Knowledge-Based Systems 000 (2017) 1–7

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Partial attribute reduction approaches to relation systems and their applications

Guilong Liu*, Zheng Hua

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

ARTICLE INFO

Article history:
Received 25 January 2017
Revised 10 October 2017
Accepted 12 October 2017
Available online xxx

Keywords:
Discernibility matrix
Lower approximation reduction
Upper approximation reduction
Relational decision system
Relation system
X-lower approximation reduction

ABSTRACT

Attribute reduction has long been an active subject of research in rough set theory, and constitutes an important step in data analysis. A relation system is an extension of a typical information system. This paper proposes the concepts of *X*-lower and -upper approximation reductions, and develops corresponding reduction algorithms for general relation systems. By using these types of reduction, we derive lower and upper approximation reductions for relation decision systems. As a special case, we obtain a reduction algorithm for the positive region for decision tables. Finally, we provide two examples from the University of California–Irvine (UCI) datasets to verify our theoretical results.

1. Introduction

Attribute reduction is the process whereby dispensable attributes are removed from a given database of knowledge while maintaining consistency. It is among the most important topics in rough set theory [1,2]. Pawlak was the first to propose the concept of attribute reduction for decision tables. Pawlak and Skowron [3,4], Skowron and Rauszer [5] proposed an algorithm for attribution reduction based on a discernibility matrix with equivalence relations. Skowron and Rauszer [6] were the first to propose the concept of the discernibility matrix. As this matrix is intuitive and easy to understand, attribute reduction based on it is efficient. Many authors [7–15] have proposed similar algorithms with an extended discernibility matrix for different types of attribute reduction. For example, Zhang et al. [16] studied distributive reduction with an appropriate discernibility matrix. Liu et al. [17] developed a unified algorithm based on invariant matrices for three types of reduction in decision tables.

An equivalence relation is too restrictive for many applications; therefore, several authors [18–20] have recently studied certain types of reduction using dominance relation-based attribute reduction. For a given decision table, Wei et al. [21] derived a compacted decision table that can preserve all information contained in the original. Yamany et al. [22] proposed an intelligent optimization method called the “flower search algorithm,” which adaptively searches for optimal attributes, for the fitness function used in rough sets-based classification. Liu et al. [23] proposed a general attribute reduction algorithm for relation decision systems. Zhang et al. [24], Zhang and Xu [25,26] extended the concepts of the lower and upper approximation reduction to ordered information systems. In 2012, Xu et al. [27] investigated upper approximation reduction in an ordered information system with fuzzy decision making. Shao et al. [28] investigated granular reduction in formal fuzzy contexts. Sun et al. [29] recently studied multi-criterion group decision making based on multi-granulation fuzzy rough sets over two universes. Ju et al. [30] considered the design of cost-sensitive rough set models using a multi-granulation strategy.

A number of researchers [31–36] have developed approximation reduction methods based on general binary relations. This paper considers this type of reduction in general relation systems. We first define the concept of *X*-lower approximation reduction and its dual reduction for relation systems. An *X*-lower approximation reduction is a partial reduction. As an application of such a reduction, we derive lower and upper approximation reduction algorithms. As a special case, we establish the relationship between positive region reduction and lower approximation reduction in a decision table.

The remainder of the paper is organized as follows: In Section 2, we recall some basic notions and notations related to relations and relation systems. In Section 3, we propose the concept of, and provide an algorithm for, *X*-lower approximation reduction. As the dual of *X*-lower approximation reduction, Section 4 consid-

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

https://doi.org/10.1016/j.knosys.2017.10.014

ers X-upper approximation reduction. Sections 5 and 6 detail lower and upper approximation reduction for relation decision systems, respectively. Section 7 shows that the positive region reduction for decision tables is a special case of lower approximation reduction and Section 8 provides two examples to verify our theoretical results. Section 9 contains the conclusions of this study.

2. Preliminaries

In this section, we rehearse some basic definitions and properties of binary relations and relation decision systems. Let $U = \{x_1, x_2, \dots, x_n\}$ be a finite set of objects called the universal set. Suppose that R is an arbitrary relation on U. Recall that the left and right R-relative sets of an element x in U are defined as

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

respectively. Based on the right R-relative set, the lower and upper approximations of $X \subseteq U$ are defined as [15]

$$\underline{R}(X) = \{x|x \in U, r_R(x) \subseteq X\} \text{ and } \overline{R}(X) = \{x|x \in U, r_R(x) \cap X \neq \emptyset\},$$

respectively.

