The Relationship Among Several Knowledge Reduction Approaches

Keyun Qin[1], Zheng Pei[2], and Weifeng Du[1]

[1] Department of Applied Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 610031, China
keyunqin@263.net
[2] College of Computers & Mathematical-Physical Science, Xihua University, Chengdu, Sichuan, 610039, China
pqyz@263.net

Abstract. This paper is devoted to the discussion of the relationship among some reduction approaches of information systems. It is proved that the distribution reduction and the entropy reduction are equivalent, and each distribute reduction is a d reduction. Furthermore, for consistent information systems, the distribution reduction, entropy reduction, maximum distribution reduction, distribute reduction, approximate reduction and d reduction are all equivalent.

1 Introduction

Rough set theory(RST), proposed by Pawlak [1], [2], is an extension of set theory for the study of intelligent systems characterized by insufficient and incomplete information. The successful application of RST in a variety of problems have amply demonstrated its usefulness. One important application of RST is the knowledge discovery in information system (decision table). RST operates on an information system which is made up of objects for which certain characteristics (i.e., condition attributes) are known. Objects with the same condition attribute values are grouped into equivalence classes or condition classes. The objects are each classified to a particular category with respect to the decision attribute value, those classified to the same category are in the same decision class. Using the concepts of lower and upper approximations in RST, the knowledge hidden in the information system may be discovered.

One fundamental aspect of RST involves the searching for some particular subsets of condition attributes. By such one subset the information for classification purpose provides is equivalent to (according to a particular standard) the condition attribute set done. Such subsets are called reducts. To acquire brief decision rules from information systems, knowledge reduction is needed.

Knowledge reduction is performed in information systems by means of the notion of a reduct based on a specialization of the general notion of independence due to Marczewski [3]. In recent years, more attention has been paid to knowledge reduction in information systems in rough set research. Many types

L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3613, pp. 1232–1241, 2005.

of knowledge reduction and their applications have been proposed for inconsistent information systems in the area of rough sets [4], [5], [6], [7], [8], [9], [10], [11], [12]. The first knowledge reduction approach due to[6] which is carry out through discernibility matrixes and discernibility functions. This kind of reduction is based on the positive region of the universe and we call it d reduction. For inconsistent information systems, Kryszkiewicz [7] proposed the concepts of distribution reduction and distribute reduction.Zhang [5] proposed the concepts of maximum distribution reduction and approximate reduction and provide new approaches to knowledge reduction in inconsistent information systems. Furthermore, some approaches to knowledge reduction based on variable precision rough set model were proposed [13]. Information entropy is a measure of information involved in a system. Based on conditional information entropy, some knowledge reduction approaches in information systems were proposed in [4].

This paper is devoted to the discussion of the relationship among some reduction approaches of information systems. It is proved that the distribution reduction and the entropy reduction are equivalent, and each distribute reduction is a d reduction. Furthermore, for consistent information systems, the distribution reduction, entropy reduction, maximum distribution reduction, distribute reduction, approximate reduction and d reduction are all equivalent.

2 Preliminaries and Notations

An information system is a quadruple $S = (U, AT \cup \{d\}, V, f)$, where

(1) U is a non-empty finite set and its elements are called objects of S.
(2) AT is the set of condition attributes and d is the decision attribute of S.
(3) $V = \cup_{q \in AT \cup \{d\}} V_q$, where V_q is a non-empty set of values of attribute $q \in AT \cup \{d\}$, called domain of the attribute q.
(4) $f : U \to AT \cup \{d\}$ is a mapping, called description function of S, such that $f(x, q) \in V_q$ for each $(x, q) \in U \times (AT \cup \{d\})$.

