Chapter 63
Attribute Reduction Based on Equivalence Classes with Multiple Decision Values in Rough Set

Dongwen Zhang, Jiqing Qiu and Xiao Li

Abstract For the attribute reduction problem of decision information systems, the concept of the equivalence class only including the condition attributes is introduced. The necessary condition of implementing attribute reduction and the attribute reduction method based on the equivalence classes with the multiple decision values are presented. After sorting the condition attributes by the cardinalities of the equivalence classes with the multiple decision value in ascending order, these ordered condition attributes are united one by one until the positive region of the united attribute subset is equal to the full region. Furthermore, if the attribute subset is independent and its indiscernibility relation is the same as the indiscernibility relation in original information system, then the subset is an attribute reduction of the information system. Finally, the experiment result demonstrates that our method is efficient.

Keywords Attribute reduction · Rough set · Equivalence class · Multiple decision values

63.1 Introduction

We have witnessed a very rapid growth of rough set theory in recent years; rough set theory has been successfully applied in such fields as knowledge discovery, decision analysis, pattern classification, fault diagnosis, etc., [1]. Attribute reduction is one of

D. Zhang (✉)
School of information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, China
e-mail: zdwwtx@hebust.edu.cn

J. Qiu · X. Li
School of Sciences, Hebei University of Science and Technology, Shijiazhuang 050018, China

the key issues in rough set theory. Reference [2] proposed a discernibility matrix method, in which any two objects determine one feature subset that can distinguish them and could obtain all attribute reducts of a given data set. According to the discernibility matrix viewpoint, Ref. [3] provided a technique of attribute reduction for interval ordered information systems, set-valued ordered information systems, and incomplete ordered information systems, respectively. The idea of attribute reduction using positive region, which remains the positive region of target decision unchanged originated by Refs. [4, 5], gave an extension of this positive region reduction for hybrid attribute reduction in the framework of fuzzy rough set. In heuristic search strategies among attribute reduction methods, some attribute significance measures such as dependency function, information gain, consistency, and other measures are employed to select a feature subset. In fact, several authors [6] have applied variants entropies, combination entropy, or mutual information to measure uncertainty of an information system and construct heuristic algorithm of attribute reduction. The information entropy [7] can measure both the uncertainty of an information system and the fuzziness of a rough decision in rough set theory.

Each of these methods is a good tool to handle data redundancy problem in real life and preserves the particular properties. However, the spatial complexity in computing discernibility matrices, the time in computing information entropy are very expensive, and the theoretical proof in entropy is extremely difficult and using intelligent optimization algorithms to solve reduction problem have to answer the question of how to select the suitable model to match the reduction problem.

In this paper, we present a novel attribute reduction algorithm based on equivalence classes with multiple decision values in rough set. The proposed algorithm fits for the decision table with discrete-valued or continuous-valued attributes. We start from equivalence classes which belong to the partition formed by single condition attribute with respect to decision attributes, union the ordered condition attribute satisfied the special conditions in every iterative round and finally acquire an attribute subset which can hold the same descriptive or classification ability as the entire set of attributes.

63.2 Preliminaries

Rough set theory deals with information represented by a table called decision system, which consists of objects and attributes. This section reviews some essential definitions of rough set that are used for attribute reduction. Detailed description and formal definitions of the theory can be found in Ref. [1].

Definition 1 Let $S = (U, A_t, V, f)$ be a decision system, also called information system, where $U = \{x_1, x_2, \ldots, x_n\}$ called universe, is a nonempty finite set of objects. $A_t = C \cup D$ is a nonempty finite set of attributes, where $C = \{a_1, a_2, \ldots, a_m\}$ is the set of condition. Attributes describing the objects and D is a set of decision attributes that indicates the classes of objects. $V = \underset{a \in A_t}{\cup} V_a$ is

the union of attribute domains, where V_a is the domain of attribute $a \in A_t$. $f : U \times A_t \rightarrow V$ is an information function where $f(x, a) \in V_a$ for every $x \in U, a \in A_t$.

