Pattern Recognition Letters 65 (2015) 81–87

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Attribute reduction approaches for general relation decision systems ☆

Guilong Liu*, Ling Li, Jitao Yang, Yanbin Feng, Kai Zhu

School of Information Science, Beijing Language and Culture University, Beijing 100083, China

ARTICLE INFO

Article history:
Received 28 January 2015
Available online 23 July 2015

Keywords:
Attribute reduction
Decision table
Discernibility matrix
Relation decision system

ABSTRACT

This paper proposes the concept of general relation decision systems and studies attribute reduction algorithms for relation decision systems, which are generalization of decision tables. In our relation decision systems, both condition and decision attribute sets consist of general binary relations. Novel attribute reduction algorithms for consistent and inconsistent relation decision systems are derived, respectively. A data set from the UCI machine learning databases is used in the empirical study, the experimental results verify the effectiveness of the proposed algorithms. The results unify the earlier attribute reduction algorithms for decision tables.

1. Introduction

Attribute reduction, originally proposed by Pawlak [12,13], is a powerful data processing technique. It plays an important role in machine learning, artificial intelligence, pattern recognition [10,27] and other fields. Reduction process removes superfluous attributes from information systems while preserving the consistency of classifications. Many works [1,5,6,9,11,13,16,25,33] have been done on attribute reduction in information systems and many attribute reduction algorithms have been developed. To speed up attribute reduction calculations, Qian et al. [15] proposed several parallel attribute reduction algorithms for large data using MapReduce. In addition, as we know, feature selection has become the focus of machine learning, since rough set theory is a tool to discover data dependencies and to reduce the number of attributes contained in a dataset using the data alone, requiring no additional information [13], it has been used as a feature selection approach with much success. That is, rough set theory provides an important feature selection approach in machine learning.

In the past, attribute reduction algorithms were proposed on approximate spaces induced by equivalence relations and complete information systems. Kryszkiewicz [7] investigated and compared five notions of attribute reduction in inconsistent systems based on equivalence relations. In recent years, more attentions have been paid to incomplete information systems [2,4,19,22–24,26,28–31,34]. Note that almost all attribute reduction algorithms need to calculate discernibility matrices [14,20], in fact, they play an important role in reduction processes. Skowron and Rauszer [20,21] first proposed to represent knowledge in the form of discernibility matrices. Their representation enables simple computation of reduction in information systems. Now the concept of discernibility matrix is extended by several authors [30,35], many types of reduction algorithms for inconsistent information systems have been proposed [7,17,18,32] based on discernibility matrices. For example, Wang et al. [30] provided a systematic study on attribute reduction with generalized rough sets [8,12]. Moreover, they gave many effective reduction algorithms. However, in their definition of relation decision systems, decision attribute set consists of equivalence relations. Clearly, the definition is restrictive since in practice, the values of decision attributes may be missing and breaking the equivalence relation.

In this paper, we propose different definitions of relation decision systems. In our definition, we do not require a decision attribute set consisting of equivalence relations. Moreover, we present novel attribute reduction algorithms for consistent and inconsistent relation decision systems, respectively. As corollaries of our algorithms, we obtain attribute reduction algorithms for decision tables. Our algorithms unify earlier reduction algorithms for decision tables.

As we know, the main obstacle in calculation reducts is the computational complexity. The calculation processes of transforming the conjunctive normal form(CNF) into the disjunctive normal form(DNF) and finding a minimum set implicants are usually time-consuming. Fortunately, Borowik and Luba [3] proposed an exact algorithm for this problem which is based on the unate complementation task. Attribute reduction calculation can be speeded up by application of our method in combination with the new exact algorithm.

The remainder of the paper is organized as follows. In Section 2, we review some basic notions of relation decision systems. In Section 3, we discuss the attribute reduction problem for consistent relation decision systems, and a simple and efficient reduction

☆ This paper has been recommended for acceptance by Egon L. van den Broek.

* Corresponding author. Tel.: +86 10 82303656.
E-mail address: liuguilong@blcu.edu.cn, liuguilong@hotmail.com (G. Liu).

http://dx.doi.org/10.1016/j.patrec.2015.06.031
0167-8655/

Table 1
An incomplete decision table.