Let $S = (U, AT \cup \{d\}, V, f)$ be an information system and $A \subseteq AT$. The discernibility relation $ind(A)$ on U derived from A, defined by $(x, y) \in ind(A)$ if and only if $\forall a \in A, f(x, a) = f(y, a)$, is an equivalent relation and hence $(U, ind(A))$ is a Pawlak approximation space. We denote by $[x]_A$ the equivalent class with respect to $ind(A)$ that containing x and U/A the set of these equivalent classes. For each $X \subseteq U$, according to Pawlak [1], the upper approximation $\overline{A}(X)$ and lower approximation $\underline{A}(X)$ of X with respect to A are defined as

$$\overline{A}(X) = \{x \in U | [x]_A \cap X \neq \emptyset\}, \quad \underline{A}(X) = \{x \in U | [x]_A \subseteq X\}. \tag{1}$$

Based on the approximation operators, Skowron proposed the concept of d reduction of an information system.

Definition 1. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system and $A \subseteq AT$. The positive region $pos_A(d)$ of d with respect to A is defined as*

$$pos_A(d) = \cup_{X \in U/d} \underline{A}(X).$$

Definition 2. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system and $A \subseteq AT$. A is called a d consistent subset of S if $pos_A(d) = pos_{AT}(d)$. A is called a d reduction of S if A is a d consistent subset of S and each proper subset of A is not a d consistent subset of S.*

All the d reductions can be carry out through discernibility matrixes and discernibility functions [6]. Let $S = (U, AT \cup \{d\}, V, f)$ be an information system and $B \subseteq AT$, $x \in U$. We introduce the following notations:

$$U/d = \{D_1, \cdots, D_r\}; \quad \mu_B(x) = (D(D_1/[x]_B), \cdots, D(D_r/[x]_B));$$
$$\gamma_B(x) = \{D_j; D(D_j/[x]_B) = max_{q \leq r} D(D_q/[x]_B)\}; \delta_B(x) = \{D_j; D_j \cap [x]_B \neq \emptyset\};$$
$$\eta_B = \frac{1}{|U|} \Sigma_{j=1}^{r} |\overline{B}(D_j)|;$$

where $D(D_j/[x]_B) = \frac{|D_j \cap [x]_B|}{|[x]_B|}$ is the include degree of $[x]_B$ in D_j.

For inconsistent information systems, Kryszkiewicz [7] proposed the concepts of distribution reduction and distribute reduction. Based on this work, Zhang [5] proposed the concepts of maximum distribution reduction and approximate reduction. Farther more, the judgement theorems and discernibility matrixes with respect to those reductions are obtained. These reductions are based on the concept of include degree.

Definition 3. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system, $A \subseteq AT$.*

(1) A is called a distribution consistent set of S if $\mu_A(x) = \mu_{AT}(x)$ for each $x \in U$. A is called a distribution reduction of S if A is a distribution consistent set of S and no proper subset of A is distribution consistent set of S.
(2) A is called a maximum distribution consistent set of S if $\gamma_A(x) = \gamma_{AT}(x)$ for each $x \in U$. A is called a maximum distribution reduction of S if A is a maximum distribution consistent set of S and no proper subset of A is maximum distribution consistent set of S.
(3) A is called a distribute consistent set of S if $\delta_A(x) = \delta_{AT}(x)$ for each $x \in U$. A is called a distribute reduction of S if A is a distribute consistent set of S and no proper subset of A is distribute consistent set of S.
(4) A is called a approximate consistent set of S if $\eta_A = \eta_{AT}$. A is called a approximate reduction of S if A is a approximate consistent set of S and no proper subset of A is approximate consistent set of S.

[5] proved that the concepts of distribute consistent set and approximate consistent set are equivalent, a distribution consistent set must be a distribute consistent set and a maximum distribution consistent set.

Let $S = (U, AT \cup \{d\}, V, f)$ be an information system, $A \subseteq AT$ and

$$U/AT = \{X_i; 1 \leq i \leq n\}, \quad U/A = \{Y_j; 1 \leq j \leq m\}, \quad U/d = \{Z_l; 1 \leq l \leq k\}.$$

The conditional information entropy $H(d|A)$ of d with respect to A is defined

$$H(d|A) = -\sum_{j=1}^{m} (p(Y_j) \cdot \sum_{l=1}^{k} p(Z_l|Y_j) log(p(Z_l|Y_j))),$$

where $p(Y_j) = \frac{|Y_j|}{|U|}$, $p(Z_l|Y_j) = \frac{|Z_l \cap Y_j|}{|Y_j|}$ and $0log0 = 0$.

