

Research Article

A Variable Precision Attribute Reduction Approach in Multilabel Decision Tables

Hua Li,^{1,2} Deyu Li,¹ Yanhui Zhai,³ Suge Wang,³ and Jing Zhang³

¹ Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, Shanxi 030006, China

² Department of Mathematics and Physics, Shijiazhuang Tiedao University, Shijiazhuang, Hebei 050043, China

³ School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China

Correspondence should be addressed to Deyu Li; lidysxu@163.com

Received 18 June 2014; Accepted 17 July 2014; Published 6 August 2014

Academic Editor: Yunqiang Yin

Copyright © 2014 Hua Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Owing to the high dimensionality of multilabel data, feature selection in multilabel learning will be necessary in order to reduce the redundant features and improve the performance of multilabel classification. Rough set theory, as a valid mathematical tool for data analysis, has been widely applied to feature selection (also called attribute reduction). In this study, we propose a variable precision attribute reduct for multilabel data based on rough set theory, called δ -confidence reduct, which can correctly capture the uncertainty implied among labels. Furthermore, judgement theory and discernibility matrix associated with δ -confidence reduct are also introduced, from which we can obtain the approach to knowledge reduction in multilabel decision tables.

1. Introduction

Conventional supervised learning deals with the single-label data, where each instance is associated with a single class label. However, in many real-world tasks, one instance may simultaneously belong to multiple class multilabel decision tables, for example, in text categorization problems, where every document may be labeled as several predefined topics, such as religion and political topics [1]; in image annotation problems, a photograph may be associated with more than one tag, such as elephant, jungle, and Africa [2]; in functional genomics, each gene may be related to a set of functional classes, such as metabolism, transcription, and protein synthesis [3]. Such data are called multilabel data.

Owing to the high dimensionality of multilabel data, feature selection in multilabel learning will be necessary in order to reduce the redundant features and improve the performance of multilabel classification. Among various feature selection approaches, rough set theory, proposed by Pawlak [4], has attracted much attention due to its special advantage, that is, the capability of studying imprecise, incomplete, or vague information without requiring prior information.

Feature selection in rough set theory is also called attribute reduction. Generally speaking, attribute reduction can be interpreted as a process of finding the minimal set of attributes that can preserve or improve one or several criteria. The minimal set of attributes is called an attribute reduct. In past few years, many researchers have done much work on attribute reduction and the summarization of important results has been done in [5, 6]. The idea of attribute reduction using positive region was first originated in [7, 8], aiming to remove redundant attributes as much as possible while retaining the so-called positive regions. Afterwards, Ziarko introduced the variable precision rough set model and β -reduct to improve the ability of modeling uncertain information [9]. Furthermore, Kryszkiewicz proposed five kinds of attribute reducts for inconsistent information systems [10] and the relationships in these five reducts and some related results are reconsidered and rectified in [11]. Applying discernibility matrix, Skowron and Rauszer [12] proposed an attribute reduction algorithm by computing disjunctive normal form, which is able to obtain all attribute reducts of a given information system. On the other hand, for obtaining a single reduct from a given information system in a relatively

short time, many heuristic attribute reduction algorithms have been developed. In order to reduce computational time, Xu et al. [13] proposed a quick attribute reduction algorithm with complexity of $\max(O(|C||U|), O(|C|^2|U/C|))$. Further, Qian et al. [14] developed a common accelerator based on four kinds of heuristic reduction algorithms to improve the time efficiency of a heuristic search process.

As far as we know, however, little work has been done on applying rough set theory to feature selection in multilabel learning. Although directly applying the existing attribute reduction methods to multilabel data is possible, it does not take into account the uncertainty conveyed by labels and thus can be enhanced further. In this paper, we propose a new attribute reduct for multilabel data, namely, δ -confidence reduct, which overcomes the limitations of existing attribute reduction methods to multilabel data. Furthermore, judgement theory and discernibility matrix associated with δ -confidence reduct are also established. These results provide approaches to knowledge reduction for multilabel data, which are significant in both the theoretic and applied perspectives.

The rest of this paper is organized as follows. Some basic notions in rough set theory are briefly reviewed in Section 2. Section 3 is devoted to introducing multilabel decision table and analyzing the limitations of the existing attribute reduction methods to multilabel data. In Section 4, the new attribute reduct, δ -confidence reduct, is proposed and the corresponding judgement theorem and discernibility matrix are also introduced. A computative example is also given to illustrate our approaches. Finally, in Section 5, we conclude the paper with a summary and outlook for further research.

2. Preliminaries

In this section, we will review several basic concepts in rough set theory.

