Neurocomputing 148 (2015) 340–353

Contents lists available at ScienceDirect

Neurocomputing

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

Forward approximation and backward approximation in fuzzy rough sets

Yi Cheng *

Department of Information Engineering, Sichuan College of Architectural Technology, Deyang 618000, China

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form
20 January 2014
Accepted 27 June 2014
Communicated by N.T. Nguyen
Available online 7 July 2014

Keywords:
Fuzzy rough sets
Rule extraction
Data mining

ABSTRACT

It is general to obtain rules by attribute reduction in fuzzy information systems. Instead of obtaining rules by attribute reduction, which may have a negative effect on inducting good rules, the objective of this paper is to extract rules without computing attribute reducts. Forward and backward approximations in fuzzy rough sets are first defined, and their important properties are discussed. Two algorithms based on forward and backward approximations, namely, mine rules based on the forward approximation (MRBFA) and mine rules based on the backward approximation (MRBBA), are proposed for rule extraction. The two algorithms are evaluated by several data sets from the UC Irvine Machine Learning Repository. The experimental results show that both MRBFA and MRBBA achieve better classification performances than the method based on attribute reduction.

1. Introduction

In the early eighties, Pawlak [1] introduced the theory of rough sets for the study of intelligent systems characterized by insufficient and incomplete information. The rough set theory describes a crisp subset by two definable subsets called lower and upper approximations. Using the lower and upper approximations, the knowledge hidden in information systems can be discovered and expressed in the form of decision rules. The classical rough set theory is used only to describe crisp sets. To describe crisp and fuzzy concepts, Dubois and Prade [2,3] extended the basic idea of rough sets to a new model called fuzzy rough sets. In fuzzy rough sets, a fuzzy similarity relation is employed to describe the degree of similarity between two objects instead of an equivalence relation used in rough sets.

Research on fuzzy rough sets include knowledge representation and knowledge reduction. There are two common approaches for knowledge representation, the constructive [4–12] and axiomatic approaches [13–20]. The objective of knowledge reduction is to reduce attributes and learn rules from samples. There are two main approaches to reducing knowledge [21]: one is attribute reduction and the other is rule extraction.

Attribute reduction based on fuzzy rough sets has been studied by some scholars [12,22–32]. Usually, reduction methods can be classified into three types: one based on the positive region [22–25], one based on the discernibility matrix [12,26–28], and the third based on entropy [29–32]. For example, Shen and Jensen [22,23] conducted pioneering studies on attribute reduction based on a positive region. However, the dependency degree of a selected reduct may be larger than that of the entire attribute set, which means that less attributes can offer better approximations [12]. This is unreasonable because more attributes will offer better approximations in a rough set framework [12]. Moreover, only one reduct can be obtained. Thus it is unclear which attribute in the reduct is indispensable, i.e., the core of the reduct is unknown [12]. Furthermore, the time complexity of the algorithm often increases exponentially with increasing samples and attributes [24]. After claiming that Shen's algorithm [23] may not be convergent on many real data sets due to its poorly designed termination criteria, Bhatt and Gopal [24] developed Shen's algorithm by improving the definition of the lower approximation on a compact computational domain. However, they still compute the positive region using the same method proposed in Ref. [23]; thus the dependency degree of a selected reduct may still be larger than that of the entire attribute set [12]. Tsang et al. [12,26] proposed an algorithm using a discernibility matrix to compute all attribute reducts. However, the computation complexity is NP-hard, so they used a heuristic algorithm to find a close-to-minimal reduct instead of all reducts [27]. In addition, the definition of fuzzy similarity relation is faulty. For example, if R is a fuzzy similarity relation defined in Refs. [12,26], then $R(0.1, 0.1)=1$, $R(0.1, 0.11)=0.1$, $R(0.9, 0.91)=0.9$. It is clear that $R(0.1, 0.11)=0.1$ is

* Correspondence address: Department of Information Engineering, Sichuan College of Architectural Technology, Deyang 618000, China.

http://dx.doi.org/10.1016/j.neucom.2014.06.062
0925-2312/

not reasonable according to general knowledge. Hu and Yu [29,30] proposed an attribute reduction method based on information entropy. The attribute reduction concept is not constructed using the existing fuzzy approximation operators [8], and studying the structure of attribute reduction is difficult [33]. Each attribute reduction method has its characteristics; at the same time, each one also has some flaws.

Methods of rule extraction based on fuzzy rough sets are relatively less researched [21,23,25,34–36]. Attribute reduction usually serves as a preparatory step before rule extraction, whose objective is to reduce attributes and thus reduce the complexity of the rule extraction process. A representative work is found in Refs. [23,25]. Shen and Jensen [23,25] reduced attributes based on the positive region and then extracted rules using an existing fuzzy rule induction algorithm (RIA). Attribute reduction is applied as a pretreatment to RIA. However, the flaws of attribute reduction may have a negative effect on the induction of good rules. Other rule extraction methods are summarized as follows. Wang and Hong [34] proposed first transforming the fuzzy values to crisp values and then computing the corresponding reducts and core. Some information hidden in the fuzzy values, such as partial ordering relation and membership degree, is lost. Hong [34] proposed an algorithm to produce a set of fuzzy rules from noisy quantitative training data by applying the variable precision rough set model. However, only a set of maximal general fuzzy rules can be found for an approximate coverage of training samples. Tsang et al. [36] presented a rule extraction method based on fuzzy rough sets. The essence is to obtain rules by attribute value reduction since each object can be seen as an original fuzzy decision rule in a fuzzy decision table. Wang et al. [21] proposed new definitions of fuzzy lower and upper approximations by considering the similarity between two objects. Based on the new fuzzy similarity relation, the close-to-minimal rule set is found. However, the time complexity of the method increases with the square of the size of the universe. Moreover, Wang [21] only applied the method to the information systems with both fuzzy condition attributes and crisp decision attributes. Few discussions focus on the information systems with both fuzzy condition and decision attributes.

