IJSTE - International Journal of Science Technology & Engineering | Volume 3 | Issue 01 | July 2016
ISSN (online): 2349-784X

An Efficient Approach for Search using Spatial Decision Tree

A. Valarmathi
Assistant Professor
Department of Computer Science & Engineering
Shrimati Indira Gandhi College, Trichy

S. Geetha
Assistant Professor
Department of Computer Science & Engineering
Shrimati Indira Gandhi College, Trichy

Abstract

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes. Given a raster spatial framework, as well as training and test sets, the spatial decision tree learning (SDTL) problem aims to minimize classification errors. The SDTL problem has many applications. In the field of remote sensing, a large amount of images of the earth surface are collected. SDTL can be used to classify remote sensing images into different land cover types. Related work relies on local tests and cannot adequately model the spatial autocorrelation effect, resulting in salt-and-pepper noise. A focal-test-based spatial decision tree (FTSDT), in which the tree traversal direction of a sample is based on both local and focal (neighborhood) information is proposed in this paper. The experimental results show that the proposed system outperforms the existing methods.

Keywords: Query processing, nearest neighbor, spatial data mining, focal-test-based spatial decision tree

I. INTRODUCTION

Decision trees are simple predictive models which map input attributes to a target value using simple conditional rules. Trees are commonly used in problems whose solutions must be readily understandable or explainable by humans, such as in computer-aided diagnostics and credit analysis.

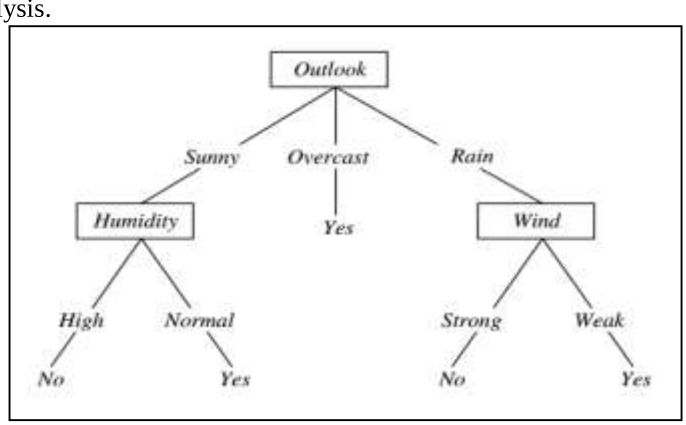

Fig. 1:

Decision Trees give a direct and intuitive way for obtaining the classification of a new instance from a set of simple rules. Because of this characteristic, decision trees find wide use in situations in which the interpretation of the obtained results and of the reasoning process is crucial, such as in computer-aided diagnostics (CAD) and in financial credit analysis. Consumer credit analysis is an interesting example because, in many countries, one cannot simply reject credit without giving a justification, justification of which is trivial to extract from a decision tree.

II. LEARNING DECISION TREES

Decision trees can be simply drawn by hand based on any prior knowledge the author may have. However, their real power becomes apparent when trees are learned automatically, through some *learning algorithm*.

Given a spatial raster framework, as well as training and test sets, the spatial decision tree learning (SDTL) problem aims to find a decision tree model. Existing work includes non-spatial decision trees (e.g., ID3, C4.5 and CART) and spatial entropy or information gain based decision trees. Non-spatial decision trees make the assumption of learning samples and ignore the spatial autocorrelation effect. Spatial entropy or information gain based decision trees use spatial autocorrelation level as well as information gain to select candidate tree node tests. However, when all the candidate tree node tests have poor spatial autocorrelation, the spatial entropy or information gain based decision tree will still select one of them.