Definition 2 In a given decision system $S = (U, A_t, V, f)$, an indiscernibility relation with respect to $A \subseteq A_t$ is defined as $IND(A) = \{(x, y) \in U \times U | \forall a \in A[f(x, a) = f(y, a)]\}$.

In rough set, an indiscernibility relation $IND(A)$ is always deemed as the knowledge in S. Obviously, an indiscernibility relation $IND(A)$ is reflexive, symmetric and transitive, and thus is an equivalence relation on U and $IND(A) = \underset{a \in A}{\cap} IND(\{a\})$. The equivalence relation $IND(A)$ induces a partition of U, which is denoted by U/A, where an element of U/A is called an equivalence class or elementary set.

Consider a partition $U/D = \{D_1, D_2, \ldots, D_k\}$ of the universe U with respect to the decision attribute D and another partition $U/A = \{U_1, U_2, \ldots, U_r\}$ defined by a set of condition attributes A. The equivalence classes induced by the partition U/A are the basic blocks to construct the Pawlak rough set approximations.

Definition 3 Given any subsets $A \subseteq C$ and $X \subseteq U$. The A-lower and A-upper approximations set of set X is defined, respectively, as follows:

$$A_-(X) = \cup\{ Y \subseteq U/A | Y \subseteq X\} \tag{63.1}$$

$$A^-(X) = \cup\{ Y \subseteq U/A | Y \cap X \neq \varnothing\} \tag{63.2}$$

$A_-(X)$ is the set of all objects from U which can be classified certainly as elements of X employing the set of attributes A. $A^-(X)$ is the set of objects of U which can be possibly classified as elements of X using the set of attributes A.

Definition 4 The A-lower approximation of X is usually called the positive region, written $POS_A(X)$, that is

$$POS_A(X) = A_-(X) \tag{63.3}$$

$POS_A(X)$ indicates the union of all the equivalence classes in U/A and each element can induce a certain decision.

Definition 5 For a given decision system $S = (U, A_t, V, f)$ and the equivalence relations P, Q defined on U, if all attributes on P are necessary for Q, and then P is independent to Q.

From those related conceptions, one could conclude that attribute reduction based on the rough set theory conceptually requires keeping the relation of object discernibility. Therefore, the reduction process has to implicitly or explicitly utilize the decision information of information system.

63.3 Main Achievement

Attribute reduction is a tool to remove redundant condition attributes on condition that the classification or decision-making capacity in information systems does not change. In this paper, our attribute reduction algorithm focuses on three core issues: whether jointed condition attributes would change the condition attribute compatibility of the original decision table; on which condition jointing attributes would achieve the required reduction condition; choosing which condition attributes to joint. This paper first presents the theorem which solves the first core issue: decision value of some equivalence classes do not change after jointing condition attributes.

Definition 7 For any given subsets $A \subseteq C$ and $X \subseteq U$ if the decision values for all objects in some of equivalence classes induced by in U/A are same, these equivalence classes are called equivalence classes with certain decision value, otherwise, they are called equivalence classes with uncertain decision values.

For simplicity, we assume $D = \{d\}, V_d = \{0,1\}$ in this paper, where d is a decision attribute which describes the decision for each object. A table with multiple decision attributes can be easily transformed into a table with a single decision attribute by considering the Cartesian product of the original decision attributes.

Theorem 1 *Suppose d is the decision attribute. A' is the subset of condition attribute set $C(A' \subseteq C)$? For each of equivalence classes in $POS_{A'}(U)$, the set $A' \cup \{a\}$ will partition it, and the partitioned equivalence classes belong to $POS_{A' \cup \{a\}}(\{d\})$ partly.*

Proof Given a decision system $S = (U, C \cup D, V, f), A \subseteq C, a \in C - A'$, According to the definition in Sect. 2, let

$$POS_{A'}(U) = U/A' = \{E_1, E_2, \ldots, E_q\}, \tag{63.4}$$

$$POS_{\{d\}}(U) = U/\{d\} = \{E'_0, E'_1\}, \tag{63.5}$$

$$POS_{A' \cup \{a\}}(U) = U/(A' \cup \{a\}) = \{E''_1, E''_2, \cdots E''_r\}, \tag{63.6}$$

