

Classification Methods of Remote Sensing Image Based on Decision Tree Technologies

Lihua Jiang^{1,2}, Wensheng Wang^{1,2}, Xiaorong Yang^{1,2}, Nengfu Xie^{1,2},
Youping Cheng³

¹Agriculture Information Institute, Chinese Academy of Agriculture Sciences, Beijing,
100081, China

²Key Laboratory of Digital Agricultural Early-warning Technology, Agriculture Information
Institute, Chinese Academy of Agriculture Sciences, Beijing, 100081

³Agriculture Bureau, Huailai County, Hebei Province, 075400, China
{jianglh, wangwsh, yxr, nf.xie, youping} @caas.net.cn

Abstract. Decision tree classification algorithms have significant potential for remote sensing data classification. This paper advances to adopt decision tree technologies to classify remote sensing images. First, this paper discussed the algorithms structure and the algorithms theory of decision tree. Second, C4.5 basic theory and boosting technology are explained. The decision tree technologies have several advantages for remote sensing application by virtue of their relatively simple, explicit and intuitive classification structure.

Keywords: Decision tree; Classification; Remote sensing image

1 Introduction

Classification and Extraction of remote sensing information has been an important content in remote sensing technology field. In remote sensing classification application, traditional classification methods [1] such as supervised classification and unsupervised classification and artificial neural nets classification [2] and expert system classification are both based on spectral image features. But because image self has the phenomenon that the same thing has different spectrum, and different things have the same spectrum, the classification methods that only rely on ground spectrum features always turn up many misclassifications and omission errors [3]. Lots of study indicates that classifications combined with image spectrum information and other assistant information can improve precision of classification largely. Decision tree classification as spatial data mining and knowledge discovery [4] supervised classification method, breaks through the problem that construction of previous classification tree or classification rule always take advantage of ecology and remote sensing knowledge ex-ante certainty and the results always closely related with experience and professional knowledge [5]. It obtains classification rules by means of decision study process and needn't satisfy normal distribution. It can use

earth knowledge in GIS database to help classify and improves precision of classify. At present, decision tree classification [7] has applied in remote sensing image information extraction and land utilization coverage classify. In American, USGS and EPA etc. departments have united taken out USA land coverage database plan and decision tree classification technology has not only applied in land classify but also urban density information and crown layer density information extraction. The land classify precision has reached 73%-77% and urban density information extraction precision has reached from 83% to 91% and tree crown precision has reached 78%-93% [8]. Mapping efficiency has improved 50% and can satisfied with large scale land classify data production requirements.

Decision tree study method is one of data mining methods to work out classify problem in practical application [9]. It can reason classify rules of decision tree form of expression. The great virtue of decision tree is that study process needn't users know a lot of background knowledge. As long as trained examples can expressed by "property - result" and use this algorithm to learn. Classify knowledge obtained by decision tree is easy to express and apply. At present, foreigner scholars have already used decision tree to obtain knowledge and applies in spatial analysis and study process [10].

2 Decision tree arithmetic

Decision tree is a method which can inductive learn by training samples and build up decision tree or decision rule and then use decision tree or decision rule to classify data. Decision tree is a tree construction. It is composed of a root node, a series of internal nodes and leaf nodes. Every node can have only one father node and two or more child nodes. Nodes are connected with each other by branches. Every internal node correspond a test properties or properties group and every side correspond every possible value of property. Leaf node correspond a class property value and different leaf node can correspond the same class property value. Decision tree can not only be expressed by tree, but also a team of IF-THEN production rules. Every road from root to leaf correspond one rule and the condition of rule is to option of all nodes property values. And result of rule is class property of leaf node in the road. Compared with decision properties, rules are more simple and convenient to understand, use and mend and can make up the base of expert system. So rules are used more and more in actual application. This paper mainly introduces a widely used in remote sending application arithmetic-classify and regression tree and another decision tree arithmetic C4.5.

2.1 Classify and Regression Tree (CART)

Classification and regression tree is a common tree growth algorithm. It is presented by Breiman etc. [11] and it is a supervised classify. It trains the samplings to construct binary tree and decode to classify. The feature is to take advantage of the binary tree-structured fully, in other words, root node includes all samplings. Root node is divided up into two child node in definite divide rules. This process repeats again in

child node and becomes a regression progress until the child node can not be divided into two child nodes. Train of thought of construction a CART is that based on the whole sampling data, build up a multilevel and multi-leaf nodes tree to reflect relations between nodes and then cut the tree to build up a series of child trees and select appropriate tree in order to classify the data. In details, the process includes building up a tree and pruning a tree.

