seven classification features, including 6 that from NDVI time series and 1 that from multi-spectral information, were taken as input bands of classification. Through the performing of CART algorithm, a decision tree with 29 leaf nodes was created.

MAXIMUM LIKELIHOOD ALGORITHM

The Maximum Likelihood (ML) based on statistical theory is a classic supervised classification algorithm and often used in the classification of remote sensing images. With the ML algorithm, probability density functions are built for each class based on the training data's spectral values. During classification all unclassified pixels are assigned class membership based on the relative likelihood (probability) of that pixel occurring within the probability density function of each class (Lillesand *et al*., 2004). We performed the ML algorithm to land use/cover classification using the same training samples and input bands of classification as the CART algorithm.

ACCURACY EVALUATION OF CLASSIFICATION

The accuracy evaluation of the classification results based on CART and ML algorithms was conducted according to a reference data set, which was created from actual ground data and visual interpretation of some remote sensing data from SPOT5 in 2008. A total of 6,237 pixels from land use/cover map that account for 0.2% of study area were randomly selected for evaluation. As shown in Table 3, the overall accuracy of classification of CART algorithm was 80.5%, which is 2.9% higher than that of the ML. Except for forest, accuracy of each land use/cover category with CART algorithm outperforms that with ML algorithm. CART algorithm was more effective than ML in this study.

Table 3: Comparison of classification accuracy

	Classification algorithm		Classification data	
Parameters	CART	ML	Employed PC1	Unemployed PC1
Construction	80.2	78.4	80.2	74.9
Water	93.7	93.3	93.7	93.7
Garden	75.9	73.4	75.9	72.6
Forest	89.5	90.2	89.5	87.7
Double cropping cropland	80.6	78.0	80.6	80.1
Single cropping cropland	78.4	76.2	78.4	77.5
Else land	62.3	55.7	62.3	48.6
Overall accuracy	80.5	77.6	80.5	76.9

Otherwise, to measure the performance of PC1, we implemented another CART classification that unemployed PC1 to compare with that employed PC1. As shown in Table 3, PC1 with multi-spectral information can significantly improve the classification accuracy, particularly, construction and else land, whose classification accuracy have increased 5.3 and 13.7%, respectively. The overall classification accuracy increased 3.6% when PC1 participated in classification.

CONCLUSION

Although MODIS data only has intermediate spatial resolution, its merits of high time resolution and multi-spectral enables that MODIS NDVI time series can be used to reflecte surface vegetation activity and useful classification informations refered to vegetation changes can be extracted. Complemented with multi-spectral information, a high accurate classification can be achieved. In this study, the overall classification accuracy is 80.5%, indicating that it is totally feasible to use MODIS multi-temperal and multi-spectral informations for monitoring the land use/cover changes of Beijing.

Results show that the first principal component (PC1) of the multi-spectral MODIS image of March obtained by Principal Component Analysis (PCA) can significantly improve the classifications of land types with sparse vegetation, which remedy the shortage of NDVI time series; CART algorithm can not only display classification rules, which is beneficial to acquisition, accumulation and transfer of classification knowledge, but also obviously outperform the classic maximum likelihood in the classification of land use/cover of Beijing using MODIS. These methods and results can be used as a reference of land use/cover monitoring of Beijing or other regions.

REFERENCES

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone, 1984. Classification and Regression Trees. 1st Edn., Wadsworth Int. Group, Belmont, California, USA., ISBN: 978-0534980542, Pages: 368.

Chen, J., P. Jonsson, M. Tamura, Z. Gu, B. Matsushita and L. Eklundh, 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sens. Environ., 91: 332-344.

Fraser, R.H., A. Abuelgasim and R. Latifovic, 2005. A method for detecting large-scale forest cover change using coarse spatial resolution imagery. Remote Sens. Environ., 95: 414-427.

Friedl, M.A. and C.E. Brodley, 1997. Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ., 61: 399-409.

Friedl, M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang and D. Muchoney *et al*., 2002. Global land cover mapping from MODIS: Algorithms and early results. Remote Sensing Environ., 83: 287-302.

Lillesand, T., R. Kiefer and J. Chipman, 2004. Remote Sensing and Image Interpretation. 5th Edn., John Wiley and Sons, New York, USA.

Myneni, R.B., R. Ramakrishna, R. Nemani and S.W. Running, 1997. Estimation of global leaf area index and absorbed PAR using radiative transfer models. IEEE Trans. Geosci. Remote Sens., 35: 1380-1393.

Rasmussen, M.S., 1998. Developing simple, operational, consistent NDVI-vegetation models by applying environmental and climatic information: Part I. Assessment of net primary production. Int. J. Remote Sens., 19: 97-117.

Reed, B.C., J.F. Brown, D. VanderZee, T.R. Loveland and J.W. Merchant *et al*., 1994. Measuring phenological variability from satellite imagery. J. Vegetation Sci., 5: 703-714.

Tingting, L. and L. Chuang, 2010. Study on extraction of crop information using time-series MODIS data in the Chao Phraya Basin of Thailand. Adv. Space Res., 45: 775-784.

Wardlow, B.D. and S.L. Egbert, 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. Remote Sens. Environ., 112: 1096-1116.

Xu, M., P. Watanachaturaporn, P.K. Varshney and M.K Arora, 2005. Decision tree regression for soft classification of remote sensing data. Remote Sens. Environ., 97: 322-336.

Yohannes, Y. and J. Hoddinott, 1999. Classification and Regression Trees: An Introduction. International Food Policy Research Institute, USA.