Wang et al. [35] proposed the concept of relation decision systems, and we generalize this concept such that the decision attribute no longer needs to be an equivalence relation [37].

Definition 2.1 [35,37]**.** Let $U = \{x_1, x_2, \dots, x_n\}$ be a finite universal set and A be a family of binary relations on U. Then, (U, A) is called a relation system. In addition, if $A = C \cup D$ and $C \cap D = \emptyset$, $(U, C \cup D)$ is called a relation decision system, C is called the condition attribute set and D the decision attribute set. If $R_C = \cap_{R \in C} R \subseteq R_D = \cap_{d \in D} d$, $(U, C \cup D)$ is said to be consistent; otherwise, $(U, C \cup D)$ is said to be inconsistent.

The following proposition is elementary, and hence we omit a proof:

Proposition 2.1. *Let (U, A) be a relation system and $X, Y \subseteq U$.*

(1) *If $B \subseteq A$, then $R_A \subseteq R_B$ and $\underline{R_B}(X) \subseteq \underline{R_A}(X)$.*
(2) *If $X \subseteq Y$, then $\underline{R_A}(X) \subseteq \underline{R_A}(Y)$.*
(3) $\underline{R_A}(X) \cup \underline{R_A}(Y) \subseteq \underline{R_A}(X \cup Y)$.

3. The *X*-lower approximation reduction for relation systems

Let (U, A) be a relation system. For any given subset $X \subseteq U$, we consider a reduction type that keeps the lower approximation $\underline{R_A}(X)$ unchanged. We now provide its definition.

Definition 3.1. Let (U, A) be a relation system. For a given subset, $X \subseteq U$ and $\emptyset \neq B \subseteq A$, B is called the X-lower approximation reduction of (U, A) if B satisfies the following conditions:

(1) $\underline{R_A}(X) = \underline{R_B}(X)$.
(2) If $B' \subset B$, $\underline{R_A}(X) \neq \underline{R_{B'}}(X)$.

Since $\underline{R_A}(U) = U$, each singleton set $\{R_i\}$ $(R_i \in A)$ is a U-lower approximation reduction of (U, A). Thus, we assume that $X \neq U$ throughout this paper.

Let (U, A) be a relation system, where $U = \{x_1, x_2, \dots, x_n\}$ and $A = \{R_1, R_2, \dots, R_m\}$. For subset $X \subseteq U$, we define the discernibility matrix $M = (m_{ij})_{s \times t}$ as follows:

$$m_{ij} = \begin{cases} \{R_l | (x_i, x_j) \notin R_l\}, & \text{if } x_i \in \underline{R_A}(X) \text{ and } x_j \notin X \\ \emptyset, & otherwise \end{cases} .$$

where $s = |\underline{R_A}(X)|$ and $t = |X^C|$ are the cardinalities of sets $\underline{R_A}(X)$ and $X^C = U - X$, respectively. The computational complexity of the discernibility matrix $M = (m_{ij})_{s \times t}$ is $O(st)$.

Now we provide the X-lower approximation reduction algorithm. We use the following lemma:

Lemma 3.1. *Let (U, A) be a relation system and $X \subseteq U$. If $x_i \in \underline{R_A}(X)$ and $x_j \notin X$, $m_{ij} \neq \emptyset$.*

Proof. Suppose that $x_i \in \underline{R_A}(X)$, $x_j \notin X$, if $m_{ij} = \emptyset$; then, $(x_i, x_j) \in R_i$ for each $R_i \in A$. Thus, $(x_i, x_j) \in R_A$ and $x_j \in r_A(x_i)$. Since $x_i \in \underline{R_A}(X)$, $x_j \in r_A(x_i) \subseteq X$. This is contradictory with the assumption. □

The X-lower approximation reduction algorithm is based on the following theorem:

Theorem 3.1. *Let (U, A) be a relation system, and $X \subseteq U$. Then, the following conditions are equivalent:*