Based on conditional information entropy, Wang [4] proposed the concept of entropy reduction for information systems.

Definition 4. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system and $A \subseteq AT$. A is called an entropy consistent set of S if $H(d|A) = H(d|AT)$. A is called an entropy reduction of S if A is an entropy consistent set of S and non proper subset of A is entropy consistent set of S.*

3 The Relationship Among Knowledge Reduction Approaches

In this section, we discuss the relationship among knowledge reduction approaches. In what follows we assume that $S = (U, AT \cup \{d\}, V, f)$ is an information system, $A \subseteq AT$ and

$$U/AT = \{X_i; 1 \leq i \leq n\}, \quad U/A = \{Y_j; 1 \leq j \leq m\}, \quad U/d = \{Z_l; 1 \leq l \leq k\}.$$

Theorem 5. *Let $Y_j = \cup_{t \in T_j} X_t$, $1 \leq j \leq m$, where T_j is an index set. For each $1 \leq l \leq k$,*

$$|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) \leq \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}).$$

Proof. Let $T_{j1} = \{X_t; t \in T_j, Z_l \cap X_t \neq \emptyset\}$. If $T_{j1} = \emptyset$, then $Z_l \cap X_t = \emptyset$ for each $t \in T_j$ and hence $Z_l \cap Y_j = \emptyset$, the conclusion holds. If $T_{j1} \neq \emptyset$, by $lnx \leq x - 1$, it follows that

$$\begin{aligned}
&|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) - \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}) \\
&= \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) - \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}) \\
&= \sum_{t \in T_{j1}} |Z_l \cap X_t| log(\frac{|Z_l \cap Y_j||X_t|}{|Y_j||Z_l \cap X_t|}) \\
&\leq \sum_{t \in T_{j1}} |Z_l \cap X_t| (\frac{|Z_l \cap Y_j||X_t|}{|Y_j||Z_l \cap X_t|} - 1) loge \\
&= \frac{loge}{|Y_j|} \sum_{t \in T_{j1}} (|Z_l \cap Y_j||X_t| - |Y_j||Z_l \cap X_t|) \\
&= \frac{loge}{|Y_j|} (\sum_{t \in T_{j1}} (|Z_l \cap Y_j||X_t| - \sum_{t \in T_{j1}} |Y_j||Z_l \cap X_t|) \\
&= \frac{loge}{|Y_j|} (|Z_l \cap Y_j|(|Y_j| - \sum_{t \in T_j - T_{j1}} |X_t|) - |Y_j||Z_l \cap Y_j|) \leq 0.
\end{aligned}$$

Theorem 6. *$H(d|AT) = H(d|A)$ if and only if*

$$|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) = \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}),$$

for each $1 \leq l \leq k$ and $1 \leq j \leq m$, where $Y_j = \cup_{t \in T_j} X_t$, $1 \leq j \leq m$, and T_j is an index set.

Proof.

$$H(d|AT) = -\sum_{i=1}^{n}(p(X_i) \cdot \sum_{l=1}^{k} p(Z_l|X_i) log(p(Z_l|X_i)))$$

$$= -\frac{1}{|U|} \sum_{l=1}^{k} \sum_{i=1}^{n} (p(X_i)|Z_l \cap X_i| log(\frac{|Z_l \cap X_i|}{|X_i|}),$$

$$H(d|A) = -\sum_{j=1}^{m}(p(Y_j) \cdot \sum_{l=1}^{k} p(Z_l|Y_j) log(p(Z_l|Y_j)))$$

$$= -\frac{1}{|U|} \sum_{l=1}^{k} \sum_{j=1}^{m} (p(X_i)|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}).$$

The sufficiency is trivial because each A equivalent class is just a union of some AT equivalent classes.