A decision table is an information system $S = (U, A \cup D)$ with $A \cap D = \emptyset$, where $U = \{x_1, x_2, \dots, x_n\}$ is a nonempty, finite set of objects called universe; $A = \{a_1, a_2, \dots, a_p\}$ is a nonempty, finite set of condition attributes; $D = \{d_1, d_2, \dots, d_q\}$ is a nonempty, finite set of decision attributes. Each nonempty subset $B \subseteq A$ determines an indiscernibility relation in the following way:

$$R_B = \{(x, y) \in U \times U : a(x) = a(y), \forall a \in B\}. \quad (1)$$

The indiscernibility relation R_B partitions U into some equivalence classes given by $U/R_B = \{[x]_B : x \in U\}$, where $[x]_B$ denotes the equivalence class determined by x with respect to B ; that is,

$$[x]_B = \{y \in U : (x, y) \in R_B\}. \quad (2)$$

Let $X \subseteq U$ and $B \subseteq A$. One can define a lower approximation of X and an upper approximation of X by

$$\underline{R}_B(X) = \{x \in U : [x]_B \subseteq X\} = \bigcup \{[x]_B : [x]_B \subseteq X\}, \quad (3)$$

$$\overline{R}_B(X) = \{x \in U : [x]_B \cap X \neq \emptyset\} = \bigcup \{[x]_B : [x]_B \cap X \neq \emptyset\}, \quad (4)$$

respectively. The lower approximation is called the positive region of X and denoted alternatively as $\text{POS}_B(X)$. If $\underline{R}_B(X) \neq \overline{R}_B(X)$, then X is called a rough set.

Attribute reduct is one of the most important topics in rough set theory, which aims to delete the irrelevant or redundant attributes while retaining the discernible ability of original attributes. Among many attribute reduction methods, the positive region reduct [7, 8] is a representative method.

Definition 1. Let $S = (U, A \cup D)$ be a decision table and $B \subseteq A$. B is a positive region reduct of S if and only if B satisfies the following conditions:

- (1) $\text{POS}_B(D) = \text{POS}_A(D)$,
- (2) $\text{POS}_{B'}(D) \neq \text{POS}_A(D)$ for any $B' \subseteq B$,

where $\text{POS}_B(D) = \bigcup_{i=1}^r \text{POS}_B(D_i)$ and D_i ($i = 1, 2, \dots, r$) are the equivalence classes, called decision classes generated by the indiscernibility relation $R_D = \{(x, y) \in U \times U : d(x) = d(y), \forall d \in D\}$.

3. The Multilabel Data

In this section, we first introduce the multilabel decision table and then analyze the limitations of existing attribute reduction approaches to multilabel data.

3.1. The Multilabel Data. Multilabel data can be characterized by a multilabel decision table $S = (U, A, F, L, G)$ with $A \cap L = \emptyset$, where $U = \{x_1, x_2, \dots, x_n\}$ is a nonempty finite set of objects, called universe; $A = \{a_1, a_2, \dots, a_p\}$ is a nonempty finite set of condition attributes, called condition attribute set; $F = \{a_k : U \rightarrow V_k, k = 1, 2, \dots, p\}$ is a set of information functions with respect to condition attributes and V_k is the domain of a_k ; $L = \{l_1, l_2, \dots, l_q\}$ is a nonempty finite set of q possible labels called label set; $G = \{l_k : U \rightarrow V'_k, k = 1, 2, \dots, q\}$ is a set of information functions with respect to labels and $V'_k = \{0, 1\}$ is the domain of the label l_k . If the object x is associated with label l_k , then $l_k(x) = 1$; otherwise $l_k(x) = 0$. The 5-tuple (U, A, F, L, G) can be expressed more simply as (U, A, L) if F and G are understood.

Some conventions in multilabel learning are shown as follows.

- (1) The object having no labels is irrelevant to multilabel learning and thus is not taken into account in the setting [15, 16]. Note that this convention is a prerequisite for the proposed approach, as discussed in Section 4.
- (2) Each label from L associates with at least one object in U [17].

The following example depicts a multilabel decision table in more detail.

Example 2. A multilabel decision table $S = (U, A, L)$ is presented in Table 1, which is a part of document topic classification problem. It consists of nine documents that belong to one or more of three labels: religion, science, and politics. It

TABLE 1: A multilabel decision table.

U	a	b	c	l_1	l_2	l_3
x_1	2	1	0	1	0	0
x_2	2	1	1	1	0	0
x_3	2	1	0	1	1	0
x_4	1	3	0	1	1	1
x_5	1	3	0	0	1	0
x_6	3	2	1	1	1	1
x_7	1	3	0	0	1	1
x_8	2	3	0	0	1	0
x_9	2	3	0	0	1	1

can be seen that $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$, $A = \{a, b, c\}$, and $L = \{l_1, l_2, l_3\}$. Note that each object in U is associated with at least one label from L and each label from L is associated with at least one object in U .

