Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 311–316
ISBN 978-83-60810-51-4

Metric Based Attribute Reduction in Decision Tables

Long Giang Nguyen
Institute of Information Technology, VAST
Viet nam
Email: nlgiang@ioit.ac.vn

Abstract—**In an information system, each subset of attributes determines knowledge structure on the set of objects, in which each element is an equivalence class. Thus, a metric which is defined on knowledge structures is established on the attribute sets. Once a metric is established, we can use the metric to measure attributes distance, cluster and discover important attributes. As a result, effective algorithms are constructed to solve attribute reduction in information systems. With metric on knowledge structures based on the Jaccard distance between two finite sets, this paper proposes a new method for attribute reduction in decision table. The paper proves theoretically and experimentally that this metric method is more effective than other methods based on conditional Shannon entropy.**

I. Introduction

ATTRIBUTE reduction is a core issue of rough set theory and also an essential pre-processing step in data mining. In recent years, there have been many papers about attribute reduction methods based on different views, and generally can be classified as attribute reduction method based on positive region (see [9], [10], [21]), attribute reduction method based on discernibility matrix (see [1], [3]), attribute reduction method used information entropy (see [7], [15], [16], [17], [18], [19]) and attribute reduction method based on granular computing (see [4], [5], [22]). Further research should focus on improvement of existing methods or construction of new advanced methods in order to reduce time complexity and enhance mega data applicability consequently.

Metric technique plays an important role in data mining. In recent years, many researchers have applied this technique to problem solving in data mining, especially problems for data classification which use decision tree (see [13], [14]).

In information systems, each subset of attributes determines knowledge structure on the set of objects. Each element in this knowledge structure is an equivalence class. The difference of metric from existing measures for the uncertainty in rough set theory is that metric can measure the similarity of knowledge structures. Therefore, a metric which is defined on knowledge structures is established on the attribute sets. Once a metric is established, we can use the metric to measure attributes distance, discover important attributes and solve attribute reduction problems in information systems. Recently, there have been some researches on metrics in information systems [11], [12]. However, there are only a few researches in attribute reduction with metric in rough set theory.

In this paper, we present a method to construct a metric among knowledge structures based on the Jaccard distance between two finite sets. With usage of constructed metric, we propose a new method for attribute reduction in a decision table. The paper is structured as follows. In Section 2, some basic concepts in rough set theory are briefly recalled. In Section 3, a method for constructing a metric is presented and some properties of metric are studied. In Section 4, a heuristic algorithm for attribute reduction in a decision table based on metric is proposed. In Section 5, we show an example. In Section 6, we conduct experiments to prove the effectiveness of supposed algorithm. The last section is the conclusion.

II. Basic Concepts

In this section, we introduction some basic concepts in rough set theory and some attribute reduction definitions.

An information system is defined as $IS = (U, A, V, f)$ in which A is a nonempty set of objects; U is a nonempty set of attributes, $V = \prod\limits_{a \in C \cup D}$ where V_a is the value range of attribute a, $f : U \times A \to V$ is an information function, where $\forall a \in A,\ u \in U,\ f(u, a) \in V_a$ hold.

Each subset of attributes $P \subseteq A$ determines a binary indistinguishable relation $IND(P)$ as follows

$$IND(P) = \{(u, v) \in U \times U \,|\forall a \in P, f(u, a) = f(v, a)\}$$

It can be easily shown that $IND(P)$ is an equivalence relation on the set U. The relation $IND(P)$ constitutes a partition of U, which is denoted by U/P. Any element $[u]_P = \{v \in U \,|(u, v) \in IND(P)\}$ in U/P is called the equivalent class.

Definition 2.1 [9], [10]. Let $S = (U, A, V, f)$ be an information system and $P, Q \subseteq A$.

1) U/P is the same as U/Q (denoted by $U/P = U/Q$) if and only if $\forall u \in U,\ [u]_P = [u]_Q$.

2) U/P is finer than U/Q (denoted by $U/P \preceq U/Q$) if and only if $\forall u \in U,\ [u]_P \subseteq [u]_Q$.

Property 2.1 [9], [10]. Let $S = (U, A, V, f)$ be an information system and $P, Q \subseteq A$.

1) If $P \subseteq Q$ then $U/Q \preceq U/P$.

2) For $\forall u \in U$, $[u]_{P \cup Q} = [u]_P \cap [u]_Q$.

Let $S = (U, A, V, f)$ be an information system and $B \subseteq A$, $X \subseteq U$. The sets $\underline{B}X = \{u \in U \,|[u]_B \subseteq X\}$ and $\overline{B}X =$

$\{u \in U \,|[u]_B \cap X \neq \emptyset\}$ are called the $B-lower$ and the $B-upper$ approximation of X, respectively.

If we distinguish in an information system two disjoint classes of attributes, called condition and decision attributes, respectively, then the system will be called a decision table and will be denoted by $DS = (U, C \cup D, V, f)$, where C and D are disjoint sets of condition and decision attributes, respectively.

Let $DS = (U, C \cup D, V, f)$ be a decision table, the set

$$POS_C(D) = \bigcup_{D_i \in U/D} (\underline{C}D_i)$$

is called the $C-positive$ region of D. DS is the consistent decision table if and only if $POS_C(D) = U$. Conversely, DS is the inconsistent decision table.

Definition 2.2 [9]. Let $DS = (U, C \cup D, V, f)$ be a decision table. If $R \subseteq C$ satisfies

1) $POS_R(D) = POS_C(D)$

2) $\forall r \in R, POS_{R-\{r\}}(D) \neq POS_C(D)$

then B is called a reduct of C based on positive region. Let $PRED(C)$ be the set of all reducts based on positive region.

