

Correlation of Attributes in Random Decision Information Systems and Its Application to Attributes Reduction

Xia Wu^{1,a}, Jialu Zhang^{1,b} and Jiming Zhong^{2,c}

¹College of Mathematics and Finance Xiangnan University Chenzhou, China

²College of Economic and Management, Xiangnan University Chenzhou, China

^awuxia351@163.com, ^bzjl0735@163.com, ^cjmzhongcn@163.com

Keywords: Random information systems; information entropy; relativity; knowledge reduction; algorithm.

Abstract. The conditional information entropy of a condition attributes set with respect to another condition attributes set is defined in a random decision information system, by using the conditional information entropy the correlation of a conditional attributes set relative to another condition attributes set with respect to decision attributes set, the properties of correlation are discussed, and some concepts about random decision information system, such as necessary attribute, consistent attribute and attributes nuclear, etc, are described by using the correlation. By using the correlation of attributes an attribute reduction method and an attribute reduction algorithm are proposed respectively, and an example shows that the algorithm is effective.

Introduction

A decision rule in a decision information system is a combination of conditional attribute values and decision attribute values connected by logical connectors. Decision rule extraction is the main research content of decision information system. In general, different conditional attributes play different roles in decision classification. Some conditional attributes are absolutely necessary. Removing these attributes will inevitably affect decision classification. These attributes are usually referred to as core attributes. Some conditional attributes are absolutely unnecessary for decision classification. Removing these conditional attributes does not affect decision classification. Some attributes are relatively necessary, which can be combined with core attributes to determine decision classification. Hence in order to get a simple decision rule the conditional attributes of the decision information system must be reduced before the decision rules are extracted. The essence of attribute reduction of a decision information system is to find a minimum set of attributes in a conditional attribute set. It can completely determine the classification of decision making, that is, the decision classification determined by the minimum condition attribute set is the same as the attribute set of all conditions [1-6]. It has been proved that finding the minimum reduction in an information system is a NP complete problem [7].

To use information entropy for attribute reduction is an important heuristic attribute reduction algorithm. Miao introduced shannon information entropy in an information system and proposed attribute reduction algorithm based on mutual information [8,9]. Wang proposed attribute reduction algorithm based on conditional information entropy [10,11]. Liang defined the concept of information measure in information system, and proposed an attribute reduction algorithm based on information measure [12]. Zhang studies the uncertainty measure of attribute and attribute set of random information system and random fuzzy information system, uses uncertainty measure to carry out attribute reduction [6]. The concept of attribute correlation of random information system is introduced, and attribute reduction is applied to random information system based on attribute correlation [14,15].

This paper finds the attribute reduction method of random decision information system as the research purpose. We introduce a conditional information entropy of a attribute set with respect to another attribute set in a random decision information system. Using conditional information entropy we introduce a correlation $r(B/A)$ of a conditional attribute subset B with respect to another conditional subset A about decision attribute, and discuss the nature of the correlation. Based on the correlation, the concepts of necessary attributes, consistent attributes and core attributes of decision information systems are described and depicted. We also propose a method of attribute reduction based on attribute correlation, and give an attribute reduction algorithm based on this method. A practical example is given to show that the algorithm proposed in this paper is effective.

Properties of Attribute Information Measures in Information Systems

A decision information system, which is also called a decision tables, is a quarternary form $S=(U,C\cup D,V,f)$, where U is an object set, C and D are condition and decision attribute sets respectively, $V = \mathbf{U}_{a \in C \cup D} V^{(a)}$ is a attribute value set, $V^{(a)}$ is the value set of attribute a , $f:U \times (C \cup D) \rightarrow V$ is an information function, i.e, $\forall x \in U, a \in C \cup D, f(x,a) \in V^{(a)}$. If there is a normal probability measure P on U then it is called a random decision information system, where a normal probability measure P means that $\forall x \in U, P(\{x\}) > 0, \sum_{x \in U} P(\{x\}) = 1$.

In a random decision information system, for $a \in C \cup D, V^{(a)} = \{v_1^{(a)}, v_2^{(a)}, \mathbf{L}, v_m^{(a)}\}$, we can give the following probability distribution:

$$\begin{array}{c|cccc} a & v_1^{(a)} & v_2^{(a)} & \mathbf{L} & v_m^{(a)} \\ \hline p^{(a)} & p(v_1^{(a)}) & p(v_2^{(a)}) & \mathbf{L} & p(v_m^{(a)}) \end{array}$$

where $p(v_i^{(a)}) = P(\{x | f(x,a) = v_i^{(a)}\})$ is the probability of attribution a taking value $v_i^{(a)}$.

