ARTICLE IN PRESS

JID: KNOSYS [m5G;December 21, 2016;9:41]

Knowledge-Based Systems 000 (2016) 1–7

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A general reduction algorithm for relation decision systems and its applications

Guilong Liu [a,*], Zheng Hua [a], Zehua Chen [b]

[a] School of Information Science, Beijing Language and Culture University, Beijing, 100083, China
[b] College of Information Engineering, Taiyuan University of Technology, Taiyuan, 030024, China

ARTICLE INFO

Article history:
Received 16 July 2016
Revised 26 November 2016
Accepted 30 November 2016
Available online xxx

Keywords:
Attribute reduction
Covering decision system
Decision table
Discernibility matrix
Relation decision system

ABSTRACT

This paper studies the attribute reduction problem for general relation decision systems. We propose a new discernibility matrix to solve this problem. Combining the discernibility matrix and a recently proposed fast algorithm, we propose a simple and unified attribute reduction algorithm for relation decision systems that is not contingent on the consistency of relation decision systems. We derive the reduction algorithm for the special cases of complete, incomplete, and numerical decision tables. As an application, we transform the attribute reduction of relation decision systems into one for covering decision systems. This gives a convenient and effective reduction algorithm for covering decision systems. The reduction results obtained using University of California Irvine data sets show that the proposed algorithm is simple and efficient. Moreover, the proposed algorithm enables the results of classical attribute reduction approaches to be reinterpreted, giving them far greater unification and generality.

1. Introduction

Pawlak [1] was the first to propose the concept of attribute reduction for decision tables. Attribute reduction plays an important role in many areas, including knowledge representation, decision aiding, data mining, machine learning, and pattern recognition. Decision table attribute reduction can be viewed as one of the most important applications of rough set theory in databases. The reduction process removes superfluous attributes from a decision table while preserving the consistency of its classifications. In order to obtain the minimal subset of attributes that induces the same indiscernibility relation as the whole set of attributes in an information system, Skowron and Rauszer [2,3] first proposed the concept of discernibility matrices, which transforms the discernibility function from its conjunctive normal form (CNF) into the disjunctive normal form (DNF). The minimal reduction subset of attributes can then be obtained.

However, Pawlak rough sets can only deal with complete and symbolic datasets [4]. In order to deal with different types of datasets, many extended rough set models have been proposed. There are now many different types of attribute reductions [5–9] with respect to different criteria. Jia et al. [10] gave a brief description of twenty-two kinds of existing reduction approaches, e.g., positive region reductions [3], distribution reductions [8], variable precision reductions [11,12], covering reductions [13], mutual information reductions [14], and test cost sensitive reduction [15]. It is well-known that discernibility matrix based attribute reduction is one of the most important attribute reduction methods. Finding the discernibility matrix and transforming the discernibility function from its CNF into a DNF are the two key steps of an attribute reduction algorithm. Recently, Borowik and Luba [16] proposed a fast algorithm for transforming the discernibility function from its CNF into a DNF. In addition, the concept of discernibility matrices has been extended by several authors [17,18]. Liu et al. [19] presented an efficient and quick attribute reduction algorithm based on neighborhood rough set models. Min et al. [15] considered the minimal test cost reduction problem and established a test-cost-sensitive reduction algorithm.

Wang et al. [20] first proposed the concept of relation decision systems and provided a systematic study on attribute reductions with rough set approaches [20]. However, in their definition, the decision attribute must be an equivalence relation. This requirement is still a restriction for applications. Recently, we further generalized their definition of relation decision systems so that the decision attribute no longer needs to be an equivalence relation [21]. As we point out in [21](see Theorem 4.1 and its Remark), because there are reductions that may be missed using the proposed algorithm in [21], we did not thoroughly solve the attribute reduction problem for general relation decision systems.

* Corresponding author.
E-mail address: liuguilong@blcu.edu.cn (G. Liu).

http://dx.doi.org/10.1016/j.knosys.2016.11.027

So far, for a relation decision system, there has been not any reduction algorithm that is able to find all reductions. Hence, the aim of the paper is to establish a general reduction algorithm that finds all reductions for relation decision systems. In other words, this paper proposes a general reduction algorithm for relation decision systems that can find all the reductions of a relation decision system. In addition, the algorithm can be applied in complete, incomplete, numerical, and covering decision systems. Complete decision tables, incomplete decision tables, numerical decision tables, and covering decision systems are important data types, and each of them is worth discussing independently. Much work has been done on complete and incomplete decision tables. As a derivative of our algorithm, we obtain a reduction algorithm for decision tables. Dai [4] proposed a fuzzy-tolerance rough set approach for handling this particular type of data, but, as a corollary of our algorithm, we obtain a different approach. Chen et al. first proposed the concept of covering decision systems and an attribute reduction approach for covering decision systems in [13], and their work is an interesting first attempt. Later, Wang et al. [17] continued to explore the reduction problem for covering decision systems and proposed a more efficient algorithm. Wang et al. [22] further improved their own work [13,17]. This paper extends the concept of covering decision systems and shows that the attribute reductions of covering decision systems are equivalent to those of relational decision systems.

The present paper is a continuation of [20,21]. Most of the algorithms for computing the reduction sets need the notion of discernibility matrices, so as a result, this paper proposes an improved discernibility matrix and introduces an attribute reduction algorithm for a general relation decision system. Our algorithm achieves far greater unification and generality than existing ones. In fact, the proposed algorithm unifies earlier attribute reduction algorithms for decision tables, incomplete data sets with numerical attribute values, and covering decision systems.