In next content, we introduce the attribute reduction based on Shannon entropy which is relative to research of this paper.

Let $DS = (U, C \cup D, V, f)$ be a decision table. Suppose that $U/C = \{C_1, C_2, ..., C_m\}$, $U/D = \{D_1, D_2, ..., D_n\}$. Conditional Shannon entropy of D with respect to C is defined as

$$H(D\,|\,C) = -\sum_{i=1}^{m} \frac{|C_i|}{|U|} \sum_{j=1}^{n} \frac{|C_i \cap D_j|}{|C_i|} \log_2 \frac{|C_i \cap D_j|}{|C_i|}$$

where $0.\log_2 0 = 0$.

Proposition 2.1 [19]. Let $DS = (U, C \cup D, V, f)$ be a decision table. If $Q \subseteq P \subseteq C$ then $H(D\,|Q) \geq H(D\,|P)$. The condition for equality is $\forall X_u, X_v \in U/P,\ X_u \neq X_v$, if $(X_u \cup X_v) \subseteq Y_k \in U/Q$ then $\frac{|X_u \cap D_j|}{X_u} = \frac{|X_v \cap D_j|}{X_v}$ for $\forall j \in \{1, 2, ..., n\}$.

Definition 2.3 [19]. Let $DS = (U, C \cup D, V, f)$ be a decision table and $a \in C$. If $H(D\,|C) = H(D\,|C - \{a\})$, then a is unnecessary (reducible) for D in C based on conditional Shannon entropy; else a is necessary for D in C. The set of all necessary attributes in C are called the core based on conditional Shannon entropy and denoted as $HCORE(C)$.

Definition 2.4 [19]. Let $DS = (U, C \cup D, V, f)$ be a decision table. If $R \subseteq C$ satisfies

1) $H(D\,|\,R) = H(D\,|\,C)$

2) $\forall r \in R,\ H(D\,|\,R - \{r\}) \neq H(D\,|\,C)$

then R is called a reduct of C based on conditional Shannon entropy. Let $HRED(C)$ be the set of all reducts based on conditional Shannon entropy, then $HCORE(C) = \bigcap_{R \in HRED(C)} R$ [19].

III. Construct a metric among knowledge structures

A metric on the set U is a map $d : U \times U \to [0, \infty)$ which satisfies the following conditions for any $x, y, z \in U$.

$P1)$ $d(x, y) \geq 0$, $d(x, y) = 0$ if and only if $x = y$.

$P2)$ $d(x, y) = d(y, x)$.

$P3)$ $d(x, y) + d(y, z) \geq d(x, z)$.

The condition $P(3)$ is called the triangular inequality. (U, d) is called a metric space.

In this section, we propose the method to construct a metric among knowledge structures based on the Jaccard distance between two finite sets.

A. Jaccard distance between two subsets

Definition 3.1 [2]. Let U be a finite set and $X, Y \subseteq U$. The following coefficient $D(X, Y) = 1 - \frac{|X \cap Y|}{|X \cup Y|}$ is called Jaccard distance between X and Y. The following coefficient $J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$ is called Jaccard coefficient. Jaccard coefficient measures similarity between two finite sets.

Theorem 3.1. Let U be a finite set and $\mathcal{P}(U)$ is the poset of U. Jaccard distance is really a metric on $\mathcal{P}(U)$, that is, for any three elements X, Y and Z of $\mathcal{P}(U)$, it satises the following conditions

$P1)$ $D(X, Y) \geq 0$, $D(X, Y) = 0$ if and only if $X = Y$.

$P2)$ $D(X, Y) = D(Y, X)$

$P3)$ $D(X, Y) + D(Y, Z) \geq D(X, Z)$.

Proof. Obviously, $D(X, Y)$ satises the conditions $P1)$ and $P2)$. To prove the condution $P3)$ (the triangular inequality), we proof the following equivalent inequality

$$J(X, Y) + J(Y, Z) \leq 1 + J(X, Z) \tag{3.1}$$

Suppose $|U| = N$ and $U = \{u_1, u_2, ..., u_N\}$. We can then characterize any subset $X \subseteq U$ by a N-dimensional vector $V^X = (v_1^X, v_2^X, ..., v_N^X)$ where $v_k^X = 1$ if $u_k \in X$ and $v_k^X = 0$ otherwise.

Let $V^{XY} = V^X V^Y$. Then $J(X, Y)$ can be written as

$$J(X, Y) = \frac{V^{XY}}{V^{XX} + V^{YY} + V^{XY}} \tag{3.2}$$

If $J(X, Y) \leq J(X, Z)$ or $J(Y, Z) \leq J(X, Z)$ then obviously (3.1) is satisfies. It is thus necessary to prove inequality (3.1) only for the case in which $J(X, Y) > J(X, Z)$ and $J(Y, Z) > J(X, Z)$. From equation (3.2) we get

$$V^{XY} = \frac{J(X,Y)}{1+J(X,Y)} \left(V^{XX} + V^{YY}\right) \tag{3.3}$$

It can be seen that the inequality $(V^Y - V^X)(V^Y - V^Z) \geq 0$ or $V^{YY} - V^{YZ} - V^{XY} + V^{XZ} \geq 0$ must be true since the product of the kth entries of $(V^Y - V^X)$ and $(V^Y - V^Z)$ equals either 0 or 1. Using equation (3.2) we have

$$V^{YY} - \frac{J(Y,Z)}{1+J(Y,Z)} \left(V^{YY} + V^{ZZ}\right) - \frac{J(X,Y)}{1+J(X,Y)} \left(V^{XX} + V^{YY}\right) + \frac{J(X,Z)}{1+J(X,Z)} \left(V^{XX} + V^{ZZ}\right) \geq 0$$
$$\Leftrightarrow \left(1 - \frac{J(X,Y)}{1+J(X,Y)} - \frac{J(Y,Z)}{1+J(Y,Z)}\right) V^{YY} \geq \left(\frac{J(X,Y)}{1+J(X,Y)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{XX} + \left(\frac{J(Y,Z)}{1+J(Y,Z)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{ZZ} \tag{3.4}$$