If we denote $p(v_i^{(a)}, v_j^{(b)}) = P(\{x | f(x,a) = v_i^{(a)} \text{ and } f(x,b) = v_j^{(b)}\}) (i=1,2,\mathbf{L},m, j=1,2,\mathbf{L},q)$, $p(v_j^{(b)} / v_i^{(a)}) = \frac{p(v_i^{(a)}, v_j^{(b)})}{p(v_i^{(a)})} (i=1,\mathbf{L},m, j=1,\mathbf{L},q)$, then we can give the joint probability distribution $P(ab)$ and the conditional probability distribution $P(b/a)$ of attributes a and b as follows:

	$v_1^{(b)}$	$v_2^{(b)}$	\mathbf{L}	$v_q^{(b)}$		$v_1^{(b)}$	$v_2^{(b)}$	\mathbf{L}	$v_q^{(b)}$
$v_1^{(a)}$	$p(v_1^{(a)}, v_1^{(b)})$	$p(v_1^{(a)}, v_2^{(b)})$	\mathbf{L}	$p(v_1^{(a)}, v_q^{(b)})$	$v_1^{(a)}$	$p(v_1^{(b)} / v_1^{(a)})$	$p(v_2^{(b)} / v_1^{(a)})$	\mathbf{L}	$p(v_q^{(b)} / v_1^{(a)})$
$v_2^{(a)}$	$p(v_2^{(a)}, v_1^{(b)})$	$p(v_2^{(a)}, v_2^{(b)})$	\mathbf{L}	$p(v_2^{(a)}, v_q^{(b)})$	$v_2^{(a)}$	$p(v_1^{(b)} / v_2^{(a)})$	$p(v_2^{(b)} / v_2^{(a)})$	\mathbf{L}	$p(v_q^{(b)} / v_2^{(a)})$
\mathbf{M}	\mathbf{M}	\mathbf{M}	\mathbf{O}	\mathbf{M}	\mathbf{M}	\mathbf{M}	\mathbf{M}	\mathbf{O}	\mathbf{M}
$v_m^{(a)}$	$p(v_m^{(a)}, v_1^{(b)})$	$p(v_m^{(a)}, v_2^{(b)})$	\mathbf{L}	$p(v_m^{(a)}, v_q^{(b)})$	$v_m^{(a)}$	$p(v_1^{(b)} / v_m^{(a)})$	$p(v_2^{(b)} / v_m^{(a)})$	\mathbf{L}	$p(v_q^{(b)} / v_m^{(a)})$

We can give analogously the joint probability distribution $P(AB)$ of two attribute subsets A and B and the conditional probability distribution $P(B/A)$ of B given A . The above probability distribution have the following properties: (1) If $a \in A$ then the joint probability distribution of a and A is same as of a and $A - \{a\}$, and it equal to the probability distribution of A . (2) The conditional probability distribution $P(a/A)$ of a given A is same as $P(a/(A - \{a\}))$ of a given $A - \{a\}$. (3) If the probability distribution of A is same as the B , then the joint probability distribution $P(AB)$ is same as $P(A)$ or $P(B)$.

If the attribute value is numerical, then the above probability distribution can be regarded as the probability distribution of random variables. Hence an attribute $a \in A$ determines a random variable and attribute set $Q = \{a_1, a_2, \mathbf{L}, a_q\} \subseteq C \cup D$ determines q dimensional random vector, these random variable and random vector are also denoted by a and Q respectively. Let $A \subseteq C$. Attribute set A is said consistent with the decision attribute D , if each probability terms of the probability distribution $P(D)$ of D is same as the probability term in $P(A)$, or it can be expressed as the sum of some probability terms in $P(A)$.

If the uniform probability distribution is given on the domain $U = \{x_1, x_2, \mathbf{L}, x_n\}$, then a decision information system can be regarded as a random decision information system.

Definition 2.1^[10] In a random information system $((U, P), C \cup \{d\}, V, f)$, the information entropy of attribute a is defined by

$$H(a) = - \sum_{i=1}^m p(v_i^{(a)}) \log p(v_i^{(a)})$$

The joint information entropy of attribute a and b is defined by

$$H(ab) = -\sum_{i=1}^m \sum_{j=1}^q p(v_i^{(a)}, v_j^{(b)}) \log p(v_i^{(a)}, v_j^{(b)})$$

The conditional information entropy of attribute b given a is defined by

$$H(b/a) = -\sum_{i=1}^m \sum_{j=1}^q p(v_i^{(a)}, v_j^{(b)}) \log p(v_j^{(b)} / v_i^{(a)}).$$

The average mutual information of attribute a and b is defined by

$$I(a, b) = H(b) - H(b/a).$$

Analogously, the above definition can be generalized to case of attributes set. The following properties hold for information entropy.

