Research Article

Approximate Accuracy Approaches to Attribute Reduction for Information Systems

Deshan Liu,[1] Dapeng Wang,[1] Deqin Yan,[1] and Yu Sang[2]

[1] *College of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China*
[2] *Institute of Computing Technology, Research Institute of Exploration and Development, Liaohe Oilfield, PetroChina, No. 98 Oil Street, Panjin 124010, China*

Correspondence should be addressed to Deqin Yan; yandeqin@163.com

The key problem for attribute reduction to information systems is how to evaluate the importance of an attribute. The algorithms are challenged by the variety of data forms in information system. Based on rough sets theory we present a new approach to attribute reduction for incomplete information systems and fuzzy valued information systems. In order to evaluate the importance of an attribute effectively, a novel algorithm with rigorous theorem is proposed. Experiments show the effect of proposed algorithm.

1. Introduction

There are various kinds of data in information systems, such as real value attribute, symbol attribute, and fuzzy sets attribute. The variety of data forms in information system gives a challenge to attribute algorithms. How to evaluate the importance of an attribute is a key problem for attribute reduction, and this also is hot point of study for researchers [1–6]. In [7] fuzzy discernibility matrix was studied for depicting the relationship for attributes in information system. Chen et al. [2] argued that the heuristic algorithm in [7] usually cannot find a proper reduct but an overcut-reduct or subreduct due to their stop criteria; based on the minimal elements in discernibility matrix, they proposed an attribute reduction algorithm with fuzzy rough sets. In [8], fuzzy rough sets attribute reduction was approached by a new construction of discernibility matrix. Rough set provides an important tool for data mining and knowledge discovery, and its application results attract more and more attention [1, 9–15]. Attribute reduction for information system by using discernibility matrix is an important application of rough set theorem. The concept and structure of the traditional discernibility matrix are based on attribute and whose attribute values are symbol [3, 10], which are also research hotspots in rough set's theorem and application field. However, it is difficult to apply the traditional discernibility matrix in incomplete information system, the real value attribute information system, and fuzzy information systems. Although extended rough set model and continuous attributes real value segmentation method are proposed [2, 13, 16, 17], there are also many problems related to traditional discernibility matrix. In particular, the capacity of reduction is greatly reduced because the threshold which decides equivalence of discernibility matrix is artificial. In this paper, a new concept of soft discernibility matrix is proposed, and related theorems as well as examples are given. We could make attributes reduction for incomplete information system, fuzzy information systems by soft discernibility matrix. Experiments show the effect of proposed algorithm.

2. Basic Concepts

Let $S = (U, Q, V, F)$ be an information system, where $U = \{x_1, x_2, \ldots, x_n\}$ a universe, Q is attribute set, V is attribute value set, and F is the mapping of $U \times Q \rightarrow V$. For convenient explanation, we assume that attribute sets include m condition attributes $C = \{C_1, C_2, \ldots, C_m\}$ and one decision attribute D (with values $D_1, D_2, \ldots, D_s$). Let $\{Y_1, Y_2, \ldots, Y_s\}$ be a partition equivalence class derived by decision attribute D in which

$$Y_i = \{x \in U \mid F(x, D) = D_i\}, \quad i = 1, 2, \ldots, s. \quad (1)$$

Each x_i and its attribute value are called information rule in information system; in this paper, we call it x_i rule for short. The attribute value of an attribute a ($a \in C$ or $a \in D$) under rule of x_i is denoted by $F(x_i, a)$.

If each attribute value in information systems is known, the system is called complete information system. In some cases, the information system is called incomplete information system in which some attribute values are unknown. Generally, attribute value set V is a symbol set in information system $S = (U, Q, V, F)$. The system is called fuzzy attribute information system when attribute set V is fuzzy set. The system is real value information system or continuous attribute information systems when attribute value set V is real value.

Based on rough sets [9], we give the following Definitions 1–4.

