An Attribute Reduction Approach and Its Accelerated Version for Hybrid Data

Wei Wei Jiye Liang Yuhua Qian Feng Wang
Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan, 030006, Shanxi, China
School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, Shanxi, China
Email: weiwei@sxu.edu.cn ljy@sxu.edu.cn jinchengqyh@126.com sxuwangfeng@126.com

***Abstract*—In practical issues, categorical data and numerical data usually coexist, and a unified data reduction technique for hybrid data is desirable. In this paper, an information measure is proposed for computing the discernibility power of a categorical or numeric attribute. Based on the measure, a uniform definition of significance of attributes with categorical values and numerical values is proposed. Furthermore, an algorithm to obtain an attribute reduct from hybrid data is presented, and one of its accelerated version is also constructed. Experiments show that these two algorithms can get the same reducts, and the classification accuracies of reduced datasets are similar with the ones using Hu's algorithm. However, the accelerated version consumes much less time than the original one and Hu's algorithm do.**

I. Introduction

Rough set theory was proposed by Pawlak in 1982. Recently, it has become a popular mathematical framework for pattern recognition, image processing, feature selection, neuro computing, conflict analysis, decision support, data mining and knowledge discovery process from large data sets [1], [17]. In recent years, more attention has been paid to attribute reduction in information systems and decision tables. Many types of attribute reduction techniques have been proposed in the last twenty years. Consistency of data [14], [17], dependence degree [5], information entropy [22], discernibility matrix [21] were employed to find reducts of an information system. In [10]–[12], a new uncertainty measure of information systems was proposed, which can be employed to compute an attribute reduct. In [18], [19], the combination entropy was presented, which can construct a heuristic function in a heuristic reduction algorithm in rough set theory. β-reduct proposed in [26] provides a kind of attribute-reduction methods in the variable precision rough set model. α-reduct and α-relative reduct that allow the occurrence of additional inconsistency were proposed in [16]. A new insight into the problem of attribute reduction was provided in [24], [25]. Five kinds of attribute reducts and their relationships in inconsistent systems were investigated by [7], [9].

All above reduction approaches are only valid for information systems with categorical attributes. However, categorical and numerical data usually coexists in real world databases. Some generalizations of the model were proposed to deal with this problem. Rough set theory and fuzzy set theory were putted together and rough-fuzzy sets and fuzzy rough sets were defined in [2], [13]. The properties and axiomatization of fuzzy rough set theory were analyzed in detail [15], [23], which has been applied to data reduction [3], [4]. A fuzzy-rough attribute reduction, called fuzzy-rough QUICKREDUCT algorithm, was given in [6], [20] based on fuzzy dependency function, which can measure the discernibility power of categorical attributes and nominal attributes.

In this paper, we will introduce a fuzzy information measure based on Liang's entropy, which can be used as an evaluation of the discernibility power of a categorical or numeric attribute. According to the properties of proposed information measure, adding a new condition attribute into the information system, the value of the measure will increase monotonously, which reflexes that adding information will lead to enhancement of the discernibility power of an attribute. Then we construct a hybrid attribute reduction algorithm and accelerated version based on the proposed measure, which is applicable to reduce a hybrid dataset.

The rest of the paper are organized as follows. Some preliminary concepts is reviewed in Section 2. In Section 3, an new information measure based on Liang's entropy and its properties are proposed. Section 4 gives a new reduction algorithms and an accelerated version for hybrid data, and an experimental analysis is performed, which shows the accelerated version consumes much less time than the original one and the algorithm in [4] do. Section 5 concludes the paper.

II. Preliminaries

In this section, we review some basic concepts such as fuzzy equivalence relation, fuzzy partition, fuzzy lower approximation and upper approximation.

Pawlak's rough set model can only deal with data containing categorical values. As we know the real-world applications usually contain categorical attributes and numerical attributes. A hybrid information system can be written as $(U, \widetilde{A} = A^r \cup A^c)$, where U is the set of objects, A^r is numerical attributes, and A^c is categorical attributes.

A categorical attribute can induce a crisp equivalence relation on the universe and generate a family of crisp information granules, whereas a numerical attribute will give a fuzzy equivalence relation and form a set of fuzzy information granules. As crisp information granules are a special case of

Proc. 8th IEEE Int. Conf. on Cognitive Informatics (ICCI'09)
G. Baciu, Y. Wang, Y.Y. Yao, W. Kinsner, K. Chan & L.A. Zadeh (Eds.)
978-1-4244-4642-1/09/$25.00 ©2009 IEEE

fuzzy ones, we will consider all of them as fuzzy ones in the following.

Each non-empty subset $\widetilde{C} \subseteq \widetilde{A}$ determines an fuzzy equivalence relation $R_{\widetilde{C}}$, which is corresponding to a fuzzy equivalence as follows

$$M(R_{\widetilde{C}}) = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix},$$

where $r_{ij} \in [0,1]$ is the relation value of x_i and x_j.

Furthermore, the relation $R_{\widetilde{C}}$ satisfies:

(1) Reflectivity: $R_{\widetilde{C}}(x,x) = 1$, $\forall x \in X$;
(2) Symmetry: $R_{\widetilde{C}}(x,y) = R_{\widetilde{C}}(y,x), \forall x, y \in X$; and
(3) Transitivity: $R_{\widetilde{C}}(x,z) \geq min_y\{R_{\widetilde{C}}(x,y), R_{\widetilde{C}}(y,z)\}$.