(1) $\underline{R_A}(X) = \underline{R_B}(X)$.
(2) *If $m_{ij} \neq \emptyset$, $B \cap m_{ij} \neq \emptyset$.*

Proof. (1)⇒(2): if $m_{ij} \neq \emptyset$ and $m_{ij} \cap B = \emptyset$. By the definition of m_{ij}, we assume that $x_i \in \underline{R_A}(X)$, $x_j \notin X$. $m_{ij} \cap B = \emptyset$ implies $x_i R_B x_j$; using condition (1), $x_i \in \underline{R_A}(X) = \underline{R_B}(X)$. Thus, $x_j \in r_B(x_i) \subseteq X$. This is a contradiction.

(2)⇒(1): For each subset $X \subseteq U$, $B \subseteq A$ implies $R_A \subseteq R_B$ and $\underline{R_B}(X) \subseteq \underline{R_A}(X)$. We need to show that $\underline{R_A}(X) \subseteq \underline{R_B}(X)$. Let $x_i \in \underline{R_A}(X)$, such that $r_A(x_i) \subseteq X$. We now show that $r_B(x_i) \subseteq X$.

If $x_j \notin X$, then, by the definition of m_{ij} and Lemma 3.1, $m_{ij} \neq \emptyset$ and, by condition (2), $B \cap m_{ij} \neq \emptyset$. That is, $(x_i, x_j) \notin R_l$ for some $R_l \in B$. This means that $x_j \notin r_B(x_i)$. This proves $r_B(x_i) \subseteq X$ and $\underline{R_A}(X) \subseteq \underline{R_B}(X)$. This completes the proof of the theorem. □

Corollary 3.1. *Let (U, A) be a relation system with $\emptyset \neq B \subseteq A$. For a given subset $X \subseteq U$ and $\underline{R_A}(X) \neq \emptyset$, B is an X-lower approximation reduction of (U, A) if and only if B is a minimal subset of A satisfying $B \cap m_{ij} \neq \emptyset$, for any $m_{ij} \neq \emptyset$.*

Proof. If B is an X-lower approximation reduction of (U, A), then by Definition 3.1(1), $\underline{R_A}(X) = \underline{R_B}(X)$. By using Theorem 3.1, B satisfies $B \cap m_{ij} \neq \emptyset$ for any $m_{ij} \neq \emptyset$. Using Definition 3.1(2), B is a minimal subset of A satisfying $B \cap m_{ij} \neq \emptyset$ for any $m_{ij} \neq \emptyset$.

Conversely, if B is a minimal subset of A satisfying $B \cap m_{ij} \neq \emptyset$ for any $m_{ij} \neq \emptyset$, by Theorem 3.1, $\underline{R_A}(X) = \underline{R_B}(X)$. Moreover, the minimal property of B implies Definition 3.1(2). □

According to Corollary 3.1, for any given subset $X \subseteq U$, we can give an X-lower approximation reduction algorithm for relation system (U, A) as follows:

(1) Find the discernibility matrix $M = (m_{ij})_{s \times t}$, where $s = |\underline{R_A}(X)|$ and $t = |X^C|$.
(2) Transform the discernibility function f from its CNF $f = \wedge_{m_{ij} \neq \emptyset} (\vee m_{ij})$ into a DNF [38] $f = \vee_{u=1}^{v} (\wedge B_u)$, $(B_u \subseteq A)$.
(3) $Red(A) = \{B_1, B_2, \dots, B_v\}$ and $Core(A) = \cap_{u=1}^{v} B_u$.

End the algorithm.

The example below illustrates our algorithm.

Example 3.1. Let $U = \{1, 2, 3, 4\}$ and $A = \{R_1, R_2, R_3, R_4\}$. Relations R_1, R_2, R_3, and R_4 on U are respectively defined by the following relational matrices:

$$M_{R_1} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}, \quad M_{R_2} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}, \quad M_{R_3} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

and

$$M_{R_4} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

By direct computation,

$$M_{R_C} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

If $X = \{2, 4\}$, $\underline{R_A}(X) = \{1, 2, 3\}$. The

Please cite this article as: G. Liu, Z. Hua, Partial attribute reduction approaches to relation systems and their applications, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.10.014