U	a_1	a_2	a_3	d
u_1	3	1	3	0
u_2	{1,2}	2	1	$*$
u_3	1	$*$	2	1
u_4	2	1	2	0

algorithm is given. Section 4 presents an attribute reduction for inconsistent relation decision systems. As a special case of Section 4, Section 5 proposes a very simple algorithm to compute the reduction for inconsistent decision tables. In order to verify the effectiveness of our proposed algorithms, Section 6 designs an experimental process to verify our theoretical results. Finally, Section 7 concludes the paper.

2. Preliminaries

In this section, we recall some basic definitions and properties of binary relations and relation decision systems. Let $U = \{u_1, u_2, \dots, u_n\}$ 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, then the pair, (U, R), is referred to as a generalized approximation space. Recall that the left and right R-relative sets of an element x in U are defined as

$$l_R(u) = \{v | v \in U, vRu\} \text{ and } r_R(u) = \{v | v \in U, uRv\},$$

respectively. With a binary relation R on U, for each $X \subseteq U$, $\underline{R}(X) = \{x | r_R(x) \subseteq X\}$ and $\overline{R}(X) = \{x | r_R(x) \cap X \neq \emptyset\}$ are called the lower and upper approximations of X, respectively.

As we know, a knowledge base [13] is a relation system (U, R), where U is universal set, and R is a family of equivalence relations on U. Now we introduce the concept of general relation decision systems.

Definition 2.1. Let U be a universal set. Suppose that $C = \{R_1, R_2, \dots, R_m\}$ and $D = \{d_1, d_2, \dots, d_t\}$ are two families of arbitrary binary relations on U. Then $(U, C \cup D)$ is called a relation decision system, C is called a condition attribute set, and D is called the decision attribute set. If $R_C = \cap_{i=1}^m R_i \subseteq R_D = \cap_{i=1}^t d_i$, then $(U, C \cup D)$ is called consistent; otherwise, $(U, C \cup D)$ is called inconsistent, where the sign $\cap$ is the intersection operation of binary relations.

For a relation decision system $(U, C \cup D)$, set $Pos_C(D) = \{x | x \in U, r_{R_C}(x) \subseteq r_{R_D}(x)\}$ is called the C-positive region of D.

Note that Wang et al. defined the concept of relation decision systems in [30]. However, their definition is different from Definition 2.1. Clearly, $R_C = \cap_{i=1}^m R_i \subseteq R_D = \cap_{i=1}^t d_i$ is equivalent to $r_{R_C}(u) \subseteq r_{R_D}(u)$ or $l_{R_C}(u) \subseteq l_{R_D}(u)$ for all $u \in U$.

Relation decision systems are a generalization of decision tables, since, if both C and D are families of equivalence relations, then a relation decision system $(U, C \cup D)$ is just a usual decision table. In this case, a consistent relation decision system is just a consistent decision table.

Example 2.1. The following incomplete decision table induces a relation decision system.

In Table 1, $*$ denotes missing attribute value, {1, 2} means the attribute value equals to 1 or 2. Attribute R is considered as a binary relation via:

$$u_i R u_j \Leftrightarrow R(u_i) = R(u_j) \text{ or } R(u_i) = * \text{ or } R(u_j) = *.$$

Let $U = \{u_1, u_2, u_3, u_4\}$, $C = \{a_1, a_2, a_3\}$ and $D = \{d\}$, then $(U, C \cup D)$ is a relation decision system. Moreover, it is easy to verify that $(U, C \cup D)$ is consistent.

Let $(U, C \cup D)$ be a relation decision system. Similar to the dependency of knowledge [13], formally, the dependency can be defined as follows.

Definition 2.2. Let $(U, C \cup D)$ be a relation decision system.

(1) D depends on C, denoted by $C \Rightarrow D$, if and only if $\cap_{i=1}^m R_i \subseteq \cap_{i=1}^t d_i$.
(2) C and D are equivalent, denoted by $C \equiv D$, if and only if $C \Rightarrow D$ and $D \Rightarrow C$.
(3) C and D are independent if and only if neither $C \Rightarrow D$ nor $D \Rightarrow C$ holds.

