Decision Tree Approach for Classification of Satellite Imagery

Havyas V B, Choodarathnakara A L, Thribhuvan R, Chethan K S

Abstract— Various practical systems capable of extracting descriptive decision making knowledge from data have been developed and evaluated. Techniques that represent knowledge about classification tasks in the form of decision trees are focused on. A sample of techniques is sketched, ranging from basic methods of constructing decision trees to ways of using them non-categorically. Some characteristics that suggest whether a particular classification task is likely to he amenable or otherwise to tree-based methods are discussed. Many urban land cover types show spectral similarity in remote sensing data. Further, the finer the spatial resolution of the data, the larger is the number of detectable subclasses within classes. This high within-class spectral variance of some classes results in multimodal distribution of spectra and may decrease their spectral separability. Hence, the existing traditional hard classification techniques which are parametric type do not perform well on high resolution data in the complex environment of the urban area as they expect datasets to be distributed normally. The aim of this paper is to investigate a non-parametric classifier as an alternative approach to classify an image data of a semi urban area.

Index Terms—Remote Sensing, Image Classification, Parametric Classifier, Non-parametric and Decision Tree Classifier

I. INTRODUCTION

Remote Sensing (RS) refers to the science of identification of earth surface features and estimation of their geo-biophysical properties using electromagnetic radiation as a medium of interaction. Spectral, spatial, temporal and polarization signatures are major characteristics of the sensor/target, which facilitate target discrimination. Earth surface data as seen by the sensors in different wavelengths (reflected, scattered and/or emitted) is radiometrically and geometrically corrected before extraction of spectral information. RS data, with its ability for a synoptic view, repetitive coverage with calibrated sensors to detect changes, observations at different resolutions, provides a better alternative for natural resources management as compared to traditional methods. Some of the major operational application themes, in which India has extensively used

remote sensing data, are agriculture, forestry, water resources, land use, urban sprawl, geology, environment, coastal zone, marine resources, snow and glacier, disaster monitoring and mitigation, infrastructure development, etc.

Image classification is the process of categorizing all the pixels automatically in an image into a finite number of land cover classes [12] and it is one of the most often used quantitative data analysis techniques in remote sensing to describe ground cover types or material classes. Classifiers are broadly categorized into supervised and unsupervised, hard and soft, parametric and non parametric type. Among them, the maximum likelihood classifier (MLC), belonging o the family of supervised parametric classifier is most commonly used in remote sensing because of its robustness and easy availability in almost all image processing software [1]. Also, MLC has traditionally been employed as a baseline for evaluating the accuracy of classifiers on remotely sensed data.

For classification of features in urban area, the expected spatial resolution should be at least 5m where buildings and roads can be easily distinguished [3]. Most of the materials found in the urban environment like concrete, asphalt, metal, plastic, glass, water etc. exhibit spectral similarity. Many urban land cover types such as roads, buildings, parking lots, grass, trees, shrubs and soil also show spectral similarity. In addition to the spectral similarity between land cover types, remote sensing images contain mixed pixels which make it difficult to classify a pixel as belonging to only one class [4], [6]. Therefore, the finer the spatial resolution, the larger is the number of detectable subclasses within the classes and this high within-class spectral variance of some classes may decrease their spectral separability resulting in lower classification accuracy. As a result, classifications accuracies may decrease for some classes, such as complex urban are as spatial resolution becomes finer [5].

As the existing traditional hard classification techniques are parametric type, they do not perform well on high resolution data in the complex environment of the urban area as they expect datasets to be distributed normally [2], [4]. The assumption of normal distribution of spectra is often violated especially in the complex landscapes in high-resolution data. In addition, insufficient, non-representative, or multimode distributed training samples can further introduce uncertainty to the image classification procedure. Another major drawback of the parametric classifiers lies in the difficulty of integrating spectral data with ancillary data [1], [2] like digital elevation model, slope, texture and context information, etc. Therefore, parametric classifiers fail to exploit the

best use of the information available through advanced sensor systems and various ancillary data. Hence, generating a satisfactory classified image from a high-resolution remotely sensed data is a challenge and is not as straightforward as classification of low resolution imagery (30 m or more) using traditional classifiers.

