Chapter 7.13
A Successive Decision Tree Approach to Mining Remotely Sensed Image Data

Jianting Zhang
University of New Mexico, USA

Wieguo Liu
University of Toledo, USA

Le Gruenwald
University of Oklahoma, USA

ABSTRACT

Decision trees (DT) has been widely used for training and classification of remotely sensed image data due to its capability to generate human interpretable decision rules and its relatively fast speed in training and classification. This chapter proposes a successive decision tree (SDT) approach where the samples in the ill-classified branches of a previous resulting decision tree are used to construct a successive decision tree. The decision trees are chained together through pointers and used for classification. SDT aims at constructing more interpretable decision trees while attempting to improve classification accuracies. The proposed approach is applied to two real remotely sensed image datasets for evaluations in terms of classification accuracy and interpretability of the resulting decision rules.

INTRODUCTION

Compared with statistical and neural/connectionist approaches to classification of remotely sensed image data, decision trees (DT) have several advantages. First of all, there is no presumption of data distribution in DT. Second, since DT adopts a divide-and-conquer strategy, it is fast in training and execution. Most importantly, the resulting classification rules are presented in a tree form. Paths from the root to leaf nodes can easily be transformed into decision rules (such as if $a>10$ and $b<20$ then Class 3), which is suitable for human interpretation and evaluation. In the past years, DT has gained considerable research interests in analysis of remotely sensed image data, such as automated knowledge-base building from remote sensing and GIS data (Huang & Jensen, 1997), land cover classification (Friedl

& Brodley, 1997), soil salinity analysis (Eklund, Kirkby, & Salim, 1998), change detection in an urban environment (Chan, Chan, & Yeh, 2001), building rule-based classification systems for remotely sensed images (Lawrence & Wright, 2001) and knowledge discovery from soil maps (Qi & Zhu, 2003). In particular, DT has been employed for global land cover classifications at 8km spatial resolution (De Fries, Hansen, Townshend, & Sohlberg, 1998) using NOAA AVHRR data. Interestingly, DT has also been adopted as the primary classification algorithm to generate global land cover maps from NASA MODIS data (Friedl et al., 2002) where spatial and radiometric attributes have been significantly improved.

In ideal situations, each leaf node contains a large number of samples, the majority of which belongs to one particular class called the dominating class of that leaf node. All samples to be classified that fall into a leaf node will be labeled as the dominating class of that leaf node. Thus the classification accuracy of a leaf node can be measured by the number of the actual samples of the dominating class over all the samples in its leaf node. However, when there are no dominating classes in the leaf nodes, class labels are assigned based on simple majority vote and, hence, the decision tree nodes have low classification accuracy.

While DT has gained considerable applications, the resulting decision trees from training datasets could be complex due to the complex relationship between features and classes. They are often the mixtures of the branches with high and low classification accuracies in an arbitrary manner and are difficult for human interpretation. In this study, we propose to apply DT multiple times to a training dataset to construct more interpretable decision trees while attempting to improve classification accuracy. The basic idea is to keep classification branches of a resulting decision tree that have high classification accuracy while combining samples that are classified under branches with low classification accuracy into a

new training dataset for further classifications. The process is carried out in a successive manner and we term our approach as successive decision tree (SDT). For notation purposes, we also term classic DT as CDT.

The heuristics behind the expectation that SDT can increase classification accuracy are based on the following observation. There are samples in a multi-class training dataset, although their patterns may be well perceived by human, they are small in sizes and are often assigned to various branches during the classification processes according to information entropy gain or gain ratio criteria. At some particular classification levels, the numbers of the samples may be below predefined thresholds in decision tree branches to be qualified as a decision tree leaf node with high classification accuracy, thus the splitting processes stop and they are treated as noises. On the other hand, if we combine these samples into a new dataset, since the distribution of the new dataset may be significantly different from the original one, meaningful classification rules may be derived in a new decision tree from the new dataset. By giving some samples a second chance to be correctly classified, the overall accuracy may be improved. The heuristics will be further illustrated through an example in "The Method" section.

The proposed SDT method is different from existing meta-learning approaches that are applied to DT, such as the boosting (Freund & Schapire, 1997) DT approach (Friedl, Brodley, & Strahler, 1999; Pal & Mather, 2003). Boosting DT gives higher weights to the samples that have been misclassified in a previous process but uses all the samples in all the classification processes. Boosting DT does not aim at generating interpretable decision rules. In contrast, the proposed SDT approach generates compact decision rules from decision tree branches with high classification accuracy; only samples that cannot be generalized by decision rules with high classification accuracy are combined for further classifications.

Related Content

Indexing in Data Warehouses: Bitmaps and Beyond
Karen C. Davisand Ashima Gupta (2007). *Data Warehouses and OLAP: Concepts, Architectures and Solutions (pp. 179-202).*
www.irma-international.org/chapter/indexing-data-warhousing/7621

Using Active Rules to Maintain Data Consistency in Data Warehouse Systems
Shi-Ming Huang, John Tait, Chun-Hao Suand Chih-Fong Tsai (2009). *Progressive Methods in Data Warehousing and Business Intelligence: Concepts and Competitive Analytics (pp. 252-272).*
www.irma-international.org/chapter/using-active-rules-maintain-data/28170

Data Mining in Gene Expression Analysis: A Survey
Jilin Han, Le Gruenwaldand Tyrrell Conway (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 1643-1673).*
www.irma-international.org/chapter/data-mining-gene-expression-analysis/7722

Text Content Approaches in Web Content Mining
Víctor Fresno Fernandezand Luis Magdalena Layos (2005). *Encyclopedia of Data Warehousing and Mining (pp. 1103-1108).*
www.irma-international.org/chapter/text-content-approaches-web-content/10761

An Information-Theoretic Framework for Process Structure and Data Mining
Gianluigi Greco, Antonella Guzzoand Luigi Pontieri (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 810-830).*
www.irma-international.org/chapter/information-theoretic-framework-process-structure/7676