Pattern Recognition 48 (2015) 50–59

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Fuzzy rule based decision trees

Xianchang Wang [a,b,*], Xiaodong Liu [a], Witold Pedrycz [c,d,e], Lishi Zhang [a,b]

[a] Research Center of Information and Control, Dalian University of Technology, No. 2, Linggong Road, Ganjingzi, Dalian, Liaoning 116024, PR China
[b] School of Sciences, Dalian Ocean University, Dalian 116023, PR China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada T6R 2V4 AB
[d] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[e] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Received 11 December 2013
Received in revised form
16 July 2014
Accepted 4 August 2014
Available online 14 August 2014

Keywords:
Decision tree
Fuzzy classifier
Fuzzy rules
Fuzzy confidence

ABSTRACT

This paper presents a new architecture of a fuzzy decision tree based on fuzzy rules – fuzzy rule based decision tree (FRDT) and provides a learning algorithm. In contrast with "traditional" axis-parallel decision trees in which only a single feature (variable) is taken into account at each node, the node of the proposed decision trees involves a fuzzy rule which involves multiple features. Fuzzy rules are employed to produce leaves of high purity. Using multiple features for a node helps us minimize the size of the trees. The growth of the FRDT is realized by expanding an additional node composed of a mixture of data coming from different classes, which is the only non-leaf node of each layer. This gives rise to a new geometric structure endowed with linguistic terms which are quite different from the "traditional" oblique decision trees endowed with hyperplanes as decision functions. A series of numeric studies are reported using data coming from UCI machine learning data sets. The comparison is carried out with regard to "traditional" decision trees such as C4.5, LADtree, BFTree, SimpleCart, and NBTree. The results of statistical tests have shown that the proposed FRDT exhibits the best performance in terms of both accuracy and the size of the produced trees.

1. Introduction

Decision trees are one of the most well-known methods used for extracting classification rules from data. There are several reasons behind their visibility and broad applicability. First, in many cases the accuracy of decision trees is comparable or higher than the accuracy of other classification models [1]. Second, most decision trees do not require a large number of parameters to be adjusted in their design [2]. Third, due to their intuitively appealing topology, the resulting classification models become easy to comprehend [3,4].

Decision trees are one of the most well-known classification methods, and there are many decision tree induction algorithms encountered in the literature. These traditional decision trees could generally be divided into three categories: decision trees (such as the classic ID3 [5] and C4.5 [6], and the recently proposed decision trees [7–13]), fuzzy decision trees (such as the well-known Fuzzy ID3 [14] and the recently proposed fuzzy decision trees [15–21] and oblique decision trees (such as the well-known CART [22], and other oblique decision trees [3,23–28]).

Decision trees and fuzzy decision trees grow in a top-down way when we successively partition the training data into subsets having similar or the same output (class labels). Usually, the growth of the tree terminates when all data associated with a node belong to the same class [29]. Most of the decision trees and fuzzy decision trees partition the training data into subsets by involving in this process only a single feature, thus, the boundaries of partition regions are parallel to one of the axis of the feature space (the structure of the classification regions is thus axis-parallel). When the data are more suitable to be partitioned by hyperplanes that are not axis-parallel, the decision trees and fuzzy decision trees may produce complicated structures (typically oversized) and yield inaccurate results [3]. In this case, oblique decision trees are more suitable. Most oblique decision trees are associated with a linear decision function positioned at each node [22,3,23–28]. However, Erick and Chandrika pointed out that oblique trees are difficult to interpret [28].

In light of these observations, we can draw a conclusion that axis-parallel decision trees cannot find the oblique geometric structures. On the other hand, oblique decision trees exhibit a lack of linguistic interpretation and transparency. Motivated by this, in this study, we develop a new class of decision tree – fuzzy

* Corresponding author at: Research Center of Information and Control, Dalian University of Technology, No. 2, Linggong Road, Ganjingzi, Dalian, Liaoning 116024, PR China. Tel.: +86 41184709381.
E-mail addresses: 79810442@qq.com (X. Wang), xdlious@dlut.edu.cn (X. Liu), wpedrycz@ualberta.ca (W. Pedrycz), lishizhangcc@163.com (L. Zhang).

http://dx.doi.org/10.1016/j.patcog.2014.08.001
0031-3203/

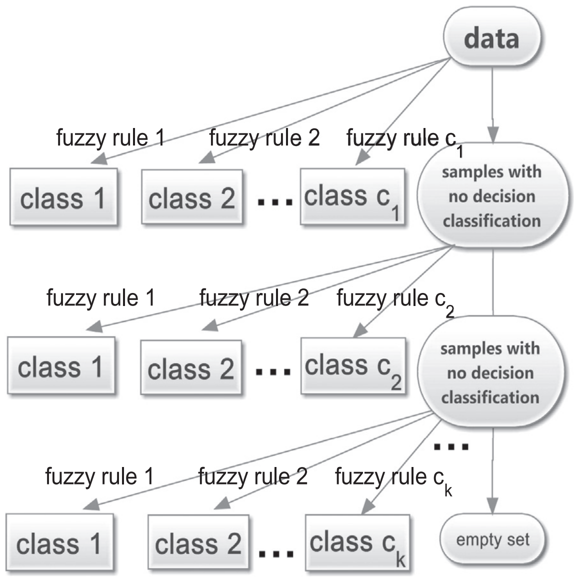

Fig. 1. An overall structure of fuzzy rule-based classification tree.

rule based decision tree (FRDT) for data with continuous features to capture the oblique geometric structure of class region and endow with a readable linguistic interpretation. The main idea is to form a node for each class at each level of the hierarchy while every node implies an oblique geometric structure represented by a fuzzy rule. The fuzzy rules are extracted by the proposed association rules extraction algorithm (AREA). The formation of these fuzzy rules is guided by a criterion of *Fuzzy Confidence*, which is used here instead of the criteria of impurity measures (such as Gain ratio [5,6] and Gini index [22]) being commonly considered in decision trees.