A relation decision system has the following properties.

Proposition 2.1. *Suppose that $(U, C \cup D)$ is a relation decision system.*

(1) $(U, C \cup D)$ is consistent if and only if $Pos_C(D) = U$.
(2) Let $V = Pos_C(D)$, then $(V, C|_V \cup D|_V)$ is consistent, where $C|_V$ is the restriction of C to V.
(3) If $B \subseteq C$, then $R_C \subseteq R_B$ and $Pos_B(D) \subseteq Pos_C(D)$.

Proof.

(1) $(U, C \cup D)$ is consistent if and only if $R_C \subseteq R_D$, if and only if $r_{R_C}(u) \subseteq r_{R_D}(u)$ for each $u \in U$, if and only if $Pos_C(D) = U$.
(2) $(V, C|_V \cup D|_V)$ is consistent because of $\{u | u \in V, r_{R_C}(u) \subseteq r_{R_D}(u)\} = V$.
(3) Since $R_C = \cap_{R \in C} R \subseteq \cap_{R \in B} R = R_B$, we have $R_C \subseteq R_B$. Suppose $u \in Pos_B(D)$, then $r_{R_B}(u) \subseteq r_{R_D}(u)$. $r_{R_C}(u) \subseteq r_{R_B}(u)$ because of $R_C \subseteq R_B$, thus $Pos_B(D) \subseteq Pos_C(D)$. □

For the sake of simplicity, we always assume $D = \{d\}$ in the sequel and $R_D = R_d$.

3. Reductions of consistent relation decision systems

In this section, we will give a reduction algorithm for consistent relation decision systems. As a special case, a simple reduction algorithm for consistent decision tables is obtained. We now introduce the definition of reductions.

Definition 3.1. Let $(U, C \cup D)$ be a consistent relation decision system and $\emptyset \neq B \subseteq C$. If $(U, B \cup D)$ is still consistent and for any subset, $B' \subset B$, $(U, B' \cup D)$ is not consistent, we say that B is a reduction of attribute set C.

For a consistent relation decision system $(U, C \cup D)$, the reduction B is the minimal subset of attributes C that keeps $(U, B \cup D)$ consistent. If C and D are families of equivalence relations, then a relation decision system, $(U, C \cup D)$, is the usual decision table [13]. For decision tables, many reduction algorithms have been given. Now we give a reduction algorithm for a general relation decision system, $(U, C \cup D)$.

Suppose that $(U, C \cup D)$ is a consistent relation decision system, $U = \{u_1, u_2, \dots, u_n\}$, $C = \{R_1, R_2, \dots, R_m\}$, and $D = \{d\}$. Consider the discernibility matrix, $(D_{ij})_{n \times n}$, and

$$D_{ij} = \begin{cases} C, & (u_i, u_j) \in R_D \\ \{R_l | (u_i, u_j) \notin R_l\}, & (u_i, u_j) \notin R_D \end{cases}.$$

We need a technical lemma.

Lemma 3.1. *Let $U = \{u_1, u_2, \dots, u_n\}$ and $(U, C \cup D)$ be a consistent relation decision system. Then $D_{ij} \neq \emptyset$ for $1 \le i, j \le n$.*

Proof. Suppose that $D_{ij} = \emptyset$ for $u_i, u_j \in U$. By the definition of D_{ij}, if $(u_i, u_j) \notin R_D$, then $(u_i, u_j) \in R_C$, this contradicts to $R_C \subseteq R_D$. □

Theorem 3.1. *Let $(U, C \cup D)$ be a consistent relation decision system, $U = \{u_1, u_2, \dots, u_n\}$, $B \subseteq C$, and $B \neq \emptyset$. Then $(U, B \cup D)$ is consistent if and only if $B \cap D_{ij} \neq \emptyset$.*

Proof. Suppose that $(U, B \cup D)$ is consistent. If $(u_i, u_j) \in R_D$, then $B \cap D_{ij} = B \cap C = B \neq \emptyset$. If $(u_i, u_j) \notin R_D$, then $u_j \notin r_{R_D}(u_i)$. Since $(U, B \cup D)$ is