International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems

ENTROPIES OF FUZZY INDISCERNIBILITY RELATION AND ITS OPERATIONS

QINGHUA HU and **DAREN YU**
Harbin Institute of Technology, Harbin, China
huqinghua@hcms.hit.edu.cn

Received (28 December 2003)
Revised (5 May 2004)

Yager's entropy was proposed to compute the information of fuzzy indiscernibility relation. In this paper we present a novel interpretation of Yager's entropy in discernibility power of a relation point of view. Then some basic definitions in Shannon's information theory are generalized based on Yager's entropy. We introduce joint entropy, conditional entropy, mutual information and relative entropy to compute the information changes for fuzzy indiscerniblity relation operations. Conditional entropy and relative conditional entropy are proposed to measure the information increment, which is interpreted as the significance of an attribute in fuzzy rough set model. As an application, we redefine independency of an attribute set, reduct, relative reduct in fuzzy rough set model based on Yager's entropy. Some experimental results show the proposed approach is suitable for fuzzy and numeric data reduction.

Keywords: discernibility power; equivalence relation; fuzzy entropy; fuzzy indiscernibility relation; relation operation; fuzzy rough set.

1. Introduction

Since Shannon introduced the concepts of information entropy and mutual information in 1940s, they have been widely applied to message coding, information compress, and pattern recognition. Entropy is used as a measure to computing the significance of an attribute in constructing decision tree [1] and reducing information system in rough set theory [2]; Entropy is used to find cuts for discretizing numeric data [3] and is used to automatically identify the number and initial locations of cluster center in fuzzy clustering [4]. Fuzzy sets play a significant role in many deployed systems because of their capability to model nonstatistical imprecision. Consequently, characterization and quantification of fuzziness are important issues that affect the management of uncertainty in many system models and designs. The entropy of a fuzzy set is a measure of fuzziness of the fuzzy set. Some general conclusions about fuzzy entropy, which are presented based on the axiom definitions of fuzzy entropy and distance measure, can refer to [5, 6, 7, 10, 11, 18].

Relation is an important concept in set theory. In classical set theory values of a relation between two elements take 1 if the relation is satisfied, otherwise 0. It's easy to introduce Shannon's entropy to measure the uncertainty of a partition induced by a classical indiscernibility defined on a set. Fuzzy indiscernibility relation extends the concept of equivalence relation to the fuzzy framework. In this case value domain of a relation is [0, 1]. In order to calculate the information of fuzzy indiscernibility relations R. Yager [8] proposed a new measure of entropy that is suitable to operate on domains over which a fuzzy equivalence relation has been defined. This measure overcomes some limitations of Shannon's one. E. Hernandez [9] presented a new interpretation in terms of the observability of the elements.

Shannon defined the joint entropy, conditional entropy and mutual information in the case where multiple random variables exist. These measures are proposed to calculate the information changes with multivariable combination. In data processing and knowledge management, there are usually multiple attributes or relations to characterize samples. Sometime we not only require analyzing one relation on a set, but also need to study the operations of multiple relations. For example, as to rough set theory and decision tree induction, usually several condition attributes and one decision attribute coexist. Each attribute can induce a relation over the element set. Yager's entropy gives us an efficient measure to compute the discernibility information in a relation, but it's a problem how to measure the change of discernibility power with relation operations. In this paper we will show a novel interpretation of Yager's entropy in view of discernibility power of a relation first, and then extend the definitions of Shannon's joint entropy, conditional entropy and mutual information based on Yager' entropy. Some properties and interpretations of the extended definitions are presented.

2. The Basic Concepts of Fuzzy Indiscernibility Relation And Yager's Entropy

2.1. *Fuzzy indiscernibility relation and its operations*

Definition 1. Let X be a set, and R is a relation defined over X, $R \to [0,1]$. For $\forall x, y \in X$, if R satisfies:

(1) Reflexivity: $R(x,x)=1$;

(2) Symmetry: $R(x,y)=R(y,x)$

we say R is a fuzzy tolerance relation over X.

Definition 2. Let X be a set, R is a relation defined over X, $R \to [0,1]$. For $\forall x, y, z \in X$, if R satisfies:

(1) Reflexivity: $R(x,x)=1$;

(2) Symmetry: $R(x,y)=R(y,x)$;

(3) T-Transitivity: $R(x,z) \ge T\{R(x,y), R(y,z)\}$

then we call R a *T*-fuzzy indiscernibility relation, where *T* is a triangle norm. R is a fuzzy equivalence relation if the property of transitivity (*T*-transitivity) is defined in terms of minimum t-norm.

We will focus on the operations and properties of fuzzy indiscernibility relations in this paper.