3.2. *The Limitations of Existing Attribute Reduction Approaches to Multilabel Data.* In this section, we mainly analyze the limitations of existing attribute reduction approaches to multilabel data.

For a multilabel decision table $S = (U, A, L)$, each label attribute can be viewed as a binary decision attribute and then form an indiscernibility relation R_L as follows:

$$R_L = \{(x, y) \in U \times U : l(x) = l(y), \forall l \in L\}. \quad (5)$$

R_L partitions U into a family of equivalence classes given by $U/R_L = \{D_1, D_2, \dots, D_r\}$. In this case, most existing attribute reduction approaches can be directly applied to multilabel data. Here we consider, for instance, positive region reduct, to delete redundant condition attributes in multilabel decision tables. The following example illustrates this process.

Example 3. For the multilabel decision table $S = (U, A, L)$ given by Table 1, we can conclude that

$$\begin{aligned} \frac{U}{R_A} &= \{X_1, X_2, X_3, X_4, X_5\} \\ &= \{\{x_1, x_3\}, \{x_2\}, \{x_4, x_5, x_7\}, \{x_6\}, \{x_8, x_9\}\}; \\ \frac{U}{R_L} &= \{D_1, D_2, D_3, D_4, D_5\} \\ &= \{\{x_1, x_2\}, \{x_3\}, \{x_4, x_6\}, \{x_5, x_8\}, \{x_7, x_9\}\}. \end{aligned} \quad (6)$$

Thus, we have $POS_A(D) = \{x_2, x_6\} = X_2 \cup X_4$. It means that the other equivalence classes X_1, X_3 , and X_5 in U/R_A are all uncertain with respect to the label set L . For instance, consider the equivalence class $X_1 = \{x_1, x_3\}$. Notice that x_1 and x_3 are indiscernible with respect to A while their respect label sets $\{l_1\}$ and $\{l_1, l_2\}$ are discernible with respect to L .

This means X_1 is uncertain with respect to the label set L . Furthermore, we can calculate that

$$\begin{aligned} \frac{U}{R_{\{a,c\}}} &= \{Y_1, Y_2, Y_3, Y_4\} \\ &= \{X_1 \cup X_5, X_2, X_3, X_4\} \\ &= \{\{x_1, x_3, x_8, x_9\}, \{x_2\}, \{x_4, x_5, x_7\}, \{x_6\}\}; \\ \frac{U}{R_{\{b,c\}}} &= \{Z_1, Z_2, Z_3, Z_4\} \\ &= \{X_1, X_2, X_3 \cup X_5, X_4\} \\ &= \{\{x_1, x_3\}, \{x_2\}, \{x_4, x_5, x_7, x_8, x_9\}, \{x_6\}\}. \end{aligned} \quad (7)$$

Since X_1, X_3 , and X_5 are all uncertain with respect to L , they can be safely merged without any information loss. In other words, removing the attribute b or a is valid from the perspective of rough sets. Moreover, one can check that no more attributes can be removed from either $\{a, c\}$ or $\{b, c\}$; so $\{a, c\}$ and $\{b, c\}$ are both positive region reducts.

However, notice that all objects in X_1 must be associated with label l_1 and may be associated with label l_2 in the probability of $1/2$ and must not be associated with label l_3 , whereas all objects in X_5 must not be associated with label l_1 and must be associated with label l_2 and may be associated with label l_3 in the probability of $1/2$. Thus, the uncertainty of X_1 and X_5 is different, and the equivalence class Y_1 , the union of X_1 and X_5 , cannot preserve the uncertainty conveyed by labels. This implies that $\{a, c\}$ is not an appropriate attribute reduct. Similarly, $\{b, c\}$ is also not an appropriate attribute reduct for multilabel data.

Through the above analysis, we know that some positive region reducts cannot preserve uncertainty implied among labels for multilabel data. In fact, since the computation of positive region reduct has to refer to the indiscernibility relation R_L , the uncertainty conveyed by labels may be not analyzed thoroughly. Furthermore, note that the uncertainty characterized by R_L is also considered by the other existing attribute reduction methods; so they have the same limitations for multilabel data like positive region reduct. Thus it is necessary to reconsider attribute reduction method for multilabel data.

4. The New Attribute Reduction Approach in Multilabel Data Decision Tables

In this section, we will introduce a new attribute reduct referred to as δ -confidence reduct and show some advantages of δ -confidence reduct in unraveling the uncertainty of multilabel data. Moreover, judgement theory and discernibility matrix associated with δ -confidence reduct are also established.

4.1. *δ -Confidence Reduct in Multilabel Decision Tables.* First, we present the following definition.