Proposition 2.2^[16] (1) $H(B/A) \leq H(B)$, and if A and B are independent then the equality holds;

(2) $H(AB) = H(A) + H(B/A) = H(A) + H(B) - I(A, B)$;

(3) $I(A, B) = H(B) - H(B/A) \geq 0$;

(4) $I(A, B) = H(A) + H(B) - H(AB)$.

Proposition 2.3^[16] Let $((U, P), C \cup \{d\}, V, f)$ be a random information system, $Q = \{a_1, a_2, \dots, a_q\}$, $S = \{b_1, b_2, \dots, b_s\} \subseteq C \cup \{d\}$. If the probability distribution of Q and S satisfy the following condition:

$$P\{a_i = v_{ik}^{(a_i)} \mid a_i \in Q\} = \sum P\{b_j = v_{jl}^{(b_j)} \mid b_j \in S\},$$

Then $H(Q) \leq H(S)$. And if the right of above equation has at least two non-zero items, then $H(Q) < H(S)$.

Corollary 2.5^[16] Let $((U, P), C \cup \{d\}, V, f)$ be a random information system. If $Q, S \subseteq C \cup \{d\}, Q \subseteq S$ and the probability distributions of Q and S are different, then $H(Q) < H(S)$.

Example 2.6 We consider the following the car information system.

Table 1

U	a_1	a_2	a_3	a_4	d
u_1	low	cro.	gas.	low	poor
u_2	low	spa.	die.	mid.	good
u_3	hig.	cro.	die.	hig.	good
u_4	hig.	cro.	die.	mid.	poor
u_5	low	spa.	gas.	mid.	good

where a_1 = "price", a_2 = "internal design", a_3 = "engine", a_4 = "acceleration performance", hig.=high, cro.=crowd, spa.=spacious, gas.=gasoline, die.=diesel, mid.=middle. A normal probability distribution P on U as follows: $P(\{u_1\}) = 1/4$, $P(\{u_2\}) = 1/4$, $P(\{u_3\}) = 1/8$, $P(\{u_4\}) = 1/8$, $P(\{u_5\}) = 1/4$.

Using the probability distribution of P , we can compute the probability distribution of attributes. For example, the probability distribution $P(a_2)$ of attribute a_2 as shown in Table 2.

Table 2

	p (cro.)	p (spa.)
P	$\frac{1}{2}$	$\frac{1}{2}$

The joint probability distribution of attributes $\{a_2, a_3\}, \{a_1, a_2, a_3\}$ as shown in Table 3 and Table 4.

Table 3

	p (cro.,gas.)	p (cro., die.)	p (spa.,gas.)	p (spa., die.)
$p(a_2, a_3)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Table 4

	$p(\text{cro.}, \text{gas.}, \text{poor})$	$p(\text{spa.}, \text{die.}, \text{good})$	$p(\text{cro.}, \text{die.}, \text{good})$	$p(\text{cro.}, \text{die.}, \text{poor})$	$p(\text{spa.}, \text{gas.}, \text{good})$
$p(a_2, a_3, d)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

The conditional probability distribution $p(a_3/a_2)$, $p(d/a_2a_3)$ as shown in Table 5 and Table 6.

Table 5

	$p(\text{gas.}/\text{cro.})$	$p(\text{die.}/\text{cro.})$	$p(\text{gas.}/\text{spa.})$	$p(\text{die.}/\text{spa.})$
p	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Table 6

	$p(\text{poor}/(\text{cro.}, \text{gas.}))$	$p(\text{good}/(\text{spa.}, \text{die.}))$	$p(\text{good}/(\text{cro.}, \text{die.}))$	$p(\text{poor}/(\text{cro.}, \text{die.}))$	$p(\text{good}/(\text{spa.}, \text{gas.}))$
p	1	1	$\frac{1}{2}$	$\frac{1}{2}$	1

By computing, we have $H(a_2)=1$, $H(a_2a_3)=2$, $H(a_3/a_2)=2$, $H(d/a_2a_3)=0.25$

Attribute Correlation Measurement of Random Decision Information System

This section introduces a correlation measurement for a conditional attribute subset with respect to another conditional attribute subset about decision attribute in a random decision information system, and then by using correlation measurement to describe the necessary attributes, attribute importance, attribute reduction and its core attributes, etc.

