Applied Soft Computing 27 (2015) 322–331

Contents lists available at ScienceDirect

Applied Soft Computing

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

Flow-based tolerance rough sets for pattern classification

Yi-Chung Hu*

Department of Business Administration, Chung Yuan Christian University, Chung Li Dist., Taoyuan, Taiwan, R.O.C.

ARTICLE INFO

Article history:
Received 10 January 2013
Received in revised form 25 October 2014
Accepted 22 November 2014
Available online 2 December 2014

Keywords:
Tolerance rough set
Classification
Preference information
Pairwise comparison
MCDA

ABSTRACT

Rough set theory is a useful mathematical tool for pattern classification to deal with vagueness in available information. The main disadvantage of rough set theory is that it cannot handle continuous attributes. Although various discretization methods have been proposed to deal with this problem, discretization can result in information loss. It has been found that tolerance rough sets with a tolerance relation can operate effectively on continuous attributes. A tolerance relation is related to a similarity measure which is commonly defined by a simple distance function to measure the proximity of any two patterns distributed in feature space. However, for a simple distance measure, it oversimplifies the criteria aggregation resulting from not considering attribute weights, and it is not a unique way of expressing the preference information on each attribute for any two patterns. This paper proposes a flow-based tolerance rough set using flow, which represents the intensity of preference for one pattern over another, to measure similarity between two patterns. To yield high classification performance, a genetic-algorithm-based learning algorithm has been designed to determine parameter specifications and generate the tolerance class of a pattern. The proposed method has been tested on several real-world data sets. Its classification performance is comparable to that of other rough-set-based methods.

1. Introduction

Pawlak [39,40] introduced rough set theory to approximate a vague concept (e.g., X) in the universe U in terms of a pair of precise sets which are known as upper and lower approximations. Undoubtedly, rough set theory is a useful technique for analysis of vague concepts in the field of multiple attributes decision analysis (MCDA) [6–8,41,42,52]. However, traditional rough set-based methods are restricted by the requirement that all quantitative attributes must be discrete [38]. Discretization is usually performed before these methods are used. However, the main disadvantage of discretization is information loss. The tolerance rough set (TRS), which was further developed on the basis of rough set theory, was found to handle continuous attributes effectively [24,38,50]. A number of researchers [24–26,29,35,38,50,58] have addressed applications of TRS to pattern classification, such as handwritten numeral characters, remote sensing data and land cover, by treating each class in a classification problem as a concept in a given decision table. TRS indeed plays an important role in pattern recognition. In a traditional TRS, the tolerance classes are determined using a tolerance relation, which is commonly defined by a simple distance measure [47] which indicates the proximity of any two patterns distributed in feature space. Moreover, a similarity threshold must be specified to determine the required level of similarity between any two patterns.

The problem addressed by this paper is that, although the use of a simple distance measure for estimating the similarity is simple enough for the traditional TRS, it oversimplify the criteria aggregation resulting from not considering attribute weights, and it is not a unique way of expressing the preference information on each attribute for any two patterns. In other words, the simple distance measure may not be an appropriate choice to measure the similarity for TRS. For MCDA classification methods, the outranking relation theory (ORT), first established by Roy [45], with pairwise judgments on each attribute have gained more attention [14,15]. The outranking relations can provide the preference information among patterns by using pairwise comparisons. This provides a motivation for using preference for one pattern over another to measure the similarity instead of a distance function.

Among outranking methods, the well-known Preference Ranking Organization METHods for Enrichment Evaluations (PROMETHEE) methods introduced by Brans, Marechal and Vincke [3,4,10,11,51] is an effective method that can be used to measure the strength of the preference for one pattern over another by estimating the outranking (leaving flow) and the outranked character (entering flow) of each pattern. This paper contributes to propose a novel flow-based TRS (FTRS) which uses preference information expressed by flows among patterns to measure the similarity

* Tel.: +886 3 2655130.
E-mail address: ychu@cycu.edu.tw

http://dx.doi.org/10.1016/j.asoc.2014.11.021
1568-4946/

between any two patterns. For the FTRS, when the net flow of one pattern is sufficiently close to that of another pattern to fall below a given similarity threshold, the former (latter) can be included within the tolerance class for the latter (former). After the tolerance classes for all patterns have been determined, a classification procedure can be used to assign each pattern to a class. To construct a classifier with high classification performance, because genetic algorithms (GA) are a powerful search and optimization method [17,31,44], a genetic-algorithm-based method has been developed here that automatically determines the relative weight of each attribute and a similarity threshold that yields high classification performance.