If R and S are two fuzzy indiscernibility relations over X, some operators are defined as:

(1) Union: $(R \bigcup S)(x, y) = \max\{R(x, y), S(x, y)\}, \forall x, y \in X$

(2) Intersection: $(R \bigcap S)(x, y) = \min\{R(x, y), S(x, y)\}, \forall x, y \in X$

(3) Containment: $R \subseteq S \Rightarrow R(x, y) \leq S(x, y), \forall x, y \in X$

A relation $R \subseteq X \times X$ can be denoted by a relation matrix M_R. Values of The diagonal of the matrix M_R are all 1 if R is reflexive, i.e. $R_{ii} = 1$. The matrix is symmetric if relation R satisfies symmetry, i.e. $R_{ij} = R_{ji}$. The matrix is called Max-min transitive if $R_{ij} \geqslant \vee_k (R_{ik} \wedge R_{kj})$, where $\vee$ and $\wedge$ stand for max and min, respectively.

Definition 3. Given a finite set $X = \{x_1, x_2, \cdots, x_n\}$, P is the probability distribution on X and R is a fuzzy indiscernibility or equivalence relation on X. Yager's entropy is defined as:

$$H_P(R) = \sum_x - p(x) \log_2 \pi(x) \tag{1}$$

where

$$\pi(x_i) = \sum_x p(x) R(x, x_i) \tag{2}$$

3. A Novel Interpretation of Yager's Entropy

In this section we present a novel interpretation of Yager's entropy as a measure of discernibility power of an indiscernibility relation. It's shown how Yager's entropy of an indiscernibility relation gives the discernibility power of a relation matrix.

Given a set $X = \{x_1, x_2, \cdots, x_n\}$, R is a fuzzy indiscernibility relation on X. It can be denoted as a relation matrix as follows:

$$R = \begin{pmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ R_{n1} & \cdots & R_{nn} \end{pmatrix}$$

R_{ij} is the indiscernibility or equivalence degree between x_i and x_j with respect to relation R. As we know that the larger R_{ij} is, the more x_i and x_j are indiscernible. If $R_{ij} > R_{ik}$, we say x_i and x_j are more indiscernible than x_i and x_k. And if $S_{ij} > R_{ij}$, we say that with respect to relation S, x_i and x_j are more indiscernible than relation R.

Definition 4. Let R be a fuzzy indiscernibility relation over a set X, P the probability distribution on X. $\forall x_i \in X$. We define expected indiscernibility degree of x_i to all $x \in X$ with respect to R as follows:

$$\pi(x_i) = \sum_{x \in X} p(x) R(x, x_i)$$

Proposition 1. $\forall x \in X : 0 \le \pi(x) \le 1$.

Corollary 1. $0 \le \sum_{x \in X} \pi(x) \le Card\ (X)$

Definition 5. The information quantity of discernibility degree of x_i is defined as

$$C(x_i) = -\log_2 \pi(x_i)\ . \tag{3}$$

It's easy to show that the larger $\pi(x_i)$ is, the more indiscernible x_i is, and the less $C(x_i)$ is, which shows that the measure $C(x_i)$ describes the discernibility degree of x_i to all elements in set *X* with respect to relation *R*.

Proposition 2. $\forall x_i \in X : 0 \le C(x_i) \le -\log_2 p(x_i)$

Definition 6. Given an indiscernibility relation *R* and a probability distribution *P* on X. The Yager's entropy of the pair <*R, P*> is defined as

$$H_P(R) = \sum_{x \in X} p(x) C(x) = -\sum_{x \in X} p(x) \log_2 \pi(x) \tag{4}$$

Yager's entropy measures the average discernibility power of the set with respect to relation R. It is easy to show that the entropy increases monotonously with the discernibility power of relation R.
As to set $X = \{x_1, x_2\}$, given a probability distribution $p(x_1) = p(x_2) = 1/2$, R is a indiscernibility relation defined on X:

$$R = \begin{pmatrix} 1 & R_{12} \\ R_{21} & 1 \end{pmatrix}$$

Then Yager's entropy of relation R varying with R_{12} or R_{21} is shown in figure 1.

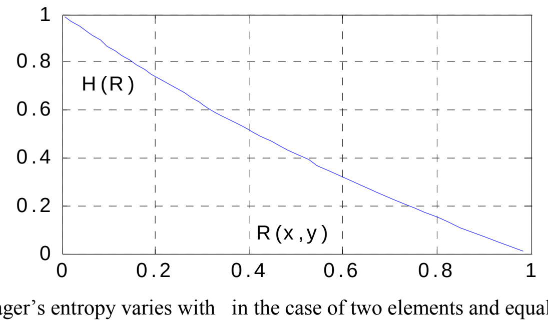

Fig. 1. Yager's entropy varies with in the case of two elements and equal probability.