Definition 4. Let $S = (U, A, L)$ be a multilabel decision table, where $A = \{a_1, a_2, \dots, a_p\}$ and $L = \{l_1, l_2, \dots, l_q\}$. For each label l_i , one defines the label decision set as the collection of all possible objects having the label:

$$H_i = \{x \in U : l_i(x) = 1\}. \tag{8}$$

Considering Convention 1 of multilabel learning, one has $\cup_{i=1}^q H_i = U$; that is, H_1, H_2, \dots, H_q form a cover of U .

In the following, we present a particular function to characterize the uncertainty implied among labels.

Definition 5. Let $S = (U, A, L)$ be a multilabel decision table, let $P(L)$ be the power set of label set L , and let H_1, H_2, \dots, H_q be q label decision sets. Given a subset $B \subseteq A$ and $\delta \in [0, 1]$, one defines a δ -confidence label function $\eta_B^\delta : U \rightarrow P(L)$, as follows:

$$\eta_B^\delta(x) = \begin{cases} \left\{ l_i \in L : \frac{|[x]_B \cap H_i|}{|[x]_B|} \geq \delta \right\} & \text{if } 0 < \delta \leq 1; \\ \left\{ l_i \in L : \frac{|[x]_B \cap H_i|}{|[x]_B|} > 0 \right\} & \text{if } \delta = 0. \end{cases} \tag{9}$$

The δ -confidence label function $\eta_B^\delta(x)$ is the collection of the labels that associate with at least $\delta\%$ objects in $[x]_B$. In other words, $\eta_B^\delta(x)$ is the collection of the labels which associate with each object in $[x]_B$ by at least $\delta\%$ confidence level.

Example 6. Consider the multilabel decision table $S = (U, A, L)$ given by Table 1. If $\delta = 0.6$, then the 0.6-confidence label function with respect to attribute set A can be calculated that

$$\begin{aligned} \eta_A^{0.6}(x_1) &= \eta_A^{0.6}(x_3) = \{l_1\}, \\ \eta_A^{0.6}(x_2) &= \{l_1\}, \\ \eta_A^{0.6}(x_4) &= \eta_A^{0.6}(x_5) = \eta_A^{0.6}(x_7) = \{l_2, l_3\}, \\ \eta_A^{0.6}(x_6) &= \{l_1, l_2, l_3\}, \\ \eta_A^{0.6}(x_8) &= \eta_A^{0.6}(x_9) = \{l_2\}. \end{aligned} \tag{10}$$

We have the following property.

Theorem 7. Let $S = (U, A, L)$ be a multilabel decision table, $B, C \subseteq A$. Then

- (1) if $0 \leq \delta_1 \leq \delta_2 \leq 1$, then $\eta_B^{\delta_2}(x) \subseteq \eta_B^{\delta_1}(x)$;
- (2) if $B \subseteq C$, then $\eta_B^1(x) \subseteq \eta_C^1(x) \subseteq \eta_C^0(x) \subseteq \eta_B^0(x)$;
- (3) for any $x \in U$, $\eta_B^0(x) \neq \emptyset$;
- (4) if $[x]_B = [y]_B$, then $\eta_B^\delta(x) = \eta_B^\delta(y)$.

Proof. (1) Let $l_i \in \eta_B^{\delta_2}(x)$. Then we have $|[x]_B \cap H_i| / |[x]_B| \geq \delta_2$. Note that $\delta_2 \geq \delta_1$; thus $|[x]_B \cap H_i| / |[x]_B| \geq \delta_1$. It means that $l_i \in \eta_B^{\delta_1}(x)$. Therefore $\eta_B^{\delta_2}(x) \subseteq \eta_B^{\delta_1}(x)$.

(2) Since $B \subseteq C$, we have $[x]_C \subseteq [x]_B$.

If $l_i \in \eta_B^1(x)$, then $[x]_B \subseteq H_i$. Since $[x]_C \subseteq [x]_B$, we have $[x]_C \subseteq H_i$. Thus $l_i \in \eta_C^1(x)$. Therefore $\eta_B^1(x) \subseteq \eta_C^1(x)$.

According to Theorem 7 (1), we have $\eta_C^1(x) \subseteq \eta_C^0(x)$.

If $l_j \in \eta_C^0(x)$, then $[x]_C \cap H_j \neq \emptyset$. Since $[x]_C \subseteq [x]_B$, we have $[x]_B \cap H_j \neq \emptyset$. That means $l_j \in \eta_B^0(x)$. Therefore $\eta_C^0(x) \subseteq \eta_B^0(x)$.

(3) If there exists $x \in U$ such that $\eta_B^0(x) = \emptyset$, then $[x]_B \cap H_i = \emptyset$, $i = 1, 2, \dots, q$. Thus $[x]_B \cap (H_1 \cup H_2 \cup \dots \cup H_q) = \emptyset$. Since $\cup_{i=1}^q H_i = U$, we have $[x]_B \cap (H_1 \cup H_2 \cup \dots \cup H_q) = [x]_B \cap U \neq \emptyset$. It is a contradiction.

