Data & Knowledge Engineering 84 (2013) 1–25

Contents lists available at SciVerse ScienceDirect

Data & Knowledge Engineering

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

Extraction of fuzzy rules from fuzzy decision trees: An axiomatic fuzzy sets (AFS) approach

Xiaodong Liu [a,b], Xinghua Feng [a], Witold Pedrycz [c,*]

[a] Research Center of Information and Control Dalian University of Technology, Dalian 116024, PR China
[b] Department of Mathematics Dalian Maritime University, Dalian 116026, PR China
[c] Department of Electrical and Computer Engineering University of Alberta, Edmonton, Canada T6G 2G7, Department of Electrical and Computer Engineering Faculty of Engineering, King Abdulaziz University Jeddah, 21589, Saudi Arabia and Systems Research Institute, Polish Academy of Sciences Warsaw, Poland

ARTICLE INFO

Article history:
Received 17 March 2009
Received in revised form 4 December 2012
Accepted 4 December 2012
Available online 14 December 2012

Keywords:
Fuzzy decision trees
Fuzzy rules
AFS fuzzy logic
Knowledge representation
Comparative analysis

ABSTRACT

In this study, we introduce a new type of coherence membership function to describe fuzzy concepts, which builds upon the theoretical findings of the Axiomatic Fuzzy Set (AFS) theory. This type of membership function embraces both the factor of fuzziness (by capturing subjective imprecision) and randomness (by referring to the objective uncertainty) and treats both of them in a consistent manner. Furthermore we propose a method to construct a fuzzy rule-based classifier using coherence membership functions. Given the theoretical developments presented there, the resulting classification systems are referred to as AFS classifiers. The proposed algorithm consists of three major steps: (a) generating fuzzy decision trees by assuming some level of specificity (detailed view) quantified in terms of threshold; (b) pruning the obtained rule-base; and (c) determining the optimal threshold resulting in a final tree. Compared with other fuzzy classifiers, the AFS classifier exhibits several essential advantages being of practical relevance. In particular, the relevance of classification results is quantified by associated confidence levels. Furthermore the proposed algorithm can be applied to data sets with mixed data type attributes. We have experimented with various data commonly present in the literature and compared the results with that of SVM, KNN, C4.5, Fuzzy Decision Trees (FDTs), Fuzzy SLIQ Decision Tree (FS-DT), FARC-HD and FURIA. It has been shown that the accuracy is higher than that being obtained by other methods. The results of statistical tests supporting comparative analysis show that the proposed algorithm performs significantly better than FDTs, FS-DT, KNN and C4.5.

1. Introduction

There have been numerous approaches to the extraction of classification rules from numeric data [1–11]. One of the quite often used alternatives is to construct a decision tree and afterwards extract rules from this tree [12,1,2]. Due to the nature of continuous attributes as well as various facets of uncertainty present in the problem one has to take into consideration, there has been a visible trend to cope with the factor of fuzziness when carrying out learning from examples in tree induction. In a nutshell, this trend gave rise to the generalizations known as fuzzy decision trees, cf. [13–26,1,12].

Fuzzy decision trees grow in a top-down way, recursively partitioning the training data into segments with similar or the same outputs. Various approaches to the generation of fuzzy decision trees have been suggested by many authors (e.g., [13–26]). Fuzzy

* Corresponding author. Tel.: +1 780 492 4661; fax: +1 780 492 1811.
E-mail address: wpedrycz@ualberta.ca (W. Pedrycz).

0169-023X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.datak.2012.12.001

decision trees encountered in the literatures can be categorized into several types depending upon the nature of the splitting mechanism being in their design:

