Pixel Based Classification of Multispectral Remote Sensed Data using Decision Tree Classifier

Shivraj S N[1], Suresh M N[2], Yeshwanth M[3], Prakash Naik[4], *UG Student Dept. of ECE, AIET*

Abstract:

Remote sensing is the science and art of obtaining information about an object through the analysis of data acquired by a device that is not in contact with the object. Remotely sensed data can be of many forms, including variations in force distribution, acoustic wave distribution or electromagnetic energy distributions and can be obtained from a variety of platforms, including satellite, airplanes, remotely pilot vehicles, handheld radiometers or even bucket trucks. They may be gathered by different devices, including sensors, film camera, digital cameras, video recorders. Our eyes acquire data on variations in electromagnetic radiations. Instruments capable of measuring electromagnetic radiation are called sensors. Sensors can be differentiated in two main groups: Passive sensors: without their own source of radiation. They are sensitive only to radiation from a natural origin. Active sensors: which have a built in source of radiation. Examples are Radar and Lidar systems. In this project, an attempt has been made to develop a decision tree classification algorithm for remotely sensed satellite data using the separability matrix of the spectral distributions of probable classes in respective bands. The spectral distance between any two classes is calculated from the difference between the minimum spectral value of a class and maximum spectral value of its preceding class for a particular band. The decision tree is then constructed by recursively partitioning the spectral distribution in a Top-Down manner. Using the separability matrix, a threshold and a band will be chosen in order to partition the training set in an optimal manner. The classified image is compared with the image classified by using classical method Maximum Likelihood Classifier (MLC)

---

1. Introduction

Image classification is one of the primary tasks in geocomputation, that being used to categorize for further analysis such as land management, potential mapping, forecast analysis and soil assessment etc. Image classification is method by which labels or class identifiers are attached to individual pixels on basis of their characteristics. These characteristics are generally measurements of their spectral response in various bands. Traditionally, classification tasks are based on statistical methodologies such as Minimum Distance-to-Mean (MDM), Maximum Likelihood Classification (MLC) and Linear Discrimination Analysis (LDA). These classifiers are generally characterized by having an explicit underlying probability model, which provides a probability of being in each class rather than simply a classification. The performance of this type of classifier depends on how well the data match the predefined model. If the data are complex in structure, then to model the data in an appropriate way can become a real problem.

In order to overcome this problem, non-parametric classification techniques such as Artificial Neural Network (ANN) and Rule-based classifiers are increasingly being used. Decision Tree classifiers have, however, not been used widely by the remote sensing community for land use classification despite their non-parametric nature and their attractive properties of simplicity, flexibility, and computational efficiency in handling the non-normal, non-homogeneous and noisy data, as well as non-linear relations between features and classes, missing values, and both numeric and categorical inputs. In this, an attempt has been made to develop a decision tree classification algorithm specifically for the classification of remotely sensed satellite data using the separability matrix of spectral distributions of probable classes. The computational efficiency is measured in terms of computational complexity measure. The proposed algorithm is coded in Visual C++ 6.0 language to develop user-friendly software for decision tree classification that requires a bitmap image of the area of interest as the basic input. For the classification of the image, the training sets are chosen for different classes and accordingly spectral separability matrix is obtained.

2. Motivation:

Global change research poses significant challenges to the scientific community. Physical and biological scientists have grappled with the challenges of data requirements for a decade or more and have identified the utility of satellite remote sensors as major sources of consistent, continuous data for atmospheric, ocean, and land studies at a variety of spatial and temporal scales. An extensive body of literature within numerous natural science disciplines documents the development or potential for satellite sensor data analysis techniques to identify environmental attributes and monitor physical and biological processes relevant to global change research.

Satellite sensor data have proven useful to the atmospheric and ocean sciences communities. While social scientists may have little involvement in the scientific study of the biological, physical, and chemical processes being addressed within these communities, human dimensions interests are associated with the causes of the perturbations to atmospheric and ocean systems being studied and in the resultant health and socioeconomic effects. In earth survey assigned role as special responsibility to ensure that the country's domain explored and mapped suitably, provide base information for expeditious and integrated development and ensure that all resource contribute with their full of measures to progress, security of our country.

3. Issues:

The classification of remote sensed data with better classification performance is a challenging task from very long time and it is a goal of today's image classification. Overlap especially when the number of classes is large, can cause the number of terminals to be much larger than the number of actual classes and thus increase the search time and memory space requirements.

Errors may accumulate from level to level in a large tree, that one cannot simultaneously optimize both the accuracy and the efficiency finally, there may be difficulties involved in designing an optimal DTC. The performance of a DTC strongly depends on how well the tree is designed.