Proposition 3. Given a set X and its probability distribution P, the Yager's entropy hits the maximum if all the elements are completely discernible with respect to relation R, and the maximum of entropy is

$$H_P(R) = -\sum_{x\in X} p(x)\log_2 p(x)$$

Proposition 4. Let R_1, R_2 be two fuzzy indiscernibility relations on X and P a probability distribution over X. Then

$$R_1 \subseteq R_2 \Rightarrow H_P(R_1) \geq H_P(R_2)\ .$$

Note that it doesn't holds: $H_P(R_1) \geq H_P(R_2) \Rightarrow R_1 \subseteq R_2$.

Proposition 5. Let R be a crisp equivalence (indiscernibility) relation on X, $\overline{P}$ is the induced probability distribution by the quotient set X/R. P is a probability distribution over X and $\forall x \in X$, $p(x) = 1/card(X)$. Shannon's entropy is denoted by $H(P)$. Then $H(\overline{P}) = H_P(R)$.

4. Some Extensions Based on Yager's Entropy

Operations of relations are very useful in some cases. For example, indiscernibility relation is a coral concept in rough set theory, where relations and their operations are used to define knowledge and granularity of knowledge. In this section we will introduce some important definitions, such as joint entropy, conditional entropy and mutual information in relation operations point of view. The properties of these measures are analyzed.

4.1. *Joint entropy, conditional entropy and their properties*

There is more than one attribute in information systems in most cases. Each attribute will induce an indiscernibility relation. Since some indiscernibility relations are usually combined to partition the set of discourse, we should introduce some novel measures to calculate the entropy of operations.

Definition 7. Let R_1, R_2 be two indiscernibility relations on X, P a probability distribution on X. Joint entropy of R_1 and R_2 is defined as

$$H_P(R_1R_2) = H_P(R_1 \cap R_2)\ . \tag{5}$$

Joint entropy of two relations embodies the discernibility power of the joint relation. As we know that the discernibility of objects increases with addition of some new relations or knowledge.

Proposition 6. $H_P(R_1R_2) \geq \max\{H_P(R_1), H_P(R_2)\}$

Proposition 7. *If* $R_1 \subseteq R_2$, $H_P(R_1R_2) = H_P(R_1)$

Proposition 8. $H_P(R_1R_2) \le H_P(R_1) + H_P(R_2)$

Proposition 6 shows that the Yager's entropy increases monotonously with the plus of new relations, which can be interpreted that with the addition of indiscernibility relations on X, discernibility power of the joint relations gets stronger for we can get a finer partition by introducing a new attribute. Proposition 7 shows the addition of some relations will not bring increment of discernibility power because the information implying in these relations has been implied in the others.

Definition 8. Let R_1, R_2 be two indiscernibility relations on X, P a probability distribution on X. Conditional entropy of R_2 with respect to R_1 is defined as

$$H_P(R_2 \mid R_1) = H_P(R_1R_2) - H_P(R_1)\,. \tag{6}$$

Accordingly, conditional entropy of R_1 conditioned to R_2 can be defined as

$$H_P(R_1 \mid R_2) = H_P(R_1R_2) - H_P(R_2)\,. \tag{7}$$

Proposition 9. $H_P(R_1 \mid R_2) \le H_P(R_1),\ H_P(R_2 \mid R_1) \le H_P(R_2)$

Proposition 10. *If* $R_1 = R_2,\ H_P(R_1 \mid R_2) = H_P(R_2 \mid R_1) = 0$

So *If* $R_1 = R_2,\ H_P(R_1R_2) = H_P(R_1) = H_P(R_2)$

As to set $X = \{x_1, x_2\}$, given probability distribution $p(x_1) = p(x_2) = 1/2$. R_1 and R_2 are two relations defined on X:

$$R_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},\ R_2 = \begin{bmatrix} 1 & R_{12} \\ R_{21} & 1 \end{bmatrix}$$

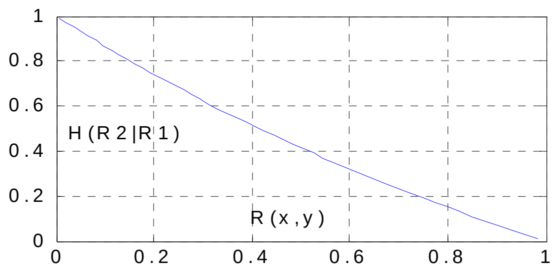

Fig. 2. Yager's conditional entropy varies with R_{12}

Yager's conditional entropy $H_P(R_2 \mid R_1)$ varying with R_{12} or R_{21} is shown in figure 2. From the figure we have the increment of discernibility power decreases with weakening of inconsistency of two relations.