Definition 1. Let $X \subseteq U$ be subset of universe and $P \subseteq C$, the lower approximation value of X about P is

$$P\_X = \{x \in U \mid [x]_P \subseteq X\}, \tag{2}$$

and $[x]_P$ is the set which is made up of equivalence class elements in U and equivalence relation P.

Definition 2. Let U be a universe, P and Q are equivalence relation clusters in U, and the P positive universe of Q is defined as $POS_P(Q) = \cup_{X \in U/Q} P\_(X)$.

Definition 3. For $P \subseteq C$, approximation accuracy of partition $\{Y_1, Y_2, \ldots, Y_k\}$ is

$$\gamma_P = \sum_{i=1}^{k} \frac{\text{card}(P\_Y_i)}{\text{card}(U)}, \tag{3}$$

where card( ) is cardinal number of set.

Definition 4. For $P \subseteq C$, if $\gamma_P = \gamma_C$ and there does not exist a R being $R \subset P$ and $\gamma_R \neq \gamma_P$, P is called one attribute reduction of C.

The whole intersection set of attribute reduction of C is called core of C, which is denoted as Core(C).

Definition 5 (see [10]). Given an information system S, the elements of discernibility matrix $M = (m_{ij})$ are

$$m_{ij} = \{a \in C \mid F(x_i, a) \neq F(x_j, a)\},$$
$$\text{if } F(x_i, D) \neq F(x_j, D), \tag{4}$$
$$m_{ij} = \Phi, \quad \text{if } F(x_i, D) = F(x_j, D),$$

where Φ is empty set.

Paper [10] pointed out that if an element in discernibility matrix includes only one attribute, this attribute is core.

3. Soft Discernibility Matrix

3.1. Probability Discernibility Matrix. Suppose that $S = (U, Q, V, F)$ is incomplete information system, V is attribute value set, $*$ represents uncertain attribute value in information system, and v is the total number of types of attribute value in the system. The matrix which is constructed by incomplete information system is called the probability discernibility matrix.

Definition 6. Given incomplete information system S, we denote

$$m_1 = \{a \in C \mid F(x_i, a) \neq F(x_j, a)\}. \tag{5}$$

If one of $F(x_i, a)$ and $F(x_j, a)$ is $*$,

$$m_2 = \left\{\left(1 - \frac{1}{v}\right) a \in C\right\}. \tag{6}$$

If both $F(x_i, a)$ and $F(x_j, a)$ are $*$,

$$m_3 = \left\{\left(1 - \frac{1}{v^2}\right) a \in C\right\}. \tag{7}$$

The elements in $M = (m_{ij})$ probability discernibility matrix are

$$m_{ij} = m_1 \cup m_2 \cup m_3,$$
$$m_{ij} = \Phi, \quad \text{if } F(x_i, D) = F(x_j, D), \tag{8}$$

where Φ is empty set.

In probability differential matrix, m_{ij} is an element set which is composed of the attribute symbols being with coefficient or weight.

Given an attribute $a \in P \subseteq C$, for any two rules x_i and x_j,

if $F(x_i, a)$ and $F(x_j, a)$ are known (namely, being certain value or symbol), we denote $a \in P_1$;

if one of $F(x_i, a)$ and $F(x_j, a)$ is $*$, we denote $a \in P_2$;

if both $F(x_i, a)$ and $F(x_j, a)$ are $*$, we denote $a \in P_3$.

Definition 7. Given incomplete information system S, the similarity of rule x_i and rule x_j in the set of $P \subseteq C$ is defined as follows.

For every attribute, $a \subseteq P_1$ satisfy $F(x_i, a) = F(x_j, a)$ or P_1 is empty set; let $s_0 = 1$; otherwise let

$$s_0 = 0, \qquad s_{ij} = \frac{s_0}{v^{k+2m}}, \tag{9}$$

where $k = \text{card}(P_2)$, $m = \text{card}(P_3)$.

Suppose the subscript set of Y_i which belongs to one partition of U derived from equivalence class of D is δ_i; if $t \in \delta_i$, we denote $\bar{s}_t = \sum_{j \notin \delta_i} s_{tj}$. If $\bar{s}_t > 1$, we let $\bar{s}_t = 1$.

