Available online at www.sciencedirect.com

ScienceDirect

Applied Soft Computing 7 (2007) 1135–1143

Applied Soft Computing

www.elsevier.com/locate/asoc

Short communication

Uncertainty measures for fuzzy relations and their applications

Daren Yu, Qinghua Hu *, Congxin Wu

Harbin Institute of Technology, Harbin 150001, PR China

Received 13 April 2006; received in revised form 6 August 2006; accepted 17 October 2006
Available online 21 November 2006

Abstract

Relations and relation matrices are important concepts in set theory and intelligent computation. Some general uncertainty measures for fuzzy relations are proposed by generalizing Shannon's information entropy. Then, the proposed measures are used to calculate the diversity quantity of multiple classifier systems and the granularity of granulated problem spaces, respectively. As a diversity measure, it is shown that the fusion system whose classifiers are of little similarity produces a great uncertainty quantity, which means that much complementary information is achieved with a diverse multiple classifier system. In granular computing, a ''coarse–fine'' order is introduced for a family of problem spaces with the proposed granularity measures. The problem space that is finely granulated will get a great uncertainty quantity compared with the coarse problem space. Based on the observation, we employ the proposed measure to evaluate the significance of numerical attributes for classification. Each numerical attribute generates a fuzzy similarity relation over the sample space. We compute the condition entropy of a numerical attribute or a set of numerical attribute relative to the decision, where the greater the condition entropy is, the less important the attribute subset is. A forward greedy search algorithm for numerical feature selection is constructed with the proposed measure. Experimental results show that the proposed method presents an efficient and effective solution for numerical feature analysis.

Keywords: Fuzzy similarity relation; Fuzzy equivalence relation; Uncertainty measure; Entropy

1. Introduction

Randomness of events, fuzziness of human's perception and lingual representation, and limitation of available knowledge, decision making with uncertainty becomes one of the most significant challenging tasks in artificial intelligence society [10]. Probability theory, fuzzy set theory [29], rough set theory [13], granular computing [33,35] and computing with words [32] were proposed to deal with this kind of problems. Just as the great statistician Rao said: uncertain knowledge + uncertainty measure = useful knowledge [15], uncertainty measures play an important role in artificial intelligence and reasoning with uncertainty for it is impossible to avoid uncertainty in decision making. Since Shannon introduced the information entropy to measure the uncertainty or the information quantity of a family of random events [16], a series of measures were proposed to compute the fuzziness of a fuzzy system, a fuzzy partition and a fuzzy random system [1,12,26,27]. These measures have been widely used in machine learning [14], attribute reduction [17–19], pattern recognition [22] and data clustering [23].

Relations, as a fundamental concept in mathematics, represent the connections of a set of elements in the domain. The two definitions of the set and relation constitute the basis of modern mathematics. Additionally, relations have been applied to discretize real-valued data, fuzzy clustering, uncertainty reasoning and decision. Equivalence relations [13], similarity relations, neighborhood relations [24], dominance relation [3] are the foundations of a sequence of rough set models. Zhang recently proposed a fuzzy reasoning model under the fuzzy equivalence relation and the fuzzy quotient space structure [34]. Fuzzy equivalence relations are used to discretize real-valued attributes and fast hierarchical clustering [8].

In the classical set theory, the relations take values in the set {0, 1}; however, the value domain was generalized to the interval [0, 1] in the fuzzy theory. The fuzziness of relations is the essential characteristic in some cases, such as the equivalence relation of a group of people in terms of their appearances, the similarity relation of a set of documents based on their meaning and the diversity of a family of classifiers in multiple classifier systems. Due to the important role of fuzzy relations in fuzzy logic and granular computing, some

* Corresponding author.
E-mail address: huqinghua@hcms.hit.edu.cn (Q. Hu).