- Fuzzy ID3 [27–29,7,14,15,23]
  The fuzzy ID3 generalizes ID3, which was initially proposed by Quinlan in the Boolean case [30]. The algorithms in this category apply fuzzy sets to describe (quantify) attributes and then use the ID3 approach to construct the decision tree. Fuzzy entropy, information gain or gain ratio are used as a measure of attribute selection.
- Yuan and Shaw's Fuzzy decision tree (FDT) [31,32,18]
  This approach was introduced by Yuan and Shaw [18]. Being different from the fuzzy ID3, the tree uses the minimal ambiguity (nonspecificity) of a possibility distribution to select attributes for splitting. The attribute is represented through fuzzy sets before constructing the decision tree, and then a measure of a minimal ambiguity is considered to guide attribute selection in the tree building.
- Gini index based FDT [33,34,2,17,19]
  Chandra and Varghese proposed several ways on how to improve the performance of the SLIQ (Supervised Learning in Quest) decision trees [35,36,2,17]. In [2,17], algorithms to generate the fuzzy decision trees using a fuzzy Gini index were presented. In these algorithms, the attribute is not encoded in terms of fuzzy sets before constructing a decision tree. To select the best attribute for splitting, attribute values are fuzzified based on the split-point value.
- Wang's FDT [12,16]
  The approach proposed by Wang et al. uses a measure namely maximum classification importance of attribute contributing to its consequent to select the expanded attributes encoded in terms of fuzzy sets.
- Normalized fuzzy Kolmogorov–Smirnov based FDT [21,22,24]
  Boyen et al. proposed an induction of fuzzy decision trees where fuzzy sets are constructed automatically during the growth of the tree and a normalized fuzzy Kolmogorov–Smirnov discrimination quality measure selects the attribute used in the node splitting.

There are also some other selection methods that emerged in the design of decision trees in the tree building. For example, Pedrycz and Sosnowski introduced algorithms [37,13] using fuzzy granulation for partitioning the input space and constructing the fuzzy decision tree. Unlike the “standard” decision tree considering one attribute at a time to partition the training samples at each node, the fuzzy granulation consider all features to partition the training data. The Fuzzy CHAID algorithm, proposed by Fowdar et al. [38] generates fuzzy trees for both classification and regression problems from pre-generated CHAID decision trees using the Pearson Chi-squared test.

Nevertheless most of these algorithms are variations of the generic algorithmic framework presented above. The applications of the fuzzy decision trees on interval-valued data [20] and multi-valued and multi-labeled [39] have been proposed as well. Combining other data mining techniques with fuzzy decision trees, several hybrid approaches were proposed such as neuro-fuzzy algorithms [40,1] and multiple fuzzy decision trees formed with the aid of rough sets [41,42].

An interesting commonality occurring across most of the existing methods is worth emphasizing: the algorithms require some knowledge about membership functions of the linguistic values of the attributes as well as specific aggregation operations (such as *t*-norms) before any optimization technique can be utilized. It becomes apparent that to a significant extent the obtained fuzzy decision trees are pre-determined by the membership functions of the fuzzy terms and the fuzzy logic operators. Besides, like in the “standard” decision trees, the class label of the terminal node is determined by the label of the majority of the training samples positioned at the node. The difference of the membership degrees and the disproportion between each class are ignored.

In this paper, we use the AFS fuzzy set theory, which facilitates an important step on how to convert the information in databases into the membership functions and their fuzzy logic operations, by taking both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account, and fuzzy entropy as an attribute selection to generate the AFS classifier (decision tree). A crucial threshold being a part of the method affects the tree structure and helps cope with the level of specificity that is being captured by the tree and a leaves labeling method is given. To offer a thorough comparative analysis, we experimented with the algorithm using a number of well known data sets coming from the UCI Repository of Machine Learning data [43]. The ensuing comparative analysis involves some other types of classifiers such as SVM [44], KNN [45], C4.5 [46], Fuzzy Decision Trees (FDTs) [18], Fuzzy SLIQ Decision Tree (FS-DT) [17], FARC-HD [47] and FURIA [48]. The main features of the proposed AFS classifier (decision tree) which distinct it from the fuzzy decision trees can be highlighted as follows:

- The AFS fuzzy sets with their underlying logic operations can eliminate potential subjective bias encountered in the “conventional” fuzzy decision tree and resulting from the use of subjectively formed membership functions.
- The tree structure comes with a great deal of flexibility supplied by the variable threshold is affected by the values of the crucial threshold (δ), which can effectively control the level of detail captured by the tree.
- The relevance of classification results can be explicitly quantified by associated confidence levels.

The paper is organized as follows. In Section 2, we recall the basic notions and properties of the AFS theory that are essential in the framework of our investigations on fuzzy rule extraction. In Section 3, we discuss the coherence membership functions of fuzzy concepts. In Section 4, we introduce an algorithm for generating fuzzy rules from AFS decision trees. Section 5 presents a suite of numeric experiments and offers some comparative analysis. Conclusions are presented in Section 6.