(4) It is straightforward by the definition of $\eta_B^\delta(x)$ and $[x_B] = [y]_B$. □

Now we define the consistent δ -confidence set using δ -confidence label function. Furthermore, we present the definition of new attribute reduct.

Definition 8. Let $S = (U, A, L)$ be a multilabel decision table and $B \subseteq A$. If $\eta_B^\delta(x) = \eta_A^\delta(x)$, for all $x \in U$, one says that B is a consistent δ -confidence set of S . If B is a consistent δ -confidence set and no proper subset of B is a consistent δ -confidence set, then B is called a δ -confidence reduct of S .

A δ -confidence reduct is the minimal set of condition attributes that preserves the invariances of the δ -confidence label function of all objects in U .

Example 9 (continued from Example 6.). For the multilabel decision table $S = (U, A, L)$ given by Table 1, we have

$$\begin{aligned} \eta_{\{a\}}^{0.6}(x_1) &= \{l_1, l_2\} \neq \eta_A^{0.6}(x_1), \\ \eta_{\{b\}}^{0.6}(x_8) &= \{l_2, l_3\} \neq \eta_A^{0.6}(x_8), \\ \eta_{\{c\}}^{0.6}(x_6) &= \{l_1\} \neq \eta_A^{0.6}(x_6), \\ \eta_{\{a,c\}}^{0.6}(x_1) &= \{l_2\} \neq \eta_A^{0.6}(x_1), \\ \eta_{\{b,c\}}^{0.6}(x_8) &= \{l_2, l_3\} \neq \eta_A^{0.6}(x_8), \\ \eta_{\{a,b\}}^{0.6}(x) &= \eta_A^{0.6}(x) \quad \text{for any } x \in U. \end{aligned} \tag{11}$$

Therefore, we obtain the unique 0.6-confidence reduct: $\{a, b\}$.

Considering Example 3, however, we know that $\{a, c\}$ and $\{b, c\}$ are two positive region reducts for the same multilabel decision table. We think δ -confidence reduct is more appropriate for multilabel data than positive region reduct. This is because δ -confidence label function can more reasonably characterize the uncertainty implied among labels than the indiscernibility relation R_L .

Note that the uncertainty characterized by R_L is also considered by the other existing attribute reduction methods. Therefore, for multilabel data, δ -confidence reduct has significant advantages when compared with existing attribute reduction methods.

4.2. Discernibility Matrix of δ -Confidence Reduct. This section provides a discernibility matrix approach [12] to obtain

all δ -confidence reducts. Firstly, we present the judgement theorem of consistent δ -confidence set.

Theorem 10 (judgement theorem of consistent δ -confidence set). *Let $S = (U, A, L)$ be a multilabel decision table, $B \subseteq A$ and $\delta \in [0, 1]$. Then the following conditions are equivalent:*

- (1) B is a consistent δ -confidence set;
- (2) for any $x, y \in U$, if $\eta_A^\delta(x) \neq \eta_A^\delta(y)$, then $[x]_B \cap [y]_B = \emptyset$.

Proof. (1) \Rightarrow (2). If there exist $x, y \in U$ such that $[x]_B \cap [y]_B \neq \emptyset$, then $[x]_B = [y]_B$. By Theorem 7(4), we have $\eta_B^\delta(x) = \eta_B^\delta(y)$. Note that B is a consistent δ -confidence set; we have $\eta_B^\delta(x) = \eta_A^\delta(x)$ and $\eta_B^\delta(y) = \eta_A^\delta(y)$. Therefore $\eta_A^\delta(x) = \eta_A^\delta(y)$.

(2) \Rightarrow (1). Since $B \subseteq A$, it is easy to verify that $\mathfrak{F}([x]_B) = \{[y]_A : [y]_A \subseteq [x]_B\}$ forms a partition of $[x]_B$.

For any $x \in U$, if $[y]_A \subseteq [x]_B$, then $[x]_B \cap [y]_B \neq \emptyset$. By the assumption we obtain $\eta_A^\delta(x) = \eta_A^\delta(y)$.

Let $l_i \in \eta_A^\delta(x)$. Then for all $[y]_A \in \mathfrak{F}([x]_B)$, we have $l_i \in \eta_A^\delta(y)$; that is to say, $|\{[y]_A \cap H_i\}|/|[y]_A| \geq \delta$.