The relation $R_{\widetilde{C}}$ partitions U into some fuzzy equivalence classes given by $U/R_{\widetilde{C}} = \{[x_i]_{R_{\widetilde{C}}}\}_{i=1}^n$, just $U/\widetilde{C} = \{[x_i]_{\widetilde{C}}\}_{i=1}^n$, where $[x_i]_{R_{\widetilde{C}}}$ denotes the fuzzy equivalence class determined by x_i as to a fuzzy attribute set $\widetilde{C}$.

Definition 1. [3] Let $S = (U, \widetilde{A})$ be a hybrid information system, $\widetilde{C} \subseteq \widetilde{A}$. The fuzzy equivalence class $[x_i]_{R_{\widetilde{C}}}$ induced by the attribute set $\widetilde{C}$ is defined as

$$[x_i]_{\widetilde{C}} = \frac{r_{i1}}{x_1} + \frac{r_{i2}}{x_2} + \cdots + \frac{r_{in}}{x_n},$$

where " + " means the union.

Obviously, $[x_i]_{\widetilde{C}}$ is a fuzzy information granule. It is easy to find that the definition of fuzzy equivalence classes is a natural extension of crisp one. Furthermore, the cardinality $[x_i]_{\widetilde{C}}$ is defined as

$$|[x_i]_{\widetilde{C}}| = \sum_{j=1}^{n} r_{ij}.$$

Definition 2. [3] Let $S = (U, \widetilde{A})$ be a hybrid information system, $\widetilde{C} \subseteq \widetilde{A}$, $X \subseteq U$ a crisp subset of objects. The lower and upper approximation of X can be defined as

$$\underline{\widetilde{C}}(X) = \{x_i \mid [x_i]_{\widetilde{C}} \subseteq X,\ x_i \in U\},\ and$$

$$\overline{\widetilde{C}}(X) = \{x_i \mid [x_i]_{\widetilde{C}} \cap X \neq \emptyset,\ x_i \in U\},$$

where $[x_i]_{\widetilde{C}} \subseteq X$ means $\mu_{[x_i]_{\widetilde{C}}}(x_i) \leq \mu_X(x_i)$, $\forall x_i \in U$, $[x_i]_{\widetilde{C}} \cap X \neq \emptyset$ denotes $\min\{\mu_{[x_i]_{\widetilde{C}}}(x_i), \mu_X(x_i)\} \neq 0$, $\emptyset = \frac{0}{x_1} + \frac{0}{x_2} + \cdots + \frac{0}{x_n}$.

III. A NEW INFORMATION MEASURE FOR HYBRID DATA

In this section, we will propose a new information measure to evaluate the discernibility power of a hybrid attribute set. Sequently, some useful property about this measure is given, which is the fundament of the algorithms in the following section.

Definition 3. Let $S = (U, \widetilde{A})$ be a hybrid information system, then the information measure of hybrid attribute set $\widetilde{A}$ is defined as

$$E(\widetilde{A}) = \frac{1}{n}\sum_{i=1}^{n}(1 - \frac{|[x_i]_{\widetilde{A}}|}{n}),$$

where n denotes the cardinality of the universe U.

Theorem 1. Let $S = (U, \widetilde{A})$ be a hybrid information system. If $\widetilde{A}$ is a categorical attribute set, then the information quantity of attribute set $\widetilde{A}$ degenerates into

$$E(\widetilde{A}) = \sum_{k=1}^{m} \frac{|X_k|}{n}(1 - \frac{|X_k|}{n}),$$

where $X_k \in U/\widetilde{A}$ and $U/\widetilde{A} = \{X_1, X_2, \cdots, X_m\}$.

Proof. From $|[x_i]_{\widetilde{A}}| = \sum_{j=1}^{n} r_{ij} = |X_k|$, $x_i \in X_k$, we have that

$$\begin{aligned} E(\widetilde{A}) &= \frac{1}{n}\sum_{i=1}^{n}(1 - \frac{|[x_i]_{\widetilde{A}}|}{n}) \\ &= \frac{1}{n}\sum_{k=1}^{m}\sum_{x_i \in X_k}(1 - \frac{|[x_i]_{\widetilde{A}}|}{n}) \\ &= \frac{1}{n}\sum_{k=1}^{m}\sum_{x_i \in X_k}(1 - \frac{|X_k|}{n}) \\ &= \sum_{k=1}^{m}\frac{|X_k|}{n}(1 - \frac{|X_k|}{n}). \end{aligned}$$

This completes the proof. □

In the follows, we define the joint entropy and condition entropy, respectively.

Definition 4. Let $S = (U, \widetilde{A})$ be a hybrid information system, $\widetilde{A}$ hybrid attribute set, and $\widetilde{C}, \widetilde{D}$ two disjoin subsets of $\widetilde{A}$. $[x_i]_{\widetilde{C}}$ and $[x_i]_{\widetilde{D}}$ are fuzzy equivalence classes containing x_i generated by $\widetilde{C}$ and $\widetilde{D}$, respectively. The joint entropy of $\widetilde{C}$ and $\widetilde{D}$ is defined as

$$E(\widetilde{C}, \widetilde{D}) = \frac{1}{n}\sum_{i=1}^{n}(1 - \frac{|[x_i]_{\widetilde{C}} \cap [x_i]_{\widetilde{D}}|}{n}).$$

Definition 5. Let $S = (U, \widetilde{C} \cup \widetilde{D})$ be a hybrid decision table, $\widetilde{C}$ a hybrid condition attribute set and $\widetilde{D}$ a decision attribute set. The conditional entropy of $\widetilde{D}$ with respect to $\widetilde{C}$ is defined as

$$E(\widetilde{D}|\widetilde{C}) = \frac{1}{n^2}\sum_{i=1}^{n}(|[x_i]_{\widetilde{C}}| - |[x_i]_{\widetilde{C}} \cap [x_i]_{\widetilde{D}}|).$$