Definition 8. For $P \subseteq C$ and a partition $\{Y_1, Y_2, \ldots, Y_s\}$, the approximate accuracy of P can be defined as follows:

$$\lambda_P = \sum_{i=1}^{s} \frac{\text{card}(P\_Y_i)}{\text{card}(U)}, \tag{10}$$

where $\text{card}(P\_Y_i) = \text{card}(\delta_i) - \sum_{t \in \delta_i} \bar{s}_t$.

Definition 9. For $P \subseteq C$, if $\lambda_P = \lambda_C$ and there does not exist $R \subset P$ satisfying $\lambda_R = \lambda_C$, then P is called one attribute reduction of C. All the intersection of attribute reduction of C is called core of C and is denoted as Core(C).

In probability discernibility matrix, if there is only one attribute in an element, the included attribute is called single attribute. If there are two or more attributes, the included attributes are called multiattribute items.

Theorem 10. *The single attribute whose coefficient in probability discernibility matrix is 1 is core attribute of corresponding incomplete information system.*

Proof. In probability discernibility matrix, single attribute is the attribute that under which two different rules (with different decision attribute value) have different attribute value but have the same attribute value under the other condition attributes. We can conclude from Definitions 6 and 8 that if single attributes are deleted, λ_P must be reduced. And then according to Definition 9, we can conclude that the attribute must be core attribute. So the theorem is correct. □

In information system, the importance of one attribute could be determined by its impact to approximate accuracy of corresponding information systems. The attribute which has larger impact on approximate accuracy is more important than that one which has smaller impact. The attributes which could be reduced are called unimportant attribute.

Theorem 11. *In probability discernibility matrix,*

(1) *if two attributes are single attribute, the attribute whose sum of the coefficients is larger is more important;*

(2) *if an attribute is always being with another attribute whose coefficient is 1 in same element set then the attribute must be unimportant.*

Proof. (1) If two attributes are single attribute, from Definitions 7 and 8, we can see that their impact on λ_P is determined by 6 the sum of coefficient in the single attribute.

The larger the sum of coefficient is, the more impact on λ_P is, so it is more important.

(2) If an attribute is always being with another attribute whose coefficient is 1 in same element set then the attribute is certainly or possibly with different values under different rules. We can conclude form Definition 7 that $s_0 = 0$, so $s_{ij} = 0$. We know that the attribute has no impact on λ_P according to Definitions 7 and 8, so this attribute is unimportant. □

3.2. Fuzzy Discernibility Matrix. For convenience of explanation, we suppose that each attribute has single value for each rule in the fuzzy information system.

For fuzzy attribute information systems, the concept of equivalence class is no longer applicable, so traditional method could not be used to construct discernibility matrix. In fact, equivalence class represents a correlationship. If the correlation is seen as similarity, then the equivalence can be seen as the association with the value of 1; therefore, the equivalence is a special case of similarity. Each rule is viewed as one fuzzy set constructed by each condition attribute value when we are constructing fuzzy discernibility matrix and investigate similarities between fuzzy set.

Definition 12. Given fuzzy information system S, the element in fuzzy discernibility matrix $M = (m_{ij})$ is defined as follows:

$$m_{ij} = \left\{ f_{i,j,1}C_1, f_{i,j,2}C_2, \ldots, f_{i,j,m}C_m \right\}, \tag{11}$$

where $f_{i,j,k} = S_{i,j}^k - S_{i,j}$ $(k = 1, 2, \ldots, m)$. $S_{i,j}$ is fuzzy similarity of x_i and x_j in the case that $F(x_i, D) \neq F(x_j, D)$. $S_{i,j}^k$ is fuzzy similarity in the case that C_k attribute is deleted.

From Definition 12 we can conclude that each element in fuzzy discernibility matrix is condition attribute set, and it corresponds to a fuzzy set $F = \{f_{i,j,1}, f_{i,j,2}, \ldots, f_{i,j,m}\}$.

Here we denote $f_k = \sum_{i,j} f_{i,j,k}$ as the sum of coefficients of C_k.