Necessity: For each $1 \leq l \leq k$ and $1 \leq j \leq m$, by Theorem 5,

$$|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) \leq \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}),$$

and hence

$$\sum_{j=1}^{m} |Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) \leq \sum_{i=1}^{n} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}).$$

If there exists $1 \leq l \leq k$ and $1 \leq j \leq m$ such that

$$|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) < \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}).$$

consequently,

$$\sum_{j=1}^{m} |Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) < \sum_{i=1}^{n} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}).$$

and hence $H(d|AT) < H(d|A)$, a contradiction.

Theorem 7. *$A \subseteq AT$ is a distribution reduction of S if and only if A is an entropy reduction of S.*

Proof. It needs only to prove that A is a distribution consistent set if and only if A is an entropy consistent set.

Sufficiency: Assume that A is a distribution consistent set. For each $1 \leq l \leq k$ and $1 \leq j \leq m$, let $Y_j = [x]_A$. We notice that $J_{x,A} = \{[y]_{AT}; [y]_{AT} \subseteq [x]_A\}$ is a partition of $[x]_A$. By $|Z_l \cap [x]_A| = \sum_{[y]_{AT} \in J_{x,A}} |Z_l \cap [y]_{AT}|$, it follows that

$$\frac{|Z_l \cap [y]_{AT}|}{|[y]_{AT}|} = \frac{|Z_l \cap [y]_A|}{|[y]_A|} = \frac{|Z_l \cap [x]_A|}{|[x]_A|},$$

for each $[y]_{AT} \in J_{x,A}$ and hence

$$|Z_l \cap [x]_A| log(\frac{|Z_l \cap [x]_A|}{|[x]|_A}) = \sum_{[y]_{AT} \in J_{x,A}} |Z_l \cap [y]_{AT}| log(\frac{|Z_l \cap [y]_{AT}|}{|[y]_{AT}|}),$$

it follows by Theorem 6 that $H(d|AT) = H(d|A)$ and A is an entropy consistent set.

Necessity: Assume that A is an entropy consistent set. For each $x \in U$, $J_{x,A} = \{[y]_{AT}; [y]_{AT} \subseteq [x]_A\}$ forms a partition of $[x]_A$. Let $J = J_{x,A} - \{[x]_{AT}\}$. For each $1 \leq l \leq k$, by Theorem 6, it follows that

$$|Z_l \cap [x]_A| log(\frac{|Z_l \cap [x]_A|}{|[x]_A|}) = \sum_{[y]_{AT} \in J_{x,A}} |Z_l \cap [y]_{AT}| log(\frac{|Z_l \cap [y]_{AT}|}{|[y]_{AT}|}).$$

and hence

$$\sum_{[y]_{AT} \in J_{x,A}} |Z_l \cap [y]_{AT}| log(\frac{|Z_l \cap [x]_A|}{|[x]_A|}) = \sum_{[y]_{AT} \in J_{x,A}} |Z_l \cap [y]_{AT}| log(\frac{|Z_l \cap [y]_{AT}|}{|[y]_{AT}|}),$$

$$\sum_{[y]_{AT} \in J_{x,A}} |Z_l \cap [y]_{AT}| log(\frac{|Z_l \cap [x]_A||[y]_{AT}|}{|[x]_A||Z_l \cap [y]_{AT}|}) = 0,$$

