Compacted decision tables based attribute reduction

Wei Wei [a,b], Junhong Wang [a], Jiye Liang [a,*], Xin Mi [a], Chuangyin Dang [b]

[a] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China
[b] Department of System Engineering and Engineering Management, City University of Hong Kong, Hong Kong

ABSTRACT

This paper first points out that the reducts obtained from a simplified decision table are different from those obtained from its original version, and from a simplified decision table, we cannot obtain the reducts in the sense of entropies. To solve these problems, we propose the compacted decision table that can preserve all the information coming from its original version. We theoretically demonstrate that the order preserving of attributes' inner significance and outer significance in the sense of positive region and two types of entropies after a decision table is compacted, which ensures that the reducts obtained from a compacted decision are identical to those obtained from its original version. Finally, several numerical experiments indicate the effectiveness and efficiency of the attribute reduction algorithms for a compacted decision table.

1. Introduction

In many practical applications, the dimensions of data sets (the number of attributes) are becoming higher and higher [1,8,24,31]. For these high-dimensional data, attributes irrelevant to recognition tasks may deteriorate the performance of learning algorithms, and result in the high computing cost [11,33]. Therefore, feature selection has become an important preprocessing step in pattern recognition, data mining and machine learning [9,36].

Among existing feature selection algorithms, supervised feature selection algorithms are commonly employed to process the data with class labels, in which there are some representatives, such as feature selection algorithm with feature selection algorithm based on mRMR [32], sparsity-inducing norms [14], feature selection algorithm based on t-test [44,45], feature subset selection algorithm with ordinal optimization [5] and feature selection algorithm based on neighborhood multi-granulation fusion [25]. For the investigation of feature selection, one of critical issues is how to select feature subset, and filters, wrappers and embedded methods have been generally recognized as the most popular methods to solve the issue [2,8]. In filters methods [16,17], the selection of feature subsets has nothing to do with the chosen learning machine. In wrappers methods [18], the selection of feature

subsets depends on the learning machine that scores subsets of feature according to their predictive power. In embedded methods [8,18], the selection of feature subsets, which is a part of training process, is embedded in learning machines. General speaking, filters and embedded methods are more efficient than wrappers methods, and the wrappers and embedded methods are more effective than filters methods [8].

Attribute reduction is an important research area in rough set theory [4,28–30]. From the perspective of feature selection, attribute reduction is a specific kind of supervised feature selection method which adopts filters method. In recent years, researchers have introduced a lot of attribute reduction algorithms. Skowron and Rauszer [40], based on discernibility matrix, proposed an attribute reduction algorithm, by which all reducts can be obtained. Hu and Cercone [10] introduced discernibility matrix into decision tables. Ye and Chen [57] found out that only the reducts for a consistent decision table can be obtained by the method in [10], and proposed a modified discernibility matrix that is suitable for an inconsistent decision table. Yang [54], through considering the discernibility information in the consistent and inconsistent parts of a decision table respectively, proposed another decision-relative discernibility matrix, by which the time of computing reducts is significantly reduced. Wei et al. [51] proposed two discernibility matrices in the sense of Shannon entropy and complement entropy, which efficiently expands the application range of attribute reduction methods based on discernibility matrix. However, the problem of finding all reducts via using these discernibility matrices has been proved to be NP-hard [52,56].

* Corresponding author. Tel./fax: +86 0351 7018176.
   E-mail addresses: weiwei@sxu.edu.cn (W. Wei), wjhwjh@sxu.edu.cn (J. Wang), ljy@sxu.edu.cn (J. Liang), 707264128@qq.com (X. Mi), mecdang@cityu.edu.hk (C. Dang).

To solve the above problem, researchers introduced heuristic search strategy into the algorithms of finding reducts, which remarkably lessens their computational burden. Hu and Cercone [10] proposed a heuristic attribute reduction algorithm, in which the positive region is utilized to evaluate attribute significance and stop criterion. Slezak [38,39] first introduced an attribute reduction algorithm in the sense of Shannon entropy. Wang et al. [46,47] further improved the kind of algorithms in the sense of shannon entropy. Sequently, Liang et al. [20–22,49], through introducing complement entropy to assess attribute significance and stop criterion, defined a new type of attribute reduction algorithms: the one based on complement entropy. To deal with hybrid data with numerical and categorical attributes, the attribute reduction algorithms based on fuzzy rough set and rough fuzzy set were proposed in [3,12,13,37,41,50]. Additionally, plenty of attribute reduction methods were introduced to process incomplete data [26,27]. Yao and Zhao proposed the attribute reduction methods in decision-theoretic rough sets [55], which can achieve the objective of minimize the cost of decisions [15]. Although these heuristic algorithms have speed up the process of finding reducts, the attribute reduction algorithms are still inefficient to deal with large data.

To further improve heuristic attribute reduction algorithms, Qian et al. [34] proposed a acceleration mechanism, in which the useless objects for finding reducts is progressively deleted in each iteration. The similar idea in [34] was developed to deal with incomplete data sets and hybrid data sets [35,48]. However, in [34,35,48], only the useless objects are gradually deleted from data sets. In fact, the number of attributes also largely affects the efficiency of attribute reduction algorithm. Based on this consideration, Liang et al. [23] developed a more effective attribute reduction algorithm, in which both the useless objects and the irrelevant attributes are progressively removed from data sets in the process of finding reducts.