The main procedure of the proposed FRDT can be highlighted as follows (refer to Fig. 1). At the root node of the tree, a single fuzzy rule is extracted by AREA to mine a major oblique geometric structure for every class to form a "pure" (homogeneous) leaf node. The samples which are not covered by these fuzzy rules of first layer are arranged in an additional "impure" node (including a mixture of data coming from different classes). If the impure node is not empty, the growth of the FRDT is realized by expanding the impure node as illustrated in Fig. 1. In this procedure some new fuzzy rules are formed for the impure node. In the sequel, these new fuzzy rules are employed to produce new pure nodes, meanwhile a new impure node is added to collect all patterns for which class assignment cannot be realized by using these new rules existing at this layer of the tree. The process is repeated until the termination criterion (such as the newly added impure node becomes empty) has been satisfied.

The proposed FRDT is different from the axis-parallel decision trees and the oblique decision trees in the sense that the former just considers only a single feature at each node, while the latter takes into account a hyperplane as a decision function. FRDT exhibits a number of visible differences. First of all, at each non-terminal node of the proposed tree, fuzzy rules are employed to describe some oblique geometric structures and to extract some pure leaf nodes. Obviously, this gives rise to a completely new geometry of the partition of the feature space which becomes quite different from the one associated with the "traditional" decision trees. Then, an additional impure node is added for those samples that cannot be assigned to the classes by fuzzy rules of each layer of the tree. The growth of the FRDT tree is realized by just expanding the single added impure node at each layer, while in case of the traditional trees, we have to check each node and expand all non-impure nodes. Thus, the tree of FRDT has only a single trunk (namely the tree of FRDT has only one non-leaf node at each layer), and save great more searching time. Finally, searching of the fuzzy rule realized by the proposed AREA algorithm is guided by *Fuzzy Confidence*, which is used instead of those impurity measures (such as Gain ratio and Gini index) in the "traditional" decision trees.

The proposed FRDT arises as a new architecture of the fuzzy decision tree based on fuzzy rules. The additional impure node is added for those samples that cannot be assigned to the classes with the defined fuzzy rules at every layer of the tree. Some new fuzzy rules are formed on the additional impure node. The samples located at the impure node give rise to some fuzzy rules for the new leaf nodes. These leaf nodes do not affect the new fuzzy rules in the next layer. The same holds for the successive layers. Along with the growth of the tree, more new fuzzy rules are formed for the samples in the impure nodes. Thus new fuzzy rules are determined for the added impure node as the growth of the FRDT tree and the fuzzy rules control the expanding of the FRDT tree progresses.

It has been found experimentally that the proposed FRDTs do not require pruning. A series of numeric studies are reported for UCI machine learning data sets. A comparative analysis is completed for other trees such as C4.5, LADtree, BFTree, SimpleCart, and NBTree. The results of statistical tests have shown that the proposed FRDT exhibits the best performance in terms of both accuracy and the size of the produced trees.

The paper is organized as follows: Section 2 introduces how to generate fuzzy rules and provides some illustrative examples. Section 3 outlines an overall architecture of FRDT and the algorithm of FRDT. In Section 4, we evaluate the proposed FRDT by running it on both synthetic data set and benchmark data sets, and offer a thorough parametric analysis of the tree. Section 5 concludes this paper.

2. Generation of fuzzy rules

The crucial design problem of FRDT is how to extract the major geometric structures with linguistic interpretation. In this paper, a geometric structure with linguistic interpretation is corresponding to the antecedent clauses of a fuzzy "if-then" rule, the antecedent part of fuzzy rule consists of fuzzy number, the inferred class is the consequence of the fuzzy rule. First of all, let us look at a way of forming fuzzy numbers.

2.1. Formation of fuzzy numbers

The data can be represented in the form $X=[x_{ij}]$ being an $n \times m$ matrix. Each column of X corresponds to a given feature (variable), and each row corresponds to a pattern (data point). Let $f_j (j=1,2,\ldots,m)$ denote the j-th column (feature) of X, $x_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$ $(i=1,2,\ldots,n)$ denote the i-th pattern, x_{ij} be the j-th feature value of x_i, and $y_i \in \{1,2,\ldots,c\}$ be a class label of x_i, c be the number of classes, X contain c classes $X_l (l=1,2,\ldots,c)$.

Quite commonly, asymmetric trapezoidal and triangular forms of fuzzy numbers [30,31] are studied. The membership functions of these fuzzy numbers are shown in Fig. 2. Here a one-dimensional feature space is described in terms of four fuzzy sets, "*big*", "*medium big*", "*medium small*", and "*small*". We assume that the number of fuzzy numbers defined for the j-th feature f_j is equal to the number of classes c. Let $f_{k,j} \in F$ denote the k-th $(k=1,2,\ldots,c)$ fuzzy number formed for the j-th feature f_j. Each triangular fuzzy number $f_{k,j}$ is characterized by three parameters $ms_{k-1,j}$, $ms_{k,j}$ and $ms_{k+1,j}$ as shown in Fig. 2. Each trapezoidal fuzzy number $f_{k,j}$ is characterized by two parameters. If $k=1$, trapezoidal fuzzy