Definition 9. Let R_1 , R_2 be two indiscernibility relations on X , P a probability distribution on X . Mutual information of two relations is defined as

$$I_P(R_2, R_1) = H_P(R_2) - H_P(R_2 \mid R_1) \tag{8}$$

or

$$I_P(R_1, R_2) = H_P(R_1) - H_P(R_1 \mid R_2) . \tag{9}$$

It's easy to show

$$I_P(R_2, R_1) = I_P(R_1, R_2) = H_P(R_1) + H_P(R_2) - H_P(R_1 R_2) .$$

Proposition 11. $0 \le I_P(R_1, R_2) \le \min\{H_P(R_1),\ H_P(R_2)\}$

The connections among $H_P(R_1)$, $H_P(R_2)$, $H_P(R_1R_2)$ $H_P(R_1 \mid R_2)$, $H_P(R_2 \mid R_1)$ and $I_P(R_1, R_2)$ can be demonstrated as figure 2.

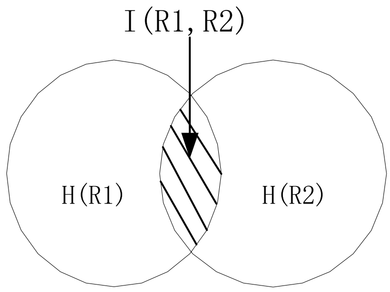

Fig. 3. Relation chart of entropy

Yager's entropy gives the information of discernibility power of a relation. Conditional entropy reflexes the increment of discernibility power by introducing a new relation after one relation has been known, and mutual information is the common part of discernibility power of two relations.

4.2. *Relative entropy*

In general, condition attributes and decision attributes coexist in an information system. The mutual information is used to measure the significance of an attribute conditioned to the decision attribute or the category label. Here we will introduce a new definition—relative entropy to compute the information induced by union operators and the interpretation is given from attribute significance point of view.

Definition 9. Let R_1, R_2 are two indiscernibility relations on *X*, *P* the probability distribution. $R = R_1 \cup R_2$, then the relative entropy of relation R_1 relative to R_2 is defined as

$$H_P(R_1;\ R_2) = H_P(R_2) - H_P(R) . \tag{10}$$

Accordingly, the entropy of relation R_2 relative to R_1 is defined as

$$H_P(R_2;\ R_1) = H_P(R_1) - H_P(R)\ . \tag{11}$$

Remark: $H_P(R_2;\ R_1)$ measures the inconsistency of two partitions between R_1 and R_2. In this definition there exits a strategy of rewards and punishment by the union operator. As we know, Yager's entropy presents the information of average discernibility power of a relation. Union operator makes relation $R = R_1 \bigcup R_2$ more indiscernible because of augments of relation values. Suppose R_1 is induced by a condition attribute and R_2 a decision attribute. On one hand, if the elements $x_i, x_j \in X$ are more discernible as to relation R_2 than to R_1, that is $R_{ij}^1 > R_{ij}^2$, then the entropy $H_P(R_1 \bigcup R_2)$ will be less than $H_P(R_2)$. That's to say, $H_P(R_1 \bigcup R_2)$ has been punished because of the weak discernibility power, therefore $H_P(R_1; R_2) > 0$. On the other hand, if the elements $x_i, x_j \in X$ are not more discernible as to relation R_2 than to R_1. In other word, $R_{ij}^1 \le R_{ij}^2$, then the entropy $H_P(R_1 \bigcup R_2)$ will equal to $H_P(R_2)$ and relative entropy $H_P(R_1; R_2) = 0$ that's to say, $H_P(R_1 \bigcup R_2)$ does not receive any reward although R_1 has more discernibility power than R_2. In a word, if elements are more discernible as to the decision attribute than a condition attribute set, relative entropy will be punished; otherwise relative entropy doesn't get reward. These strategies of rewards and punishment are the same as the idea of relative reduct in rough set theory.