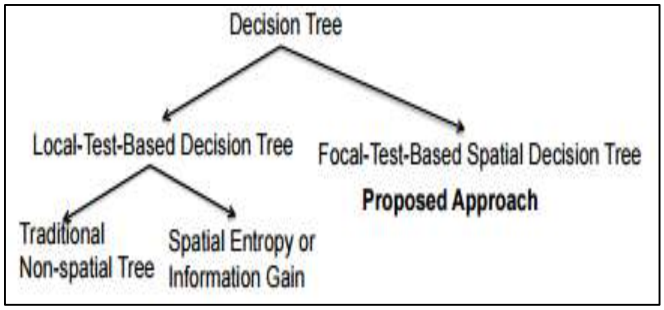

Fig. 2: Decision Tree

In summary, both of these existing approaches use local-test-based decision nodes (i.e., decision tree node testing each learning sample independently), and thus cannot adequately model spatial autocorrelation in the test phase, leading to salt-and-pepper noise. In contrast to existing work, proposed a focal-test-based spatial decision model, whereby the tree traversal direction of learning sample is based on not only local but also focal (neighborhood) properties of features.

Contributions: There are three main contributions of the paper: (1) Define a focal-test-based spatial decision tree (FTSDT) model; (2) Proposed learning algorithms (including a training algorithm and a prediction algorithm) for the FTSDT model; (3) Evaluate the proposed approach using experiments.

III. EXISTING SYSTEM

Given a spatial raster framework, as well as training and test sets, the spatial decision tree learning (SDTL) problem aims to find a decision tree model that minimizes classification errors as well as salt-and-pepper noise.Classification results by existing decision tree classifiers predicted maps exhibit poor appearance accuracy with high levels of salt-and-pepper noise, when Compared with ground truth classes.

A. Drawbacks of Existing System:

A key challenge in the SDTL problem is that learning samples show spatial autocorrelation in class labels.Incorporating focal (i.e., neighborhood) information increases both the number and the complexity of candidate tree node tests

IV. PROPOSED SYSTEM

Defined a focaltest-based spatial decision tree (FTSDT) model, whereby the tree traversal direction of a learning sample is based on not only local but also focal (neighborhood) properties of features. Proposed FTSDT learning algorithms and evaluated the classification performance of the proposed approach on real world remote sensing data sets.

More specifically, the following additional contributions are made:

Add a new design decision in the FTSDT model to allow focal function computation with adaptive neighborhoods (i.e., FTSDT-adaptive).Design a refined algorithm (FTSDT-Refined) that reuses focal values across candidate thresholds and prove its correctness.

A. Advantages of Proposed System:

Compared with previous FTSDT with fixed neighborhoods (i.e.,FTSDT-fixed), the new design decision can adjust the neighborhood shape to avoid over-smoothing in wedge-shaped areas.

The refined algorithm improves computational scalability

V. PROJECT DESCRIPTION

- Spatial Database Creation Module
- FTSDT-Train
  - Node Split
  - Focal Function
- FTSDT-Prediction
  - Decision Tree
- Map View

A. Spatial Database Creation Module:

A spatial raster framework F is a tessellation of a 2-D plane into a regular grid. On a spatial raster framework, there may exist a set of explanatory feature maps, as well as a class label map. A spatial neighborhood relationship describes the range of dependency between spatial locations. It is commonly represented as a W-matrix, whose element Wi,j has a non-zero value when locations i and j are neighbors, and a zero value otherwise.

B. FTSDT Train:

FTSDT-Train learns an FTSDT classifier from training samples. It includes two sub-routines (Node-Split and Focal Function).
If the training samples are less than the minimum tree node size, or all the class labels are identical, a leaf labelled with the majority class will be returned. The process enumerate through every candidate feature f, every neighborhood size s, and every candidate threshold d to select the best setting for a model tree node. NodeSplit subroutine is used to split training samples. Finally update the current best candidate test. Create an internal node with the best test, split the training samples into two subsets accordingly, recursively call FTSDT-Train on each subset, and return the internal node.