If a conditional attribute set C is consistent with respect to decision attribute d , we call it as consistent decision information system. For $A \subseteq C$, $a \in A$, if the conditional probability distribution $P(\{d\}/A - \{a\})$ and $P(\{d\}/A)$ are the same, then a is called not necessary with respect to the decision attribute d , otherwise it is called necessary. If each attribute in A is necessary, then A is called a necessary attribute set. The set of all the necessary attributes of C is called the core of C and is denoted by $Core(C)$.

Definition 3.1 Let $((U, P), C \cup \{d\}, V, f)$ be a random information system. $A \subseteq C$ is called is a relative reduction of C with respect to decision attributes d if and only if the conditional probability distribution $P(\{d\}/A)$ and $P(\{d\}/C)$ are same.

Obviously, the relative reduction of C is not unique, and the intersection of all reduction of C is the core $Core(C)$.

Proposition 3.2 Let $((U, P), C \cup \{d\}, V, f)$ be a random information system, $A \subseteq C$. If the probability distribution of A and $A - \{a\}$ satisfy: a term (or some terms) in probability distributions of $A - \{a\}$ can be represented as the sum of some terms in probability distribution terms of A , and the rest terms in the probability distribution of $A - \{a\}$ and A are the same, then $H(\{d\}/A) \leq H(\{d\}/A - \{a\})$.

Proof We assume that $v^{(a)} = \{v_1^{(a)}, v_2^{(a)}\}$ and there only is a term in probability distribution of $A - \{a\}$ can be represented as the sum of some terms in probability distribution of A . We simple denote

$$[v^{(b)}]_{b \in A} = \{x \mid f(x, b) = v^{(b)}, b \in A\}, \quad (v^{(d)}, [v^{(b)}]_{b \in A}) = \{x \mid f(x, d) = v^{(d)}, f(x, b) = v^{(b)}, b \in A\},$$

$$(v^{(d)}, [v^{(b)}]_{b \in A}, v_1^{(a)}) = \{x \mid f(x, d) = v^{(d)}, f(x, b) = v^{(b)}, b \in A, f(x, a) = v_1^{(a)}\}, \quad (v_1^{(a)}) = \{x \mid f(x, a) = v_1^{(a)}\}.$$

then we have

$$P([v^{(b)}]_{b \in A}) = P([v^{(b)}]_{b \in A}, v_1^{(a)}) + P([v^{(b)}]_{b \in A}, v_2^{(a)}).$$

Hence

$$H(\{d\}/A - \{a\}) - H(\{d\}/A)$$

$$\begin{aligned}
 &= -\sum_{v^{(d)}} \sum_{v^{(b)}} P(v^{(d)}, [v^{(b)}]_{b \in A - \{a\}}) \log \frac{P(v^{(d)}, [v^{(b)}]_{b \in A - \{a\}})}{P(v^{(d)}, [v^{(b)}]_{b \in A - \{a\}})} + \sum_{v^{(d)}} \sum_{v^{(b)}} P(v^{(d)}, [v^{(b)}]_{b \in A}) \log \frac{P(v^{(d)}, [v^{(b)}]_{b \in A})}{P(v^{(d)}, [v^{(b)}]_{b \in A})} \\
 &= \sum_{v^{(d)}} P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}) \log \frac{P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}})}{P([v_{j_0}^{(b)}]_{b \in A - \{a\}})} - \sum_{v^{(d)}} P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_1^{(a)}) \log \frac{P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_1^{(a)})}{P([v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_1^{(a)})} \\
 &\quad - \sum_{v^{(d)}} P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_2^{(a)}) \log \frac{P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_2^{(a)})}{P([v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_2^{(a)})} \\
 &= \sum_{v^{(d)}} P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}) \log \frac{P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}})}{P([v_{j_0}^{(b)}]_{b \in A - \{a\}})} - \sum_{v^{(d)}} P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}) P\left(\frac{v_1^{(a)}}{(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}})}\right) \log \frac{P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_1^{(a)})}{P([v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_1^{(a)})} \\
 &\quad - \sum_{v^{(d)}} P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}) P\left(\frac{v_2^{(a)}}{(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}})}\right) \log \frac{P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_2^{(a)})}{P([v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_2^{(a)})}.
 \end{aligned}$$

Because $P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_1^{(a)}) \leq P([v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_1^{(a)})$ and $P(v^{(d)}, [v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_2^{(a)}) \leq P([v_{j_0}^{(b)}]_{b \in A - \{a\}}, v_2^{(a)})$, the last two terms in the above equation are non-negative. Thus $H(\{d\}/A - \{a\}) - H(\{d\}/A) \geq 0$, and if the probability distributions of A and $A - \{a\}$ are the same, the above equation holds.