It is evident that $V^{XX} \geq V^{XY}$, by applying equation (3.3) we obtain $V^{XX} \geq \frac{J(X,Y)}{1+J(X,Y)} \left(V^{XX} + V^{YY}\right)$ or

$$V^{XX} \geq J(X, Y)\, V^{XY} \tag{3.5}$$

From $J(X, Y) - J(X, Z) > 0$ by assumption, we have $\frac{J(X,Y)}{1+J(X,Y)} - \frac{J(X,Z)}{1+J(X,Z)} > 0$. So from (3.5) we have

$\left(\frac{J(X,Y)}{1+J(X,Y)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{XX} \geq$
$J(X,Y)\left(\frac{J(X,Y)}{1+J(X,Y)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{YY}$ (3.6)

An analogous argument leads to

$\left(\frac{J(Y,Z)}{1+J(Y,Z)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{ZZ} \geq$
$J(Y,Z)\left(\frac{J(Y,Z)}{1+J(Y,Z)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{YY}$ (3.7)

From (3.4), (3.6), (3.7) we have

$\left(1 - \frac{J(X,Y)}{1+J(X,Y)} - \frac{J(Y,Z)}{1+J(Y,Z)}\right) V^{YY} \geq$
$J(X,Y)\left(\frac{J(X,Y)}{1+J(X,Y)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{YY} +$
$J(Y,Z)\left(\frac{J(Y,Z)}{1+J(Y,Z)} - \frac{J(X,Z)}{1+J(X,Z)}\right) V^{YY}$ (3.8)

If $V^{YY} = 0$ then obviously (3.1) is satisfies. Suppose that $V^{YY} \neq 0$, from (3.8) we have

$1 \geq \frac{J(X,Y)+J(X,Y)^2}{1+J(X,Y)} + \frac{J(Y,Z)+J(Y,Z)^2}{1+J(Y,Z)} - J(X,Z)\left(\frac{J(X,Y)+J(Y,Z)}{1+J(X,Z)}\right)$

$\Leftrightarrow J(X,Y) + J(Y,Z) \leq 1 + J(X,Z)$ which is inequality (3.1) we want to prove.

B. Metric between two knowledge structures and its properties

Let $IS = (U, A, V, f)$ be an information system. For each $P \subseteq A$, $K(P) = \{[u_i]_P \,|u_i \in U\}$ is called the knowledge structure on U. $K(P)$ consists of $|U|$ elements, each element is a equivalence class in U/P. Let $\mathcal{K}(U)$ be the set of all knowledge structures on U. From Theorem 3.1, the following theorem constructs a metric between two knowledge structures $K(P)$ and $K(Q)$.

Theorem 3.2. The map $d : \mathcal{K}(U) \times \mathcal{K}(U) \to [0, \infty)$ defined by $d(K(P), K(Q)) = 1 - \frac{1}{|U|}\sum_{i=1}^{|U|} \frac{\left|[u_i]_P \cap [u_i]_Q\right|}{\left|[u_i]_P \cup [u_i]_Q\right|}$ is a metric on $\mathcal{K}(U)$.

Proof. (1) $d(K(P), K(Q)) = 1 - \frac{1}{|U|}\sum_{i=1}^{|U|} \frac{\left|[u_i]_P \cap [u_i]_Q\right|}{\left|[u_i]_P \cup [u_i]_Q\right|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\left(1 - \frac{\left|[u_i]_P \cap [u_i]_Q\right|}{\left|[u_i]_P \cup [u_i]_Q\right|}\right)$. By applying the Definition 3.1 to $[u_i]_P$ and $[u_i]_Q$, we have $1 - \frac{\left|[u_i]_P \cap [u_i]_Q\right|}{\left|[u_i]_P \cup [u_i]_Q\right|} \geq 0$. Therefore, $d(K(P), K(Q)) \geq 0$. $d(K(P), K(Q)) = 0$ if and only if $\left|[u_i]_P \cap [u_i]_Q\right| = \left|[u_i]_P \cup [u_i]_Q\right| \Leftrightarrow [u_i]_P \cap [u_i]_Q = [u_i]_P \cup [u_i]_Q \Leftrightarrow [u_i]_P = [u_i]_Q$ for $u_i \in U$, that is $K(P) = K(Q)$.

(2) From the definition of $d(K(P), K(Q))$, it is clear that $d(K(P), K(Q)) = d(K(Q), K(P))$ for any $K(P), K(Q) \in \mathcal{K}(U)$.

(3) From Theorem 3.1, it is easy to see that $d(K(P), K(Q)) + d(K(Q), K(R)) \geq d(K(P), K(R))$.

From (1), (2), (3) we conclude that $d(K(P), K(Q))$ is a metric on $\mathcal{K}(U)$.