On the contrary, non-parametric classifiers are independent of the properties of the distribution of data. With nonparametric classifiers, the assumption of the normal distribution of the dataset is not required. No statistical parameters are needed to separate image classes. Nonparametric classifiers are thus suitable for the incorporation of non-spectral data into a classification procedure. Among the most commonly used non-parametric classification approaches are neural networks, decision trees, support vector machines, and expert systems [1].

II. IMAGE CLASSIFICATION

Image classification is a particular case of Pattern Recognition. The overall objective of the classification process is to automatically classify all pixels in an image into land cover classes based on the predefined classification model. The term pattern in case of image classification refers to the set of radiance measurements obtained in the various wavelength bands for each pixel. There are numerous classification algorithms. This paper gives a brief introduction of the most popular classifiers in the field of remote sensing. Classifiers are described under broad categories such as supervised and unsupervised classifiers, parametric and non parametric, fuzzy classifiers and knowledge base classifiers.

A. Supervised Classification

In this type of classification the image analyst supervises the pixel categorisation process by specifying to the algorithm specific information of the various land cover types present in a scene. To do this, representative sample site of known cover type, called training areas, are used to compile a numerical interpretation key that describes the spectral attributes for each feature type of interest. Reflectance value of each pixel in the image is then compared numerically to each category in the interpretation key labelled with the name of the category it looks most like. Steps involved in supervised classification are,

- Training stage
- Classification stage
- Accuracy assessment

B. Unsupervised Classification

Unsupervised ones do not utilize training data as the basis for classification. These classifiers try to aggregate reflectance value of pixels in the feature space into well separated clusters. Clusters are considered as classes. Once the spectral grouping has been done, the analyst identifies the obtained classes to some form of reference data. There are numerous clustering algorithms that can be used to determine the natural spectral clusters present in the image. The most common algorithm is "K-means". In this approach user has to define the number of clusters or classes to be located in the image. The algorithm automatically locates the centre means of various clusters present in the image and

each pixel in the image is then assigned to the cluster whose mean is closest. After all pixels have been classified, revised mean vectors for each of the cluster is computed. The whole process is repeated again until there is no further change in the location of class means vectors

C. Parametric Classifier

Parametric classification algorithms assume that the observed measurement vectors Xc obtained for each class in each spectral band during the training phase of the supervised classification follow some statistical distribution such as Gaussian distribution (Jensen, 1996). The major parametric classifiers under this category are minimum distance, Mahalanobis distance, and maximum likelihood classifier. Maximum likelihood gives better accuracy than others and frequently used in the remote sensing image classification. Therefore Maximum likelihood algorithm is described here as a representative of parametric classifiers.

D. Maximum likelihood classification

The MLC quantitatively evaluates both the variance and covariance of the category spectral response pattern when classifying an unknown pattern. An assumption is made that the distribution of the training set is Gaussian. Under this assumption, the distribution of a training set of a class can be completely described by the mean vector and covariance matrix. Given these parameters, we may compute the statistical probability of a given pixel being a member of a particular class.

E. Non-parametric Classifier

A non- parametric classifier is not based on statistics, therefore, it is independent of the properties of the data. Non-Parametric classification algorithm does not take into account the distribution of the training set. They do not require that that the observed measurement vectors Xc obtained for each class in each spectral band during the training phase of the supervised classification should follow Gaussian distribution. Best known classifiers in this category are parallelepiped, decision tree and neural network. A brief introduction of decision tree is given in the next section.

III. DECISION TREE CLASSIFIER

Decision tree is one of the inductive learning algorithms that generate a classification tree to classify the data. It is based on the "divide and conquer" strategy. The classification tree is made by recursive partitioning of the feature space, based on a training set. At each branching, a specific decision rule is implemented, which may involve one or more combinations of the attribute inputs or features (Quinlan, 1993).