that is,

$$\begin{aligned}
&-|Z_l \cap [x]_{AT}| log(\frac{|Z_l \cap [x]_A||[x]_{AT}|}{|[x]_A||Z_l \cap [x]_{AT}|}) \\
&= \sum_{[y]_{AT} \in J} |Z_l \cap [y]_{AT}| log(\frac{|Z_l \cap [x]_A||[y]_{AT}|}{|[x]_A||Z_l \cap [y]_{AT}|}) \\
&\leq \sum_{[y]_{AT} \in J} |Z_l \cap [y]_{AT}| (\frac{|Z_l \cap [x]_A||[y]_{AT}|}{|[x]_A||Z_l \cap [y]_{AT}|} - 1) loge \\
&= \frac{loge}{|[x]_A|} \sum_{[y]_{AT} \in J} (|Z_l \cap [x]_A||[y]_{AT}| - |[x]_A||Z_l \cap [y]_{AT}|) \\
&= \frac{loge}{|[x]_A|} (|Z_l \cap [x]_A|(|[x]_A| - |[x]_{AT}|) - |[x]_A|(|Z_l \cap [x]_A| - |Z_l \cap [x]_{AT}|)) \\
&= \frac{loge}{|[x]_A|} (|[x]_A||Z_l \cap [x]_{AT}| - |[x]_{AT}||Z_l \cap [x]_A|).
\end{aligned}$$

It follows that

$$ln(\frac{|Z_l \cap [x]_A||[x]_{AT}|}{|[x]_A||Z_l \cap [x]_{AT}|}) \geq \frac{|Z_l \cap [x]_A||[x]_{AT}|}{|[x]_A||Z_l \cap [x]_{AT}|} - 1.$$

By $lna \leq a - 1$ for each $a > 0$,

$$ln(\frac{|Z_l \cap [x]_A||[x]_{AT}|}{|[x]_A||Z_l \cap [x]_{AT}|}) = \frac{|Z_l \cap [x]_A||[x]_{AT}|}{|[x]_A||Z_l \cap [x]_{AT}|} - 1,$$

and hence

$$\frac{|Z_l \cap [x]_A||[x]_{AT}|}{|[x]_A||Z_l \cap [x]_{AT}|} = 1,$$

because $a = 1$ is the unique root of $lna = a - 1$, that is

$$\frac{|[x]_{AT} \cap Z_l|}{|[x]_{AT}|} = \frac{|[x]_A \cap Z_l|}{|[x]_A|}$$

and A is a distribution consistent set.

Theorem 8. *A is an entropy consistent set if and only if*

$$\frac{|[x]_{AT} \cap [x]_d|}{|[x]_{AT}|} = \frac{|[x]_A \cap [x]_d|}{|[x]_A|}$$

for each $x \in U$.

Proof. By Theorem 7, the necessity is trivial.

Sufficiency: We prove that

$$|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) = \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}),$$

for any $1 \leq l \leq k$ and $1 \leq j \leq m$ and finish the proof by Theorem 6, where $Y_j = \cup_{t \in T_j} X_t$.

If $Z_l \cap Y_j = \emptyset$, then $Z_l \cap X_t = \emptyset$ for each $t \in T_j$ and the conclusion holds.

If $Z_l \cap Y_j \neq \emptyset$, suppose that $x \in Z_l \cap Y_j$, it follows that $Z_l = [x]_d$ and $Y_j = [x]_A$. Let $J_{x,A} = \{[z]_{AT}; [z]_{AT} \subseteq [x]_A\}$. Consequently,

$$\begin{aligned}
& |Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) - \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}) \\
&= |[x]_d \cap [x]_A| log(\frac{|[x]_d \cap [x]_A|}{|[x]_A|}) - \sum_{[z]_{AT} \in J_{x,A}} |[x]_d \cap [z]_{AT}| log(\frac{|[x]_d \cap [z]_{AT}|}{|[z]_{AT}|}) \\
&= \sum_{[z]_{AT} \in J_{x,A}} |[x]_d \cap [z]_{AT}| log(\frac{|[x]_d \cap [x]_A|}{|[x]_A|}) - \sum_{[z]_{AT} \in J_{x,A}} |[x]_d \cap [z]_{AT}| log(\frac{|[x]_d \cap [z]_{AT}|}{|[z]_{AT}|}) \\
&= \sum_{[z]_{AT} \in J_{x,A}} |[x]_d \cap [z]_{AT}| log(\frac{|[x]_d \cap [x]_A||[z]_{AT}|}{|[x]_A||[x]_d \cap [z]_{AT}|}).
\end{aligned}$$