For the decision table $DS = (U, C \cup D, V, f)$, the following proposition constructs the formula of metric between two knowledge structures $K(P)$ and $K(P)$ based on equivalence classes of partitions U/C and U/D.

Proposition 3.1. Let $DS = (U, C \cup D, V, f)$ be a decision table. Assume that $U/C = \{C_1, C_2, ..., C_n\}$ and $U/D = \{D_1, D_2, ..., D_m\}$. Then

$$d_J(K(C), K(C \cup D)) = 1 - \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{|C_i \cap D_j|^2}{|U|\,|C_i|}$$

Proof. Suppose that $C_i \cap D_j = \{u_{i1}, u_{i2}, ..., u_{is_i}\}$ where $|C_i \cap D_j| = s_i$ and $|D_j| = t_j$, then $\sum_{i=1}^{n} s_i = t_j$ and $\sum_{j=1}^{m} t_j = |U|$. We have

$C_i \cap D_j = [u_{i1}]_C \cap [u_{i1}]_D = [u_{i2}]_C \cap [u_{i2}]_D = ... = [u_{is_i}]_C \cap [u_{is_i}]_D$, i.e., $|C_i \cap D_j| = |[u_{i1}]_C \cap [u_{i1}]_D| = |[u_{i1}]_C \cap [u_{i1}]_D| = ... = |[u_{is_i}]_C \cap [u_{is_i}]_D| = s_i$.

$\frac{|C_i \cap D_j|^2}{|C_i|} = \frac{|C_i \cap D_j||C_i \cap D_j|}{|C_i|} = \frac{\left|[u_{i1}]_C \cap [u_{i1}]_D\right|}{\left|[u_{i1}]_C\right|} + \frac{\left|[u_{i2}]_C \cap [u_{i2}]_D\right|}{\left|[u_{i2}]_C\right|} + ... + \frac{\left|[u_{is_i}]_C \cap [u_{is_i}]_D\right|}{\left|[u_{is_i}]_C\right|} = \sum_{k=1}^{s_i} \frac{\left|[u_{ik}]_C \cap [u_{ik}]_D\right|}{\left|[u_{ik}]_C\right|}$.

Hence, $\sum_{i=1}^{n} \frac{|C_i \cap D_j|^2}{|C_i|} = \sum_{i=1}^{n}\sum_{k=1}^{s_i} \frac{\left|[u_{ik}]_C \cap [u_{ik}]_D\right|}{\left|[u_{ik}]_C\right|} = \sum_{i=1}^{t_j} \frac{\left|[u_i]_C \cap [u_i]_D\right|}{\left|[u_i]_C\right|} \Leftrightarrow \sum_{j=1}^{m}\sum_{i=1}^{n} \frac{|C_i \cap D_j|^2}{|C_i|} = \sum_{j=1}^{m}\sum_{i=1}^{t_j} \frac{\left|[u_i]_C \cap [u_i]_D\right|}{\left|[u_i]_C\right|} = \sum_{i=1}^{|U|} \frac{\left|[u_i]_C \cap [u_i]_D\right|}{\left|[u_i]_C\right|} \Leftrightarrow 1 - \frac{1}{|U|}\sum_{j=1}^{m}\sum_{i=1}^{n} \frac{|C_i \cap D_j|^2}{|C_i|} = 1 - \frac{1}{|U|}\sum_{i=1}^{|U|} \frac{\left|[u_i]_C \cap [u_i]_{C \cup D}\right|}{\left|[u_i]_C \cup [u_i]_{C \cup D}\right|} = d_J(K(C), K(C \cup D))$.

Proposition 3.2. Let $DS = (U, C \cup D, V, f)$ be a decision table. If $Q \subseteq P \subseteq C$ then $d(K(Q), K(Q \cup D)) \geq d(K(P), K(P \cup D))$.

Proof. From $Q \subseteq P$ we have $U/P \preceq U/Q$. Without loss of generality, suppose that $U/P = \{X_1, X_2, ..., X_m\}$ and $U/Q = \{X_1, X_2, ..., X_{u-1}, X_{u+1}, ..., X_{v-1}, X_{v+1}, ..., X_m, X_u \cup X_v\}$ is the partition generated by merging the two equivalence class X_u and X_v in U/P. Suppose that $U/D = \{D_1, D_2, ..., D_n\}$. We have

$d(K(Q), K(Q \cup D)) - d(K(P), K(P \cup D)) =$

$= \sum_{i=1,\, i\neq u,\, i\neq v}^{m}\sum_{j=1}^{n} \frac{|X_i \cap D_j|^2}{|U||X_i|} + \sum_{j=1}^{n} \frac{|X_u \cap D_j|^2}{|U||X_u|} + \sum_{j=1}^{n} \frac{|X_v \cap D_j|^2}{|U||X_v|} - \sum_{i=1,\, i\neq u,\, i\neq v}^{m}\sum_{j=1}^{n} \frac{|X_i \cap D_j|^2}{|U||X_i|} - \sum_{j=1}^{n} \frac{|(X_u \cup X_v) \cap D_j|^2}{|U||X_u \cup X_v|}$

$= \sum_{j=1}^{n} \frac{|X_u \cap D_j|^2}{|U||X_u|} + \sum_{j=1}^{n} \frac{|X_v \cap D_j|^2}{|U||X_v|} - \sum_{j=1}^{n} \frac{(|X_u \cap D_j| + |X_v \cap D_j|)^2}{|U|(|X_u| + |X_v|)}$