The practical advantages of decision tree classifiers over the traditional statistical classifier are:

- DTs are non-parametric in nature and are not constrained by lack of knowledge of class distributions.
- DTs can be trained quickly and take less computational time.
- DTs can handle high dimensional data and represent

the acquired knowledge in tree form which is intuitive and generally easy to assimilate.

- They are simple and able to handle missing/noisy data.
- Decision trees cam be applied in many different situations like exploring a large dataset to pick out useful variables and predicting future states of different variables in an industrial process.

A decision tree is composed of a root node, a set of interior nodes, and terminal nodes, called "leaves". The root node and interior nodes, referred to collectively as non-terminal nodes, are linked into decision stages. The terminal nodes represent final classification. The classification process is implemented by a set of rules that determine the path to be followed, starting from the root node and ending at one terminal node, which represents the label for the object being classified. At each non-terminal node, a decision has to be taken about the path to the next node. Figure 1 illustrates a simple decision tree using pixel reflectance as input.
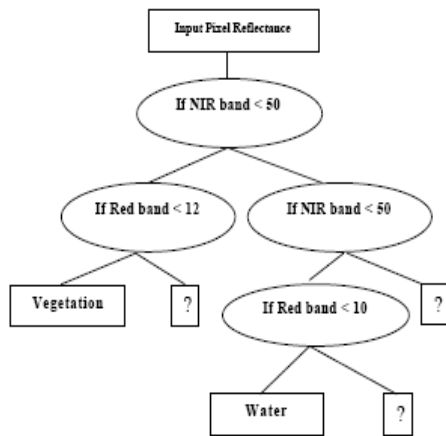

Fig 1: Example of Decision Tree (Source: Pal et al, 2001)

A. Decision Tree Algorithm

Step1: Let T be the set of training instances.
Step2: Choose an attribute that best differentiates the instances in T.
Step3: Create a tree node whose value is the chosen attribute. Create child links from this node where each link represents a unique value for the chosen attribute. Use the child link values to further subdivide the instances into subclasses.
Step4: For each subclass created in step 3.

If the instances in the subclass satisfy predefined criteria or if the set of remaining attribute choices for this path is null, specify the classification for new instances following this decision path. If the subclass does not satisfy the criteria and there is at least one attribute to further subdivide the path of the tree, let T be the current set of subclass instances and return to step 2. It is obvious that if the rules are not complete after tracing through the decision tree, some pixels will remain unclassified. Therefore the efficiency and performance of this approach is strongly affected by tree structure and choice of features selected for training.

What made the decision tree classifiers so popular is that the construction of the classifier does not require any domain knowledge or parameter setting, and therefore it is appropriate for exploratory knowledge discovery. Decision

trees can handle high dimensional data and represent the acquired knowledge in tree form which is intuitive and generally easy to assimilate by humans [8]. As they are considered to be a nonparametric method, they have no assumptions about the distribution of data. Decision trees are capable of handling datasets that may have errors and/ or have missing values. The most important advantage of decision tree is the explanation capability by extracting classification rules directly from the tree [9]. Originally, it has been studied in the fields of decision theory and statistics and now found to be effective in other disciplines such as data mining, machine learning, and pattern recognition. Decision tree induction algorithms have also been used extensively as a supervised approach to classification in many application areas such as medical diagnosis, manufacturing and production, financial analysis, astronomy, radar signal classification, character recognition, remote sensing, expert systems, speech recognition and molecular biology [10][11][12].