Proposition 3.3 Let $((U, P), C \cup \{d\}, V, f)$ be a random decision information system, $A \subseteq C$. If the conditional attribute set A is consistent with respect to decision attribute d , then $a \in A$ is not necessary with respect to the decision attribute d if and only if $H(\{d\}/A) = H(\{d\}/A - \{a\}) = 0$.

Proof Suppose that the probability distribution of A is $P(A) = \{p_1, p_2, \dots, p_n\}$. Because A is consistent with respect to decision attribute d , a term in $P(d)$ equals to a term in $P(A)$ or equals to the sum of some terms in $P(A)$. Hence $P(A) = P(A \cup \{d\})$.

Necessity. If $a \in A$ is not necessary with respect to the decision attribute d , then the conditional probability distribution $P(d/A - \{a\})$ and $P(d/A)$ are the same. Hence the terms in $P(d/A - \{a\}) = P(d/A)$ are all equal to 1. Therefore $H(\{d\}/A) = H(\{d\}/A - \{a\}) = 0$.

Sufficiency. Assume that $a \in A$ is necessary with respect to the decision attribute d . Then the conditional probability distribution $P(d/A - \{a\})$ and $P(d/A)$ are different, it means that there is at least one termsuch that

$$P\left(\frac{v_{i_0}^{(d)}}{\{v_{j_0}^{(a_j)} \mid a_j \in A - \{a\}\}}\right) = \frac{P\{x \mid f(x, d) = v_{i_0}^{(d)}, f(x, a_j) = v_{j_0}^{(a_j)}, a_j \in A - \{a\}\}}{P\{x \mid f(x, a_j) = v_{j_0}^{(a_j)}, a_j \in A - \{a\}\}}$$

does not equal to 1. Hence $H(\{d\}/A - \{a\}) \neq 0$, and it is in contradiction to $H(\{d\}/A - \{a\}) = 0$. Hence $P(d/A - \{a\})$ are the same $P(d/A)$. Therefore a is not necessary with respect to d .

Definition 3.4 Let $((U, P), C \cup \{d\}, V, f)$ be a random decision information system, $A, B \subseteq C$. Denote

$$I(\{d\}, B/A) = H(\{d\}/A) - H(\{d\}/(A \cup B))$$

Define a correlation $r(B/A)$ of two attribute subsets A and B with respect to the decision attribute d as follows:

$$r(B/A) = \frac{I(\{d\}, B/A)}{H(\{d\}/A)},$$

and if $H(\{d\}/A) = 0$, then set $r(B/A) = 0$.

Theorem 3.5 Let $((U, P), C \cup \{d\}, V, f)$ be a random decision information system, $A, B \subseteq C$. Then

- (1) $0 \leq r(B/A) \leq 1$.
- (2) If $A \subseteq B$ then $r(B/A) = 0$.
- (3) $r(B/A) = 1$ if and only if the probability distribution of $A \cup B$ and d are the same.

Proof (1) By Proposition 2.3 we have $I(\{d\}, B/A) \geq 0$, then $r(B/A) \geq 0$. By Proposition 2.3 we also have $H(\{d\}/A) \geq H(\{d\}/(A \cup B))$, hence

$$I(\{d\}, B/A) = H(\{d\}/A) - H(\{d\}/A \cup B) \leq H(\{d\}/A).$$

Therefore $r(B/A) = \frac{I(\{d\}, B/A)}{H(\{d\}/A)} \leq 1$.

(2) If $B \subset A$, then $H(\{d\}/A) = H(\{d\}/(A \cup B))$. Thus $I(\{d\}, B/A) = 0$, so $r(B/A) = 0$.

(3) If the probability distribution of $P(A \cup B)$ and $P(d)$ are the same, then the joint probability distribution $P(A \cup B \cup \{d\})$ and the probability distribution $P(d)$ are the same. Hence the conditional probability distribution $P(d/A \cup B)$ taking value 0 or 1. This shows that $H(\{d\}/A \cup B) = 0$, so $I(\{d\}, B/A) = H(\{d\}/A)$. Therefore $r(B/A) = 1$.