$= \frac{1}{|U|}\sum_{j=1}^{n}\left(\frac{|X_u \cap D_j|^2}{|X_u|} + \frac{|X_v \cap D_j|^2}{|X_v|} - \frac{(|X_u \cap D_j| + |X_v \cap D_j|)^2}{|X_u| + |X_v|}\right)$

$= \frac{1}{|U|}\sum_{j=1}^{n}\left(\frac{|X_u \cap D_j|^2|X_v|^2 - 2|X_u \cap D_j||X_v \cap D_j||X_u||X_v| + |X_v \cap D_j|^2|X_u|^2}{|X_u||X_v|(|X_u| + |X_v|)}\right)$

$= \frac{1}{|U|}\sum_{j=1}^{n} \frac{(|X_u \cap D_j||X_v| - |X_v \cap D_j||X_u|)^2}{|X_u||X_v|(|X_u| + |X_v|)} \geq 0$

Hence, $d(K(Q), K(Q \cup D)) \geq d(K(P), K(P \cup D))$. $d(K(Q), K(Q \cup D)) = d(K(P), K(P \cup D)) \Leftrightarrow |X_u \cap D_j|\,|X_v| - |X_v \cap D_j|\,|X_u| = 0 \Leftrightarrow |X_u \cap D_j|\,|X_v| = |X_v \cap D_j|\,|X_u| \Leftrightarrow \frac{|X_u \cap D_j|}{|X_u|} = \frac{|X_v \cap D_j|}{|X_v|}$.

Therefore, the condition for equality is $\forall X_u, X_v \in U/P$, $X_u \neq X_v$ if $(X_u \cup X_v) \subseteq Y_k \in U/Q$ then $\frac{|X_u \cap D_j|}{X_u} = \frac{|X_v \cap D_j|}{X_v}$ for $\forall j \in \{1, ..., n\}$.

Proposition 3.2 indicates that the bigger the attribute set P is, the smaller the metric $d\left(K\left(P\right),K\left(P\cup D\right)\right)$ is, and vice versa. In other words, the metric $d\left(K\left(P\right),K\left(P\cup D\right)\right)$ decreases monotonically because the granularity of information becomes smaller through finer partitions. From the conditions for equality of Proposition 3.2 and Proposition 2.1, we have the following proposition.

Proposition 3.3. Let $DS=(U,C\cup D,V,f)$ be a decision table, $Q\subseteq P\subseteq C$. $d\left(K\left(Q\right),K\left(Q\cup D\right)\right)=d\left(K\left(P\right),K\left(P\cup D\right)\right)$ if and only if $H\left(D\left|Q\right.\right)=H\left(D\left|P\right.\right)$.

IV. An attribute reduction algorithm based on metric

Definition 4.1. Let $DS=(U,C\cup D,V,f)$ be a decision table and $a\in C$. If $d(K\left(C-\{c\}\right),K\left(C-\{c\}\cup D\right))=d(K\left(C\right),K\left(C\cup D\right))$, then a is unnecessary (reducible) for D in C; else a is necessary for D in C. The set of all necessary attributes in C are called the core based on metric and denoted as $MCORE(C)$.

Definition 4.2. Let $DS=(U,C\cup D,V,f)$ be a decision table. If $R\subseteq C$ satisfies

1) $d\left(K\left(R\right),K\left(R\cup D\right)\right)=d\left(K\left(C\right),K\left(C\cup D\right)\right)$

2) $\forall r\in R$, $d(K\left(R-\{r\}\right),K\left(R-\{r\}\cup D\right))\neq d(K\left(C\right),K\left(C\cup D\right))$

then R is called a reduct of C based on metric. Let $MRED(C)$ be the set of all reducts based on metric.

From Proposition 3.3, we can conclude that the core and the reduct based on metric are the same as those based on Shannon entropy. Therefore, $MCORE\left(C\right)=HCORE\left(C\right)$ and $MRED\left(C\right)=HRED\left(C\right)$.

Definition 4.3. Let $DS=(U,C\cup D,V,f)$ be a decision table and $B\subset C$. The significance of attribute $b\in C-B$ is defined as $SIG_B\left(b\right)=d\left(K\left(B\right),K\left(B\cup D\right)\right)-d\left(K\left(B\cup\{b\}\right),K\left(B\cup\{b\}\cup D\right)\right)$ where $[u_i]_\emptyset=U$ for any $u_i\in U,\ i\in\{1,...,|U|\}$.

Definition 4.3 implies that the significance of attribute $b\in C-B$ is measured by the changes of the metric $d\left(K\left(B\right),K\left(B\cup D\right)\right)$ when b is added to B, the bigger value of $SIG_B\left(b\right)$ is, the more important attribute b is. This significance of attribute is a attribute selection criterion in our heuristic algorithm for attribute reduction.

According to Definition 2.4, the core is unique, and it is the intersection of all reducts. Thus, it can be taken as the beginning point of finding the reduct. In this paper, in order to find the reduct, first, we compute the core; then the most important attribute is chosen from searching space and added into the core. The above processes are done until we get the reduct.

Algorithm 4.1. Finding the core based on metric.

Input: A decision table $DS=(U,C\cup D,V,f)$.

Output: The core $MCORE(C)$.