Theorem 2. Let $S = (U, \widetilde{C} \cup \widetilde{D})$ be a hybrid decision table, $\widetilde{C}$ a hybrid condition attribute set and $\widetilde{D}$ a decision attribute set. The condition entropy and joint entropy satisfy

$$E(\widetilde{D}|\widetilde{C}) = E(\widetilde{C}, \widetilde{D}) - E(\widetilde{C}).$$

Theorem 3. Let $S = (U, \widetilde{C} \cup \widetilde{D})$ be a hybrid decision table, $\widetilde{C}$ a hybrid condition attribute set and $\widetilde{D}$ a decision attribute set. If $\widetilde{C}_1 \subseteq \widetilde{C}_2 \subseteq \widetilde{C}$, then

$$E(\widetilde{D}|\widetilde{C}_1) \geq E(\widetilde{D}|\widetilde{C}_2).$$

Proof. Let $[x_i]_{\widetilde{C}_1}$ and $[x_i]_{\widetilde{C}_2}$ be fuzzy equivalence classes containing x_i generated by $\widetilde{C}_1$ and $\widetilde{C}_2$, respectively. $r^{\widetilde{C}}_{ij}$ denotes the element of ith row and jth column in the fuzzy equivalence matrix $M(R_{\widetilde{C}})$ corresponding to fuzzy attribute set $\widetilde{C}$.

If $r^{\widetilde{C}_2}_{ij} = r^{\widetilde{C}_1}_{ij}$, $\forall i, j \leq n$, then $E(\widetilde{D}|\widetilde{C}_1) = E(\widetilde{D}|\widetilde{C}_2)$. Otherwise, without any of generality, we suppose that $r^{\widetilde{C}_2}_{uv} < r^{\widetilde{C}_1}_{uv}$, $u, v \leq n$. There are four different cases to be discoursed as follows:

(1) $r^{\widetilde{C}_2}_{uv} < r^{\widetilde{C}_1}_{uv} \leq r^{\widetilde{D}}_{uv}$. From the condition, we have $|[x_u]_{\widetilde{C}_1}| - |[x_u]_{\widetilde{C}_2}| = r^{\widetilde{C}_1}_{uv} - r^{\widetilde{C}_2}_{uv}$, and $|[x_u]_{\widetilde{C}_1} \cap [x_u]_{\widetilde{D}}| - |[x_u]_{\widetilde{C}_2} \cap [x_u]_{\widetilde{D}}| = r^{\widetilde{C}_1}_{uv} - r^{\widetilde{C}_2}_{uv}$. Then $|[x_u]_{\widetilde{C}_1}| - |[x_u]_{\widetilde{C}_2}| - (|[x_u]_{\widetilde{C}_1} \cap [x_u]_{\widetilde{D}}| - |[x_u]_{\widetilde{C}_2} \cap [x_u]_{\widetilde{D}}|) = 0$. Similarly, $|[x_v]_{\widetilde{C}_1}| - |[x_v]_{\widetilde{C}_2}| - (|[x_v]_{\widetilde{C}_1} \cap [x_v]_{\widetilde{D}}| - |[x_v]_{\widetilde{C}_2} \cap [x_v]_{\widetilde{D}}|) = 0$. Therefore, $E(\widetilde{D}|\widetilde{C}_1) - E(\widetilde{D}|\widetilde{C}_2) = 0$, i.e., $E(\widetilde{D}|\widetilde{C}_1) = E(\widetilde{D}|\widetilde{C}_2)$.

(2) $r^{\widetilde{D}}_{uv} \leq r^{\widetilde{C}_2}_{uv} < r^{\widetilde{C}_1}_{uv}$. Similarly, one has $|[x_u]_{\widetilde{C}_1}| - |[x_u]_{\widetilde{C}_2}| = r^{\widetilde{C}_1}_{uv} - r^{\widetilde{C}_2}_{uv}$, and $|[x_u]_{\widetilde{C}_1} \cap [x_u]_{\widetilde{D}}| - |[x_u]_{\widetilde{C}_2} \cap [x_u]_{\widetilde{D}}| = r^{\widetilde{D}}_{uv} - r^{\widetilde{D}}_{uv} = 0$. Then $|[x_u]_{\widetilde{C}_1}| - |[x_u]_{\widetilde{C}_2}| - (|[x_u]_{\widetilde{C}_1} \cap [x_u]_{\widetilde{D}}| - |[x_u]_{\widetilde{C}_2} \cap [x_u]_{\widetilde{D}}|) > 0$. Similar to the above result, $|[x_v]_{\widetilde{C}_1}| - |[x_v]_{\widetilde{C}_2}| - (|[x_v]_{\widetilde{C}_1} \cap [x_v]_{\widetilde{D}}| - |[x_v]_{\widetilde{C}_2} \cap [x_v]_{\widetilde{D}}|) > 0$. Hence, $E(\widetilde{D}|\widetilde{C}_1) - E(\widetilde{D}|\widetilde{C}_2) > 0$, i.e.$E(\widetilde{D}|\widetilde{C}_1) > E(\widetilde{D}|\widetilde{C}_2)$.