Assume that $u \in [x]_d \cap [z]_{AT}$, it follows that $u \in [x]_A$ and hence $[x]_d = [u]_d$, $[z]_{AT} = [u]_{AT}$ and $[x]_A = [u]_A$, consequently,

$$\frac{|[x]_d \cap [x]_A||[z]_{AT}|}{|[x]_A||[x]_d \cap [z]_{AT}|} = \frac{|[u]_d \cap [u]_A||[u]_{AT}|}{|[u]_A||[u]_d \cap [u]_{AT}|} = 1,$$

and hence

$$|Z_l \cap Y_j| log(\frac{|Z_l \cap Y_j|}{|Y_j|}) = \sum_{t \in T_j} |Z_l \cap X_t| log(\frac{|Z_l \cap X_t|}{|X_t|}).$$

Theorem 9. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system and $A \subseteq AT$. If A is a distribute consistent set of S, then A is a d consistent set of S.*

Proof. Let $A \subseteq AT$ be a distribute consistent set of S and $U/d = \{D_1, D_2, \cdots, D_r\}$. For each $x \in pos_{AT}(d)$, it follows that $[x]_{AT} \subseteq [x]_d$ and hence $\delta_{AT}(x) = \{D_j; D_j \cap [x]_{AT} \neq \emptyset\} = \{[x]_d\} = \delta_A(x)$. Assume that $[x]_d = D_j$, it follows that $D_l \cap [x]_A = \emptyset$ for each $l \leq r, l \neq j.$, that is $[x]_A \subseteq D_j = [x]_d$ and $x \in \underline{A}([x]_d) \subseteq pos_A(d)$, it follows that $pos_{AT}(d) \subseteq pos_A(d)$.

$pos_A(d) \subseteq pos_{AT}(d)$ is trivial.

4 Knowledge Reduction for Consistent Information Systems

An information system $S = (U, AT \cup \{d\}, V, f)$ is called to be consistent, if $[x]_{AT} \subseteq [x]_d$ for each $x \in U$. In this section, we discuss knowledge reductions for consistent information systems. We will prove that the concepts of distribution reduction, approximate reduction and d reduction are equivalent for consistent information systems.

Theorem 10. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system. S is consistent if and only if $pos_{AT}(d) = U$.*

Proof. If S is consistent, then $[x]_{AT} \subseteq [x]_d$ for each $x \in U$ and hence $x \in \underline{AT}([x]_d) \subseteq \cup_{X \in U/d} \underline{AT}(X) = pos_{AT}(d)$, that is $pos_{AT}(d) = U$.

If $pos_{AT}(d) = U$, then $x \in pos_{AT}(d)$ for each $x \in U$ and hence $x \in \underline{AT}([x]_d)$, that is $[x]_{AT} \subseteq [x]_d$.

Theorem 11. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system. S is consistent if and only if $\delta_{AT}(x) = \{[x]_d)\}$ for each $x \in U$.*

Proof. If S is consistent, then $[x]_{AT} \subseteq [x]_d$ for each $x \in U$ and hence $\delta_{AT}(x) = \{[x]_d)\}$.

If $\delta_{AT}(x) = \{[x]_d)\}$ for each $x \in U$, then $[x]_{AT} \cap [y]_d = \emptyset$ for each $[y]_d \neq [x]_d$, that is $[x]_{AT} \subseteq [x]_d$ and S is consistent.

Theorem 12. *Let $S = (U, AT \cup \{d\}, V, f)$ be an information system. S is consistent if and only if $H(d|AT) = 0$.*

Proof. Assume that

$$U/AT = \{X_1, X_2, \cdots, X_n\}, \quad U/d = \{Y_1, Y_2, \cdots, Y_m\}.$$

It follows that

$$\begin{aligned} H(d|AT) &= -\sum_{i=1}^{n}(p(X_i) \cdot \sum_{j=1}^{m} p(Y_j|X_i)log(p(Y_j|X_i))) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|Y_j \cap X_i|}{|U|}log(\frac{|Y_j \cap X_i|}{|X_i|}). \end{aligned}$$

If S is consistent, then for each $i(1 \leq i \leq n)$, there exists unique $j(1 \leq j \leq m)$ such that $X_i \subseteq Y_j$, and hence $\frac{|Y_j \cap X_i|}{|X_i|} = 1$ or $\frac{|Y_j \cap X_i|}{|X_i|} = 0$, consequently, $H(d|AT) = 0$.

