Applied Soft Computing 26 (2015) 235–243

Contents lists available at ScienceDirect

Applied Soft Computing

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

An improved attribute reduction scheme with covering based rough sets

Changzhong Wang[a], Mingwen Shao[b,*], Baiqing Sun[c], Qinghua Hu[d]

[a] Department of Mathematics, Bohai University, Jinzhou 121000, PR China
[b] College of Information Science and Technology, Shihezi University, Shihezi, Xinjiang 832003, PR China
[c] Economy and Management School, Harbin Institute of Technology, Harbin 150001, PR China
[d] School of Computer Science and Technology, Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:
Received 28 October 2012
Received in revised form 3 September 2014
Accepted 7 October 2014
Available online 17 October 2014

Keywords:
Covering based rough set
Attribute reduction
Discernibility matrix

ABSTRACT

Attribute reduction is viewed as an important preprocessing step for pattern recognition and data mining. Most of researches are focused on attribute reduction by using rough sets. Recently, Tsang et al. discussed attribute reduction with covering rough sets in the paper (Tsang et al., 2008), where an approach based on discernibility matrix was presented to compute all attribute reducts. In this paper, we provide a new method for constructing simpler discernibility matrix with covering based rough sets, and improve some characterizations of attribute reduction provided by Tsang et al. It is proved that the improved discernibility matrix is equivalent to the old one, but the computational complexity of discernibility matrix is relatively reduced. Then we further study attribute reduction in decision tables based on a different strategy of identifying objects. Finally, the proposed reduction method is compared with some existing feature selection methods by numerical experiments and the experimental results show that the proposed reduction method is efficient and effective.

1. Introduction

Attribute reduction and feature selection have become one of the important steps for machine learning tasks. Classical rough set theory [21] is a mathematical tool for handling data sets with imprecision and uncertainty. It can be employed to study attribute reduction and feature selection in information systems [2,4,6–8,10–17]. Equivalence relations are the mathematical basis for the rough set theory. On the basis of equivalence relations, samples of a universe can be partitioned into exclusive equivalence classes, which form basic information granules. By using basic information granules one can approximate arbitrary subset of the universe. The main idea of rough sets is to remove redundant information in data and to make correct decision or classification. Rough set theory has attracted wide attention in both the theory and its applications [9–19,23–26,29–39].

However, equivalence relation in classical rough set theory is limited in practice, as it only deals with discrete variables. There are large amount of continuous data in real-life applications. For example, one can be faced with a lot of numerical data in performance analysis and equipment condition monitoring and diagnosis in power systems [22]. When dealing with such numerical attributes by using classical rough sets, numerical attributes are often discretized into symbol-type attributes as a pretreatment [20]. This type of conversion can bring information loss, thus affecting the accuracy of extracted rules [7]. In order to solve this problem, scholars have proposed a series of extensions of the rough set model [1,3,5,18,24,27,31] and presented some important feature selection criterions [4,6–8,10–16,23,26]. For example, Jensen and Shen [7] presented fuzzy-rough reduct algorithm based on Max-Dependency criterion. Pradipta Maji proposed maximum relevance and maximum significance criterion of feature selection based on fuzzy and rough sets [10–16,23]. These generalized methods of rough sets have been applied successfully to feature selection of continuous data [7,10–16]. Another important extension of rough sets is covering based rough sets. In Ref. [24], Pomykala first introduced the concepts of covering lower and upper approximation operators in rough approximation space. This innovation is based on the substitution of indiscernibility relations by coverings. Afterward, many authors investigated properties of covering approximation operators [19,28,32,37–39]. However, few people employ covering based rough sets to make research on attribute reduction. In [28] a pioneering work related to attribute reduction with covering based rough sets was conducted, where the authors constructed discernibility matrix, analyzed its some important properties, and developed an approach to compute all the reducts. Their main idea of attribute reduction is that each object is associated with a neighborhood, and then the identification between any two objects is performed by distinguishing the relationships of their neighborhoods. A major drawback of the strategy is that the constructed formula for computing discernibility matrix is very complicated, and so cannot be easily applied to practice.