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2006.10.004

researchers have begun to discuss how to measure and utilize the uncertainty of fuzzy relations. Yager introduced an uncertainty measure for similarity relations and discussed its application to questionnaire design [25]. Hernandez and Recasens extended Yager's work and presented the formulae of joint entropy and conditional entropy based on the measure, then used the measures to learn fuzzy decision trees from a set of data samples [4]. In [6] Hu and Yu redefined the joint entropy and conditional entropy based on Yager's work. He extended measures and then successfully used them to reduce hybrid attribution [8] and measure uncertainty of fuzzy probability approximation spaces [9]. In [11] Mi et al. introduced an uncertainty measure for partition based fuzzy-rough set model.

In this paper, we will further generalize the notions given in [25,6] and present more general uncertainty measures for fuzzy binary relations. We compare the proposed measures with some existing functions and present some potential applications. The rest of the paper is organized as follows. Section 2 reviews the basic notions about crisp and fuzzy relations. Some novel uncertainty measures are presented in Section 3. Then, two applications (diversity for multiple classifiers systems and granularity for granular computing) of the proposed measures are given in Sections 4 and 5, respectively. Section 6 presents a greedy algorithm for numerical feature selection and experimental results. Finally, Section 7 concludes the paper.

2. Preliminary notions on crisp and fuzzy relations

Let X, Y be two sets, and any subset of the Cartesian product $X \times Y$ is called a binary relation, denoted by R. $\forall x \in X, y \in Y$, if $\langle x, y\rangle \in R$, we say that x and y satisfy relation R, marked with xRy; otherwise, $\langle x, y\rangle \notin R$, we say x, y do not satisfy relation R. Specially, if $X = Y$, R is called a binary relation on X.

Let $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_m\}$, $R = X \times Y$.

$$r_{ij} = \begin{cases} 1, \langle x_i, y_j\rangle \in R, & i = 1, \ldots, n, \quad j = 1, \ldots, m; \\ 0, \langle x_i, y_j\rangle \notin R, & i = 1, \ldots, n, \quad j = 1, \ldots, m. \end{cases}$$

A binary relation can be represented as a matrix. And the matrix $M_R = (r_{ij})_{n\times m}$ is called the relation matrix of R.

Let R be a relation on a non-empty set X. $\forall x, y, z \in X$, if the relation satisfies: (1) reflexivity: xRx; (2) symmetry: $xRy \Rightarrow yRz$, R is called a similarity relation. Furthermore, if R satisfies transitivity: $xRy, yRz \Rightarrow xRz$, then we say that R is an equivalence relation. In this paper, similarity relations are denoted as S and equivalence relations E.

Given a set X and a similarity relation S on X, $\forall x \in X$, a set $[x]_S$, called similarity class of x induced by relation S, is associated, where $[x]_S = \{x_i \mid \langle x_i, x\rangle \in S, x_i \in X\}$. E is an equivalence on X, corresponding, $\forall x \in X$, we define an equivalence subset of x with respect to relation E, denoted by $[x]_E$, where $[x]_E = \{x_i \mid \langle x_i, x\rangle \in E, x_i \in X\}$.

Let X be a set, and $\coprod$ is a family of sets constituting with non-empty subset of X, namely, $\coprod = \{\coprod_\lambda \mid \coprod_\lambda \neq \emptyset, \coprod_\lambda \subseteq X, \lambda \in \aleph\}$, where $\aleph$ is a subscript set. If $\forall x \in X$, there is $\lambda_x \in \aleph$ so that $x \in \coprod_{\lambda_x}$, we call the set family is a cover of X. Its easy to show $\cup_{\lambda\in\aleph}\coprod_\lambda = X$ if $\coprod$ is a cover of X. What's more, for any λ_1, λ_2 in $\aleph$, if we have $\coprod_{\lambda_1} \cap \coprod_{\lambda_2} = \emptyset$, then $\coprod$ is a partition of X. Given any set X, we say $\coprod = \{X\}$ is the coarsest partition and $\Pi = \{\{x_i\} \mid x_i \in X\}$ is the finest partition. Let $\coprod_1 = \{\coprod_\lambda \mid \lambda \in M\}$ and $\Pi = \{\Pi_\gamma \mid \gamma \in N\}$ be two partitions of set X. $\forall \lambda_0 \in M$, $\exists \gamma_0 \in N$, so that $\coprod_{\lambda_0} \subseteq \prod_{\gamma_0}$, then we say the partition $\coprod$ is finer than Π, denoted by $\coprod \prec \Pi$. Let $\coprod_1 = \{\coprod_\lambda \mid \lambda \in M\}$ and $\Pi = \{\Pi_\gamma \mid \gamma \in N\}$ be two partitions of set X, then