Example Given a set $X = \{x_1, x_2, x_3\}$, $P(x_i) = 1/3$, $i = 1, 2, 3$. R_d is a relation induced by a decision attribute and R_1, R_2 are induced by two condition attributes.

$$R_d = \begin{bmatrix} 1 & 0.9 & 0 \\ 0.9 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ R_1 = \begin{bmatrix} 1 & 0.1 & 0 \\ 0.1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ R_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.9 \\ 0 & 0.9 & 1 \end{bmatrix}.$$

$$\begin{aligned} H_P(R_1; R_d) &= H_P(R_d) - H_P(R_d \bigcup R_1) \\ &= 0 \end{aligned}$$

$$\begin{aligned} H_P(R_2; R_d) &= H_P(R_d) - H_P(R_d \bigcup R_2) \\ &= 0.9676 - 0.4900 \\ &= 0.4776 \end{aligned}.$$

From the results we can see that R_1 has more discernibility power than R_d, $H_P(R_1, R_d)$ doesn't get any reward and the value is 0. R_2 has the same discernibility power as R_d, but their discernibility power are not consistent, so a punishment is performed.

Proposition 12. $H_P(R_1;\ R_2) \ge 0$, if and only if $R_1 \subseteq R_2$, $H_P(R_1;\ R_2) = 0$.

Assuming R_1 is an indiscernibility relation by a decision attribute, and R_2 is an indiscernibility relation by a condition attribute, then the relative entropy $H_P(R_2;\ R_1)$ measures the inconsistency of two partitions between R_1 and R_2.

As to set $X = \{x_1, x_2\}$, given probability distribution $p(x_1) = p(x_2) = 1/2$. R and R_d are two relations defined on X:

$$R = \begin{bmatrix} 1 & R_{12} \\ R_{21} & 1 \end{bmatrix}, \ R_d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Yager's conditional entropy $H_P(R;R_d)$ varying with R_{12} or R_{21} is shown in figure 3. I can see relative entropy increases monotonously with R_{12} from the curve, which shows relative entropy measures the inconsistence between R and R_d.

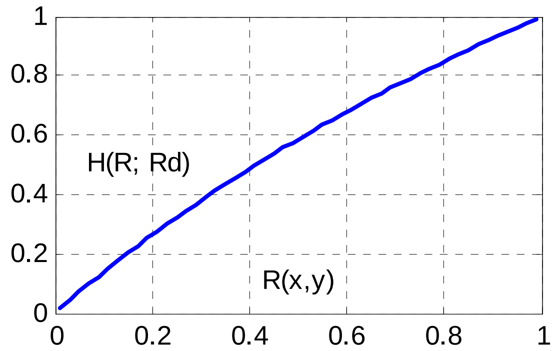

Fig. 4. Relative entropy $H_P(R;R_d)$ varies with R_{12}

Definition 11. Let X be the set of discourse, P the probability distribution and $R_1, R_2, R_3 \subseteq X \times X$ three indiscernibility relations, $R = R_1 \cap R_2 \bigcup R_3$, then the relative joint entropy R_2, R_1 to R_3 denoted as $H_P(R_1R_2;\ R_3)$ is defined as

$$H_P(R_1R_2;\ R_3) = H_P(R_3) - H_P(R). \tag{12}$$

Proposition 13. $H_P(R_1R_2;\ R_3) \le \min\{H_P(R_1, R_3), H_P(R_2, R_3)\}$.

Proposition 13 shows that discernibility power of the condition attribute set relative to the decision attribute gets stronger monotonously by introducing an unseen condition attribute. In other words, the inconsistency between condition attributes and decision attribute decreases.

Proposition 14. The minimum of $H_P(R_1R_2;R_3)$ is $H_P(R_1R_2;R_3) = 0$.

Definition 12. Let X be the universe of discourse, P the probability distribution and $R_1, R_2, R_3 \subseteq X \times X$ three fuzzy indiscernibility relations, $R = R_1 \cap R_2 \bigcup R_3$, then the relative conditional entropy R_1 to R_2, denoted by $H_P(R_1 \mid R_2;R_3)$, is defined as

$$H_P(R_1 \mid R_2;R_3) = H_P(R_2;R_3) - H_P(R_1R_2;R_3). \tag{13}$$

$H_P(R_1 \mid R_2;R_3)$ is the increment of consistency by introducing R_1. It can be used as the relative significance of a condition attribute.

Proposition 15. $H_P(R_1 \mid R_2;R_3) \ge 0$, if and only if $R_1 \subseteq R_2 \Rightarrow H_P(R_1 \mid R_2;R_3) = 0$.

5. An Application: Fuzzy Rough Set Model Based on Yager's Entropy

As we have stated in section 3, fuzzy entropy introduced by R. Yager measures the discernibility power of a relation on an object set. The entropy, which is brought by a fuzzy equivalence relation with less similarity between elements, will be larger. Proposition 6 shows entropy of the joint relation R of R_1, R_2 is larger than that of R_1 or R_2. The increment of entropy can be interpreted as the enhancement of discernibility power by introducing an unseen relation or attribute. So Yager's fuzzy entropy can be adapted to measuring the significance of attributes in fuzzy rough set model [12-17].