From the viewpoint of classical rough sets, two objects can be distinguished if one object does not belong to the neighborhood of another one. Since covering based rough sets are an extension of classical rough sets, we can use the similar strategy to distinguish objects with different decision values. In this paper, we present a simple approach to distinguish two objects using covering base rough sets and reconstruct discernibility matrix of attribute reduction with a simpler formula. The reconstruction is consistent to the viewpoint of identifying objects in classical rough sets. Compared with the approach in [20], the computational complexity of the improved approach is relatively lower. We improve some important properties of attribute reduction proposed in [28]. Furthermore, we study attribute reduction

* Corresponding author. Tel.: +86 15063946809.
E-mail address: mingwenshao@gmail.com (M. Shao).

http://dx.doi.org/10.1016/j.asoc.2014.10.006
1568-4946/

of decision tables based on the improved technique of identifying objects. Finally, experimental results show that the improved approach is feasible and valid.

The remainder of this paper is structured as follows. In Section 2, we recall and define some basic notions related to covering based rough sets. In Section 3, we reconstruct the discernible approach to attribute reduction with covering based rough sets and then develop an approach to attribute reduction in decision tables. The improved method in this paper is compared with the old one [28] and some existing feature selection methods by some numerical experiments in Section 4. Section 5 concludes the paper with a summary.

2. Some notions related to covering based rough sets

Attribute reduction is an important application field of rough set theory. However, in real world there are lots of data sets that cannot be handled well by classical rough sets. In light of this, similarity relation rough sets [27], dominance rough sets [5], and even neighborhood rough sets [6,31] were developed one by one. All these models induce coverings of a universe, instead of partitions, and thus can be categorized into covering rough sets, which are more general than classical rough sets and can handle more complex tasks.

Granulating information in data sets is the basis of rough set theory. The granulated information forms elementary information granules to approximately describe arbitrary concepts in approximation spaces. Covering rough set theory employs the notion of coverings to granulate information in data sets.

Definition 2.1. [24] Let U be a nonempty and finite set of objects, where $U = \{x_1, x_2, \ldots, x_n\}$ is called a universe of discourse. $C = \{K_1, K_2, \ldots, K_m\}$ is a family of nonempty subsets of U and $\bigcup_{i=1}^{m} K_i = U$. We say C is a covering of U, K_i is a covering element, and the ordered pair (U, C) is a covering approximation space.

Definition 2.2. [28] Suppose U is a finite universe and $C = \{K_1, K_2, \ldots, K_m\}$ is a covering of U. For any $x \in U$, let $C(x) = \cap\{K_j \in C : x \in K_j\}$, then $Cov(C) = \{C(x) : x \in U\}$ is also a covering of U, we call it the induced covering of C

$C(x)$ is the minimal descriptive subset containing x. This means $C(x)$ cannot be written as the union of other elements in $Cov(C)$. Thus $C(x)$ can be seen as the information granule of x with respect to C and $Cov(C)$ can be viewed as a set of information granules. These information granules are minimal covering elements associated with objects. For any $x, y \in U$, $y \in C(x)$ if and only if $C(y) \subseteq C(x)$. So if $y \in C(x)$ and $x \in C(y)$, then $C(x) = C(y)$. The relationships between information granules have the following properties.

(1) Reflexivity: $\forall x \in U, x \in C(x)$.
(2) Anti-symmetry: if $y \in C(x)$ and $x \in C(y)$, then $C(x) = C(y)$.
(3) Transitivity: $\forall x, y, z \in U$, if $x \in C(y)$ and $y \in C(z)$, then $x \in C(z)$.

In classification and regression learning, we are usually confronted with the task of approximating some concepts with provided knowledge. With information granules in covering approximation spaces, any concepts can be approximated.

Definition 2.3. [28] Let (U, C) be a covering approximation space. $X \subseteq U$ is an arbitrary subset of the universe. The covering lower and upper approximations of X are defined as $\underline{C}(X) = \{x : C(x) \subseteq X\}$, $\overline{C}(X) = \{x : C(x) \cap X \neq \varnothing\}$.

Definition 2.4. [28] Suppose U is a finite universe and $\Delta = \{C_i : i = 1, 2, \ldots, m\}$ is a family of coverings of U. For any $x \in U$, let $\Delta(x) = \cap\{C_i(x) \in Cov(C_i) : i = 1, 2, \ldots, m\}$, then $Cov(\Delta) = \{\Delta(x) : x \in U\}$ is also a covering of U, we call it the induced covering of Δ.

