

Research and application of Object-oriented remote sensing image classification based on decision tree

Peng Dou^{1,3} Liang Zhai¹ Huiyong Sang¹ Wenhan Xie²

1. Key Laboratory of Geo-informatics of NASG, Chinese Academy of Surveying and mapping, Beijing, China

2. Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying and mapping, Beijing, China

3. Lanzhou Jiaotong University, Lanzhou, China

Abstract—The paper mainly introduced the basic principle and method of Object-oriented classification technology and made studies and research on decision tree classification technology. In addition, the remote sensing image was classified with Object-oriented way by using C5.0 algorithm, and compared with the result based on pixels classification and the result of SVM Object-oriented classification, which has improved that the accuracy and effectiveness of decision tree classification can be used in Object-oriented remote sensing image classification.

Keywords: remote sensing image classification; object-oriented; decision tree classification; C5.0 algorithm

I. INTRODUCTION

With the development of remote sensing technology, remote sensing images had been widely used in city planning, land use investigation, environmental governance, emergency service etc. Remote sensing image records and reflects the electromagnetic strength of the ground features, through the interpretation of remote sensing image the classes' attributes and distribution characteristics of ground objects could be identified. The traditional classification of remote sensing images usually used the information of spectrum to extract the features, this is the first level of understanding about remote sensing images. As the lack of spatial information and the correlation between pixels, the problem of "the same light with different spectrum" and "the same spectrum with different object" is difficult to solve, and the accuracy is also been limited[1]. The high resolution images has more information about the spatial, the geometry structure of features and the grains, the adjacent pixels on it have high correlation each other so the traditional technology could not meet the higher classification requirements. Rise the understanding of the remote sensing images to the second level, in the process of classification the texture[2], shape, size and space factors of the features would be made full account, and the pixels would be divided into homogeneous plaque which have the same characteristics. Based on this factor, object-oriented method could be used to make classification strategy and create rules. This technology had been more widely used in high-resolution image classification and realized by various classifiers. For example the SVM classifier in ENVI_EX, the nearest classification and the membership function classification in eCognition etc. The decision tree classification is a supervised classification technology which based on spatial

data mining and knowledge discovery, it can extract features automatically and merge the knowledge out of the image together. Accordingly it could be used to classify the remote sensing images combine with some knowledge and spatial texture information.

II. OBJECT-ORIENTED REMOTE SENSING IMAGE CLASSIFICATION

A. The Concepts of object-oriented Classification

The traditional per-pixel classification focuses on a single pixel, but in a remote sensing image it contains a lot of geographical entities, those entities are clusters which have the same attributes and adjacent each other. On this basis, from a user's perspective, the images could split into plaques which represent geographical objects, considering the shape, texture and spectral information of targets, this kind of classification is called object-oriented classification. As the objects which are constituted by a plurality of adjacent pixels contain more information, such as spectrum, shape, size, structure, spatial location and distribution, the object-oriented technology could realize the classification of remote sensing image and extract information of the features in a higher level and reduce the loss of information. In the process of object-oriented classification, two factors are important, one is how to segment the images to make the objects have more information, another one is how to extract the characteristics rapidly and make the computers know what kind of objects in the images and how to classify and get, which use the knowledge of people's cognition about ground clutter.

B. The General Process of object-oriented Classification

The object-oriented remote sensing image classification could be divided into several stages: (1) Data pre-processing, including image radiometric correction, image enhancement, as well as other categories of auxiliary materials finishing. (2) Image segmentation, split the image into a series of meaningful plaques and each plaque corresponding to geographical entities in the image. (3) Calculated properties, select the characteristics of the target objects and make combination. (4) Create classification rules using expert knowledge and other information for classification. (5) Extract information and make of map or other products[3].

III. DECISION TREE CLASSIFICATION

A. Principles of Decision Tree Classification

The decision tree is a tree structure that is similar to flowcharts, of which each internal node is a test on an attribute and the branches can be considered a output of the test, while ,the leafs of the tree expresses classes or classes distribution. Decision tree classification technology merely based on spatial data mining and knowledge discovery, it's a supervised learning method which put emphasis on logic. Unlike previous classifications made by tree structure or rules, decision tree classification technology requests users less experience and lower level of professional knowledge and it also broads the types of data processing. Using decision tree in remote sensing image classification not only makes full use of the information the remote sensing image contained ,but also makes the best of the geographical knowledge stored in GIS database and other auxiliary information to improve classification accuracy[4]. Beyond that, there are several advantages about the decision tree classification as follows[5]: (1)The decision tree has better flexibility and robustness and can be used for the data with a complex spatial distribution and the multi-source data with different statistic distribution and scale.(2)The rules created by the decision tree is simple to understand ,computers can make calculations whit it easily. The advantages of the decision tree to process massive amounts of data are obviously, especially for the remote sensing images.(3)The impact made by noises in the training set can be reduced by a few of sample data.(4) Data preparation is often simple, without to standardized the data.