Conversely, if $r(B/A) = 1$, then $H(\{d\}/A \cup B) = 0$. By the definition of conditional entropy we have that $p(a_i, d_j) = 0$ or $p(a_i, d_j) = p(d_j)$ for each attribute value a_i of $A \cup B$ and each attribute value d_j of d . This shows that the probability distribution of $A \cup B$ and d are the same.

Proposition 3.6 If $H(\{d\}/A) = H(\{d\}/A \cup \{a\})$, then $P(d/A) = P(d/A \cup \{a\})$.

Proof By Proposition 3.2 we have $H(\{d\}/A \cup \{a\}) \leq H(\{d\}/A)$, and only if the he probability distribution of A and $A \cup \{a\}$ are the same, then the equation $H(\{d\}/A) = H(\{d\}/A \cup \{a\})$ holds. Hence if $H(\{d\}/A) = H(\{d\}/A \cup \{a\})$, then $P(A) = P(A \cup \{a\})$, and so $P(A \cup \{d\}) = P(A \cup \{a\} \cup \{d\})$. Therefore $P(d/A) = P(d/A \cup \{a\})$.

Theorem 3.7 Suppose that $((U, P), C \cup \{d\}, V, f)$ is a random decision information system, $A \subseteq C$, then

(1) $a \in A$ is unnecessary if and only if $r(\{a\}/(A - \{a\})) = 0$.

(2) A is consisten with respect to decision attribute d , if and only if any $a \in A$, $r(\{a\}/(A - \{a\})) = 1$.

(3) $Core(C) = \{a \in C \mid r(\{a\}/(C - \{a\})) > 0\}$.

Proof (1) If a is unnecessary, then the conditional probability distributions $P(d/A - \{a\})$ are the sames $P(d/A)$. Then $H(\{d\}/A - \{a\}) = H(\{d\}/A)$. Hence $I(\{d\}, \{a\}/A - \{a\}) = 0$. Thus $r(A - \{a\}, \{a\}) = 0$.

Conversely, let $r(\{a\}/A - \{a\}) = 0$. Then $I(\{d\}, \{a\}/A - \{a\}) = 0$. This means that $H(\{d\}/A - \{a\}) = H(\{d\}/A)$. By Proposition 3.6 we have $P(d/A) = P(d/A \cup \{a\})$. Thus a is unnecessary.

(2) Because A is consisten with respect to decision attribute d , a term in $P(d)$ equals to a term in $P(A)$ or equals to the sum of some terms in $P(A)$. Hence $P(A) = P(A \cup \{d\})$. Then

$$H(\{d\}/A) = -\sum P(A \cup \{d\}) \log \frac{P(A \cup \{d\})}{P(A)} = 0.$$

Conversely, let $r(\{a\}/A - \{a\}) = 1$. Then $H(\{d\}/A) = -\sum P(A \cup \{d\}) \log \frac{P(A \cup \{d\})}{P(A)} = 0$. This means that each term in above equation is 0. Hence $P(A) = P(A \cup \{d\})$. Therefore A is consisten with respect to decision attribute d .

(3) By the mean of $Core(C)$ and (2), we can obtain it.

Remark 3.8 According to the above conclusion, we have the following analysis: the correlation $r(\{a\}/(A - \{a\}))$ depicted the similarity degree between the joint probability distribution of attribute set $A \cup \{d\}$ and A . When the value of correlation $r(\{a\}/(A - \{a\}))$ increases from 0 to 1, the ratio of the joint probability $P(A \cup \{d\})/P(A)$ increases from $P((A - \{a\}) \cup \{d\})/P(A - \{a\})$ to 1.

Example 3.9 We consider the car information system as shown in Example 2.6. $r(\{a_4\}/C - \{a_4\}) = \frac{H(\{d\}/C - \{a_4\}) - H(\{d\}/C)}{H(\{d\}/C - \{a_4\})} = 1$, $r(\{a_3\}/\{a_4\}) = \frac{H(\{d\}/\{a_3\}) - H(\{d\}/\{a_3, a_4\})}{H(\{d\}/\{a_3\})} = 0.2367$, $r(\{a_1\}/\{a_3, a_4\}) = 1$.

By Theorem 3.7, the below conclusion is obtained.

Theorem 3.10 Let $((U, P), C \cup \{d\}, V, f)$ be a random decision information system. For $A \subseteq C$, if

(1) $H(\{d\}/A) = H(\{d\}/C)$,

(2) $\forall a \in A$, $r(\{a\}/(A - \{a\})) > 0$,

then A is a reduction of C .

Application of Attribute Correlation to Attribute Reduction

In this section, we use correlation to proceed attribute reduction in a random decision information systems.