Clearly, $\Delta(x)$ is the intersection of all the covering elements including x in all coverings, and so the minimal descriptive set containing x in $Cov(\Delta)$. Similarly, $\Delta(x)$ can be viewed as the information granule of x with respect to Δ and $Cov(\Delta)$ can be viewed as a set of information granules with respect to Δ If every covering in Δ is a partition, then $Cov(\Delta)$ is also a partition and $\Delta(x)$ is the equivalence class containing x. Each information granule in $Cov(\Delta)$ cannot be written as the union of other granules. For any $x, y \in U$, $y \in \Delta(x)$ if and only if $\Delta(y) \subseteq \Delta(x)$. So if $y \in \Delta(x)$ and $x \in \Delta(y)$, then $\Delta(x) = \Delta(y)$. The relationships between information granules in $Cov(\Delta)$ also have such properties as reflexivity, anti-symmetry and transitivity. Let $X \subseteq U$, the lower and upper approximations of X with respect to Δ are defined as follows:

$$\underline{\Delta}(X) = \{x \in U : \Delta(x) \subseteq X\}, \ \overline{\Delta}(X) = \{x \in U : \Delta(x) \cap X \neq \varnothing\}.$$

Definition 2.5. Let U be a universe and $\Delta = \{C_i : i = 1, 2, \ldots, m\}$ a family of coverings on U. Then (U, Δ) is called a covering information system; Δ is called a conditional covering (attribute) set.

Definition 2.6. Let (U, Δ) be a covering information system and $C_i \in \Delta$. C_i is called superfluous in Δ if $Cov(\Delta - \{C_i\}) = Cov(\Delta)$, i.e., $(\Delta - \{C_i\})(x) = \Delta(x)$ for any $x \in U$. Otherwise, C_i is called indispensable in Δ. For any subset $P \subseteq \Delta$. P is called a reduct of Δ if each element in P is indispensable in P and $Cov(\mathrm{P}) = Cov(\Delta)$. The collection of all indispensable elements in Δ is called the core of Δ, denoted as $Core(\Delta)$.

Definitions 2.5 and 2.6 are natural extensions of the corresponding concepts in classical rough set theory by substituting equivalence relations with coverings. It can be seen from the two definitions that the purpose for reducing conditional covering set is to find a minimal covering subset that keeps original information granularity invariant.

3. Attribute reduction based on discernibility matrix

In this section, we first develop some theorems to describe discernibility between objects. Then, we reconstruct the discernibility matrix of information systems based on covering and improve some characterizations of basic properties of attribute reduction. Finally, we examine how to find a relative reduct from a given decision table.

Let $U = \{x_1, x_2, \ldots, x_n\}$ be a universe, $\Delta = \{C_1, C_2, \ldots, C_m\}$ be a family of coverings of U. For any $x_i, x_j \in U$, if $x_j \notin C_k(x_i)$, then we say x_i and x_j can be distinguished by C_k. This statement accords to the corresponding views in classical rough sets.

Proposition 3.1. Let $\Delta = \{C_i : i = 1, 2, \ldots, m\}$ be a family of coverings on U, $P \subseteq \Delta$. Then $Cov(P) = Cov(\Delta)$ if and only if $\Delta(x) = \mathrm{P}(x)$ for all $x \in U$. □

The proposition presents an equivalence condition to judge whether two coverings are equal and shows the fact that two coverings are equal if and only if their induced granularities are equal.

Theorem 3.2. Let $\Delta = \{C_1, C_2, \ldots, C_m\}$ be a family of coverings of U. Then C_i is an indispensable covering if and only if there exist $x, y \in U$, such that $y \notin \Delta(x)$ and $y \in \{\Delta - \{C_i\}\}(x)$.

Proof. Straightforward. □

The above theorem implies that an indispensable covering can be characterized by the discernibility between objects. That is to say, C_i is an indispensable covering if and only if there exist $x, y \in U$, such that y not belonging to the neighborhood of x with respect to Δ implies y belonging to the neighborhood of x with respect to $\Delta - \{C_i\}$. This implies that C_i is a sole covering that can distinguish the two objects.