(3) $r^{\widetilde{C}_2}_{uv} \leq r^{\widetilde{D}}_{uv} < r^{\widetilde{C}_1}_{uv}$. It follows from the existing condition that $|[x_u]_{\widetilde{C}_1}| - |[x_u]_{\widetilde{C}_2}| = r^{\widetilde{C}_1}_{uv} - r^{\widetilde{C}_2}_{uv}$, and $|[x_u]_{\widetilde{C}_1} \cap [x_u]_{\widetilde{D}}| - |[x_u]_{\widetilde{C}_2} \cap [x_u]_{\widetilde{D}}| = r^{\widetilde{D}}_{uv} - r^{\widetilde{C}_2}_{uv}$. Then $|[x_u]_{\widetilde{C}_1}| - |[x_u]_{\widetilde{C}_2}| - (|[x_u]_{\widetilde{C}_1} \cap [x_u]_{\widetilde{D}}| - |[x_u]_{\widetilde{C}_2} \cap [x_u]_{\widetilde{D}}|) > 0$. Similarly, $|[x_v]_{\widetilde{C}_1}| - |[x_v]_{\widetilde{C}_2}| - (|[x_v]_{\widetilde{C}_1} \cap [x_v]_{\widetilde{D}}| - |[x_v]_{\widetilde{C}_2} \cap [x_v]_{\widetilde{D}}|) > 0$. Thus, $E(\widetilde{D}|\widetilde{C}_1) - E(\widetilde{D}|\widetilde{C}_2) > 0$, i.e.$E(\widetilde{D}|\widetilde{C}_1) > E(\widetilde{D}|\widetilde{C}_2)$.

(4) $r^{\widetilde{C}_2}_{uv} < r^{\widetilde{D}}_{uv} \leq r^{\widetilde{C}_1}_{uv}$. From the condition, we have $|[x_u]_{C_1}| - |[x_u]_{C_2}| = r^{C_1}_{uv} - r^{C_2}_{uv}$, and $|[x_u]_{C_1} \cap [x_u]_D| - |[x_u]_{C_2} \cap [x_u]_D| = r^D_{uv} - r^{C_2}_{uv}$. Then $|[x_u]_{C_1}| - |[x_u]_{C_2}| - (|[x_u]_{C_1} \cap [x_u]_D| - |[x_u]_{C_2} \cap [x_u]_D|) > 0$. In like manner, $|[x_v]_{C_1}| - |[x_v]_{C_2}| - (|[x_v]_{C_1} \cap [x_v]_D| - |[x_v]_{C_2} \cap [x_v]_D|) > 0$. Therefore, $E(D|C_1) - E(D|C_2) > 0$, i.e.$E(D|C_1) > E(D|C_2)$.

This completes the proof. □

Theorem 3 states that adding a novel condition attribute into the information system, the condition entropy value will increase monotonously, which reflexes that adding information will lead to enhancement of the discernibility power.

Definition 6. [3] Let $S = (U, \widetilde{C} \cup D)$ be a hybrid decision table, $\widetilde{C}$ a hybrid condition attribute set and D a crisp decision attribute set. $U/D = \{Y_1, Y_2, \cdots, Y_m\}$ is a partition of U generated by D. Then, the lower and upper approximations of the decision D are defined as

$$\underline{C}D = \cup_{j=1}^{m} \{x_i \mid [x_i]_B \subseteq Y_j, x_j \in U\};$$

$$\overline{C}D = \cup_{j=1}^{m} \{x_i \mid [x_i]_B \cap Y_j \neq \emptyset, x_i \in U\}.$$

The lower approximation of decision D also called positive region, denoted as $POS_B(D)$. Let $U' = U - POS_{\widetilde{C}}(D)$, $E_U(D|\widetilde{C})$ denote the condition entropy D with respect to $\widetilde{C}$ in the universe U and $E_{U'}(D|\widetilde{C})$ denote the one in the universe U'. When the universe U becomes U', the change mechanism of conditional entropy will be explicitly explained by following Theorem 4.

Theorem 4. Let $S = (U, \widetilde{C} \cup D)$ be a hybrid decision table, $\widetilde{C}$ a hybrid condition attribute set and D a decision attribute set. If $U' = U - POS_{\widetilde{C}}(D)$, then

$$E_U(D|\widetilde{C}) = \frac{|U'|^2}{n^2} E_{U'}(D|\widetilde{C}).$$

Proof. For similarity, without any of generality, we suppose that $x_1, \ldots, x_p$ belong the positive region and $U' = \{x_{p+1}, \ldots, x_n\}$. From the definition of positive region, we have $r^{\widetilde{C}}_{ij} < r^D_{ij}$, $\forall j \leq n$, if $x_i \in POS_{\widetilde{C}}(D)$. Furthermore, the relationship $E_U(D|\widetilde{C})$ and $E_{U'}(D|\widetilde{C})$ can be obtained as follows

$$\begin{aligned}
E_U(D|\widetilde{C}) &= \frac{1}{n^2}\sum_{i=1}^{n}(|[x_i]_{\widetilde{C}}| - |[x_i]_{\widetilde{C}} \cap [x_i]_D|) \\
&= \frac{1}{n^2}\sum_{i=1}^{p}\sum_{j=1}^{n}(r^{\widetilde{C}}_{ij} - \min\{r^{\widetilde{C}}_{ij}, r^D_{ij}\}) \\
&+ \frac{1}{n^2}\sum_{i=p+1}^{n}\sum_{j=1}^{p}(r^{\widetilde{C}}_{ij} - \min\{r^{\widetilde{C}}_{ij}, r^D_{ij}\}) \\
&+ \frac{1}{n^2}\sum_{i=p+1}^{n}\sum_{j=p+1}^{n}(r^{\widetilde{C}}_{ij} - \min\{r^{\widetilde{C}}_{ij}, r^D_{ij}\}) \\
&= \frac{1}{n^2}\sum_{i=p+1}^{n}\sum_{j=p+1}^{n}(r^{\widetilde{C}}_{ij} - \min\{r^{\widetilde{C}}_{ij}, r^D_{ij}\}) \\
&= \frac{|U'|^2}{n^2} E_{U'}(D|\widetilde{C}).
\end{aligned}$$

This completes the proof. □

Definition 7. Let $S = (U, \widetilde{C} \cup D)$ be a hybrid decision table, $\widetilde{C}$ a hybrid condition attribute set, D a decision attribute set, and $\widetilde{B} \subseteq \widetilde{C}$. The significance $\widetilde{a} \in \widetilde{B}$ is defined as

$$Sig_U(\widetilde{a}, \widetilde{B}, D) = E(D|\widetilde{B}) - E(D|\widetilde{B} \cup \{\widetilde{a}\}).$$

According to Theorem 4, we immediately get Corollary 1 and Corollary 2.