In this section we will propose a novel definition of independency, reduct and relative reduct in fuzzy rough set framework based on Yager's entropy and its extensions.

Definition 13. Given an information system (*IS*) $< U, A, V, f >$, U is a finite object set, P is a probability distribution on U . A is the attribute set charactering the objects. We can the information system a decision table if $A = C \bigcup D$ and $C \bigcap D = \varnothing$, where C is condition attribute set and D the decision attribute set.

Definition 14. $B \subseteq A$ is a subset of attributes, $B = \{a_1, a_2, \cdots, a_k\}$. M_B is the fuzzy relation matrix by an indiscernibility relation, $a \in B$. Then significance of the attribute a is defined as

$$SIG(a \mid B - a) = H_P(a \mid B - a) = H_P(B) - H_P(B - a) . \tag{14}$$

If $SIG(a \mid B - a) = 0$, we say attribute a is *redundant* or *superfluous*. Otherwise the attribute a is *indispensable*. If $\forall a \in B$: a is indispensable, we say B is *independent*.

Definition 15. Given an IS $< U, A, V, f >$, P is the probability distribution, $A = C \bigcup D$, C is the set of condition attributes and D the set of decision attributes. Here we just consider the Multiple-inputs and single-output case, hence $D = \{d\}$. $B \subseteq C$ is a subset of attributes. The relative significance of attribute $a \in B$ is defined as

$$SIG(a \mid B - a; D) = H_P(a \mid B - a; D) = H_P(B - a; D) - H_P(B; D) \tag{15}$$

If $SIG(a \mid B - a; D) = 0$, we say a is *D-superfluous*, otherwise the attribute a is *D-indispensable*. If $\forall a \in B$: a is *D-indispensable*, we say B is *D-independent*.

Definition 16. Given an $IS < U, A, V, f >$, P is the probability distribution of U. $B \subseteq A$ is a subset of attributes, $B = \{a_1, a_2, \cdots, a_k\}$. We say B is a *reduct* if B satisfies:

(1) $H_P(B) = H_P(A)$;

(2) $\forall a \in B$: $H_P(a \mid B - a) > 0$

The first condition guarantees that the discernibility power of a reduct is equal to that of entire attributes and the second one shows attribute subset B is independent.

Usually there are several reducts for an information system.

Definition 17. *Core* is the intersection of *reduct*:.

$$Core(C) = \bigcap_i red_i$$

Definition 18. Given an IS $< U, A, V, f >$, $A = C \bigcup D$, C is the condition attribute set and D is the decision attribute. P is the probability distribution on U. $B \subseteq C$ is a subset of condition attributes, $B = \{a_1, a_2, \cdots, a_k\}$. We say B is a *relative D-reduct* if B satisfies:

(1) $H_P(B; D) = H_P(A; D)$;

(2) $\forall a \in B$: $H_P(a \mid B - a; D) > 0$

Relative D-reduct denotes as $D - red$. There are usually several $D - red$ for a decision system.

Definition 19. *Relative core* of a decision system is the intersection of all *relative D-reduct*:

$$D - core = \bigcap_i D - red_i$$

6. Illustrative Examples

In this section we will show how to compute Yager's entropy of an indiscernibility relation and its operations. And illustrate the meaning of attribute's significance.

Example 1.

Given a set $X = \{x_1, x_2, x_3\}$. R_1, R_2, R_3 are fuzzy indiscernibility relations on X as following:

$$R_1 = \begin{bmatrix} 1 & 0.9 & 0 \\ 0.9 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ R_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.8 \\ 0 & 0.8 & 1 \end{bmatrix},\ R_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.9 \\ 0 & 0.9 & 1 \end{bmatrix}.$$

P is the probability distribution on X:

$$P(x_1) = P(x_2) = P(x_3) = 1/3.$$

Then

$$R_4 = R_1 \bigcap R_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ R_5 = R_1 \bigcap R_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let us compute the following Yager's entropy $H_P(R_1), H_P(R_2)$.

$$H_P(R_1) = -\sum_{x \in X} p(x) \log_2 \pi_{R_1}(x),$$

$$= 0.9676$$

$$H_P(R_2) = -\sum_{x\in X} p(x)\log_2 \pi_{R_2}(x) \\ = 1.0196 .$$

The joint entropy of relations R_1 and R_2 is

$$H_P(R_1R_2) = H_P(R_1 \cap R_2) \\ = -\sum_{x\in X} P(x)\log_2 \pi_{R_3}(x) . \\ = 1.5850$$

As has been pointed before, Yager's entropy can be interpreted as the discernibility power of a relation. The entropy of R_1 is less than that of R_2, so X is less discernible with respect o relation R_1. $H_P(R_1R_2)$ is larger than $H_P(R_1)$ or $H_P(R_2)$ because the introduction of a new attribute will enhance the discernibility power.