If $H(d|AT) = 0$, then

$$-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|Y_j \cap X_i|}{|U|}log(\frac{|Y_j \cap X_i|}{|X_i|}) = 0,$$

it follows that

$$\sum_{j=1}^{m}\frac{|Y_j \cap X_i|}{|U|}log(\frac{|Y_j \cap X_i|}{|X_i|}) = 0,$$

for each $i(1 \leq i \leq n)$, that is there exists $j(1 \leq j \leq m)$ such that $X_i \subseteq Y_j$, and S is consistent.

Theorem 13. *Let $S = (U, AT \cup \{d\}, V, f)$ be a consistent information system and $A \subseteq AT$.*

(1) A is an entropy consistent set if and only if $S' = (U, A \cup \{d\}, V, f)$ is consistent.
(2) A is a approximate consistent set if and only if $S' = (U, A \cup \{d\}, V, f)$ is consistent.
(3) A is a positive domain consistent set if and only if $S' = (U, A \cup \{d\}, V, f)$ is consistent.

By this Theorem, for consistent information systems, the concepts of distribution reduction, entropy reduction, maximum distribution reduction, distribute reduction, approximate reduction and d reduction are all equivalent.

Acknowledgements

The authors are grateful to the referees for their valuable comments and suggestions. This work has been supported by the National Natural Science Foundation of China (Grant No. 60474022).

References

1. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Science. **11** (1982) 341–356
2. Pawlak, Z.(ed.): Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Boston (1991)
3. Marczewski.: A General Scheme of Independence in Mathematics. Bulletin de L Academie Polonaise des Sciences–Serie des Sciences Mathematiques Astronomiques et Physiques. **6** (1958) 731–736
4. Wang, G, Y., Yu, H., Yang, D, C.: Decision Table Reduction Based on Conditional Information Entropy. Chinese Journal of Computers (in Chinese). **25** (2002) 759–766
5. Zhang, W, X., Mi, J, S., Wu, W, Z.: Knowledge Reductions in Inconsistent Information Systems. Chinese Journal of Computers (in Chinese). **26** (2003) 12–18
6. Skowron, A., Rauszer, C.: The Discernibility Matrices And Functions in Information System. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, Dordrecht (1992)
7. Kryszkiewicz.: Comparative Study of Alternative Type of Knowledge Reduction in Inconsistent Systems. International Journal of General Systems. **16** (2001) 105–120
8. Beynon, M.: Reducts within The Variable Precision Rough Set Model: A Further Investigation. European Journal of Operational Research. **134** (2001) 592–605
9. Quafatou, M.: RST: A Generalization of Rough Set Theory. Information Sciences. **124** (2000) 301–316
10. Zheng, P., Keyun, Q.: Obtaining Decision Rules And Combining Evidence Based on Modal Logic. Progress in Natural Science (in chinese). **14** (2004) 501–508
11. Zheng, P., Keyun, Q.: Intuitionistic Special Set Expression of Rough Set And Its Application in Reduction of Attributes. Pattern Recognition and Artificial Intelligence(in chinese). **17** (2004) 262–266
12. Slowinski, R., Zopounidis, Dimitras, A. I.: Prediction of Company Acquisition in Greece by Means of The Rough Set Approach. European Journal of Operational Research. **100** (1997) 1–15
13. Jusheng, M., Weizhi, W., Wenxiu, Z.: Approaches to Knowledge Reduction Based on Variable Precision Rough Set Model. Information Sciences. **159** (2004) 255–272