$$U = \{\coprod_\lambda \cap \prod_\gamma | \coprod_\lambda \in \coprod, \prod_\gamma \in \prod, \coprod_\lambda \cap \prod_\lambda \neq \emptyset\}$$

is also a partition of X. Moreover, $U \prec \coprod$, $U \prec \Pi$.

In the classical set theory, the relation R takes values in {0, 1}. In this case the relation matrix is a Boolean matrix. In the fuzzy set theory, relations take values in the interval [0, 1], namely $r \in [0, 1]$; the grades of the relations mean the strength that the elements satisfy the relation. In the fuzzy logic context, the properties of fuzzy relations are defined as (1) reflexivity: $R(x, x) = 1$; (2) symmetry: $R(x, y) = R(y, x)$; (3) T-transitivity: $R(x, z) \geq \mathrm{MAX}_y(R(x, y) \wedge R(y, z))$, where $\wedge$ is a t-norm. In this paper, we consider ''min'' as the t-norm. The relation is called a fuzzy similarity relation if it satisfies the conditions of reflexivity and symmetry. The relation is called a T-indistinguishability relation or a fuzzy equivalence relation if a similarity relation satisfies the t-transitivity.

Remark. In some literatures, fuzzy equivalence relations are also called similarity relations [15,20]. Here, we will use the four concepts: similarity relation, fuzzy similarity relation, equivalence relation and fuzzy equivalence relation for distinguishing.

If R and S are two fuzzy relations on X, some operators are defined as

(1) Union: $(R \cup S)(x, y) = \max\{R(x, y), S(x, y)\}, \forall x, y \in X$;
(2) Intersection: $(R \cap S)(x, y) = \min\{R(x, y), S(x, y)\}, \forall x, y \in X$;
(3) Containment: $R \subseteq S \Rightarrow R(x, y) \leq S(x, y), \forall x, y \in X$.

Given a fuzzy relation R on X, $\forall \alpha \in [0, 1]$, the α-cuts R_α of the fuzzy relation is a crisp relation, where

$$R_\alpha(x, y) = \begin{cases} 1, & R(x, y) \geq \alpha, \\ 0, & R(x, y) < \alpha. \end{cases}$$

R is a fuzzy equivalence relation if and only if the α-cuts R_α of R is a crisp equivalence relation for all $\alpha \in [0, 1]$. Given a finite set X and a fuzzy equivalence relation R, the fuzzy equivalence class $[x_i]_R$ of $x_i \in X$ is a fuzzy subset, where $[x_i]_R$ is defined as

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

where $r_{ij} = R(x_i, x)$. The fuzzy equivalence class is a fuzzy information granule, the elements in the class are fuzzy indiscernible with x_i; r_{ij} means the grade how the two elements are equivalent or indiscernible. The family of the fuzzy equivalence classes $[x_i]_R$, written as $X/R = \{[x_i]_R \mid x_i \in R\}$, is called a fuzzy quotient set of X induced by R.

In order to distinguish fuzzy sets and fuzzy relations from crisp sets and relations, we will write crisp sets as A, B, X, etc., and crisp relations as R, S, etc. Fuzzy sets and fuzzy relations are written as $\tilde{A}$, $\tilde{B}$, $\tilde{X}$ and $\tilde{R}, \tilde{S}$, etc.