Therefore we have that

$$\begin{aligned} \frac{|[x]_B \cap H_i|}{|[x]_B|} &= \frac{\sum \{ |[y]_A \cap H_i| : [y]_A \in \mathfrak{F}([x]_B) \}}{|[x]_B|} \\ &= \sum \left\{ \frac{|[y]_A \cap H_i|}{|[y]_A|} \cdot \frac{|[y]_A|}{|[x]_B|} : [y]_A \in \mathfrak{F}([x]_B) \right\} \\ &\geq \delta \cdot \sum \left\{ \frac{|[y]_A|}{|[x]_B|} : [y]_A \in \mathfrak{F}([x]_B) \right\} \\ &= \delta. \end{aligned} \tag{12}$$

As a result, $l_i \in \eta_B^\delta(x)$.

On the other hand, we assume that $l_i \in \eta_B^\delta(x)$; however, $l_i \notin \eta_A^\delta(x)$. For any $[y]_A \in \mathfrak{F}([x]_B)$, we have $\eta_A^\delta(x) = \eta_A^\delta(y)$; hence $l_i \notin \eta_A^\delta(y)$; that is to say, $|\{[y]_A \cap H_i\}|/|[y]_A| < \delta$. Since $[x]_B = \cup\{[y]_A : [y]_A \in \mathfrak{F}([x]_B)\}$, we have

$$\begin{aligned} \frac{|[x]_B \cap H_i|}{|[x]_B|} &= \frac{\sum \{ |[y]_A \cap H_i| : [y]_A \in \mathfrak{F}([x]_B) \}}{|[x]_B|} \\ &= \sum \left\{ \frac{|[y]_A \cap H_i|}{|[y]_A|} \cdot \frac{|[y]_A|}{|[x]_B|} : [y]_A \in \mathfrak{F}([x]_B) \right\} \\ &< \delta \cdot \sum \left\{ \frac{|[y]_A|}{|[x]_B|} : [y]_A \in \mathfrak{F}([x]_B) \right\} \\ &= \delta. \end{aligned} \tag{13}$$

Therefore $l_i \notin \eta_B^\delta(x)$, which is a contradiction.

Thus we conclude that $\eta_A^\delta(x) = \eta_B^\delta(x)$ for any $x \in U$. According to Definition 8, we have that B is a consistent δ -confidence set. \square

Theorem 10 provides an approach to judge whether a subset of attributes is a consistent δ -confidence set in multilabel decision tables. Now we present a method for computing all δ -confidence reducts. First, we give the following notion.

Definition 11. Let $S = (U, A, L)$ be a multilabel decision table and $U/R_A = \{X_1, X_2, \dots, X_m\}$. One denotes

$$\Delta^\delta = \{([x]_A, [y]_A) : \eta_A^\delta(x) \neq \eta_A^\delta(y)\}. \tag{14}$$

By $a_k(X_i)$ the value of a_k with respect to the objects in X_i . Define

$$\begin{aligned} M^\delta(X_i, X_j) &= \begin{cases} \{a_k \in A : a_k(X_i) \neq a_k(X_j)\}, & (X_i, X_j) \in \Delta^\delta; \\ A, & (X_i, X_j) \notin \Delta^\delta. \end{cases} \end{aligned} \tag{15}$$

Then $M^\delta(X_i, X_j)$ is called δ -confidence discernibility attribute sets. And $\mathbb{M}^\delta = (M^\delta(X_i, X_j), i, j \leq m)$ is called the δ -confidence discernibility matrix.

For the δ -confidence discernibility matrix, we have the following property.

Theorem 12. *The discernibility matrix $\mathbb{M}^\delta = (M^\delta(X_i, X_j), i, j \leq m)$ satisfies the following properties:*

- (1) \mathbb{M}^δ is a symmetric matrix; that is, for any $i, j \leq m$, $M^\delta(X_i, X_j) = M^\delta(X_j, X_i)$;
- (2) elements in the main diagonals are all A ; that is, for any $i \leq m$, $M^\delta(X_i, X_i) = A$;
- (3) for any $i, s, j \leq m$, $M^\delta(X_i, X_j) \subseteq M^\delta(X_i, X_s) \cup M^\delta(X_s, X_j)$.

Proof. The proofs of (1) and (2) are straightforward. We only need to prove (3). If there exists $a_k \in A$ such that $a_k \in M^\delta(X_i, X_j)$ but $a_k \notin M^\delta(X_i, X_s) \cup M^\delta(X_s, X_j)$, that is, $a_k \notin M^\delta(X_i, X_s)$ and $a_k \notin M^\delta(X_s, X_j)$, then according to Definition 11, we have $a_k(X_i) = a_k(X_s)$ and $a_k(X_s) = a_k(X_j)$. Thus $a_k(X_i) = a_k(X_j)$; that is, $a_k \notin M^\delta(X_i, X_j)$, a contradiction. \square

In the following, we establish some connections between consistent δ -confidence set and discernibility matrix.