For any $E_j(j \in \{1, 2, \ldots, n\})$, there must exist an equivalence class $E''_s, s \in \{1, 2, \cdots r\}$ in the partition $POS_{A' \cup \{a\}}(U)$ and a equivalence class $E'_k(k \in \{0, 1\})$, such that $E''_i \subset E_j, E_j \subset E'_k$ furthermore, $E''_i \subset E'_k$ Hence, we have $E''_i \subset POS_{A' \cup \{a\}}(U)$. The theorem holds.

This theorem shows that equivalence classes induced by condition attribute subset could be segmented and after merging the condition attributes the segments have the same decision value as every subset of the equivalence classes do. This theorem provides a theoretical basis for jointing condition attributes.

Theorem 2 will answer the key problem on which condition combining attributes could achieve the required reduction condition.

Theorem 2 *for a given decision system* $S = (U, C \cup D, V, f)$, *assume* $A_i, A_k \in C$,

$E_{i,q_j} \in U/A_i, E_{k,q_j} \in U/A_k, (j = 0, 1), i, k = 1, 2, \ldots m$ *whose decision value is 0 or 1*, E_{i,q_x}, E_{k,q_x} *be the uncertain equivalence class formed by condition attribute* A_i, A_k *with respect to the decision attribute respectively, when the condition* $E_{i,q_x} \cap E_{k,q_x} = \varnothing$ *and* $i \neq k$ *are satisfied, we can achieve the reduction ensuring the compatibility.*

Proof Take the two uncertain equivalence classes E_{i,q_x} and E_{k,q_x} for example, $E_{i,q_x} \cap E_{k,q_x} = \varnothing$ shows that although the decision value of the uncertain equivalence classes in the partition induced by A_i condition attribute is uncertain, the decision value of these classes can be confirmed in the partition induced by A_k condition attribute. In term of theorem 1, the decision value of the equivalence classes in the partition $U/\{a_i, a_k\}$ is confirmed. So we get the reduction ensuring the compatibility. Theorem 2 is proved.

Next we propose attribute reduction algorithm based on uncertain equivalence classes in rough set. The algorithm starts from the uncertain equivalence class with lest cardinality in a single condition attribute and jointing condition attributes one by one by step until the reduction condition in theorem 2 is satisfied.

In the rough set model, attribute reduction algorithms mainly deal with categorical data. Thus, we can recode the symbol attributes with a set of consecutive nature numbers. In this paper, we focus on discussing a decision table with a set of integral numbers and apply an efficient sort algorithm for computing equivalence classes, positive regions [8].

Algorithm 1 Computing U/A algorithm

Input: a decision information system S, $U = \{x_1, x_2, \ldots, x_n\}$, $A = \{a_1, a_2, \ldots a_h\}$ $\subseteq C$
Output: U/A
Begin

Step 1: To every $a_i (i = 1, 2, \ldots h)$ denote, the maximum value on attribute a_i by M_i initialize every element in array Order with 1, 2… n, respectively. and all the elements in array TempOrder with 0.

Step 2: Step 2: For $i = 1$ to h do

  For $k = 1$ to M_i do
  {Count[k] = 0;}
  For $j = 1$ to n do
  {Count $[a_i(x_j)]$ ++;TempOrder[j] = Order[j];}//count $[a_i(x_j)]$ now contains the number of elements equal to $a_i(x_j)$
  For $k = 2$ to M_i do
  {Count[k] = count $[k-1]$ + count[k]};
  //count[k]now contains the number of elements less than or equal to k
  For $k = n$ to 1 do

{Order [count [$a_i(x_j)$]] = TempOrder[j]; count [$a_i(x_j)$] −;}

Step 3: let the object sequence from step 2 be { $x'_1, x'_2, \ldots, x'_n$ } ; $t = 1$, $B[t] = x'_1$
For $j = 2$ to n do
If $a_i(x'_j) = a_i(x'_{j-1})$ for all $a_i \in A(i = 1, 2, \cdots h)$, then

$$B[t] = B[t] \cup \{x'_j\};$$

Else {t = t+1; $B[t] = x'_j$;
End

Algorithm 2 Attribute reduction algorithm.
Input: Give a decision information system $S = (U, C \cup D, V, f), C = \{a_1,$
$a_2, \ldots a_m\}$.
Output: Output the reduction of information system S.
Begin