3. Uncertainty measures for fuzzy relations

In this section, we will introduce a novel uncertainty measure for fuzzy binary relations.

3.1. The case without probability distributions

Definition 1. Given a finite set X and a fuzzy binary relation $\tilde{R}$ on X, we define a fuzzy class $[x_i]_{\tilde{R}}$ of $x_i \in X$ induced by the relation $\tilde{R}$, where $[x_i]_{\tilde{R}} = (r_{i1}/x_1) + (r_{i2}/x_2) + \cdots + (r_{in}/x_n)$. $[x_i]_{\tilde{R}}$ is a fuzzy set, and the fuzzy cardinal number of $[x_i]_{\tilde{R}}$ is defined as [31]:

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

For a finite set X, $\forall x_j \in X$ we have $r_{ij} \leq 1$, then the cardinality of $[x_i]_{\tilde{R}}$ is also finite and $|[x_i]_{\tilde{R}}| \leq n$.

Definition 2. Let X be a finite set and $\tilde{R}$ a fuzzy relation on X. The fuzzy relation class of x is $[x_i]_{\tilde{R}}$, then, the expected cardinality of $[x_i]_{\tilde{R}}$ is computed as

$$\overline{\mathrm{Card}}([x_i]_{\tilde{R}}) = \frac{|[x_i]_{\tilde{R}}|}{|X|} = \frac{\sum_{j=1}^{n} r_{ij}}{n}$$

The expected cardinality $\overline{\mathrm{Card}}([x_i]_{\tilde{R}}) \leq 1$, which can be understood as the ratio of $[x_i]_{\tilde{R}}$ in X.

Definition 3. The uncertainty quantity of the fuzzy relation class $[x_i]_{\tilde{R}}$ is defined as

$$U([x_i]_{\tilde{R}}) = -\log_2 \overline{\mathrm{Card}}([x_i]_{\tilde{R}}).$$

As $\overline{\mathrm{Card}}([x_i]_{\tilde{R}}) \leq 1$, we have $U([x_i]_{\tilde{R}}) \geq 0$ and, $U([x_i]_{\tilde{R}})$ decreases monotonously with the increase of $\overline{\mathrm{Card}}([x_i]_{\tilde{R}})$.

Definition 4. Given a finite set X and A fuzzy relation $\tilde{R}$ on X, we calculate the average uncertainty quantity $H(\tilde{R})$ of the fuzzy relation with

$$H(\tilde{R}) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \overline{\mathrm{Card}}([x_i]_{\tilde{R}}).$$

The average uncertainty quantity of the fuzzy relation R on X is a mapping $H : (X, \tilde{R}) \rightarrow \mathfrak{R}^{+}$, where $^{+}$ is the domain of non-negative real numbers. With the mapping we form an order to compare the fuzzy relations with respect to the uncertainty quantity. Note that the uncertainty quantity is not only a function of the fuzzy relation R, but also related to the set X.

We define $H(\tilde{R}) = 0$ if $X = \varnothing$.

Proposition 1. *Given a non-empty and finite set X and a relation R on X, if $\forall x, y \in X$, $R(x, y) = 1$, then we have*

$$H(\tilde{R}) = 0.$$

Proposition 2. *Let $\tilde{R}_1$ and $\tilde{R}_2$ be two fuzzy relations on a non-empty and finite set X, we have*

$$\tilde{R}_1 \subseteq \tilde{R}_2 \Rightarrow H(\tilde{R}_1) \geq H(\tilde{R}_1).$$

Proposition 3. *Let $\tilde{R}_1$ and $\tilde{R}_2$ be two fuzzy relations on a non-empty and finite set X, we have*

$$H(\tilde{R}_1 \cap \tilde{R}_2) \geq \max(H(\tilde{R}_1), H(\tilde{R}_2)),$$

$$H(\tilde{R}_1 \cup \tilde{R}_2) \geq \min(H(\tilde{R}_1), H(\tilde{R}_2)).$$

3.2. The case with probability distributions

The probability measure of fuzzy sets was introduced by Zadeh [30]. It is calculated as membership-weighted sum of the probabilities of the elements in the fuzzy set. Fuzzy relation classes are fuzzy sets induced by the fuzzy relation. Therefore, similarly, the probability of a fuzzy relation class can be defined as follows.