Then the conditional entropies $H_P(R_1 \mid R_2)$ and $H_P(R_2 \mid R_1)$ are

$$H_P(R_1 \mid R_2) = H_P(R_1R_2) - H_P(R_2) \\ = 1.5850 - 1.0196 \\ = 0.5654 ,$$

$$H_P(R_2 \mid R_1) = H_P(R_1R_2) - H_P(R_1) \\ = 1.5850 - 0.9676 \\ = 0.6174 ,$$

and

$$I_P(R_1;R_2) = I_P(R_2;R_1) \\ = H_P(R_1) + H_P(R_2) - H_P(R_1R_2) \\ = 0.9676 + 1.0196 - 1.5850 \\ = 0.4022 .$$

Given an indiscernibility relation R_d induced by a decision attribute:

$$R_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let us compute the relative entropy of R_1 and R_2 relative to R_d respectively.

$$H_P(R_1;R_d) = H_P(R_d) - H_P(R_1 \cup R_d) \\ = 1.5850 - 0.9676 \\ = 0.6174$$

$$H_P(R_2;R_d) = H_P(R_d) - H_P(R_2 \cup R_d) \\ = 1.5850 - 1.0196 \\ = 0.5654 .$$

The relative entropy measures the inconsistency between relations by condition attributes and decision attribute. The entropy value of R_1 relative to R_d is larger than that of R_2, so R_2 has more consistency to R_d.

The joint relative entropy of R_1 and R_2 relative to R_d is

$$\begin{aligned} H_P(R_1R_2;R_d) &= H_P(R_d) - H_P(R_1 \cap R_2 \cup R_d) \\ &= 1.5850 - 1.5850 \\ &= 0 \end{aligned},$$

The joint relative entropy of R_1 and R_3 relative to R_d is

$$\begin{aligned} H_P(R_1R_3;R_d) &= H_P(R_d) - H_P(R_1 \cap R_3 \cup R_d) \\ &= 1.5850 - 1.5850 \\ &= 0 \end{aligned}$$

which shows the joint relation of R_1 and R_2 has the same partition as the relation R_d, and so does the joint relation of R_1 and R_3. $\{R_1, R_2\}$ and $\{R_1, R_3\}$ are two relative D-reducts. R_1 is the relative core.

Example 2.

In order to test the method, a greedy algorithm is designed to compute reduct for hybrid data. Two approaches have been compared. One is to discretize the numeric attribute in advance, and then classical rough set reduction is performed on it. The other is to construct a fuzzy relation matrix for each numeric attribute and reduction based on fuzzy rough set approach is taken.

An experiment was performed on a real-word data set *Car* downloaded from the UCI repository. There are 946 objects with 18 numeric condition attributes *A1*, *A2*, ⋯, *A18* and a nominal decision attribute *d* with values *Opel, Saab, Bus* and *Van*. Original rough set approach and approach we present are tested on the data set, respectively.

Pawlak's rough set model just works in nominal attribute case. Hence, discretization is usually used as preprocessing on the data, and then a discretized information system will be induced. Here the value domain of each attribute was partitioned into four intervals. And then we computed reducts with classical rough set model. We get a reduct:

{A1, A2, A3, A4, A5, A6, A8, A9, A10, A11, A13, A14, A17, A18}.

The result shows that the superfluous attribute subset is

{A7, A12, A15, A16}

In second case, we first normalize the data to [0, 1], and then introduce triangle function to compute the fuzzy similarity relation between objects. Max-min transformation is performed to induce a fuzzy indiscerniblity relation from a similarity one. The Yager's fuzzy entropy and its extensions are used to measure discernibility of the relations.

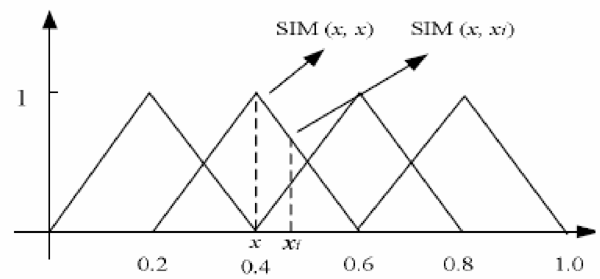

Fig.5. Triangle function similarity computation

Reduction algorithm give the result:

{A1, A2, A3, A4, A5, A6, A8, A10, A13, A15, A16, A17, A18}.