The rest of the paper is organized as follows. Sections 2 and 3 briefly introduce rough sets and TRS with a traditional similarity measure respectively. Section 4 presents the proposed FTRS. Section 5 describes the GA-based learning algorithm for the proposed FTRS-based classifier (FTRSC). Section 6 reports the experimental results of the application of the proposed method to some real-world data sets. Several rough-set-based classification methods presented by Skowron et al. [49] are taken into account. The results show that the proposed FTRSC with subset and concept approximations performs well in comparison with traditional TRSC. Section 7 presents the discussion and conclusions.

2. Rough sets

Rough set theory can deal with vagueness and uncertainty in decision making. Let $\mathbf{S} = (U, A \cup D)$ be a decision table, where U is a non-empty set of finite elements, A is a non-empty set of finite attributes, and D is a non-empty set of finite decision classes. Each attribute $a \in A$ defines an information function f_a: $U \to V_a$, where V_a is the set of values of a. For any $P \subseteq A$, an indiscernibility relation Ind(P) can be defined as follows:

$$\mathrm{Ind}(P) = \{(\mathbf{x}_i, \mathbf{x}_j) \in U^2 | f_i(a) = f_j(a), \forall a \in P\} \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are indiscernible when $(\mathbf{x}_i, \mathbf{x}_j) \in \mathrm{Ind}(P)$. Some equivalence classes or elementary sets are generated by Ind(P). The elementary set of a pattern $\mathbf{x}$ is represented by $[\mathbf{x}]_P$. Any finite union of elementary sets is called a P-definable set [19]. For pattern classification, a concept X consists of elements that have the same class label, so that $X \in U/D$.

Sometimes, $X \subseteq U$ is not P-definable. In other words, there exist elements in the same elementary set which have different class labels, so that X is a vague concept. In this case, X can be approximated by a pair of precise concepts [39,52] using the P-upper approximation, $\bar{P}X$, and the P-lower approximation, $\underline{P}X$, as follows:

$$\bar{P}X = \{\mathbf{x} | \mathbf{x} \in U, [\mathbf{x}]_P \cap X \neq \phi\} \tag{2}$$

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

where $\underline{P}X \subseteq \bar{P}X$ and $\underline{P}X$ consists of elements that certainly belong to X, whereas $\bar{P}X$ consists of elements that possibly belong to X. The tuple $\langle \underline{P}X, \bar{P}X \rangle$ composed of the lower and upper approximations is called a rough set. $\underline{P}X$ and $\bar{P}X$ are so-called traditional singleton approximations. When $\bar{P}X = \underline{P}X$, X is precise with respect to P (i.e., X is definable); when $\bar{P}X \neq \underline{P}X$, X is rough with respect to P (i.e., X is undefinable). Moreover, the vagueness of X can be described by the accuracy of the rough set representation of X:

$$\alpha_P(X) = \frac{|\underline{P}X|}{|\bar{P}X|} \tag{4}$$

where $0 \le \alpha_P(X) \le 1$. $\alpha_P(X)$ provides an indication of how closely the rough set approximates X. $\alpha_P(X) = 1$ means that this concept can be approximated without any uncertainty using the granulation of rough set theory. A vague concept has the boundary region $\mathrm{BND}_P(X)$, consisting of elements that cannot be categorized into the concept with certainty, where $\mathrm{BND}_P(X)$ is defined as:

$$\mathrm{BND}_P(X) = \bar{P}X - \underline{P}X \tag{5}$$

The degree of inclusion of $\mathbf{x}$ within X with respect to P can be defined by a rough membership function as:

$$\mu_X^P(\mathbf{x}) = \frac{|[\mathbf{x}]_P \cap X|}{|[\mathbf{x}]_P|} \tag{6}$$

where $\mu_X^P(\mathbf{x}) \in [0, 1]$ and $|[\mathbf{x}]_P|$ denotes the cardinality of $[\mathbf{x}]_P$. Undoubtedly, the value of the rough membership function of each pattern in $\underline{P}X$ is 1, that of patterns in $\bar{P}X$ lies in the interval (0, 1], and that of patterns in $\mathrm{BND}_P(X)$ lies in the interval (0, 1). Decision rules induced from the lower approximation of the concept are called certain rules, whereas those induced from the upper approximation of the concept are called possible rules [18]. The set of rules can be used for classification, but this is beyond the scope of this paper.

Because rough set theory is unable to deal with real-valued data, a discretization procedure is usually performed before using it. Discretization is the process of converting continuous attributes into discrete attributes. Although many discretization methods have been proposed [5,12,20], the use of such methods can result in information loss. Furthermore, there is no optimal discretization method for all decision problems [20]. Attention has also been focused on TRS because a TRS can handle real-valued attributes by defining a suitable similarity relation for each attribute.

3. Tolerance rough sets with traditional similarity measures

In this section, the TRS with a traditional similarity measure and a classification procedure proposed by Kim and Bang [25] for TRS are introduced. $\mathbf{x}_i\ R_a\ \mathbf{x}_j$ denotes that $\mathbf{x}_i$ and $\mathbf{x}_j$ are similar with respect to attribute a, where R_a is a tolerance relation with respect to attribute a. A standard similarity measure $S_a(\mathbf{x}_i, \mathbf{x}_j)$ with respect to R_a can be defined by a simple distance function in [47]:

$$S_a(\mathbf{x}_i, \mathbf{x}_j) = 1 - \frac{|a(\mathbf{x}_i) - a(\mathbf{x}_j)|}{\max_a - \min_a} \tag{7}$$

where $a(\mathbf{x}_i)$ and $a(\mathbf{x}_j)$ are attribute values of $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively in V_a, and $\max_a$ and $\min_a$ denote the maximum and minimum values respectively of the domain interval of attribute a. Of course, the same definition can be used for all attributes [38]. The relation between R_a and $S_a(\mathbf{x}_i, \mathbf{x}_j)$ is as follows:

$$\mathbf{x}_i\ R_a\ \mathbf{x}_j \Leftrightarrow S_a(\mathbf{x}_i, \mathbf{x}_j) \ge \tau_a \tag{8}$$

where $\tau_a \in [0,1]$ is the similarity threshold of attribute a. For A, an overall similarity measure $S_A(\mathbf{x}_i, \mathbf{x}_j)$ can be defined as:

$$S_A(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{a \in A} S_a(\mathbf{x}_i, \mathbf{x}_j)}{|A|} \tag{9}$$

The global tolerance relation R_A is related to $S_A(\mathbf{x}_i, \mathbf{x}_j)$ as:

$$\mathbf{x}_i\ R_A\ \mathbf{x}_j \Leftrightarrow S_A(\mathbf{x}_i, \mathbf{x}_j) \ge \tau \tag{10}$$

where $\tau \in [0,1]$ is a global similarity threshold based on all attributes. In contrast to Ind(P), which is an equivalence relation, a tolerance relation has the reflexive and symmetric properties but not the transitivity property.

A tolerance class $TC(\mathbf{x}_i)$ of $\mathbf{x}_i$ can be generated for a certain τ by considering the patterns that have a tolerance relation with $\mathbf{x}_i$ as:

$$TC(\mathbf{x}_i) = \{\mathbf{x}_j \in U | \mathbf{x}_i\ R_A\ \mathbf{x}_j\} \tag{11}$$