Step 1: $L = \phi; M = \phi;$//L, M is variables//.
Step 2: by algorithm 1, acquire $E_{i,0}, E_{i,1}, E_{i,qx}, POS_{a_i}(U), i = 1, 2, \ldots m,$
Step 3: Sort condition attributes in ascending order of the cardinality of $E_{i,qx}, i = 1, 2, \cdots m$ and stored sorted attribute's ID in array sort-attar [].
Step 4: M = sort-attar [1];
Step 5:

$$j = 2$$

Step 6: While ($card(E_{M,qx}) \neq 0$) and $j <= length(sort\_attr[])$, $K = card(E_{M,qx})$, $L = M \cup sort\_attr[j]$; if $card(E_{L,qx}) < K$ then $M = L; K = card(E_{L,qx});$ $j = j+1;$
Step 7: if M is independent, turn to step 8, otherwise, For $\forall a_i \in M$, if $IND(M) <> IND(M - \{a_i\})$, then $M = M - \{a_i\}$; $IND(M) = IND(M - \{a_i\})$; turn to step 7.
Step 8: if$IND(M) = IND(C)$, so the set M is just an attribute reduction of the condition attributes set C

Otherwise $P = IND(C) - IND(M)$, for any $\bar{a} = C - M$, if $P/(M \cup \{\bar{a}\}) = \phi$, then turn to step 4, otherwise by algorithm 1, acquire $E_{a',qx}, \forall a' \in M \cup \{\bar{a}\}$, and search for min $(E_{a',qx})$ denoted as $a_{\min}$. Let $M = M \cup a_{\min}$, turn to step 7
End

So the set M is just an attribute reduction of the condition attributes set C and a reduction of information system S is also obtained.

In the algorithm, we only consider the judgment $card(E_{L,qx}) < card(E_{M,qx})$ and do not consider the issue $card(E_{M,qx}) \leq card(E_{L,qx})$ precisely. In fact, when the judgment is $card(E_{M,qx}) = card(E_{L,qx})$, the candidate attributes are not unique and we could obtain multiple solutions.

63.4 Experimental Analysis

In this section, we will demonstrate the performance of our algorithm given in the above section. The data set from Ref. [9] is a turbine fault diagnosis example. In this example, there are 11 continuous attributes and 21 objects. We collect 15 objects, ID denoted by 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 14, 15, 17, 19, 20 as the training set, and the remaining 6 objects, ID denoted by 1, 7, 11, 16, 18, 21 as the test set. We use our algorithm on the training set for attribute reduction and verify its result.

Step 1: Calculate $E_{i,0}, E_{i,1}, E_{i,qx}, POS_{a_i}(U), i = 1, 2, \cdots 11$, then we get Table 63.1, where a_i is the condition attribute, $j = 0, 1$ corresponds to the equivalence class whose decision value is 0 or 1, $j = X$ corresponds to the equivalence classes whose decision value is uncertain, the data in Table 63.1 corresponds to the objects' IDs and the IDs in a {} are in a equivalence class.

Step 2: Sort condition attributes in ascending order of the cardinality of $E_{i,qx}, i = 1, 2, \ldots 11$, acquire, $Sort\_attr[] = [a_9, a_7, a_4, a_2, a_1, a_8, a_{11}, a_6, a_3, a_5, a_{10}]$

Step 3: Let M = Sort-attar [1]