2.1.1 Tree growth.

A discrimination of tree nodes is named a branch and it corresponds to a subset of training samples. Branch of root node corresponds to the whole training samplings. Thereafter discrimination is process of partition training samplings. So process of building up a tree is querying property to produce partition rules. In this paper, CART adopts a index called “node impurity level”: $i(N)$ presents the impurity level of node N . When mode data in node come from the same category, $i(N)=0$; when categories of data distributes evenly, $i(N)$ will be very big. Partition rules are produces based on minimal value of impurity level function. Here two impurity functions are introduced.

(1) “Entropy Impurity”, is also called information impurity:

$$i(N) = -\sum_j P(w_j) \log_2 P(w_j) \quad (1)$$

Including, $P(w_j)$ is the calculus of probability accounting for w_j mode sampling data that node N belonged to of the whole samplings. According to the characters of entropy, if all mode data come from the same category, the impurity level is zero; or else impurity level is more than zero; when all categories data appears with the same calculus of probability, entropy is the maximum.

(2) VAR Impurity—“Gini impurity level”.According to node samplings come from different categories and it is related with total distribution variance, below formula is put forward.

$$i(N) = -\sum_{i \neq j} P(w_i)P(w_j) = 1 - \sum_j P^2(w_j) \quad (2)$$

Meaning of “Gini impurity level” is to represent error rate of category making in node N .

When given a tree which has grown to node N and the node is attribute queried, a visible heuristic train of thought is to select the query whose impurity level drops fastest. The impurity level drops can be noted:

$$\Delta i(N) = i(N) - P_L i(N_L) - (1 - P_L) i(N_R) \quad (3)$$

Including, N_L and N_R are separately left node and right node; $i(N_L)$, $i(N_R)$ are separately impurity level. P_L is the probability that when query T is adopted, the tree grows from N to N_L . And the optimum query value S is the maximum value of $\Delta i(T)$.

2.1.2 Tree pruning

If we persist in building up trees until all the leaf nodes reach minimal impurity level. The data will be fitted excessively and classify tree will degenerate to a convenient lookup table. It is maybe not good for noise signal lamp generalization performance of bigger Bayes error. On the contrary, if the branches stop too early, error of training sampling will be not small enough resulted in category performance is very poor. A main stopping branch method is pruning, at the same can prevent the tree growing too gigantic. In this paper, below index is adopted to reach the aim.

$$\text{cost} = \alpha \bullet \text{size} + \sum_{\text{leaf node}} i(n) \quad (4)$$

cost is represented cost function of tree weighting error probability and complexity penalty summation. size is represented leaf node quantity to weight complexity of

tree classifier. α is represented complexity index. $\sum_{\text{leaf node}} i(n)$ is represented summation of impurity level of all of leaf nodes to show the uncertainty of adopting this classification tree to classify training samplings.

According to formula (4), tree pruning can be completed by below two steps:

- (1) In all the brother leaf nodes, compare cost after combined leaf nodes.
- (2) Delete the leaf node that cost reduces the most. If cost has not reduced, nothing will be done.

Repeat above pruning process until pruning can not go on.

In pruning process, training error deduces with leaf nodes increasing; testing error deduces at the beginning and reaches minimum and then gradual roll up affected by training samplings. Take use of independent data to test, and select the subtree which has the minimum test error as decision tree. This paper adopts a heuristics verification technique —cross validation to select the best tree: 10-fold cross validation. The training samplings are divided into ten subsets which are equality in number and disjoint with each other. Classifier will train the data 10 times and every time nine groups data subsets are trained and the test one is as validation set to estimate testing error. Estimated testing error is the average value of the ten groups.

2.2 C4.5 basic theory

C4.5 is another widely applied signal decision tree building up method, and is adopts information gain ratio to classifier. It uses training group to select the properties whose information capture rate is the largest and information gain in not less than all properties average value as tree nodes. Take every possible value as a branch of node and recursively builds up a decision tree. Entropy impurity level function in CART is adopted in building up a tree. The information gain is equivalent to “impurity level decreasing value” in CART. In addition, index of capture rate is added in order to wipe off influence of high branch property. At the same time, capture rate take leaf node count and size of every node after every partition into account. Consideration objects mainly are every partition but not information content in category.

Termination condition is that properties of records in subset are the same or no property can be divided.