B. C5.0 Decision tree Classification

C5.0 is one of the most forward algorithm about decision tree classification, it's a member of Clementine's decision tree model and the initial prototype of it is the famous ID3 lgorithm, which was come forward by J.R Quinlan in 1979.In 1993 Quinlan developed the C4.5 algorithm based on the ID3,the algorithm used information gain ratio as a standard to select the split attribute, which overcame the shortage of tending to the attributes with more values when use information gain in ID3 to select the split attribute.C5.0 algorithm was established on the bases of C4.5 and used the maximum information gain ratio as attributes selection criterion. The best optimal threshold value could be find to decide the classes with cutting and merging the leafs of the decision tree. Inherited all the characteristics of C4.5, C5.0 algorithm used boosting technology and EBP pruning strategies to improve the precision.

IV. EXPERIMENT

A. Data Preparation

The image segmentation in the experiment was made by the software ENVI_EX and the images were classified using the C5.0 classification program which was developed by C++ and ArcEngine .After the accuracy evaluation and comparative

analysis, it was proved that the object-oriented classification based on C5.0 algorithm was effective. The remote sensing image used in this test is the pretreated World View image, of which the spatial resolution is 0.5m (panchromatic band) and 2m (multispectral), as shown in Figure 1. According to actual situations, six kinds of ground information would be get in the experimental area, such as woodland, grassland, bare land, cultivated land, artificial cover (residential areas, roads, etc.) and water. Before the classification sample data must be gathered, it can be realized using ArcGIS to create a point vector file , add the CLASS field, the digits 1-6 respectively represents woodland, grassland, bare land, cultivated land, artificial cover and water. The sampling should be done compared with the remote sensing image and the distribution of sample data points should be even and representative, 30-40 points would be the best choice for each category of features.



Fig. 1 World view panchromatic remote sensing image

B. The process of experiment

(1)Image segmentation. In the experiment, the soft ENVI_EX was used to segment the image, the segment scale was 40 and the parameter for merge was 60. (2)Generate training set. This step was to use some data made a model to find the rules.(3) Generate rule set. Using the program of C5.0, the rules for classification would be created as shown in Figure 2.(4)Execute classification. Took the rule set file and segmented vector file as the input file and execute classification, the final result as shown in Figure 3..

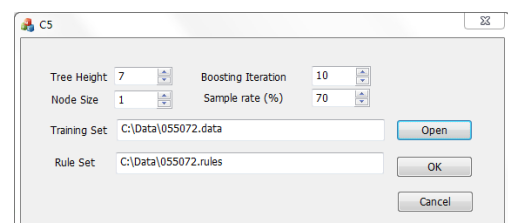


Fig. 2 Create rule set

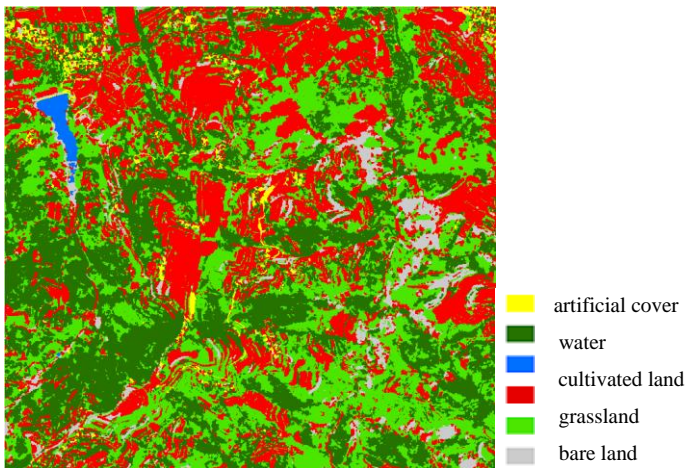


Fig. 3 Classification results

C. Precision evaluation

In the experiment, the comparison analysis between object-oriented remote sensing image classification technology and traditional per-pixel classification technology was made (chose the maximum likelihood classifier to classify the image with per-pixel classification technology), the tow results as show in Figure 4 and Figure 5. The object-oriented remote sensing image classification used tow algorithm to realize, they are C5.0 algorithm and SVM respectively, the results as show in Figure 5 and Figure 6. The precision index can be find in Table1, Table2 and Table3.