1. $MCORE\left(C\right)=\emptyset$;
2. Calculate $d\left(K\left(C\right),K\left(C\cup D\right)\right)$;
3. For each $c\in C$
4. Begin
5. Calculate $d(K\left(C-\{c\}\right),K\left(C-\{c\}\cup D\right))$;
6. If $d(K\left(C-\{c\}\right),K\left(C-\{c\}\cup D\right))\neq d(K\left(C\right),K\left(C\cup D\right))$ then $MCORE\left(C\right):=MCORE\left(C\right)\cup\{c\}$;
7. End;
8. Return $MCORE(C)$;

With usage of the method for calculating the partition U/C in [6], the time complexity to calculate $d\left(K\left(C\right),K\left(C\cup D\right)\right)$ is $O\left(|C|\,|U|\right)$. Consequently, the time complexity of Algorithm 4.1 is $O\left(|C|^2\,|U|\right)$.

Let $DS=(U,C\cup D,V,f)$ be a decision table, $R\subset C$ and $a\in C-R$. Suppose that $U/R=\{R_1,R_2,...,R_k\}$, $U/R\cup\{a\}=\left\{R_1^1,R_2^1,...,R_m^1\right\}$. According to Definition 4.3, the significance of the attribute $a\in C-R$ is $SIG_R\left(a\right)=$

$$\frac{1}{|U|}\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left|R_i^1\cap D_j\right|^2}{\left|R_i^1\right|}-\frac{1}{|U|}\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{|R_i\cap D_j|^2}{|R_i|}\qquad(4.1)$$

To calculate $SIG_R\left(a\right)$, we must calculate U/R and $U/R\cup\{a\}$. This paper proposes an effictive algorithm to calculate $U/R\cup\{a\}$ when U/R has already calculated. The algorithm as follows.

Algorithm 4.2. Calculating $U/R\cup\{a\}$.

Input: $U/R=\{R_1,R_2,...,R_k\}$.

Output: $U/R\cup\{a\}$.

1. $TMP=\emptyset$;
2. For each $R_i\in U/R$ do
3. Begin
4. Calculate $R_i/\{a\}$;
5. $TMP=TMP\ \cup\ R_i/\{a\}$;
6. End;
7. Return (TMP);

With usage of the methods for calculating the partition U/C in [6], the time complexity to calculate $R_i/\{a\}$ is $O\left(|R_i|\right)$. Therefore, the time complexity of Algorithm 4.2 is $\sum_{i=1}^{k}O\left(|R_i|\right)=O\left(|U|\right)$.

Algorithm 4.3. Finding a reduct based on metric.

Input: A decision table $DS=(U,C\cup D,V,f)$.

Output: One reduct R of C.

1. Using Algorithm 4.1, we get $MCORE(C)$;
2. $R=MCORE(C)$;

// add the most important attributes little by little to the core

3. While $d\left(K\left(R\right),K\left(R\cup D\right)\right)\neq d\left(K\left(C\right),K\left(C\cup D\right)\right)$
4. Begin
5. For each $a\in C-R$
6. Begin
7. Calculate $d\left(K\left(R\cup\{a\}\right),K\left(R\cup\{a\}\cup D\right)\right)$;
8. Calculate $SIG_R\left(a\right)=d\left(K\left(R\right),K\left(R\cup D\right)\right)-d\left(K\left(R\cup\{a\}\right),K\left(R\cup\{a\}\cup D\right)\right)$;
9. End;
10. Select $a_m\in C-R$ such that $SIG_R\left(a_m\right)=\underset{a\in C-R}{Max}\{SIG_R\left(a\right)\}$;
11. $R=R\cup\{a_m\}$;
12. Calculate $d\left(K\left(R\right),K\left(R\cup D\right)\right)$;

13. End;
//Deleting redundant attributes in R
14. $R* = R - MCORE\,(C)$;
15. For each $a \in R*$
16. Begin
17. Calculate $d\left(K\left(R - \{a\}\right), K\left(R - \{a\} \cup D\right)\right)$;
18. If $d\left(K\left(R - \{a\}\right), K\left(R - \{a\} \cup D\right)\right) = d\left(K\left(C\right), K\left(C \cup D\right)\right)$ then $R = R - \{a\}$;
19. End;
20. Return R;

Computational complexity analysis of Algorithm 4.3

Let us consider While loop from Command line 3 to Command line 13. According to the equation (4.1), to compute $SIG_R\,(a)$ we only need to compute $U/R \cup \{a\}$ because U/R have already been computed in the previous step. According to Algorithm 4.2, the time complexity to compute $U/R \cup \{a\}$ when U/R has already been calculated is $O\left(|U|\right)$, so the time complexity to compute all $SIG_R\,(a)$ is $(|C| + (|C| - 1) + ... + 1)*|U| = (|C| * (|C| - 1)/2)*|U| = O\left(|C|^2\,|U|\right)$.

The time complexity to choose maximum for significance of attribute is $|C| + (|C| - 1) + ... + 1 = |C| * (|C| - 1)/2 = O\left(|C|^2\right)$. The For loop at Command line 15 runs $|R^*|$ times, each time we have to compute $d\left(K\left(R\right), K\left(R \cup D\right)\right)$ with $O\left(|R|\,|U|\right)$ complexity. So the time complexity of Command line 15 is $O\left(|R^*|\,|R|\,|U|\right)$. Thus the time complexity of this algorithm is $O\left(|C|^2\,|U|\right)$.

With attribute reduction algorithms based on Shannon entropy, the time complexity of MIBARK Algorithm [7], CEBARKNC Algorithm and CEBARKCC Algorithm [18] is $O\left(|C|\,|U|^2 + |U|^3\right)$, $O\left(|C|^2\,|U| + |C|\,|U|^3\right)$ and $O\left(|C|\,|U|^2 + |U|^3\right)$ respectively. Therefore, Algorithm 4.3 is more effective than the above algorithms in theory.