However, all the objects in one equivalence class are dealt with one by one when running these algorithms mentioned above, though they have the same value on each condition attribute. Thus, it is obvious that the duplicated counting results in the unnecessary time-consuming. To address this issue, some researchers introduced several homomorphisms of an information system, by which a massive information system can be compacted into a relatively small-scale information system and all its reducts are unchanged under the condition of homomorphism [6,7,19,42,43]. Furthermore, to remove the redundancy of a decision table, Xu et al. [53] proposed the simplified decision table, in which all the objects in a condition equivalence class are represented by one of objects in the equivalence class. Thus, the attribute reduction algorithms based on the simplifying decision table become more efficient than the previous ones. But, it is worth noticing that for the objects in one condition equivalence class, their values on the decision attribute are possibly different. In other words, the simplification of a decision table in [53] could make a loss of the values on decision attributes. It is precisely the fault of the simplified decision table that motivates us to seek a new method which cannot only eliminate the repetition of condition attribute values, but also preserve all the information on decision attributes.

Based on the analysis mentioned above, in this paper, we first point out that the reducts obtained from a simplified decision table are different from those obtained from its original version. Then, we propose the compacted decision table, and demonstrate that the sequence preserving of inner significance and outer significance in the sense of positive region after a decision table is compacted. And then, we indicate that from a simplified decision table, the reducts in the senses of Shannon entropy and complement entropy cannot be acquired, and demonstrate that they are able

to obtained from a compacted decision table. Sequently, we design three algorithms based on the proposed compacted decision table to find the reducts in the sense of positive region, Shannon entropy and complement entropy. Finally, several numerical experiments are carried out to verify that our proposed algorithms are more efficient than the existing algorithms.

The remainder of the paper is organized as follows. In Section 2, some preliminaries about the rough set theory and attribute reduction algorithms are reviewed. In Section 3, we point out the fault of the simplified decision table, propose the compacted decision table, demonstrate sequence preserving of inner significance and outer significance in the sense of positive region, and design a new positive region attribute reduction algorithm. In Section 4, based on the proposed compacted decision table, we demonstrate the sequence preserving of inner significance and outer significance in the sense of Shannon entropy and complement entropy, and give the corresponding attribute reduction algorithms. In Section 5, several numerical experiments are carried out to indicate the effectiveness and efficiency of the proposed algorithms. Section 6 concludes the paper with some remarks.

2. Preliminaries

2.1. Rough set

An information system (also known as a data table, an attribute–value system, a knowledge representation system) is a 4-tuple $S = (U, A, V, f)$ (for short $S = (U, A)$), where U is a non-empty and finite set of objects, called a universe, and A is a non-empty and finite set of attributes, V_a is the domain of the attribute a, $V = \bigcup_{a \in A} V_a$ and $f : U \times A = V$ is a function $f(x, a) \in V_a$ for each $a \in A$ [28].

Each attribute subset $B \subseteq A$ derives an indiscernibility relation in the following way: $R_B = \{(x, y) \in U \times U | f(x, a) = f(y, a), \forall a \in B\}$, where $f(x, a)$ and $f(y, a)$ denote the values of objects x and y with respect to the attribute a, respectively. Moreover, the relation R_B partitions U into some equivalence classes given by $U/R_B = \{[x]_B | x \in U\}$, just U/B, where $[x]_B$ is the equivalence class determined by x with respect to B, i.e., $[x]_B = \{y \in U | (x, y) \in R_B\}$. Furthermore, for any $Y \subseteq U$, one defines that $(\overline{B}(Y), \underline{B}(Y))$ is the rough set of Y with respect to B, where the lower approximation $\underline{B}(Y)$ and the upper approximation $\overline{B}(Y)$ of Y [28] are described by

$$\underline{B}(Y) = \{x | [x]_B \subseteq Y\} \quad \text{and} \quad \overline{B}(Y) = \{x | [x]_B \cap Y \neq \emptyset\}.$$

The objects in $\underline{B}(Y)$ can be certainly classified as the members of Y on the basis of knowledge in B, while the objects in $\overline{B}(Y)$ can be only possibly classified as the members of Y on the basis of knowledge in B. The set $BN_B(Y) = \overline{B}(Y) - \underline{B}(Y)$ is called the B−boundary region of Y. A set is called a rough set (or a crisp set) if the boundary region is non-empty (or empty).

A classification problem can be represented by a decision table $DT = (U, C \cup D)$ with $C \cap D = \emptyset$, where an element of C is called a condition attribute, C is called a condition attribute set, an element of D is called a decision attribute, and D is called a decision attribute set. For the convenience of the latter discussion, we define $\partial([x_i]_B) = \{f(x_j, d) | x_j \in [x_i]_B)\}$, where $D = \{d\}$.

For a given decision table $DT = (U, C \cup D), B \subseteq C, U/D = \{Y_1, Y_2, \ldots, Y_n\}$, we define the lower and upper approximations of the decision attribute set D with respect to B as

$$\underline{B}D = \{\underline{B}Y_1, \underline{B}Y_2, \ldots, \underline{B}Y_n\}, \quad and \quad \overline{B}D = \{\overline{B}Y_1, \overline{B}Y_2, \ldots, \overline{B}Y_n\}.$$

Furthermore, the positive region of D with respect to B is defined as $POS_B^U(D) = \bigcup_{i=1}^n \underline{B}Y_i$, the boundary region is defined as $BND_B^U(D) = \overline{B}D - \underline{B}D$, and the negative region $NEG_B^U(D) = U - \overline{B}D$.