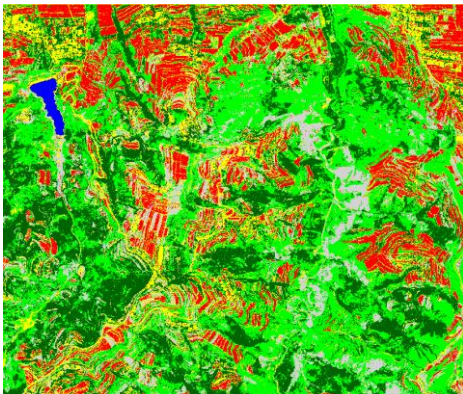


Fig. 4 Per-pixel classification

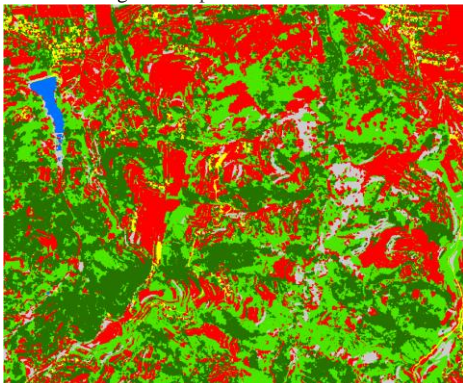


Fig. 5 Object-oriented classification(C5.0)

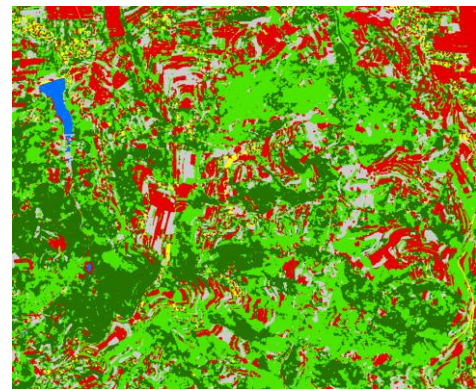


Fig. 5 Object-oriented classification(SVM)

In the comparison of the classified results which based on pixels and object-oriented, it could be discovered that some pixels belonged to residential areas were classified as roads falsely, the problem of “the same spectrum with different object” is very obviously. It was resolved by object-oriented technology, due that the object-oriented technology used both spectrum information and spatial factors, and made the result more equitable. In addition object-oriented technology could make trade-offs on some small objects through changing the segmentation scale. The per-pixel classification is sensitive to the noise [6], so the image classified seems somewhat broken. The object-oriented technology has good advantages to the continuity and integrity of the patterns and avoids the salt and pepper effect which was results in by per - pixel classification. Compare with the object-oriented classification results between C5.0 and SVM, it is obviously that the accuracy of C5.0 decision tree classification is higher than SVM for the reason that C5.0 decision tree classification used Adaboosting technology to make more classifier to classify the images and used “Weighted voting” method to get the best result for each objects. Even so, there was a small amount of cultivated land was error classified, to solve this problem, some soil data could be used as the characteristics of target objects to classify[7].

V. SUMMARY

This paper made research on object-oriented classification and using the C5.0 algorithm classified the remote sensing image based on Object-oriented technology. Compared with result of per - pixel classification and SVM Object-oriented classification, the research gets conclusion as follows:

(1)To the traditional per-pixel classification, the spectrum, shape, size, structure, space location and distribution information of the remote sensing image can be fully use in object-oriented technology and it increases the data source to classify and improves the classification accuracy.

(2)As the per-pixel classification is sensitive to the voice, it's often result in figure spot broken and “salt and pepper effect”. By using object-oriented technology to classify the remote sensing image, this problem could be solved effectively. The geographic entities keep their integrity, with the attributes, they can be stored into the GIS database easily [8].

(3)It's important to get the entities which includes more geographic information through the image segmentation in the process of remote sensing image classification, while using

what kind methods to do what kind of classification decided the accuracy and effect of the classification.