Time complexity of attribute reduction algorithms depends on time complexity of U/C calculation. Algorithm 4.3 is considered more effective than the above algorithms with the same U/C calculation [6] for the following reasons:

1) Algorithm 4.3 uses Algorithm 4.2 to compute $U/R \cup \{a\}$ based on U/R. So the time complexity is reduced.

2) To compute the metric, we need only to carry out simple arithmetic operations, while to compute the conditional entropy, we have to compute the logarithm of the frequency, a time-consuming operation.

V. Example

A decision table $DS = (U, C \cup D, V, f)$ is shown in TABLE I, where $C = \{a_1, a_2, a_3\}$, $D = \{d\}$, $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$.

We have $U/\{d\} = \{\{u_1, u_3\}, \{u_2, u_4, u_5, u_6, u_7\}\}$, $U/C = \{\{u_1, u_2\}, \{u_3, u_4, u_5\}, \{u_6\}, \{u_7\}\}$. Using Algorithm 4.1 to find the core, we have

TABLE I
THE DECISION TABLE

U	a_1	a_2	a_3	d
u_1	0	1	1	0
u_2	0	1	1	0
u_3	0	1	1	0
u_4	0	1	1	0
u_5	0	1	1	0
u_6	0	1	1	0
u_7	0	1	1	0

$d\left(K\left(C\right), K\left(C \cup \{d\}\right)\right) = 1 - \frac{1}{|U|}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|C_i \cap D_j|^2}{|C_i|} = 1 - \frac{1}{7}\left(\frac{1^2}{2} + \frac{1^2}{2} + \frac{1^2}{3} + \frac{2^2}{3} + \frac{1^2}{1} + \frac{1^2}{1}\right) = \frac{1}{3}$. Similarly,
$d\left(K\left(C - \{a_1\}\right), K\left(\left(C - \{a_1\}\right) \cup \{d\}\right)\right) = \frac{1}{3}$
$d\left(K\left(C - \{a_2\}\right), K\left(\left(C - \{a_2\}\right) \cup \{d\}\right)\right) = \frac{1}{3}$
$d\left(K\left(C - \{a_3\}\right), K\left(\left(C - \{a_3\}\right) \cup \{d\}\right)\right) = \frac{12}{35}$
Therefore, $MCORE\,(C) = \{a_3\}$.

Using Algorithm 4.3 to find a reduct, we have $R = \{a_3\}$, $d\left(K\left(\{a_3\}\right), K\left(\{a_3, d\}\right)\right) = \frac{17}{42}$. So $d\left(K\left(\{a_3\}\right), K\left(\{a_3, d\}\right)\right) \neq d\left(K\left(C\right), K\left(C \cup D\right)\right)$, go to the While loop. $SIG_{\{a_3\}}\,(a_1) = d\left(K\left(\{a_3\}\right), K\left(\{a_3, d\}\right)\right) - d\left(K\left(\{a_1, a_3\}\right), K\left(\{a_1, a_3, d\}\right)\right) = \frac{17}{42} - \frac{1}{3} = \frac{1}{14}$. Similarly, we compute $SIG_{\{a_3\}}\,(a_2) = SIG_{\{a_3\}}\,(a_1) = \frac{1}{14}$.

Because a_1 and a_2 have the same significance, so we can choose either a_1 or a_2. Suppose that we choose a_1, then $R = \{a_1, a_3\}$ and $d\left(K\left(\{a_1, a_3\}\right), K\left(\{a_1, a_3, d\}\right)\right) = d\left(K\left(C\right), K\left(C \cup \{d\}\right)\right)$. Go to the For loop.

Consider $R* = R - MCORE(C) = \{a_1\}$, $R - \{a_1\} = \{a_3\}$, $d\left(K\left(\{a_3\}\right), K\left(\{a_3, d\}\right)\right) \neq d\left(K\left(C\right), K\left(C \cup \{d\}\right)\right)$. Therefore, $R = \{a_1, a_3\}$ is a reduct of C.

VI. Experiments

The experiments on PC (Pentium Dual Core 2.13 GHz, 1GB RAM, WINXP) are performed on 8 data sets obtained from UCI Machine Learning Repository [23]. We compare CEBARKCC Algorithm (AL01 for short) with Algorithm 4.3 (AL02 for short). The results of experiments are showm in TABLE II and TABLE III, where $|U|$, $|C|$, $|R|$ are the numbers of objects, condition attributes, and after reduction respectively, and t is the time of operation (calculated by second). Condition attributes will be denoted by 1, 2,...,$|C|$.

The results show that the reduct of Algorithm 4.3 is the same as that of the CEBARKCC Algorithm. However, the time of operation in Algorithm 4.3 is faster than that in the CEBARKCC Algorithm. It means that Algorithm 4.3 is more effective. Furthermore, the reducts of these data sets in our experiments are the same as those in [15, 16, 17]. This proves result of the experiments.

VII. Conclusion

In this paper, we have proposed a method to construct a metric on the family of knowledge structures. Based on proposed metric, we have proposed a new method for attribute reduction in decision tables. The paper has proved theoretically and experimentally that this metric method is more effective than other methods based on conditional Shannon entropy.