Corollary 1. Let $S = (U, \widetilde{C} \cup D)$ be a hybrid decision table, C a hybrid condition attribute set, D a decision attribute set and $\widetilde{B} \subseteq \widetilde{C}$, the significance

$$Sig_U(\widetilde{a}, \widetilde{B}, D) = \frac{|U'|^2}{n^2} Sig_{U'}(\widetilde{a}, \widetilde{B}, D).$$

Corollary 2. Let $S = (U, \widetilde{C} \cup D)$ be a hybrid decision table, $\widetilde{C}$ a hybrid condition attribute set, D a decision attribute set, and $\widetilde{C}_1 \subseteq \widetilde{C}_2 \subseteq \widetilde{C}$. If $Sig_U(\widetilde{a}, \widetilde{C}_1, D) \leq Sig_U(\widetilde{a}, \widetilde{C}_2, D)$, then

$$Sig_{U'}(\widetilde{a}, \widetilde{C}_1, D) \leq Sig_{U'}(\widetilde{a}, \widetilde{C}_2, D).$$

Corollary 2 states that the sequence of attribute significance is unchangeable after deleting the objects in the positive region.

IV. Reduction algorithms and experimental analysis for hybrid data

In this section, we employ the attribute significance proposed in Section III to construct an algorithm and its accelerated version for attribute reduction. Some experiments are performed for comparing these two algorithms with Hu's algorithm.

First, the algorithm based on new information measure can be expressed as follows.

Algorithm 1. Attribute reduction for hybrid data based on the information measure (ARHIM).

Input: An hybrid decision table $(U, \widetilde{C} \cup D)$, $\widetilde{C} = C^c \cup C^r$, C^c and C^r are categorical and numerical attributes;
Output: One reduct RED.

Step 1: Compute the equivalence relation $R_{\widetilde{a}}$, for all $\widetilde{a} \in \widetilde{C}$;
Step 2: $RED := \emptyset$;
Step 3: $\widetilde{B} := \widetilde{C} - RED$. Compute $Sig(\widetilde{a}_i, RED, D)$ for each $\widetilde{a}_i \in \widetilde{B}$;
Step 4: Select $\widetilde{a}_k \in \widetilde{B}$ which satisfies $Sig(\widetilde{a}_k, RED, D) = \max\{Sig(\widetilde{a}_i, RED, D)\}$;
Step 5: If $Sig(\widetilde{a}_k, RED, D) > 0$, $RED := RED \cup \{\widetilde{a}_k\}$ and go to Step 3, else return RED;
Step 6: end.

The following Algorithm 2 is the accelerated version of Algorithm 1.

Algorithm 2. Accelerated attribute reduction for hybrid data based on the information measure (AARHIM).

Input: An hybrid decision table $(U, \widetilde{C} \cup D)$, $\widetilde{C} = C^c \cup C^r$, C^c and C^r are categorical and numerical attributes, respectively;
Output: One reduct RED.

Step 1: Compute the equivalence relation $R_{\widetilde{a}}$, for all $\widetilde{a} \in \widetilde{C}$;
Step 2: $RED := \emptyset$;
Step 3: $\widetilde{B} := \widetilde{C} - RED$. Compute $U' := U - POS_{\widetilde{B}}(D)$ and $Sig_{U'}(\widetilde{a}_i, RED, D)$, for each $\widetilde{a}_i \in \widetilde{B}$;
Step 4: Select $\widetilde{a}_k \in \widetilde{B}$ which satisfies $Sig_{U'}(\widetilde{a}_k, RED, D) = \max\{Sig_{U'}(\widetilde{a}_i, RED, D)\}$;
Step 5: If $Sig_{U'}(\widetilde{a}_k, RED, D) > 0$, $RED := RED \cup \{\widetilde{a}_k\}$ and go to Step 3, else return RED;
Step 6: end.

In classical rough set theory, a reduct is defined as a subset of attributes which has the same value of information measure as the full attribute set. By means of literature [4], it is not necessarily the case in the fuzzy-rough approaches. Therefore, the algorithms will stop if the condition $Sig_{U'}(\widetilde{a}_k, RED, D) \leq \lambda$ is satisfied, where λ is a degree threshold.

In order to compare ARHIM, AARHIM and the algorithm in [4] (Sequently, called as Hu's algorithm), we use three datasets (Credit, Heart and Wine) from the UCI repository of machine learning databases, whose explicit denotation is in Table 1. We can find that there are some numerical attributes in all of the datasets, and some datasets contain categorical attributes in the same time.