Theorem 13. *Let $S = (U, A, L)$ be a multilabel decision table, $B \subseteq A$ and $\delta \in [0, 1]$. Then, B is a consistent δ -confidence set if and only if $B \cap M^\delta(X_i, X_j) \neq \emptyset$ for all $(X_i, X_j) \in \Delta^\delta$.*

Proof. " \Rightarrow " For any $(X_i, X_j) \in \Delta^\delta$, there exist $x, y \in U$, such that $X_i = [x]_A$ and $X_j = [y]_A$. From the definition of Δ^δ , we have $\eta_A^\delta(x) \neq \eta_A^\delta(y)$. Since B is a consistent δ -confidence set, we have $[x]_B \cap [y]_B = \emptyset$ from Theorem 10. Therefore there exists $a_k \in B$ such that $a_k(x) \neq a_k(y)$; that is, $a_k(X_i) \neq a_k(X_j)$. Hence $a_k \in M^\delta(X_i, X_j)$; that is, $B \cap M^\delta(X_i, X_j) \neq \emptyset$.

" \Leftarrow " Let $(X_i, X_j) \in \Delta^\delta$. Since $B \cap M^\delta(X_i, X_j) \neq \emptyset$, for all $(X_i, X_j) \in \Delta^\delta$, there exists $a_l \in B$ such that $a_l \in M^\delta(X_i, X_j)$.

Then we have $a_l(X_i) \neq a_l(X_j)$; that is, $a_l(x) \neq a_l(y)$ for $[x]_A = X_i$ and $[y]_A = X_j$. It means $[x]_B \cap [y]_B = \emptyset$. We then conclude that if $(X_i, X_j) \in \Delta^\delta$, that is, $\eta_A^\delta(x) \neq \eta_A^\delta(y)$, then $[x]_B \cap [y]_B = \emptyset$. It follows from Theorem 10 that B is a consistent δ -confidence set. \square

Next we introduce the concept of discernibility function which helps us to compute δ -confidence reduct.

Definition 14. Let $S = (U, A, L)$ be a multilabel decision table, let $\delta \in [0, 1]$, and let $M^\delta = (M^\delta(X_i, X_j), i, j \leq m)$ be the δ -confidence discernibility matrix, where $A = \{a_1, \dots, a_p\}$. A δ -confidence discernibility function F_S^δ for a multilabel decision table S is a boolean function of p boolean variables $\tilde{a}_1, \dots, \tilde{a}_p$ corresponding to the attributes a_1, \dots, a_p , respectively, and is defined as follows:

$$\begin{aligned} F_S^\delta(\tilde{a}_1, \dots, \tilde{a}_p) &= \bigwedge \left\{ \bigvee M^\delta(X_i, X_j), i, j \leq m \right\} \\ &= \bigwedge \left\{ \bigvee M^\delta(X_i, X_j), (X_i, X_j) \in \Delta^\delta \right\}, \end{aligned} \quad (16)$$

where $\bigvee M^\delta(X_i, X_j)$ is the disjunction of all variables \tilde{a} such that $a \in M^\delta(X_i, X_j)$.

In the sequel we will write a_i instead of \tilde{a}_i when no confusion arises. Furthermore, according to related logical knowledge, we have the following theorem.

Theorem 15. Let $S = (U, A, L)$ be a multilabel decision table. Then an attribute subset B of A is a δ -confidence reduct of S if and only if $\bigwedge B$ is a prime implicant of S .

Theorem 15 provides a discernibility matrix based method to compute all δ -confidence reducts. The following example illustrates the validity of the approach.

Example 16. Consider the multilabel decision table given by Table 1. We have $U/R_A = \{X_1, \dots, X_5\}$, where

$$\begin{aligned} X_1 &= \{x_1, x_3\}, \\ X_2 &= \{x_2\}, \\ X_3 &= \{x_4, x_5, x_7\}, \\ X_4 &= \{x_6\}, \\ X_5 &= \{x_8, x_9\}. \end{aligned} \quad (17)$$

According to the calculation results of $\eta_A^{0.6}(x)$ in Example 6, we have

$$\begin{aligned} \Delta^{0.6} &= \{(X_1, X_3), (X_1, X_4), (X_1, X_5), (X_2, X_3), \\ &\quad (X_2, X_4), (X_2, X_5), (X_3, X_4), (X_3, X_5), (X_4, X_5)\}. \end{aligned} \quad (18)$$

TABLE 2: The δ -confidence discernibility matrix \mathcal{M} .