TABLE II
THE RESULTS OF ALGORITHM 4.3 AND CEBARKCC ALGORITHM

Seq	Data sets	$\|U\|$	$\|C\|$	AL01		AL02	
				$\|R\|$	t	$\|R\|$	t
1	Tic-tac-toe	958	9	8	8.343	8	5.937
2	Hepatitis	155	19	3	0.484	3	0.312
3	Lung-cancer	32	56	4	0.78	4	0.62
4	Automobile	205	25	6	3.921	6	2.562
5	Abalone	4177	8	3	256.12	3	127.43
6	Liver-disorders	345	6	3	0.796	3	0.531
7	Iris	150	4	3	0.93	3	0.78

TABLE III
THE REDUCTS OF ALGORITHM 4.3 AND CEBARKCC ALGORITHM

Seq	Data sets	The reducts of AL01	The reducts of AL02
1	Tic-tac-toe	$1, 2, 3, 4, 5, 7, 8, 9$	$1, 2, 3, 4, 5, 7, 8, 9$
2	Hepatitis	$2, 15, 16$	$2, 15, 16$
3	Lung-cancer	$3, 4, 9, 43$	$3, 4, 9, 43$
4	Automobile	$1, 2, 7, 14, 20, 21$	$1, 2, 7, 14, 20, 21$
5	Abalone	$2, 5, 6$	$2, 5, 6$
6	Liver-disorders	$1, 2, 5$	$1, 2, 5$
7	Iris	$1, 2, 3$	$1, 2, 3$

ACKNOWLEDGMENT

This work was funded by the Vietnam's National Foundation for Science and Technology Development NAFOSTED) via a research grant for fundamental sciences, grant number: 102.01-2010.09

REFERENCES

[1] Andrzej Skowron and Rauszer C, The Discernibility Matrices and Functions in Information Systems, Interlligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer, Dordrecht, (1992), 331-362.
[2] Deza M. M. and Deza E., Encyclopedia of Distances, Springer, 2009.
[3] Hu X. H. and Cercone N., Learning in relational databases: a rough set approach, International Journal of computational intelligence, (1995), 323-338.
[4] Li J. H. and Shi K.Q. (2006),A algorithm for attribute reduction based on knowledge granularity, Computer Applications, 26(6)(2006), 76-77.
[5] Li J. H., Lv Y. J., and Liu N. X., A Different Quantity of Partition-based Efficient Algorithm for Reduction of Attribute in Information Systems, Fourth International Conference on Fuzzy Systems and Knowledge Discovery, (2007).
[6] Lv Y. J. and Li J. H., A Quick Algorithmfor Reduction of Attribute in Information Systems, The First International Symposium on Data, Privacy, and E-Commerce (ISDPE 2007), (2007), 98-100.
[7] Miao D. Q. and Hu G. R. (1999), A heuristic algorithm for knowledge reduction, Computer Research and Development, 36(6)(1999), 681-684.
[8] Nguyen S. Hoa, Nguyen H. Son, Some Efficient Alogrithms for Rough Set Methods, Proceedings of the sixth International Conference on Information Processing Management of Uncertainty in Knowledge Based Systems, (1996), 1451 - 1456.
[9] Pawlak Z., Rough sets: Theoretical Aspects of Reasoning About Data, Kluwer Aca-demic Publishers, 1991.
[10] Pawlak Z., Rough set theory and its applications in data analysis, Cybernetics and systems 29, (1998), 661-688.
[11] Qian Y. H., Liang J. Y. and Dang C. Y., Knowledge structure, knowledge granulation and knowledge distance in a knowledge base, International Journal of Approximate Reasoning 50, (2009), 174–188.
[12] Qian Y. H., Liang J. Y., Dang C. Y., Wang F. and Xu W., Knowledge distance in information systems, Journal of Systems Science and Systems Engineering, Vol.16, (2007), 434-449.
[13] R.Lłpez de Mntaras, A distance-based attribute selection measure for decision tree induction, Machine Learning Vol. 6 (1991), 81-92.
[14] Simovici D. A., Jaroszewicz S., Generalized conditional entropy and decision trees, Proceeding of EGC, Lyon, France, (2003), 369-380.
[15] Sun L., Xu J.C and Cao X.Z, Decision Table Reduction Method Based on New Conditional Entropy for Rough Set Theory, International Workshop on Intelligent Systems and Applications, (2009), 1-4.
[16] Wang B. Y. and Zhang S. M., A Novel Attribute Reduction Algorithm Based on Rough Set and Information Entropy Theory, 2007 International Conference on Computational Intelligence and Security Workshops, IEEE CISW, (2007), 81–84.
[17] Wang C.R. and OU F.F., An Attribute Reduction Algorithm in Rough Set Theory Based on Information Entropy, 2008 International Symposium on Computational Intelligence and Design, IEEE ISCID, (2008), 3-6.
[18] Wang G. Y., Yu H. and Yang D.C., Decision table reduction based on conditional information entropy, Journal of Computers, Vol.25, No.7, (2002), 759-766.
[19] Wang G. Y., Yu H., Yang D.C. and Wu Z.F., Knowledge Reduction Based on Rough Set and Information Entropy, Proc. Of the World Multiconference on Systemics, Cybernetics and Informatics, Orlando, Florida, (2001), 555-560.
[20] Xu Z. Y., Yang B.R. and Song W., Complete attribute reduction algorithm based on Simplified discernibility matrix, Computer Engineering and Applications, Vol.42, No.26, (2006), 167–169.
[21] Ye D. Y., An improvement to Jelonek's attribution reduction algorithm, Acta Electronics Sinica, Vol. 28, No.12, (2000), 81–82.
[22] Zhao M., Luo K. and Qin Z., Algorithm for attribute reduction based on granular computing, Computer Engineering and Applications, Vol.44, No.30, (2008), 157-159.
[23] The UCI machine learning repository, http://archive.ics.uci.edu/ml/datasets.html