For constructing fuzzy similarity relation matrix, we firstly normalize the numerical attribute $\widetilde{c}_p \in C^r$ into the interval $[0, 1]$ with

$$f(x_i, \widetilde{c}_p) = \frac{f(x_i, \widetilde{c}_p)}{\max\limits_{x_j \in U}\{f(x_j, \widetilde{c}_p)\} - \min\limits_{x_j \in U}(f(x_j, \widetilde{c}_p))}.$$

Furthermore, the value of the fuzzy similarity degree $r_{ij}^{\widetilde{c}_p}$ between objects x_i and x_j with respect to a numerical attribute $\widetilde{c}_p \in C^c$ is computed as

$$r_{ij}^{\widetilde{c}_p} = \begin{cases} 1 - 4 \times |f(x_i, \widetilde{c}_p) - f(x_j, \widetilde{c}_p)|, \\ \text{if} \, |f(x_i, \widetilde{c}_p) - f(x_j, \widetilde{c}_p)| \leq 0.25; \\ \qquad 0, \; otherwise. \end{cases}$$

And, the value of the fuzzy similarity degree $r_{ij}^{\widetilde{c}_q}$ between objects x_i and x_j with a categorical attribute $\widetilde{c}_q \in C^c$ is computed as

$$r_{ij}^{\widetilde{c}_q} = \begin{cases} 1, \; f(x_i, \widetilde{c}_q) = f(x_j, \widetilde{c}_q), \; \forall \widetilde{c}_q \in \widetilde{C}^c; \\ 0, \; otherwise. \end{cases}$$

Thus, the matrix $M(R_{\widetilde{c}_k}) = (r_{ij}^{\widetilde{c}_k})_{n \times n}$ is corresponding to a fuzzy similarity relation $R_{\widetilde{c}_k}$, which is determined by a categorical or numerical attribute $\widetilde{c}_k \in \widetilde{C}$. A fuzzy similarity relation $R_{\widetilde{C}}$ derived from $\widetilde{C}$ can be obtained by the formulation $r_{ij} = \min\limits_{\widetilde{c}_k \in \widetilde{C}}\{r_{ij}^{\widetilde{c}_k}\}$, which is also corresponding to a matrix $M(R_{\widetilde{C}}) = (r_{ij}^{\widetilde{C}})_{n \times n}$. A fuzzy equivalence relation from $R_{\widetilde{C}}$ with max-min transitivity operation [8]. In practice the operation cannot be effectively conducted and we directly search reducts with a similarity relations.

To compare the consuming time resulted from our algorithms with the one performing Hu's algorithm, we divide each of these three datasets into ten parts with equal size. The first part is regarded as the 1st dataset, the combination of the first part and the second part is viewed as the 2nd dataset, the combination of the 2nd dataset and the third part is regarded as the 3rd dataset, $\cdots$, the combination of all ten parts is viewed as the 10th dataset. These datasets can be used to calculate time consumed by ARHIM, AARHIM and Hu's algorithm. These algorithms are run on a personal computer with Windows XP with Pentium D 3.4GHz CPU and 1GB memory. The software is Visual $C\#$.

Tables 2-4 show the comparisons of numbers of selected attributes, classification accuracies and consuming time through using ARHIM, AARHIM and Hu's algorithm with regard to three different datasets in Table 1, where N1 and N2 are the numbers of attributes using our algorithms (ARHIM and AARHIM) and Hu's algorithm, respectively. Accuracy 1, Accuracy 2 and Accuracy 3 are the classification accuracies with the original datasets, the reduced datasets derived from our algorithms (ARHIM and AARHIM) and the ones proceeded by Hu's algorithm, respectively, Time 1, Time 2 and Time 3 indicate the consuming time of ARHIM, AARHIM and Hu's algrithm. The classical classification learning algorithms RBF-SVM is introduced to evaluate the selected attributes. All of the classification accuracies are obtained with 10-fold cross validation. Moreover, the time consuming of reduction for the three datasets are explicitly illustrated in Figures 1-3.

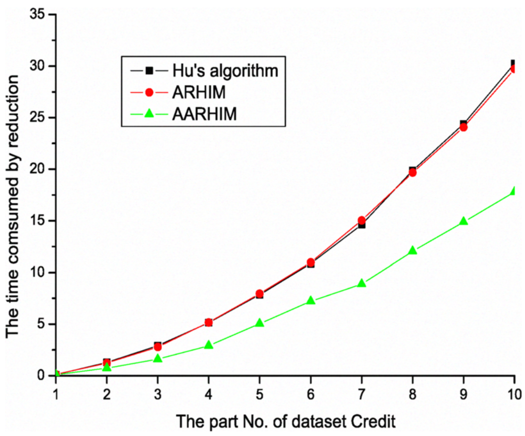

Figure 1: The time consumed by Hu's algorithm, ARHIM and AARHIM with the size of universe (data set Credit)

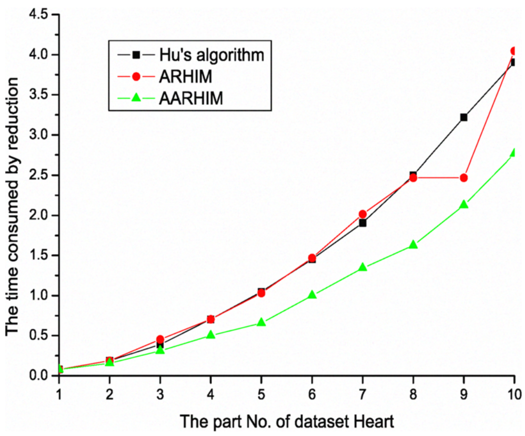

Figure 2: The time consumed by Hu's algorithm, ARHIM and AARHIM with the size of universe (data set Heart)

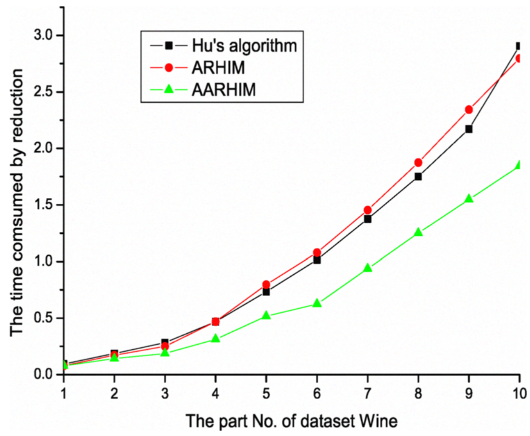

Figure 3: The time consumed by Hu's algorithm, ARHIM and AARHIM with the size of universe (data set Wine)