In paper [16], the authors give a similarity measure formula. Suppose that A, B are two fuzzy set; then similarity measure base on inclusion degree is

$$H(A, B) = \frac{1}{n} \sum \text{Count} \left(\frac{A \cap B}{A \cup B} \right), \tag{12}$$

and similarity measure base on close degree is

$$D(A, B) = \frac{\sum \text{Count}(A \cap B)}{\sum \text{Count}(A \cup B)}. \tag{13}$$

Given a fuzzy attribute information system S, the similarity of rule x_i and x_j is denoted as S_{ij}. For each $Y_i \subset U$ derived from equivalence class of D, let the subscript set of Y_i be δ_i; if $t \in \delta_i$, we denote

$$\bar{s}_t = \sum_{j \notin \delta_i} s_{tj}. \tag{14}$$

Definition 13. For $P \subseteq C$, approximation accuracy of partition $\{Y_1, Y_2, \ldots, Y_k\}$ is

$$\lambda_P = \sum_{i=1}^{k} \frac{\text{card}(P\_Y_i)}{\text{card}(U)}, \tag{15}$$

where $\text{card}(P\_Y_i) = \text{card}(\delta_i) - \sum_{t \in \delta_i} \bar{s}_t$. λ_C is called approximate accuracy of fuzzy information system.

Definition 14. An attribute $a \in C$ satisfying $\lambda_{C-\{a\}} < \lambda_C$ is called core attribute. The core attribute of system denotes Core(C).

As to attribute reduction, the importance of an attribute depends on its impact on approximate accuracy λ_C.

Theorem 15. *An attribute with larger sum of coefficient in all the elements of fuzzy discernibility matrix is more important.*

Proof. We can conclude form Definition 12 that the sum of coefficients in all elements can be described as

$$f_k = \sum_{t \subseteq U} \bar{s}_t^k - \bar{s}_t, \tag{16}$$

where

$$\bar{s}_t = \sum_{j \notin \delta_i} S_{t,j}, \qquad \bar{s}_t^k = \sum_{j \notin \delta_i} S_{t,j}^k. \qquad (17)$$

From Definitions 13 and 14, we can conclude that the size of approximate accuracy depends on the size of f_k. The attribute with larger coefficient has more impact on $\lambda_{C-\{C_k\}}$, so it is more important.   □

Theorem 16. *In fuzzy discernibility matrix, for an attribute, if sum of coefficient in all elements is larger than 0, then the attribute is core attribute.*

Proof. We can conclude the proof of Theorem 10 that λ_C could turns larger if C_k is reduced in the case of $f_k > 0$, and then C_k is core attribute.   □

We can make attribute reduction in fuzzy information system by use of Definitions 12–14 and Theorems 15–16.

4. Attribute Reduction

To illustrate the application of soft discernibility matrix, we give two application examples of attribute reduction in incomplete information and fuzzy information system.

4.1. Reduction for Incomplete Information System. The method utilizing probability discernibility matrix for attribute reduction in incomplete information is as follows.

The first step of attribute reduction on incomplete information system is constructing corresponding discernibility matrix according to Definition 6 and then finding out core attribute and unimportant attribute according to Theorems 10 and 11. The core attributes are retained and the unimportant attributes are deleted. As to other attributes, the most important attributes are joined into reduction set based on their impact on approximate accuracy continuously.

Table 1 is incomplete information system, in which a, b, c, and d are condition attribute and e is decision attribute.

In Table 1, the values of attribute are 0 or 1, $v = 2$. We can obtain the elements of probability discernibility matrix M based on Definition 6:

$$m_{i,i} = \Theta \quad (i = 1, 2, \ldots, 8),$$

$$m_{2,1} = m_{5,4} = m_{8,7} = \Theta,$$

$$m_{3,1} = m_{3,2} = \Theta,$$

$$m_{6,4} = m_{6,5} = \Theta,$$

$$m_{4,1} = \{b\},$$

$$m_{4,2} = \{b, kd\},$$

$$m_{4,3} = \{b, kc\},$$

$$m_{5,1} = \{c, d\},$$

$$m_{5,2} = \{kc, kd\},$$

TABLE 1: Incomplete information system.