Definition 5. Let P and $\tilde{R}$ be a probability distribution and a fuzzy relation on a given non-empty and finite set X, respectively. $\forall x_i \in X$, the expect of cardinality of the fuzzy relation class $[x_i]_{\tilde{R}}$ is defined as

$$\overline{\mathrm{Card}}([x_i]_{\tilde{R}}) = \sum_{j=1}^{n} p_j \times r_{ij}.$$

As we know $0 \leq p_j \leq 1$, $\sum_j p_j = 1$ and $r_{ij} \leq 1$, we have $\overline{\mathrm{Card}}([x_i]_{\tilde{R}}) \leq 1$. However, $\sum_i \overline{\mathrm{Card}}([x_i]_{\tilde{R}}) = 1$ does not hold in all cases. So P is not a probability distribution on $[x_i]_{\tilde{R}}$, and we have

$$1 \leq \sum_i \overline{\mathrm{Card}}([x_i]_{\tilde{R}}) \leq n.$$

Definition 6. Let P be a probability distribution on X, and $\tilde{R}$ is a fuzzy relation. The uncertainty quantity of a fuzzy relation class $[x_i]_{\tilde{R}}$ is defined as

$$U([x_i]_{\tilde{R}}, P) = -\log_2 \overline{\mathrm{Card}}([x_i]_{\tilde{R}}) = -\log_2 \sum_j p_j \times r_{ij}.$$

Definition 7. Let P and $\tilde{R}$ be a probability distribution and a fuzzy relation on a given non-empty and finite set X, respectively. We denote $H(\tilde{R}, P)$ as the expected uncertainty quantity of the fuzzy relation, where $H(\tilde{R}, P)$ is computed as

$$H(\tilde{R}, P) = -\sum_{i=1}^{n} p_i \log_2 \overline{\mathrm{Card}}([x_i]_{\tilde{R}}).$$

Proposition 4. *If $\forall x_i \in X$, $p(x_i) = 1/n$, we have*

$$H(\tilde{R}, P) = H(\tilde{R}).$$

Proposition 4 shows that $H(\tilde{R}, P)$ is a general case of $H(\tilde{R})$. If the elements are equally probable, $H(\tilde{R}, P)$ will degrade to $H(\tilde{R})$.

Proposition 5. *Let P and $\tilde{R}$ be a probability distribution and a fuzzy relation on a given non-empty and finite set X, respectively. Denoting H(P) as Shannon's uncertainty measures introduced by the probability distribution, we have*

$$H(\tilde{R},P) \leq H(P).$$

Proof.

$$\begin{aligned} &H(\tilde{R},P) - H(P) \\ &= -\sum_{i=1}^{n} p_i \log_2 \overline{\text{Card}}([x_i]_{\tilde{R}}) - \left(-\sum_{i=1}^{n} p_i \log_2 p_i\right) \\ &= -\sum_{i=1}^{n} p_i \log_2 \left(\frac{\overline{\text{Card}}([x_i]_{\tilde{R}})}{p_i}\right)\end{aligned}$$

As $\overline{\text{Card}}([x_i]_{\tilde{R}}) = \sum_{j=1}^{n} p_j \times r_{ij} = p_i + \sum_{j=1,\ldots,n, j\neq i} p_j \times r_{ij} \geq p_i$, we have $\overline{\text{Card}}([x_i]_{\tilde{R}})/p_i \geq 1$ and $-\sum_{i=1}^{n} p_i \log_2 (\overline{\text{Card}}([x_i]_{\tilde{R}})/p_i) \leq 0$.