The superfluous attribute subset is

{A7, A9, A11, A12, A14}.

We introduce support vector machine to evaluate the reducts from two algorithms. Two thirds of samples are selected as training set, and others are test set. Classification accuracy by the full data without reduction is 79.56%. The result by original rough set approach is 79.87%, while the accuracy by the approach we present is 81.45%.

From the examples we can conclude that the proposed approach is suitable for fuzzy or numeric data reduction. And the approach gets a more concise reduct while better classification performance.

7. Conclusion

Yager's entropy, a generalization of Shannon' one, has been proposed to compute the information of fuzzy indiscernibility relation. In this paper we extend some definitions in Shannon's information theory to measure operations of fuzzy indiscernibility relations based on Yager's entropy, and apply these novel measures to fuzzy rough set model. As we interpreted, Yager's entropy measures the discernibility power of a fuzzy indiscernibility relation. We introduce joint entropy, conditional entropy and relative entropy to compute the information of the relation operations. Conditional entropy and relative conditional entropy are proposed to measure the information increment, which is interpreted as the significance of an attribute in fuzzy rough set model. From the view of information theory we redefine the reduct and relative reduct. The examples show these extensions are fit for rough set model.

Some conclusions are shown as follows:

(1) Yager's entropy can been interpreted as a measure of discernibility power of a fuzzy indiscernibility relation.
(2) Conditional entropy and relative conditional entropy measure the increment of discernibility power by introducing an unseen attribute, which can be taken as a significance measure of the attribute.

(3) The independence of attributes, reduct and relative reduct are redefined based on Yager's entropy.

References

1. R. Quinlan. Induction of decision trees. Machine learning, 1(1986) 81-106
2. J. Wang, D. Miao. Analysis on attribute reduction strategies of rough set. J. of computer science and technology. 13 (1998) pp189-193
3. U. M. Fayyad, K. B Irani. Multi-interval discrezation of continuous-valued attributes for classification learning. Proceeding of 13th international joint conference on artificial intelligence. (1993) pp.1022-1027.
4. J. Yao, M. Dash, S.T. Tan, H. Liu. Entropy-based fuzzy clustering and fuzzy modeling. Fuzzy Sets and Systems. 113 (2000) pp381-388
5. N. R. Pal, J. C. Bezdek. Measuring fuzzy uncertainty. IEEE Trans. Fuzzy Systems 2 (1994) pp107-118
6. N. R. Pal, J. C. Bezdek. Quantifying different facets of fuzzy uncertainty, in D. Dubas, H. Prade (Eds.), Fundamentals of fuzzy sets, Kluwer Academic Publishers, London, (2000) pp. 459-480.
7. J. L. Fan, Y. L. Ma. Some new entropy formulas. Fuzzy sets and systems. 128 (2002) 277-284
8. R. Yager. Entropy measures under similarity relations. International J. General Systems. 20 (1992) pp341-358
9. E. Hernandez, J. Recasens. A reformulation of entropy in the presence of indistinguishability operators. Fuzzy sets and systems. 128 (2002) pp185-196
10. H.D. Cheng, Yen-Hung Chen, Ying Sun. A novel fuzzy entropy approach to image enhancement and thresholding. Signal Processing 75 (1999) pp277-301
11. Ehud Lehrer, Rann Smorodinsky. Relative entropy in sequential decision problems. Journal of Mathematical Economics 33 (2000) pp425–439
12. D. Dubois, H. Prade. Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems. 17 (1990) pp191-207
13. D. Dubois, H. Prade. Putting rough sets and fuzzy sets together. Handbook of applications and advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht, (1991) pp.89-101
14. Q. Mohamed. α-RST: a generalization of rough set theory. Information Sciences Volume: 124 (2000) Issue: 1-4, pp301-316
15. Y. Wang. Mining stock price using fuzzy rough set system, Expert Systems with Applications. 24 (2003) Issue: 1, pp13-23
16. W. Wu, J. Mi, W. Zhang. Generalized fuzzy rough sets. Information science, vol. 151 (2003) pp263-282
17. Q. Shen, A. Chouchoulas. A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems. Engineering Applications of Artificial Intelligence. 13 (2000) pp263-278
18. S. Al-sharhan, F. Karray, W. Gueaieb, O. Basir. Fuzzy entropy: a brief survey. IEEE international fuzzy sets conference (2001) pp1135-1139
19. R. Jensen, Q. Shen. Fuzzy rough sets for descriptive dimensionality reduction. Proceedings of the 11th International Conference on Fuzzy Systems (2002) pp29-34