(4)C5.0 decision tree algorithm could create classification rules rapidly, while the rules set up are automatic. The decision tree classification is adaptable and has high accuracy and it's

suited to the geographic information extraction in a large area with a rapid speed. Any way there are no special requirements to the data, so any kind of data for classification can be used sufficiently.

TABLE I. TABLE 1 CLASSIFICATION BASED ON PIXELS(MAXIMUM LIKELIHOOD)

per-pixel classification	artificial cover	woodland	water	cultivated land	grassland	bare land
Users Accuracy	81.63%	96.36%	100.00%	83.33%	92.86%	97.37%
Producers Accuracy	97.56%	92.98%	95.00%	96.77%	86.67%	74.00%
Overall Accuracy	90.18%					
Kappa Statistics	0.8799					

TABLE II. OBJECT-ORIENTED CLASSIFICATION BASED ON C5.0 ALGORITHM

SVM Object-oriented classification	artificial cover	woodland	water	cultivated land	grassland	bare land
Users Accuracy	100.00%	86.36%	100.00%	73.42%	84.62%	91.67%
Producers Accuracy	85.37%	100.00%	100.00%	93.55%	73.33%	66.00%
Overall Accuracy	85.82%					
Kappa Statistics	0.8528					

TABLE III. OBJECT-ORIENTED CLASSIFICATION BASED ON SVM

C5.0 Object-oriented classification	artificial cover	woodland	water	cultivated land	grassland	bare land
Users Accuracy	100.00%	93.44%	100.00%	92.06%	86.67%	94.00%
Producers Accuracy	87.80%	100.00%	100.00%	93.55%	86.67%	94.00%
Overall Accuracy	93.45%					
Kappa Statistics	0.92					

ACKNOWLEDGEMENTS

This work was funded by National Natural Science Foundation of China (41001299), National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2012BAH28B01), Open Research Fund of Key Laboratory of Geo-Informatics of National Administration of Surveying, Mapping and Geo-information (777131404), Basic Research Fund of Chinese Academy of Surveying and Mapping (7771312). We also would like to acknowledge every member of the Global Land Cover project team at CASM.

REFERENCE

- [1] Qingchun Ge, Linhan Zhang and Jie Yang, "Objected-oriented Classification of Remote Sensing Image Based on Decision Trees," Remote Sensing Information. Beijing. China, vol.31, pp. 86-89, February 2009
- [2] Assia Kourgli, Mounira Ouarzeddine, Youcef Oukil, and Aichouche Belhadj-Aissa, "Texture modelling for land cover classification of fully polarimetric SAR images," International Journal of Image and Data Fusion. London.UK, vol. 3, pp.129-148, June 2012
- [3] Bao Cao, Qiming Qin and Haijian Ma, "Application of Object-Oriented Approach to SPOT5 Image Classification:A Case

- Study in Haidian District,Beijing City," Geography and Geo-Information Science. Beijing.China,vol.22,pp. 86-89, March 2006
- [4] Lihua Jiang, Xiaorong Yang, " Classification Methods of Remote Sensing Imag Based on Decision Tree Technologies," Agriculture Network Information.Beijing.China, vol.10,pp. 34-42 ,October 2009
- [5] Zhihui Wang, Shiming Li and Yiwei Zhang, "Methodsological Study on the Detection of the Variations of Forest Resources Based on C5.0 Algorithm-A Case of Culai Forest in Shandon," Journal of Northwest Forestry University.Xian.China vol.26,pp.185-191, September 2011
- [6] Yunhao Cheng, Tong Feng, PeiJun Shi and Lingfei Wang, "Classification of Remot Sensing Image Based on Object Oriented and Class Rules," WUHAN UNIVERSITY JOURNAL.Wuhan.China,vol.31,pp.316-320,April 2006
- [7] Nikolaos E. Mitrakis, Giorgos Mallinis, Nikos Koutsias, and John B. Theocharis, "Burned area mapping in Mediterraneanenvironment using medium-resolutionmulti-spectral data and a neuro-fuzzy classifier," International Journal of Image and Data Fusion. London.UK, vol. 3,pp. 299-318, June 2012
- [8] Changbao Yang and Jihong Ding, "Study of Object-Oriented Based Remote Sensing Image Classification," Journal of Jilin University.Jilin.China, vol.7,pp. 642-646, July 2006