The difference compared with CART, C4.5 take advantage of statistical significant error probability technique based on branches to realize pruning. Another significant difference is that processing method to damage pattern. In training period, C4.5 has not adopted surrogate split to settle damage of categorization data, but adopts probability weighting method to deal with “property missing”.

2.3 A new technique adopted in decision tree—boosting method

In decision tree classifier designing, a boosting technology is widely used in the middle period of 1990s in machine learning field to improve classifier precision. This method can boost samplings classifier precision which is difficult to recognize. At the same time, this technology can cut down sensitivity that classifier algorithm affecting data noise and training sampling error.

Boosting is a learning method which can boost any learning algorithm precision and it can boost weak learn algorithm to strong learn algorithm. Its theory comes from probably approximately correct learning model. It can take advantage of some learning algorithm to generate a series of base classifiers. Every base classify training depends on classifier results produced by former classifier and endows failed training samplings with major weight to pay them more attention in subsequent learning. At last, classifier weights voting every base classifier and gets the last result and reduces signal classifier error and improve classify precision. Freund and Schapire brought forward the most pragmatic boosting algorithm—AdaBoost according to boosting basic theory in 1995 and widely applied.

3 Conclusion

The advantage of decision tree algorithm used in remote sensing data classify lies in that it can show the shortage of MLC algorithm when deals with complicated distribution data sets. Decision tree has better flexibility and robustness for data distribution feature and classify marking. So when remote sensing image data features distribution is very complicated or dimensions of source data have different statistical distribution and scales, decision tree classify method can obtain the best classified results.

Tree classify construction of decision classify method need not suppose some sort of parametric density distribution in advance. So the whole classify precision is superior to traditional parametric statistics classify method. But with the development of artificial intelligence technology and theory, study of remote sensing image classify has developed to a higher level. Geonomy knowledge and aid decision making of geographic information can boost precision of remote sensing image classification and information extraction and expert system is a good means to resolve this problem. So combination of decision tree and expert system based on knowledge is becoming a cause for concern.

Acknowledgements

This work is supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2009ZX03001-019-01), Special fund project for Basic Science Research Business Fee, AIIIS(Grant No. 2010-J).

References

1. Shuang Li, Shengyan Ding: Decision Tree Classify Method and Application in Earth Coverage Classify, *Remote Sensing Technology and Application*, 17(1), 6-11 (2002)
2. Laiping Luo, Huili Gong: Study and Implement of Remote Sensing Image Decision Tree Classifier, *Remote Sensing Information*, 13-16 (2006)
3. Feixue Li, Manchun Li: Remote Sensing Image Auto Classify Study Based on Combination of Artificial Neural Networks and Decision Tree, *Remote Sensing Information*, 3, 3-25 (2003)
4. Qingxiang Jiang, Huiping Liu: Use Texture Analysis to Extract TM Image Information, *Remote Sensing Journal*, 8(5), 458-464 (2004)
5. Friedl M A, Brodley C E, Strahler A H: Maximizing land Cover Classification Accuracies Produced by Decision Trees at Continental to Global Scales, *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 969-977 (1999)
6. Kaichang Di, Deren Li, Deyi Li: Remote Sensing Image Classify Study Based on Spatial Data Mining, *Wuhan Technical University of Surveying and Mapping Journal*, 125(1), 42-48 (2000)
7. McIver D K, Friedl M A: Estimating Pixel-scale land Cover Classification Confidence Using Non-parametric Machine Learning Methods, *IEEE Transaction on Geo-science and Remote Sensing*, 39, 1959-1968 (2001)
8. McIver D K, Friedl M A: Using Prior Probabilities in Decision-tree Remotely Sensed Data, *Remote Sensing of Environment*, 81, 253-261 (2002)
9. Zhan X, Sohlberg R A, Townshend J R G: Detection of Land Cover Changes Using MODIS 250 m Data, *Remote Sensing of Environment*, 83, 336-350 (2002)
10. Rogan J, Franklin J, Roberts D A: A Comparison of Methods of Monitoring Multi-temporal Vegetation Change Using Thematic Mapping Imagery, *Remote Sensing of Environment*, 80(1), 143-156 (2002)
11. Shuang Li, Erxun Zhang: Remote Sensing Image Classify Method Study Based on Decision Tree, *Territory Study and Development*, 22(1), 17-21 (2003)

Checklist of Items to be Sent to Volume Editors

1. A final PDF file
2. A copyright form, signed by one author on behalf of all of the authors of the paper