Based on Theorem 3.10, we give an algorithm for attribute reduction in a random decision information systems.

Algorithm 4.1: (Attribute reduction algorithm for a random decision information system based on attribute correlation)

Input: A random decision information system $((U, P), C \cup \{d\}, V, f)$.

Output: Relative reduction Q of conditional attribute set C of a random decision information system.

Step 1. Compute probability distribution of conditional attribute C .

Step 2. Compute the conditional information entropy $H(\{d\}/C)$ of C .

Step 3. Compute $Core(C) = \{a \in C \mid r(\{a\}/(C - \{a\})) > 0\}$.

If $H(\{d\}/Core(C)) = H(\{d\}/C)$, the algorithm terminates ($Core(C)$ is the minimum relative reduction).

Step 4. (Continue adding smaller attributes to $Core(C)$ to construct subset A of attribute set C)

$A := Core(C)$.

While $H(\{d\}/A) \neq H(\{d\}/C)$ Do

(i) For $\forall a \in C - A$, compute $r(\{a\}/A - \{a\})$,

(ii) Select the attribute a such that $r(\{a\}/A - \{a\}) = \min\{r(\{b\}/(A - \{b\})) \mid b \in C - A\}$ and $A := A \cup \{a\}$,

(iii) Compute $H(\{d\}/A)$.

Endwhile

Step 5. (Getting the relative reduction Q of attribute set C)

$A' := A - Core(C), |A'| = N$.

For $i = 1$ to N Do

(i) Delete the attribute a_i from A' ,

(ii) Compute $H(\{d\}/A' \cup Core(C))$,

(iii) If $H(\{d\}/A' \cup Core(C)) \neq H(\{d\}/C)$, $A' := A' \cup \{a_i\}$.

Endfor

Step 6. If $Q := A' \cup Core(C)$, the algorithm terminates. Q is a relative reduction of C .

By using this algorithm 4.1, the time complexity to find a reduction is $O(|C|^2 |D| |U|)$.

At Step 2, the time complexity of computing $H(\{d\}/C)$ is $O(|U| |D|)$.

At Step 3, need to compute $r(\{a\}/C - \{a\})$ for all $a \in C$ in computing $Core(C)$, the time complexity of computing $Core(C)$ is $O(|C| |U|)$.

At Step 4, the time complexity of computing all $r(\{a\}/A - \{a\}) (a \in A, A \subseteq C)$ is $O((|C| + (|C| - 1) + \dots + 1) |U|) = O(|C|^2 |U|)$

At Step 5, the time complexity of computing all $H(\{d\}/A' \cup Core(C))$ is $O(|C| |U|)$.

Hence if we ignore the time complexity of computing the probability distribution, then the time complexity of Algorithm 4.1 is $O(|C|^2 |D| |U|)$.

Example 4.2 We compute the attribute reduction of the car information system as shown in Example 2.9 by using Algorithm 4.1.

Step 1A. Computing the probability distribution of C (omitted).

Step 2A. Computing the joint information entropy of C : $H(\{d\}/C) = 0$.

Step 3A. By computing we have $r(\{a_1\}/C - \{a_1\}) = 0$, $r(\{a_2\}/C - \{a_2\}) = 0$, $r(\{a_3\}/C - \{a_3\}) = 1$, $r(\{a_4\}/C - \{a_4\}) = 1$. Hence $Core(C) = \{a_3, a_4\}$. Since $H(\{d\}/Core(C)) = 0.451$, $H(\{d\}/Core(C)) \neq H(\{d\}/C)$, go to Step 4.

Step 4A. Set $A := Core(C) = \{a_3, a_4\}$. Since $H(\{d\}/Core(C)) \neq H(\{d\}/C)$, we need to compute $r(\{a_1\}/A - \{a_1\}) = r(\{a_1\}/\{a_3, a_4\}) = 1$, $r(\{a_2\}/A - \{a_2\}) = 1$. Taking $a = a_1$, $A := A \cup \{a_1\} = \{a_1, a_3, a_4\}$. By computing we have $H(\{d\}/A) = 0$. Since $H(\{d\}/A) = H(\{d\}/C)$, go to Step 5.

Step 5A. Set $A' := A - \text{Core}(C) = \{a_1\}$. $N := |A'| = 1$. Taking $a_i = a_1$, $A' := A' - \{a_1\} = \emptyset$. By computing we have $H(\{d\}/A' \cup \text{Core}(C)) = H(\{d\}/\{a_3, a_4\}) = 0.344$. Since $H(\{d\}/A' \cup \text{Core}(C)) \neq H(\{d\}/C)$, we get $A' := A' \cup \{a_1\} = \{a_1\}$.