4. Objectives:

1. To study the performance the different classifier in the area of remote sensing.
2. To understand the issues of different classifier in classification of remotely sensed data.
3. To use the best classifier in classification of remotely sensed data to obtain better classification result.
4. To provide concurrent technological developments, society demands for more data with increasing higher stands of accuracy..

5. Literature Survey

Sharma et al [1] have proposed decision tree approach for classification of remotely sensed satellite data using open source support. They have concluded that the study demonstrated use of open source remote sensing datasets and data mining software support for Land Use and Land Cover (LULC) classification.

Open source data mining software Waikato Environment for Knowledge Analysis (WEKA) was used to develop a DTC algorithm for remotely sensed satellite data. DTC outcomes were compared with outputs of Iterative Self-Organizing Data (ISODATA) clustering and Maximum Likelihood Classifier (MLC) classifier. The comparison of the classification algorithms shows DTC to be superior to the other two. DTC performed better than both MLC and SVMs for similar kind of studies using Landsat data.

Ghose et al [2] have proposed decision tree classification of remotely sensed satellite data using spectral separability matrix. Authors have concluded that a decision tree classification algorithm for remotely sensed satellite data using separability matrix of spectral distributions of probable classes has been developed. To test and validate the proposed decision tree algorithm, the sample image taken into consideration is multi- spectral image of Mayurakshi reservoir of Jharkhand state. The proposed Decision Tree classifier can also be used for hyperspectral remote sensing data considering the best bands as input for preparing spectral class distribution. The sample image is classified by both decision tree method and maximum likelihood method and then the overall accuracy, kappa coefficients were calculated. The overall accuracy for the sample test image was found to be 98% using the decision tree method and 95% using the maximum likelihood method with kappa values 97% and 94 % respectively.

Kulkarni and Shrestha [3] have proposed that multispectral image analysis using decision trees. They have concluded that implemented decision trees using the C4.5 algorithm to classify pixels in the landsat 8 image. In addition, we extracted classification rules from the pruned decision tree and evaluated the rules by implementing the Force Information Service (FIS). One of the main concerns in decision tree classifiers is that often decision trees are over fitted, when the training dataset is noisy or contains anomalies in the form of outliers. They generated training set data by selecting training set areas from the scene. To study the effect of overfitting we added noise to the original training set data. They selected five levels for the pruned decision tree because the resulting decision tree represented all categories. The pruned tree represents the generalized version of the full decision tree classifier.

Bittencourt et al [4] have proposed a binary decision tree classifier implementing logistic regression as a feature selection and classification method and its comparison with maximum likelihood. Binary decision trees implementing traditional Likelihood Ratios (LR) present results weaker than methods based on gaussian distribution. In what concerns computer time, the approach implementing LR proved to be the more efficient one. The tree based on LR performed the classification and the feature selection procedure in about two minutes to a sample size n=200, whereas the National Aerospace Laboratory (NLR) approach needed 180 minutes. The experiments have shown that the LR approach produces accuracies lower than the Gaussian Maximum Likelihood (GML), but the results can still be considered satisfactory.

6. Conclusions

In this project, a decision tree classification algorithm for remotely sensed satellite data using separability matrix of spectral distributions of probable classes has been developed. The proposed Decision Tree classifier can also be used for hyper-spectral remote sensing data considering the best bands as input for preparing spectral class distribution. The sample image is classified by both Decision Tree method and Maximum Likelihood method and then the overall accuracy, kappa coefficients were calculated. Since the accuracy of the results depends only upon the test set chosen, the efficiency of any algorithm shall not be considered on the accuracy measure alone. Out of eleven classes considered for the sample image, many classes were found to be closely matching in both the methods. However, differences are observed in certain classes in both the methods. The classified images shall also be compared with the input image (FCC) and collecting ground truth information physically.

7. References

[1] Richa Sharma, Aniruddha Ghosh and P K Joshi, “Decision tree approach for classification of remotely sensed satellite data using open source support”, Journal of Earth System Science, pp. 1237-1247, No. 5, October 2013.

[2]. M. K. Ghose , Ratika Pradhan, Sucheta Sushan Ghose, “Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matrix”, International Journal of Advanced Computer Science and Applications, Vol. 1, No.5, November 2010.

[3]. Arun Kulkarni, Anmol Shrestha “Multispectral Image Analysis using Decision Trees”, International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.

[4]. [13] Radke Bittencourt, Denis Altieri de Oliveira Moraes, Victor Haertel, “A Binary Decision Tree Classifier Implementing Logistic Regression as a Feature Selection and Classification Method and its Comparison with Maximum Likelihood”, National council for scientific and technological development, 2007.