It is easy to find from Tables 2-4 that the classification accuracies of the datasets through using our algorithms(ARHIM and AARHIM) are similar to those employed Hu's method, which can obtain more precise classification than the original datasets in most cases. Meanwhile, we can note the proposed algorithms delete many unimportant attributes from original data, by which the number of attribute of reduced datasets is same as or a little more than the one using Hu's algorithm. Furthermore, from Tables 2-4 and Figures 1-3, we can obtain that the computing time of each of these three algorithms increases with the increase of the size of data. As one of the important advantages of the AARHIM, as shown in Figures 1-3, we see that the consuming time of the algorithm AARHIM is much less than the original counterpart and Hu's algorithm. Furthermore, the advantage of the algorithm AARHIM will be much better when the size of the dataset increases.

V. Conclusion

In this paper, a new information measure is proposed, which can evaluate the discernibility power of a categorical or numeric attribute. It can overcome the limitations of Pawlak's rough set model that just works in categorical data. Furthermore, the proposed information measure will be degraded to Liang's entropy when a dataset consists only of categorical data. Based on the measure, a representative heuristic algorithm and its accelerated version are presented, respectively. Experimental analyses on three UCI datasets show that an attribute reduct can be obtained by our two new algorithms, and the classification accuracies of reduced datasets are similar with the ones using Hu's algorithm. Above all, the accelerated version (AARHIM) consumes much less computing time than the ARHIM and Hu's algorithm.

Acknowledgement

This work was supported by the national natural science foundation of China (No. 60773133), the national high technology research and development program (No. 2007AA01Z165), the natural science foundation of Shanxi

TABLE I: Data sets description

	Data sets	Samples	Numerical features	Categorical features	Classes
1	Credit	690	6	9	2
2	Heart	270	7	6	2
3	Wine	178	13	0	3

TABLE II: Comparison of attribute number in reducts, classification accuracy and consuming time with Credit

	Feature		Accuracy			Time		
NO.	N1	N2	Accuracy1	Accuracy2	Accuracy3	Time 1	Time 2	Time3
1	1	1	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	0.1250	0.1094	0.1094
2	6	7	0.7099±0.0966	0.7462±0.1029	0.7253±0.1045	1.2344	0.7344	1.2969
3	1	1	0.7627±0.0501	0.7737±0.0268	0.7737±0.0268	2.7813	1.6094	2.9219
4	3	3	0.7992±0.1133	0.7613±0.0200	0.7613±0.0200	5.1719	2.9063	5.1406
5	3	1	0.8059±0.1718	0.8118±0.1757	0.8118±0.1757	7.9688	5.0469	7.8594
6	2	1	0.8363±0.1879	0.8412±0.1911	0.8412±0.1911	10.9688	7.2188	10.8594
7	2	2	0.8602±0.1945	0.8644±0.1968	0.8644±0.1968	15.0313	8.8906	14.6250
8	2	1	0.8318±0.2062	0.8354±0.2084	0.8354±0.2084	19.6719	12.0469	19.8750
9	2	1	0.8440±0.1716	0.8440±0.1716	0.8440±0.1716	24.0625	14.8750	24.3750
10	2	1	0.8548±0.1851	0.8548±0.1851	0.8548±0.1851	29.7188	17.7813	30.2344

TABLE III: Comparison of attribute number in reducts, classification accuracy and consuming time with Heart

	Feature		Accuracy			Time		
NO.	N1	N2	Accuracy1	Accuracy2	Accuracy3	Time 1	Time 2	Time 3
1	1	1	0.8667±0.2331	0.8833±0.1933	0.8833±0.1933	0.0781	0.0781	0.0938
2	5	4	0.7778±0.2572	0.8578±0.1657	0.8467±0.2201	0.1875	0.1563	0.1875
3	8	6	0.8432±0.1339	0.8414±0.1496	0.8414±0.1496	0.4531	0.3125	0.3906
4	6	6	0.8019±0.1629	0.8273±0.1385	0.8273±0.1385	0.7031	0.5000	0.7031
5	8	8	0.8329±0.1036	0.8483±0.1164	0.8483±0.1164	1.0313	0.6563	1.0469
6	9	8	0.8291±0.1073	0.8406±0.0996	0.8354±0.1039	1.4688	1.0000	1.4531
7	6	7	0.8152±0.0873	0.8307±0.0935	0.8307±0.0935	2.0156	1.3438	1.9063
8	8	8	0.8426±0.0488	0.8424±0.0541	0.8424±0.0541	2.4688	1.6250	2.5000
9	9	8	0.8442±0.0688	0.8524±0.0674	0.8362±0.0738	2.4688	2.1250	3.2188
10	5	6	0.8333±0.0531	0.8333±0.0531	0.8296±0.0558	4.0469	2.7656	3.9063

TABLE IV: Comparison of attribute number in reducts, classification accuracy and consuming time with Wine