As a previous work, the author had presented the forward approximation and the backward approximation in rough fuzzy sets [37]. The rule exaction methods based on the forward and backward approximations were put forward [37]. Rough fuzzy sets are special cases of fuzzy rough sets. The rule extraction methods in Ref. [37] are only effective for fuzzy information systems with crisp condition and fuzzy decision attributes. For the other fuzzy information systems, the methods appear unsuitable. Establishing a more practical model for fuzzy rule extraction in fuzzy information systems is necessary. The model should be effective for three types of fuzzy information systems, namely, (1) crisp condition and fuzzy decision, (2) fuzzy condition and crisp decision, and (3) fuzzy condition and decision. This paper intends to avoid the attribute reduction process and establish the structure of the approximation by introducing granulation order.

From the viewpoint of granular computing, a concept is described by the upper and lower approximations under static granulation in the fuzzy rough sets, as defined by Dubois and Prade [2,3]. Provided the granulation is unchangeable, it is unacceptable when the granulation is too fine or too coarse. Excessively fine granulation may increase time and cost, while an excessively coarse one may not satisfy the requirements. We consider describing a concept under dynamic granulation. This means a proper granulation family can be selected to describe a target concept according to the practical requirement. In this paper, forward and backward approximations in fuzzy rough sets are proposed based on a granulation order. Two algorithms based on forward and backward approximations, namely, mine rules based on the forward approximation (MRBFA) and mine rules based on the backward approximation (MRBBA), are proposed for rule extraction.

The rest of this paper is organized as follows. In Section 2, related notions of fuzzy rough sets are briefly introduced. Forward and backward approximations in fuzzy rough sets are first defined, and their important properties are discussed. Two algorithms based on forward and backward approximations, namely, mine rules based on the forward approximation (MRBFA) and mine rules based on the backward approximation (MRBBA), are proposed for rule extraction. In Section 3, the two algorithms MRBFA and MRBBA are compared with the algorithm in Ref. [23] and then evaluated by several data sets from the UC Irvine Machine Learning Repository (UCI). Section 4 concludes the paper.

2. Forward approximation and backward approximation

2.1. Preliminaries

In this section, we briefly introduce related discussions about fuzzy rough sets.

2.1.1. Fuzzy rough sets

An equivalence relation is a basic notion in Pawlak's rough set theory [1]. In fuzzy rough sets, a fuzzy similarity relation is used to replace the equivalence relation. Let U be a nonempty universe. A fuzzy binary relation R on U is called a fuzzy similarity relation if R satisfies reflexivity ($R(x, x)=1$), symmetry ($R(x, y)=R(y, x)$) and sup-min transitivity ($R(x, y) \geq \sup_{z \in U} \min\{R(x, z), R(z, y)\}$). Using the fuzzy similarity relation, the fuzzy equivalence class $[x]_R$ can be defined by $\mu_{[x]_R}(y)=\mu_R(x, y)$ for all $y \in U$. The family of normal fuzzy sets produced by a fuzzy partition of the universe can play the role of fuzzy equivalence classes. Consider a crisp partition $U/Q=\{\{1, 3, 6\}, \{2, 4, 5\}\}$. This contains two equivalence classes ($\{1, 3, 6\}$ and $\{2, 4, 5\}$) that can be regarded as degenerated fuzzy sets, with those elements belonging to the class possessing a membership of one, zero otherwise. For the first class, for instance, objects 2, 4 and 5 have a membership of zero. Extending this to the case of fuzzy equivalence classes is straightforward: objects can be allowed to assume membership values, with respect to any given class, in the interval [0, 1]. Therefore, U/Q is not restricted to crisp partition only; fuzzy partition is equally acceptable. The collection of all fuzzy equivalence classes can be denoted as U/R. There are some discussions on how to construct a fuzzy similarity relation by fuzzy attributes [38]. However, there is a lack of a normalized method. The objective of the paper is to introduce the rule extraction methods. For simplicity, we follow the method in Refs. [14,22,23], where a fuzzy attribute is denoted by a fuzzy similarity relation briefly.

We adopt the Cartesian product to construct fuzzy equivalence classes for multiple fuzzy attributes. In general, let C be a fuzzy condition attribute set, then $U/C= \otimes\{U/IND(\{a\}), a \in C\}$. Each set in U/C denotes a fuzzy equivalence class. For example, if $C=\{a, b\}$, $U/IND(\{a\})=\{N_a, Z_a\}$ and $U/IND(\{b\})=\{N_b, Z_b\}$, then $U/C=\{N_a \cap N_b, N_a \cap Z_b, Z_a \cap N_b, Z_a \cap Z_b\}$.

The concept of fuzzy rough sets was first proposed by Dubois and Prade [2,3]. Their idea is presented as follows:

Let U be a nonempty universe, R be a fuzzy binary relation on U and F be a fuzzy set of U. A fuzzy rough set is a pair of fuzzy sets $\underline{apr}_R(F)$ and $\overline{apr}_R(F)$ defined as

$$\mu_{\overline{apr}_R(F)}(x)=\sup_{y \in U} \min\{\mu_R(x, y), \mu_F(y)\},$$

$$\mu_{\underline{apr}_R(F)}(x)=\inf_{y \in U} \max\{1-\mu_R(x, y), \mu_F(y)\}.$$