Information Sciences 178 (2008) 3188–3202

Contents lists available at ScienceDirect

Information Sciences

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

Induction of multiple fuzzy decision trees based on rough set technique

Xi-Zhao Wang *, Jun-Hai Zhai, Shu-Xia Lu

Key Lab. of Machine Learning and Computational Intelligence, College of Mathematics and Computer Science, Hebei University, Baoding 071002, China

ARTICLE INFO

Keywords:
Learning
Rough sets
Fuzzy attribute reduct
Fuzzy decision tree induction
Fusion
Integral

ABSTRACT

The integration of fuzzy sets and rough sets can lead to a hybrid soft-computing technique which has been applied successfully to many fields such as machine learning, pattern recognition and image processing. The key to this soft-computing technique is how to set up and make use of the fuzzy attribute reduct in fuzzy rough set theory. Given a fuzzy information system, we may find many fuzzy attribute reducts and each of them can have different contributions to decision-making. If only one of the fuzzy attribute reducts, which may be the most important one, is selected to induce decision rules, some useful information hidden in the other reducts for the decision-making will be losing unavoidably. To sufficiently make use of the information provided by every individual fuzzy attribute reduct in a fuzzy information system, this paper presents a novel induction of multiple fuzzy decision trees based on rough set technique. The induction consists of three stages. First several fuzzy attribute reducts are found by a similarity based approach, and then a fuzzy decision tree for each fuzzy attribute reduct is generated according to the fuzzy ID3 algorithm. The fuzzy integral is finally considered as a fusion tool to integrate the generated decision trees, which combines together all outputs of the multiple fuzzy decision trees and forms the final decision result. An illustration is given to show the proposed fusion scheme. A numerical experiment on real data indicates that the proposed multiple tree induction is superior to the single tree induction based on the individual reduct or on the entire feature set for learning problems with many attributes.

1. Introduction

Since the concept of rough set was originally proposed by Pawlak [1] in 1982, rough set theory (RST), as a new mathematical tool for processing incomplete information, has become a popular topic for many researchers and has been applied successfully to many fields [2–5]. An excellent introduction to rough sets can be found in [6], an excellent collection of some extensions of rough sets has been given in [7]. One particular use of rough set theory is attribute reduction in databases. It means, for a given a dataset with discrete attribute values, to find a subset of the original attributes with the most informative (named reduct) and then to remove all other attributes from the dataset so that the information loss is minimum.

Values of attributes in databases are possibly crisp or fuzzy but traditional RST can only handle the crisp case. Fuzzy values are usually transformed into crisp values to handle. One approach to transformation is to preprocess the data set by discretization in which continuous values of attributes are classified into several symbols. The discretization does not consider linguistic terms with membership functions. Noting that real numbers are well ordered but after discretization the symbols are totally non-ordered, the information loss (at least the order information loss) occurs in the process of discretization. To reduce the loss of this type of information, fuzzification has been proposed to stand for the discretization. Fuzzification is first used to fuzzify real attributes for obtaining a number of linguistic terms with semi-order, and then

* Corresponding author.
E-mail addresses: wangxz@hbu.cn (X.-Z. Wang), mczjh@hbu.cn (J.-H. Zhai), mclsx@hbu.cn (S.-X. Lu).

doi:10.1016/j.ins.2008.03.021

the attribute reduction technique in RST is used to generate rules for reasoning. Combining RST with fuzzy sets has led to a hybrid soft-computing technique in which fuzzy attribute reducts can be achieved by the use of fuzzy core [8–10].