Step 6A. Let $Q := A' \cup \text{core}(C) = \{a_1, a_3, a_4\}$. Algorithm is completed, and $Q = \{a_1, a_3, a_4\}$ is a reduction of C .

Conclusion

In a random decision information system, the conditional information entropy of two attribute sets is defined, and it is depicted that an attribute $a \in A$ relative to the decision attribute d is not necessary. We use conditional information entropy to introduce the correlation $r(B/A)$ of a conditional attributes set A relative to another condition attributes set B with respect to the decision attributes set. The properties of correlation are discussed, and the concepts of necessary attributes, consistent attributes and core attributes of decision information systems are described by correlation. An attribute $a \in A$ is not necessary if and only if the correlation $r(\{a\}/A - \{a\}) = 0$. The attribute subset A is consistent with the decision attribute d if and only if its correlation is $r(\{a\}/A - \{a\}) = 1$ for any $a \in A$. The attribute a is an element of a set of kernel attributes if and only if as $r(\{a\}/A - \{a\}) > 0$. We propose a method of attribute reduction in a random decision information system based on attribute correlation, give an attribute reduction algorithm based on this method, and the time complexity of algorithm is analyzed. The listed examples show that the algorithm proposed in this paper is effective.

Acknowledgements

This work was financially supported by the Natural Science Foundation of Hunan province (No.2016JJ6138, No.2017JJ2241), Science Research Project of Education Department of Hunan Province (15C1284).

References

- [1] Z. Pawlak: Rough sets theoretical aspects of reasoning about data. Boston: Kluwer Academic Publishers, (1991).
- [2] J.Y. Liang, Z.B. Xu: The algorithm on knowledge reduction in incomplete information systems. International Journal of Uncertainty, Fuzziness and Knowledge—Based Systems, 10(1):95-103 (2002).
- [3] W.Z. Wu: Attribute reduction based on evidence theory in incomplete decision systems. Information Science, 178:1355-1371 (2008).
- [4] Q.H. Hu, X.J. Che, L. Zhang, et al: Rank entropy based decision trees for monotonic classification. IEEE Transactions on Knowledge and Data Engineering, 24:2052-2064 (2012).
- [5] W.S. Du, B.Q. Hu: Approximate distribution reduce in inconsistent interval-valued ordered decision tables. Information Science, 271:93-114 (2014).
- [6] W.S. Du, B.Q. Hu: Attribute reduction in ordered decision tables via evidence theory. Information Science, 364-365:91-110 (2016).
- [7] W.X. Zhang, Y. Liang, W.Z. Wu: Information systems and knowledge discoveries. Beijing: Science Press, (2003).
- [8] D.Q. Miao, J. Wang: An information representation of concepts and operation in rough set theory. Journal of Software, 10(2):113-116 (1999).
- [9] D.Q. Miao, G.R. Hu: A heuristic algorithm for reduction of knowledge. Journal of Computer Research and Development, 36(6):681-684 (1999).
- [10] G.Y. Wang, H. Yu, D.C. Yang: Decision table reduction based on conditional information entropy. Chinese Journal of Computers, 25(7): 759-766 (2002).

- [11]H. Yu, G.Y. Wang, Y.Y. Yao: Current research and future perspectives on decision-theoretic rough sets. *Chinese Journal of Computers*, 38(8): 1628-1639 (2015).
- [12]J.Y. Liang, W. Wei, Y.H. Qian: An incremental approach to computation of a core based on conditional entropy. *Systems Engineering Theory and Practice*, 28(4):81-89 (2008).
- [13]J.L. Zhang, X.L. Liu: Fuzzy belief measure in random fuzzy information systems and its application to knowledge reduction. *Neural Computing and Applications*, 22:1419-1431 (2013).
- [14]J.L. Zhang: Rough set model based on random fuzzy sets. *Chinese Journal Engineering Mathematics*, 22(3):323-327 (2005).
- [15]X. Wu, J.L. Zhang: Rough set models based on random fuzzy sets and belief function of fuzzy sets. *International Journal General Systems*, 41:123-141 (2012).
- [16]T. Zhou, J.L. Zhang: Relativity of attributes in random information systems and its application to knowledge reduction. *Computer Engineering and Applications*, 47(14):140-142 (2011).
- [17]H. Qu, S.B. Li: *Elements of Information Theory and application*. Beijing: Tsinghua University Press (2005).