	X_1	X_2	X_3	X_4	X_5
X_1					
X_2	a, b, c				
X_3	a, b	a, b, c			
X_4	a, b, c	a, b	a, b, c		
X_5	b	b, c	a	a, b, c	

Note that $\eta_A^{0.6}(x_1) = \eta_A^{0.6}(x_2) = \eta_A^{0.6}(x_3)$. Therefore $(X_1, X_2) \notin \Delta^{0.6}$.

We can calculate the δ -confidence discernibility matrix shown in Table 2.

Consequently, we have

$$\begin{aligned} F &= (a \vee b \vee c) \wedge (a \vee b) \wedge (a) \wedge (b) \wedge (b \vee c) \\ &= a \wedge b. \end{aligned} \quad (19)$$

By Theorem 15 we derive that $\{a, b\}$ is the unique 0.6-confidence reduct which accords with the results in Example 9.

5. Conclusion

The δ -confidence reduct presented in this paper is an attribute reduction method designed for multilabel decision tables. Compared with the existing attribute reduction methods, the δ -confidence reduct accurately characterizes uncertainty implied among labels; thus it is more appropriate for multilabel data. Moreover we proposed the corresponding discernibility matrix based method to compute δ -confidence reduct, which is significant in both the theoretic and applied perspectives. In further research, the property of δ -confidence reduct and corresponding heuristic algorithm will be considered.

Conflict of Interests

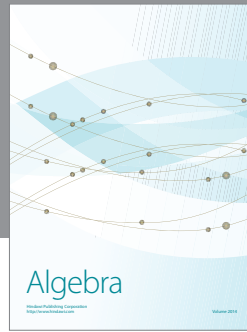
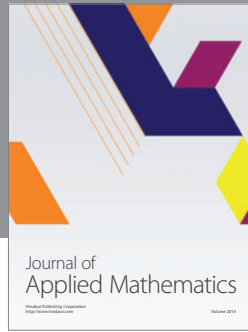
The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank the anonymous reviewers for their valuable comments and suggestions to improve the paper. The work was supported by the National Natural Science Foundation (nos. 61272095, 61175067, 71031006, 61303107, and 71201111), Shanxi Scholarship Council of China (no. 2013-014), Project supported by National Science and Technology (no. 2012BAH33B01), Shanxi Foundation of Tackling Key Problem in Science and Technology (no. 20110321027-02), and the Natural Science Foundation of Hebei Education Department (no. Z2014106).

References

- [1] R. E. Schapire and Y. Singer, "Booster: a boosting-based system for text categorization," *Machine Learning*, vol. 39, no. 2-3, pp. 135–168, 2000.
- [2] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.
- [3] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *Advances in Neural Information Processing Systems 14*, 2002.
- [4] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
- [5] K. Thangavel and A. Pethalakshmi, "Dimensionality reduction based on rough set theory: a review," *Applied Soft Computing Journal*, vol. 9, no. 1, pp. 1–12, 2009.
- [6] C. Wu, Y. Yue, M. Li, and O. Adjei, "The rough set theory and applications," *Engineering Computations*, vol. 21, no. 5, pp. 488–511, 2004.
- [7] J. W. Grzymała-Busse, *Managing Uncertainty in Expert Systems*, Kluwer Academic Publishers, 1991.
- [8] J. W. Grzymała-Busse, "LERS—a system for learning from examples based on rough sets," in *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Set Theory*, pp. 3–18, Kluwer Academic Publishers, New York, NY, USA, 1992.
- [9] W. Ziarko, "Variable precision rough set model," *Journal of Computer and System Sciences*, vol. 46, no. 1, pp. 39–59, 1993.
- [10] M. Kryszkiewicz, "Comparative study of alternative types of knowledge reduction in inconsistent systems," *International Journal of Intelligent Systems*, vol. 16, no. 1, pp. 105–120, 2001.
- [11] D. Li, B. Zhang, and Y. Leung, "On knowledge reduction in inconsistent decision information systems," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 12, no. 5, pp. 651–672, 2004.
- [12] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in *Intelligent Decision Support—Handbook of Applications and Advances of the Rough Sets Theory*, pp. 331–362, Kluwer Academic Publishers, 1992.
- [13] Z. Xu, Z. Liu, B. Yang, and W. Song, "A quick attribute reduction algorithm with complexity of $\max(o(|c||u|); o(|c|^2|u/c|))$," *Chinese Journal of Computers*, vol. 29, no. 3, pp. 391–398, 2006.
- [14] 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.
- [15] N. Ghamrawi and A. McCallum, "Collective multi-label classification," in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM '05)*, pp. 195–200, New York, NY, USA, November 2005.
- [16] S. Kiritchenko, *Hierarchical text categorization and its application to bioinformatics [Ph.D. thesis]*, Queen's University, Kingston, Canada, 2005.
- [17] J. Read, *Scalable multi-label classification [Ph.D. thesis]*, University of Waikato, Hamilton, New Zealand, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