C. Node – Split:

The Node-Split subroutine splits the training samples into two subsets based on their focal test results, and proceeds as follows: Step 1 initializes the two subsets as empty sets. Samples with node test results TRUE will be assigned to one subset and samples with test results FALSE will be assigned to the other. Then compute the focal tree node test result of each training sample and add the sample to its appropriate subset accordingly. The algorithm begins by computing local test indicators (I) of all samples. It then computes the focal function value via a Focal Function subroutine, and computes the focal test result on each sample location

D. Focal Function:

The Focal Function subroutine computes the focal function values of local test indicators in the neighborhood of a location. It has an important parameter neigh Type, whose value is 0 for a fixed neighborhood, and is 1 for an adaptive neighborhood. The idea behind an adaptive neighborhood is to utilize spatial topological relationships to select proper neighbors of the central pixel in a fixed window.

E. FTSDT Prediction:

The FTSDT-Predict algorithm uses an FTSDT to predict the class labels of test samples based on their feature values and a spatial neighborhood structure. It is a recursive algorithm. If the tree node is a leaf, then the class label of the leaf is assigned to all current samples. Otherwise, samples are split into two subsets according to the focal test results in the root node, and each subset is classified by its corresponding sub-tree.

F. Map View:

1) Google Map API:
An API is a set of functions and procedures that allow the creation of applications which access the features or data of an operating system, application, or other service. Google Maps allows you to display maps on your web site. The user can see the view of their locality by Google Map (such as map view, satellite view) .With the Google Maps API, you can customize maps, and the information on maps.
2) Finding Route:
The Google Directions API is a service that calculates directions between locations using an HTTP request. You can search for directions for several modes of transportation, include transit, driving, walking or cycling. Directions may specify origins, destinations and waypoints either as text strings (e.g. "Chicago, IL" or "Darwin, NT, Australia") or as latitude/longitude coordinates. The Directions API can return multi-part directions using a series of waypoints. This service is generally designed for calculating directions for static (known in advance) addresses for placement of application content on a map.

VI. CONCLUSION & FUTURE WORK

This project explores the spatial decision tree learning problem for raster dataset. The problem is challenging due to the spatial autocorrelation effect and computational cost. Related work is limited to using local tests in tree nodes. In contrast, proposed a focal test based spatial decision tree model and its learning algorithm. Further computational optimization and design a refined algorithm that selectively updates focal values is conducted. Both theoretical analysis and experimental evaluation show that refined algorithm is more scalable than baseline algorithm. Also design a new focal test approach with adaptive neighborhoods to avoid over-smoothing in wedge-shaped areas. Experiment results on real world data sets show that new FTSDT with adaptive neighborhoods improves classification accuracy of both the default FTSDT with fixed neighborhoods and traditional LTDT.

Future work is planned to design a spatial decision tree learning algorithm to account for another important characteristic of spatial data, i.e., heterogeneity. Also planned to work on an ensemble of spatial decision trees.

References

[1] J. R. Quinlan, C4.5: Programs for Machine Learning, San Mateo, CA, USA: Morgan Kaufmann, 1993.
[2] Z. Jiang, S. Shekhar, P. Mohan, J. Knight, and J. Corcoran, "Learning spatial decision tree for geographical classification: A summary of results, " in Proc. 20th Int. Conf. Adv. Geographic Inf. Syst., 2012, pp. 390–393.
[3] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," Remote Sens. Environ., vol. 61, no. 3, pp. 399–409, 1997.
[4] T. Yuan and W. Kuo, "Spatial defect pattern recognition on semiconductor wafers using model-based clustering and Bayesian inference," Eur. J. Oper. Res., vol. 190, no. 1, pp. 228–240, 2008.
[5] R. G. Congalton and K. Green, Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Boca Raton, FL, USA: CRC Press, 2008.
[6] D. Brownrigg, "The weighted median filter," Commun. ACM, vol. 27, no. 8, pp. 807–818, 1984.
[7] H. Hwang and R. A. Haddad, "Adaptive median filters: New algorithms and results," IEEE Trans. Image Process., vol. 4, no. 4, pp. 499–502, Apr. 1995.