Step 4: The $card(E_{M,qx}) = 2 \neq 0$. So we start the loop. When $L = M \cup a_j = \{a_9, a_7, a_2\}$, then $card(E_{L,qx}) = 0 < card(E_{M,qx}) = 1$ and $N = \{a_2\}$, the loop end the $M = M \cup N = \{a_9, a_7, a_2\}$.and $\cup\{a_i | a \in M, i = 1, 2, \ldots, card(M)\} = a_9 \cup a_7 \cup a_2$. Calculate the partition $U/(a_9 \cup a_7 \cup a_2) = \{\{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{8\}, \{9\}, \{10\} \{12\}, \{13\}, \{14\}, \{15\}, \{17\}, \{19\}, \{20\}\}$; so we achieve a reduction:$a_9 \cup a_7 \cup a_2$

Table 63.1 The result of step 1

a_i (condition attribute)	$j = 0$ (the decision value of equivalence is 0)	$j = 1$ (the decision value of equivalence is 1)	$j = X$ (the decision value of equivalence is uncertain)
$i = 1$	{8},{10},{12},{17}	{2,4}	{3,5,6},{9,13,14,15},{19,20}
$i = 2$	{6,10},{13,17},{14},{19}		{2,3,4,5},{8,9},{12,15,20}
$i = 3$	{13,14,17},{15}	{4}	{2,3,5},{6,8,9,10,12},{19,20}
$i = 4$	{6},{8,10,15},{17},{19}	{4}	{2,3,5},{9,14},{12,13,20}
$i = 5$	{14},{19}	{4}	{2,3,5},{6,8,9,10,12,13,15,17,20}
$i = 6$	{6,10},{12},{15,17}		{2,3,4,5,13,14,19,20},{8,9}
$i = 7$	{5},{10,12,14,15,17,19}	{2},{3,4},{20}	{6,8,9,13}
$i = 8$	{6},{8,10},{15},{19}	{20}	{2,3,4,5},{9,12},{13,14,17}
$i = 9$	{6},{8},{10,14,15},{12,17,19},{13}	{2},{4},{9,20}	{3,5}
$i = 10$		{4}	{2,3,5},{6,8,9},{10,12,13,14,15,17,19,20}
$i = 11$	{6,19},{8},{13,14},{17}		{2,3,4,5},{9,12},{10,15,20}

After obtaining the attribute reduction of decision table, we could test the performance of the reduction $a_9 \cup a_7 \cup a_2$.

Experimental results show that the attribute reduction algorithm is simple and easy to implement. Compared with the algorithm of Ref. [6], it has a smaller number of cuts, a smaller number of rules, and the prediction results reach the required accuracy. The main work of the next step is to improve the accuracy of classification rules and eliminate the unclassified objects.

63.5 Conclusions

This paper offers a novel algorithm of attribute reduction which is suited for the information system with discrete-valued or continuous-valued attributes. The presented algorithm gives attribute reduction in continuous data itself and achieves classification prediction on dataset. In this paper, we prove the effectiveness and efficiency of the algorithm in theory and practice.

References

1. Salamó M, López-Sánchez M (2011) Rough set based approaches to feature selection for case-based reasoning classifiers. Pattern Recogn Lett 32(15):280–292
2. Das AK, Sil J (2011) An efficient classifier design integrating rough set and set oriented database operations. Appl Soft Comput 11(8):2279–2285
3. Skowron A (1995) Extracting laws from decision tables: a rough set approach. Comput Intell 11(27):371–388
4. Shao MW, Zhang WX (2005) Dominance relation and rules in an incomplete ordered information system. Int J Intell Syst 20(14):13–27
5. Grzymala-Busse JW (1991) An algorithm for computing a single covering, vol 62(25). Kluwer Academic Publishers, Dordrecht, pp 347–351
6. Grzymala-Busse JW (1991) LERS–a system for learning from examples based on rough sets, vol 53(15). Kluwer Academic Publishers, Dordrecht, pp 643–647
7. Hu QH, Xie ZX, Yu DR (2007) Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. Pattern Recogn 40(17):3509–3521
8. Qian J, Miao DQ, Zhang ZH (2011) Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation. Int J Approximate Reasoning 52(16):212–230
9. Yang SZ, Ding H, Shi TL (1993) Diagnosis reasoning based on knowledge, vol 14(28). Tsinghua University Press, Beijing, pp 455–459