	Feature		Accuracy			Time		
NO.	N1	N2	Accuracy1	Accuracy2	Accuracy3	Time 1	Time 2	Time 3
1	4	3	0.9000±0.3162	0.9000±0.3162	0.8500±0.3375	0.0781	0.0781	0.0938
2	8	8	0.9750±0.0790	0.9750±0.0790	1.0000±0.0000	0.1719	0.1407	0.1875
3	8	8	0.9800±0.0632	0.9800±0.0632	0.9800±0.0632	0.2500	0.1875	0.2813
4	7	4	0.9857±0.0452	0.9857±0.0452	0.9900±0.0316	0.4688	0.3125	0.4688
5	5	7	0.9778±0.0468	0.9889±0.0351	0.9889±0.0351	0.7969	0.5156	0.7344
6	5	4	0.9909±0.0287	0.9832±0.0355	0.9832±0.0355	1.0781	0.6250	1.0156
7	8	4	0.9833±0.0351	0.9833±0.0351	0.9861±0.0300	1.4531	0.9375	1.3750
8	10	4	0.9867±0.0281	0.9867±0.0281	0.9933±0.0211	1.8750	1.2500	1.7500
9	7	7	0.9938±0.0198	0.9892±0.0231	0.9892±0.0231	2.3438	1.5469	2.1719
10	11	6	0.9889±0.0234	0.9833±0.0268	0.9833±0.0268	2.7968	1.8438	2.9063

province (No. 2008011038, 2009021017-1) and the foundation of key laboratory for computational intelligence and Chinese information processing, ministry of education, China.

REFERENCES

[1] J. Bazan, J. F. Peters, A. Skowron, H. S. Nguyen, and M. Szczuka, *Rough set approach to pattern extraction from classifiers*, Electronic Notes in Theoretical Computer Science, 82(3): 1-10, 2003.

[2] D. Dubois and H. Prade, *Putting fuzzy sets and rough sets together*, In: Slowiniski, R. (Ed.), Intelligent Decision Support, Dordrecht: Kluwer Academic Publisher, 1992.

[3] Q. H. Hu, Z. X. Xie, and D. R. Yu, *Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation*, Pattern Recognition, 40: 3509-3521, 2007.

[4] Q. H. Hu, D. R. Yu, and Z. X. Xie, *Information-preserving hybrid data reduction based on fuzzy-rough techniques*, Pattern Recognition Letters, 27(5): 414-423, 2006.

[5] X. H. Hu and N. Cercone, *Learning in relational databases: a rough set approach*, International Journal of Computational Intelligence, 11(2): 323-338, 1995.

[6] R. Jensen and Q. Shen, *Fuzzy-rough attribute reduction with application to web categorization*, Fuzzy Sets and Systems, 141(3): 469-485, 2004.

[7] M. Kryszkiewicz, *Comparative study of alternative type of knowledge reduction in inconsistent systems*, International Journal of Intelligent Systems, 16: 105-120, 2001.

[8] H. S. Lee, *An optimal algorithm for computing the maxmin transitive closure of a fuzzy similarity matrix*, Fuzzy Sets and Systems, 123(1): 129-136, 2001.

[9] D. Y. Li, B. Zhang, Y. Leung, and H. S. Nguyen, *On knowledge reduction in inconsistent decision information systems*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 12(5): 651-672, 2004.

[10] J. Y. Liang and Y. H. Qian, *Information granules and entropy theory in information systems*, Science in China (Series F), 51(9): 1427-1444, 2008.

[11] J. Y. Liang, Z. Z. Shi, D. Y. Li, and M. J. Wireman, *The information entropy, rough entropy and knowledge granulation in incomplete information systems*, International Journal of General Systems, 34(1): 641-654, 2006.

[12] J. Y. Liang and D. Y. L. Z Z Shi, *The information entropy, rough entropy and knowledge granulation in rough set theory*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 12(1): 37C46, 2004.

[13] T. Y. Lin, *Granular data model: semantic data mining and computing with words* in: Proceeding of IEEE Conference on Fuzzy Systems, 1141-1146, 2004.

[14] J. S. Mi, W. Z.Wu, and X. W. Zhang, *Approaches to knowledge reduction based on variable precision rough set model*, Information Sciences, 159(3-4): 255-272, 2004.

[15] N. N. Morsi and M. M. Yakout, *Axiomatics for fuzzy-rough sets*, Fuzzy Sets and Systems, 100(8): 327-342, 1998.

[16] H. S. Nguyen and D. W. Slezak, *Approximate reducts and association rules correspondence and complexity results*, Lecture Notes in Artificial Intelligence, 1711: 137-145, 1999.

[17] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, London: Kluwer Academic Publisher, 1991.

[18] Y. H. Qian and J. Y. Liang, *Combination entropy and combination granulation in incomplete system*, Lecture Notes in Computer Science, 4062: 184-190, 2006.

[19] Y. H. Qian and J. Y. Liang, *Combination entropy and combination granulation in rough set theory*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 16(2): 179-193, 2008.

[20] Q. Shen and R. Jensen, *Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring*, Pattern Recognition, 37(7): 1351-1363, 2004.

[21] A. Skowron and C. Rauszer, *The discernibility matrices and functions in information systems*, In: R. Slowinski(Eds),Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht: Kluwer Academic Publisher, 1992.

[22] G. Y. Wang, *Rough reduction in algebra view and information view*, International Journal of Intelligent Systems, 18: 679-688, 2003.

[23] W. Z. Wu and W. X. Zhang, *Constructive and axiomatic approaches of fuzzy approximation operators*, Information Sciences, 159(3-4): 233-254, 2004.

[24] Y. Y. Yao, *Decision-theoretic rough set models*, Lecture Notes in Artificial Intelligence, 4481: 1-12, 2007.

[25] Y. Y. Yao and Y. Zhao, *Attribute reduction in decisiontheoretic rough set models*, Information Sciences, 178: 3356-3373, 2008.

[26] W. Ziarko, *Variable precision rough set model*, Journal of Computer and System Science, 46(1): 39-59, 1993.