U	a	b	c	d	e
X_1	1	1	1	0	1
X_2	1	1	1	*	1
X_3	1	1	*	0	1
X_4	1	0	1	0	0
X_5	1	1	0	1	0
X_6	*	0	*	1	0
X_7	1	0	*	1	2
X_8	0	1	0	1	2

$$m_{5,3} = \{kc\},$$

$$m_{6,1} = \{ka, b, kc, d\},$$

$$m_{6,2} = \{ka, b, kc, kd\},$$

$$m_{6,3} = \{ka, b, k^2 c, d\},$$

$$m_{7,1} = \{b, kc, d\},$$

$$m_{7,2} = \{b, kc, kd\},$$

$$m_{7,3} = \{b, k^2 c, d\},$$

$$m_{7,4} = \{kc, d\},$$

$$m_{7,5} = \{b, kc\},$$

$$m_{7,6} = \{ka, k^2 c\},$$

$$m_{8,1} = \{a, c, d\},$$

$$m_{8,2} = \{a, c, kd\},$$

$$m_{8,3} = \{a, kc, d\},$$

$$m_{8,4} = \{a, b, c, d\},$$

$$m_{8,5} = \{a\},$$

$$m_{8,6} = \{ka, kc\},$$

$$(18)$$

where Θ is empty set, $k = 1/v$.

Because the discernibility matrix is symmetrical, so we only give out lower triangular elements of discernibility matrix M. We can see that a, b, and c are all single attribute according to probability discernibility matrix M, so they are core attributes. The approximate accuracy of incomplete information system condition attribute in table is $\lambda_C = (1/8)(5 + k^3)$. The approximate accuracy λ_C remains unchanged, even though attribute d is deleted. So the incomplete information system in Table 1 can be reduced as a, b, and c.

We can realize attribute reduction for fuzzy information system based on fuzzy discernibility matrix according to definitions and theorems in Section 3.2. Table 2 is a fuzzy attribute information system, in which a, b, and c are condition attributes and d is decided attribute.

TABLE 2: Fuzzy attribute information system.

U	a	b	c	d
X_1	0.1	0.5	0.8	0
X_2	0.2	0.3	0.9	0
X_3	0.8	0.6	0.7	1
X_4	0.6	0.5	0.3	1
X_5	0.4	0.4	0.6	2

4.2. Reduction for Fuzzy Information System. Given a fuzzy information system, at first similarity, S_{ij} between rules x_i and x_j is computed, and then fuzzy discernibility matrix $M = (m_{ij})$ could be determined based on Definition 12. Based on M, we construct a new discernibility matrix M^* according to following step.

Initiate m_{ij}^* the elements of M^* to be null sets.

For each element,

$$m_{ij} = \left\{ f_{i,j,1} C_1, f_{i,j,2} C_2, \ldots, f_{i,j,m} C_m \right\}. \tag{19}$$

Compute

$$\varepsilon = f_{i,j,k} - S_{i,j}, \quad (k = 1, 2, \ldots, m). \tag{20}$$

If $\varepsilon > 0$, the corresponding attribute C_k is joined into element m_{ij}^*.

The new matrix M^* has the traditional form of discernibility matrix. At last, attribute reduction of fuzzy information system could be obtained by discernibility function which is composed of logical computation among elements.

The discernibility matrix (lower triangle) could be computed based on similarity measure of closeness degree. The following elements of discernibility matrix M^* is obtained from the fuzzy attribute system Table 2 (Θ is empty set):

$$\begin{aligned} m_{i,i}^* &= \Theta, \quad (i = 1, 2, \ldots, 5), \\ m_{2,1}^* &= m_{4,3}^* = \Theta, \\ m_{3,1}^* &= m_{3,2}^* = m_{5,1}^* = m_{5,2}^* = \{a\}, \\ m_{4,1}^* &= m_{4,2}^* = \{a, c\}, \\ m_{5,3}^* &= \{a, b\}, \\ m_{5,4}^* &= \{c\}. \end{aligned} \tag{21}$$

We can conclude that attributes a, c are single attributes, so a, c are core attributes.