In RST the attribute reduct is the crucial idea. Like RST fuzzy reducts are also particularly useful in fuzzy rough set approaches [11–13]. More and more researchers have been paying attention to investigating the approaches to find fuzzy attribute reducts in fuzzy information system [3,14,15]. The indiscernibility matrix was generalized to fuzzy rough set, by calculating the relative indiscernibility degree of fuzzy condition attribute set with respect to fuzzy decision attributes using fuzzy indiscernibility matrix, and several algorithms for finding fuzzy attribute reducts were given in [14]. Dependency degree was extended in [15] to fuzzy rough set, which is based on fuzzy positive region defined by the fuzzy equivalence class. By calculating fuzzy dependency degree of Q (fuzzy condition attributes) on P (fuzzy decision attribute), a fuzzy attribute reduct can be found, and an algorithm for finding all fuzzy attribute reducts was given in [15]. Using the algorithm in [15], the most important fuzzy attribute subsets (reducts) could be found to extract decision rules, but some more important subsets (reducts) could be discarded. Because every fuzzy attribute reduct may make significant and different contributions to decision-making, if only one of them is selected to induce decision rules, even if it is the most important one, partial useful information could be lost. In order to make full use of the information provided by every fuzzy attribute reduct, in this paper, we present a novel induction of multiple fuzzy decision trees based on the fuzzy rough set technique, the induction approach consist of three stages. First we find several fuzzy attribute reducts using an improved version based on the algorithm in [15], then a fuzzy decision tree for each fuzzy attribute reduct is generated by fuzzy ID3 algorithm and therefore a number of fuzzy decision trees are obtained. Finally we use the method of multiple classifiers fusion based on fuzzy integral to combine the outputs of the multiple fuzzy decision trees into a final result.

The rest of this paper is organized as follows. In Section 2, some notions and methods related to rough attribute reduction and fuzzy rough attribute reduction are introduced and an improved algorithm for finding reducts is given. In Section 3, the method of single fuzzy decision tree induction is introduced. In Section 4, the method of induction of multiple fuzzy decision trees based on fuzzy integral and fuzzy rough set technique is presented and an example is provided to illustrate our method. An experiment is conducted on a real dataset to verify the effectiveness of the proposed method in Section 5. Finally, Section 6 concludes this paper.

2. Rough set and fuzzy rough set attribute reduction

In this section, we recall some notions and methods related to rough attribute reduction and fuzzy rough attribute reduction, present the improved algorithm. Throughout this paper, we confine ourselves to the consideration only on the finite universe of discourse.

2.1. Rough set attribute reduction

An information system can be represented as a triple $\langle U, A, \{V_a\}_{a\in A}\rangle$, where U is set objects (the universe of discourse), A is set of attributes and V_a is a domain of the attribute a. In a decision system, $A = \{C \cup D\}$, where C is the set of conditional attributes and D is the set of decision attributes. With any $P \subseteq A$ there is an associated equivalence relation$IND(P)$ (called indiscernability relation)

$$IND(P) = \{(x,y) \in U \times U | a(x) = a(y) \text{ for all } a \in P\} \tag{1}$$

The partition of U, generated by $IND(P)$ is denoted by U/P and can be calculated as follows:

$$U/P = \otimes\{a \in P | U/IND\{a\}\} \tag{2}$$

where $A \otimes B = \{X \cap Y |\ \forall X \in A, \forall Y \in B, X \cap Y \neq \phi\}$, the equivalence classes of the P-indiscernibility relation are denoted $[x]_P$.

Let $X \subseteq U$, the P-lower approximation of a set can be defined as:

$$\underline{P}X = \cup\{Y \in U/P, Y \subseteq X\} = \{[x]_P | [x]_P \subseteq X\} \tag{3}$$

Let P and Q be equivalence relations over U, then the P-positive region of Q, denoted by $POS_P(Q)$ can be defined as follows:

$$POS_P(Q) = \cup_{X\in U/Q}\underline{P}X \tag{4}$$

The P-positive region of Q contains all objects of the universe U that can be properly classified into classes of U/Q using the knowledge expressed by the classification U/P.

An important concept in rough set techniques is dependence between attributes; dependency can be defined as follows: For P, $Q \subseteq A$, Q dependency on P in a degree k $(0 \leqslant k \leqslant 1)$, denoted $P \Rightarrow_k Q$,

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \tag{5}$$

where $|S|$ denotes the cardinality of set S.

By calculating the change in dependency when an attribute is removed from the set of considered conditional attributes, a measure of the significance of the attribute can be obtained. Given P, Q and an attribute $x \in P$, the significance of attribute x upon Q is defined by