Decision makers need to make predictions- whether the U.S. dollar will rise in the short term, whether a patient will benefit from a surgical procedure, whether it will rain tomorrow. One sound basis for such predictions is an extrapolation of past, known cases. The science of statistics provides a range of tools for this purpose, usually based on the idea of fitting a particular class of models to the data and then hypothesizing that future event will conform to the fitted model. Researchers in artificial intelligence (AI) have long been interested in the same task, usually from a less model driven standpoint. The basic scenario for this branch of earning is one in which an intelligent agent, shown a collection of case studies of some activity, employs inductive inference to derive useful information about that activity. As with their statistical counterparts, many early learning programs were concerned with finding appropriate values for numeric parameters, a line of research well summarized in [13].

B .Accuracy Assessment

Classification process is not complete until its accuracy is assessed. Accuracy assessment can be performed by comparing two sources of information (Jensen, 1996):

- Remote-sensing derived classification data and
- Reference test data

The relationship of these two sets is summarized in an error matrix where columns represent the reference data while rows represent the classified data. An error matrix is a square array of numbers laid out in rows and columns that expresses the number of sample units assigns to a particular category relative to the actual category as verified in the field.

The accuracy of a classification has traditionally been measured by the overall accuracy. The overall accuracy of classification is obtained by dividing the sum of the correctly classified pixels (i.e. summed up values on the major diagonal of the error matrix) by the total number of pixels classified of the reference points. The kappa statistic, also called *KHAT* value, is a measure of how well the classification agrees with the reference data. It is also a measure of overall accuracy [13] and most commonly

employed to evaluate the performance of a classifier.

But the overall accuracy alone gives no insight into how well the classifier is performing for each of the different classes. In particular, a classifier might perform well for a class which accounts for a large proportion of the test data being correctly classified and also called a measure of omission error. It is obtained by dividing the total number of correct pixels in a category by the total number of pixels of that category as derived from the reference data [14]. User's accuracy can be obtained by dividing the total number of correct pixels in a category by the total number of pixels that were classified in that category and also called a measure of commission error. It is an indicative of the probability that a pixel classified on the image actually represents that category on the ground.

IV. CONCLUSION

Only the most commonly affecting issues which show influence on the performance of the decision tree are dealt with here.

- Since decision trees make no prior assumptions about the data and are non parametric in nature, the size of the training data largely affects the performance of the classifier.
- Small variations in the data an cause very different looking trees, and may also cause classification trees unstable. Further, noise in data causes over fitting of the tree model.
- Improper tree pruning and attribute selection criterion may generate a very large and complex model, and it can make the rule sets difficult to interpret, understand or justify.

Since each leaf leads to one classification rule the decision tree is a function of how many terminal nodes or leaves exists on the decision tree.

REFERENCES

1. Navalgund, R. R., "Remote sensing: Basics and applications". *Resonance*, 2001, 6, 51–60.
2. Patel, N. K., Medhavy, T. T., Patnaik, C. and Hussain, A., "Multi-temporal ERS-1 SAR data for identification of rice crop". *J. Indian Soc. Remote Sensing*, 1995, 23, 33–39.
3. Panigrahy, S., Chakraborty, M., Sharma, S. A., Kundu, N., Ghose, S. C. and Pal, M., "Early estimation of rice acre using temporal ERS-1 synthetic aperture radar data – A case study for Howrah and Hooghly districts of West Bengal", India. *Int. J. Remote Sensing*, 1997, 18, 1827–1833
4. Srivastava, H. S., Patel, P. and Navalgund, R. R., "Incorporating soil texture in soil moisture estimation from extended low-1 beammode Radarsat-SAR data". *Int. J. Remote Sensing*, 2006, 27, 2587– 2598.
5. Jayaraman, V., Srivastava, S. K., Kumaran Raju, K. and Rao, U.R., "Total solution approach using IRS-1C and IRS-P3: A perspective of multi-resolution data fusion and improved vegetation indices". *IEEE Trans. Geosci. Remote Sensing*, 2000, 38, 587–604
6. Navalgund, R. R., Parihar, J. S., Ajai and Nageshwar Rao, P. P., "Crop inventory using remotely sensed data". *Curr. Sci.*, 1991, 61, 162–171.
7. Vasudevan, B. G., Gohil, B. S. and Agarwal, V. K., "Backpropagation neural network based retrieval of atmospheric water vapour and cloud liquid water from IRS-P4 MSMR". *IEEE Trans Geosci. Remote Sensing*, 2004, 42(5), 985–990.
8. C. Apte and S. Weiss, "Data mining with decision trees and decision rules," *Future Generation Computer Systems*," pp. 197-210, Nov. 1997
9. Selim Aksoy, K. Koperski, Carsten Tusk, and Giovanni Marchisio, "Interactive training of advanced classifiers for mining remote sensing image archives," *KDD'04*, Seattle, Washington, USA, August 22-25,2004.