The discernibility matrix M^* could be reduced as (a, c) based on difference function $f(S) = (a) \wedge (c) \wedge (a \vee c) \wedge (a \vee b)$.

5. Conclusion

In this paper, a new concept of soft discernibility matrix is proposed; related theorems as well as corresponding application are given. Because soft discernibility matrix could be used in the attribute reduction of incomplete information system, fuzzy information system, it could provide important research basis for attribute reduction of real attribute information system; the method proposed in this paper is of wide application significance in practice. The new concept and related method for attribute reduction are a development to the study for information system and data mining.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] W. Xu, Y. Li, and X. Liao, "Approaches to attribute reductions based on rough set and matrix computation in inconsistent ordered information systems," *Knowledge-Based Systems*, vol. 27, pp. 78–91, 2012.

[2] D. Chen, L. Zhang, S. Zhao, Q. Hu, and P. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 385–389, 2012.

[3] W. Wei, J. Liang, J. Wang, and Y. Qian, "Decision-relative discernibility matrices in the sense of entropies," *International Journal of General Systems*, vol. 42, no. 7, pp. 721–738, 2013.

[4] J. W. Grzymala-Busse and M. Hu, "A comparison of several approaches to missing attribute values in data mining," in *Rough Sets and Current Trends in Computing*, pp. 378–385, Springer, Berlin, Germany, 2000.

[5] T.-P. Hong, T.-T. Wang, S.-L. Wang, and B.-C. Chien, "Learning a coverage set of maximally general fuzzy rules by rough sets," *Expert Systems with Applications*, vol. 19, no. 2, pp. 97–103, 2000.

[6] R. Jensen and Q. Shen, "Fuzzy-rough attribute reduction with application to web categorization," *Fuzzy Sets and Systems*, vol. 141, no. 3, pp. 469–485, 2004.

[7] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824–838, 2009.

[8] E. C. C. Tsang, D. Chen, D. S. Yeung, X.-Z. Wang, and J. W. T. Lee, "Attributes reduction using fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1130–1141, 2008.

[9] Z. Pawlak, "Rough set theory and its applications to data analysis," *Cybernetics and Systems*, vol. 29, no. 7, pp. 661–688, 1998.

[10] X.-H. Hu and N. Cercone, "Learning in relational databases: a rough set approach," *Computational Intelligence*, vol. 11, no. 2, pp. 323–338, 1995.

[11] Y. E. Dy and Z. J. Chen, "A new discernibility matrix and the computation of a core," *Acta Electronica Sinica*, vol. 30, no. 7, pp. 1086–1088, 2002.

[12] H. L. Zwng, *Rough Sets Theorem and Applications*, Chongqing University Press, Chongqing, China, 1998.

[13] H. Xie, H.-Z. Cheng, and D.-X. Niu, "Discretization of continuous attributes in rough set theory based on information entropy," *Chinese Journal of Computers*, vol. 28, no. 9, pp. 1570–1574, 2005.

[14] Y. Yao and Y. Zhao, "Attribute reduction in decision-theoretic rough set models," *Information Sciences*, vol. 178, no. 17, pp. 3356–3373, 2008.

[15] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "Positive approximation: an accelerator for attribute reduction in rough set theory," *Artificial Intelligence*, vol. 174, no. 9-10, pp. 597–618, 2010.

[16] C.-T. Su and J.-H. Hsu, "An extended Chi2 algorithm for discret-
     ization of real value attributes," *IEEE Transactions on Knowledge
     and Data Engineering*, vol. 17, no. 3, pp. 437–441, 2005.

[17] X. Liu and H. Wang, "A discretization algorithm based on a
     heterogeneity criterion," *IEEE Transactions on Knowledge and
     Data Engineering*, vol. 17, no. 9, pp. 1166–1173, 2005.