The remainder of the paper is organized as follows. In Section 2, we review some basic concepts and properties of binary relations and relation decision systems. In Section 3, we propose a new discernibility matrix and establish an attribute reduction algorithm for an arbitrary relation decision system $(U, C \cup D)$. In Section 4, as a special case, we obtain a simple reduction algorithm for complete decision tables. Section 5 gives a reduction algorithm for incomplete decision tables, Section 6 deals with numerical decision tables, and Section 7 is an application of Section 3 to covering decision systems. We show that the attribute reduction of a covering decision system can be transformed into that of a relation decision system. This gives a reduction algorithm for covering decision systems. In Section 8, we present a case study to verify our theoretical results. Finally, Section 9 concludes the paper.

2. Preliminaries

In this section, we recall some basic definitions and properties of binary relations and relational decision systems. Let $U = \{x_1, x_2, \ldots, x_n\}$ be a finite set of objects called the universal set and $P(U)$ be the power set of U. Suppose that R is an arbitrary relation on U. The left and right R-relative sets of an element x in U are defined as [21,23]

$$l_R(x) = \{y | y \in U, yRx\} \text{ and } r_R(x) = \{y | y \in U, xRy\},$$

respectively.

Recall the following terminology: (1) R is reflexive if $x \in r_R(x)$ for each $x \in U$; (2) R is symmetric if $l_R(x) = r_R(x)$ for each $x \in U$; (3) R is transitive if $x \in r_R(y)$ implies that $r_R(x) \subseteq r_R(y)$; and (4) R is an equivalence relation if R is reflexive, symmetric, and transitive.

Definition 2.1. Let U be a universal set and A be a family of arbitrary binary relations on U, then (U, A) is called a relation system [20]. In addition, if $A = C \cup D$, and $C \cap D = \emptyset$, then $(U, C \cup D)$ is called a relation decision system [21], C is called a condition attribute set, and D is called the decision attribute set. If $R_C = \cap_{R \in C} R \subseteq R_D = \cap_{d \in D} d$, then $(U, C \cup D)$ is called consistent; otherwise, $(U, C \cup D)$ is called inconsistent.

One of the most important instances of relation decision systems is the decision table. That is, both C and D consist of equivalence relations on U. Thus, relation decision systems are a generalization of decision tables. Additionally, an incomplete decision table can be viewed as a relation decision system.

For the sake of simplicity, we always assume $D = \{d\}$ in the sequence. In fact, if $D = \{d_1, d_2, \ldots, d_t\}$, then we replace $d_1, d_2, \ldots, d_t$ with $d = \cap_{i=1}^{t} d_i$. From now on, let $(U, C \cup D)$ be a relation decision system, $C = \{R_1, R_2, \ldots, R_m\}$, $R_C = \cap_{i=1}^{m} R_i$, and $D = \{d\}$.

In order to describe the consistent part of a relation decision system $(U, C \cup D)$, we suppose that

$$U_{CD} = \{x | r_{R_C}(x) \subseteq r_d(x)\},$$

and we call U_{CD} the consistent part of $(U, C \cup D)$. Note that if d is an equivalence relation on U, then U_{CD} is just the positive region $Pos_C(D)$ [20]. If $U_{CD} = \emptyset$, then $(U, C \cup D)$ is called totally inconsistent.

3. Attribute reduction algorithm for relation decision systems

This section presents a very simple and subtle reduction algorithm for relation decision systems. The algorithm unifies earlier attribute reduction approaches for the usual decision tables. Let $(U, C \cup D)$ be a relation decision system. Note that if $(U, C \cup D)$ is totally inconsistent, then each singleton set $\{R\} (R \in C)$ is a reduction of C. Hence, from now on, we always assume $U_{CD} \neq \emptyset$.

Generally, an attribute reduction algorithm consists of the following steps

(1) Find a discernibility matrix. This is a key step. Generally, there are different matrices for different types of reductions.
(2) Transform the discernibility function f from its CNF into a DNF. This step is usually time-consuming; fortunately, Borowik and Luba [16] proposed a fast algorithm to complete this transformation.
(3) Obtain all attribute reduction sets.

Definition 3.1. [21] Let $(U, C \cup D)$ be a relation decision system and $\emptyset \neq B \subseteq C$. Set B is called the attribute reduction of C if B satisfies the following conditions

(1) $U_{CD} = U_{BD}$;
(2) For any $\emptyset \neq B' \subset B, U_{CD} \neq U_{B'D}$.

In order to obtain an attribute reduction algorithm, we define the discernibility matrix $M = (c_{ij})_{s \times n}$ for a relation decision system $(U, C \cup D)$ as follows:

$$c_{ij} = \begin{cases} \{R_l | (x_i, x_j) \notin R_l\}, & x_i \in U_{CD}, (x_i, x_j) \notin d \\ \emptyset, & \text{otherwise} \end{cases},$$

where $s = |U_{CD}|$ is the cardinality of U_{CD}. The computational complexity of discernibility matrix $M = (c_{ij})_{s \times n}$ is $O(ns)$ for $s \leq n$. Because inconsistent objects are not compared to each other, this saves some time. Thus, the discernibility matrix is simpler than existing ones. We further need a technical lemma.

Lemma 3.1. *Let $(U, C \cup D)$ be a relation decision system with $D = \{d\}$, if $x_i \in U_{CD}$, $x_j \in U$ and $(x_i, x_j) \notin d$, then $c_{ij} \neq \emptyset$.*

Proof. If $c_{ij} = \emptyset$, then $x_i R_l x_j$ for each $R_l \in C$. This means that $x_j \in r_{R_C}(x_i)$. Because $r_{R_C}(x_i) \subseteq r_d(x_i)$, we have $x_j \in r_d(x)$ and $(x_i, x_j) \in d$. This is a contradiction. □