and this will bias the overall accuracy, despite low class accuracies for other classes. To avoid such a bias, it is important to consider the individual class accuracy under producer's accuracy and user's accuracy. Producer's accuracy is a measure of the probability of a reference pixel

10. Jiawei Han and Micheline Kamber, "*Data Mining Concepts and Techniques*", 2nd ed., Elsevier, 2006.
11. S. Rasoul Safavian and David Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man Cybnet*., vol. 21, no. 3, pp. 660-674, May/ June 1991
12. Lior Rokach and Oded Maimon, "Top-down induction of decision trees classifiers-A survey," *IEEE Trans. Syst. Man Cybnet*., part C, vol. 35, no. 4, pp. 476-487, Nov. 2005.
13. R. G. Congalton and A. Roy Mead, "A review of three discrete multivariate analysis techniques used in assessing the accuracy of remotely sensed data from error matrix," IEEE Trans. Geosci. Remote Sensing, vol. GE 24, no. 1, pp. 169-174, Jan. 1986.
14. R. G. Congalton, "Considerations and techniques for assessing the accuracy of remotely sensed data," Geosci. Remote Sens. Symp., vol. 3, Jul. 10-14, 1989, pp. 1847-1850.

AUTHOR PROFILE

Havyas V B was born in Virajpet Taluk, Kodagu District of Karnataka State in INDIA. He received the B.E. degree in Electronics & Communication Engineering in and M.Tech. Degree in VLSI & Embedded Systems from Visvesvaraya Technological University, Belgaum, INDIA. He is presently working as Guest Faculty in the Department of Electronics & Communication Engineering, Government Engineering College, Kushalnagar of Karnataka State, INDIA. His areas of interest are Pattern Recognition and Satellite Data Processing.

A. L. Choodarathnakara was born in Hassan District of Karnataka State in INDIA. He received the B.E. degree in Electronics & Communication Engineering in the year 2002 and M.Tech. degree in Digital Electronics & Communication Systems in the year 2008, both from the Malnad College of Engineering, Hassan affiliated to Visvesvaraya Technological University, Belgaum, INDIA. He is an Assistant Professor of Electronics & Communication Engineering Department at Government Engineering College, Kushalnagar of South Kodagu District in Karnataka. His subjects of interest include Natural Language Processing, Language Learning Technologies, Text, Speech, Image, Signal & Remote Sensed Data Processing. Asst. Prof. Choodarathnakara A. L. is a Life Member of the Indian Society of Technical Education (ISTE-43168), Indian Society of Remote Sensing (ISRS-3533) and Indian Society of Geomatics (ISG-1258).

Thribhuvan R was born in Hassan District of Karnataka State in INDIA. He received the B.E. degree in Electronics & Communication Engineering in and M.Tech. Degree from Visvesvaraya Technological University, Belgaum, Karnataka State, INDIA.

Chethan K S received the B.E. degree in Electronics & Communication Engineering and M.Tech. Degree in Digital Electronics & Communication Systems from the Malnad College of Engineering, Hassan affiliated to Visvesvaraya Technological University, Belgaum, INDIA. He is presently working as Assistant Professor in the Department of Electronics & Communication Engineering, PESIT, Bangalore of Karnataka State, INDIA. His areas of interest are Pattern Recognition and Medical Image Processing.