Therefore, $H(\tilde{R},P) \leq H(P)$. □

Proposition 6. *Let $\tilde{R}$ be a reflexive fuzzy relation on X and P be a probability distribution, where $P_k = 1$ for some element x_k and $P_i = 0$ for $i \neq k$. Then, $H(\tilde{R},P) = 0$.*

Proposition 7. *Given a probability distribution P and a fuzzy relation $\tilde{R}$ on a non-empty and finite set X, if $\forall x, y \in X$, $R(x, y) = 1$, then, we have $H(\tilde{R},P) = 0$.*

Proof.

$$H(\tilde{R},P) = -\sum_{i=}^{n} p_i \log_2 \overline{\text{Card}}([x_i]_{\tilde{R}}) = -\sum_{i=1}^{n} p_i \log_2 \sum_{j=1}^{n} p_j \times r_{ij}$$

$$\because \forall i, j, r_{ij} = 1$$

$$\therefore H(\tilde{R},P) = -\sum_{i=1}^{n} p_i \log_2 \sum_{j=1}^{n} p_j = -\sum_{i=1}^{n} p_i \log_2 1 = 0$$

□

Proposition 8. *Given two fuzzy relations $\tilde{R}_1, \tilde{R}_2$ and a probability distribution P on X, we have*

(1) $\tilde{R}_1 \subseteq \tilde{R}_2 \Rightarrow H(\tilde{R}_1,P) \geq H(\tilde{R}_2,P)$;
(2) $\tilde{R} = \tilde{R}_1 \cap \tilde{R}_2 \Rightarrow H(\tilde{R},P) \geq \max(H(\tilde{R}_1,P), H(\tilde{R}_2,P))$;
(3) $\tilde{R} = \tilde{R}_1 \cup \tilde{R}_2 \Rightarrow H(\tilde{R},P) \geq \min(H(\tilde{R}_1,P), H(\tilde{R}_2,P))$.

Proof. (1) $H(\tilde{R},P) = -\sum_{i=1}^{n} p_i \log_2 \overline{\text{Card}}([x_i]_{\tilde{R}})$. *If* $\tilde{R}_1 \subseteq \tilde{R}_2$, *then* $\overline{\text{Card}}([x_i]_{\tilde{R}_1}) \leq \overline{\text{Card}}([x_i]_{\tilde{R}_1})$.

$$-\sum_{i=1}^{n} p_i \log_2 \overline{\text{Card}}([x_i]_{\tilde{R}_1}) \geq -\sum_{i=1}^{n} p_i \log_2 \overline{\text{Card}}([x_i]_{\tilde{R}_2}).$$

We have $H(\tilde{R}_1,P) \geq H(\tilde{R}_2,P)$ (2) and (3) share the similar proof as (1). □

The uncertainty quantity $H(\tilde{R},P)$ measures the uncertainty of a general fuzzy binary relation in a probability space. It is a compound index of uncertainty of fuzziness and randomness. This measure forms a basis for comparing the strength of the relation.

As fuzzy similarity and equivalence relations are widely applied in the applications, fuzzy similarity relations are used in text categorization, multiple classifier systems and hyperspectral image fusion [19–21]. Fuzzy equivalence relations are the basis of fuzzy quotient reasoning [35], fuzzy rough set and fuzzy rule extraction, and we will discuss them in detail below.

4. Measuring diversity for information fusion

In multiple sensor fusion, linear correlation coefficient is usually used to measure the correlation of different sensor data [20,21]. As there are n sensors we can get a $n \times n$ matrix of correlation coefficient $M(R) = (r_{ij})_{n\times n}$, where

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)^2}\sqrt{\sum_{k=1}^{n}(x_{jk} - \bar{x}_j)^2}}.$$

According the properties of correlation coefficient, we have

(1) $r_{ii} = 1$;
(2) $r_{ij} = r_{ji}$;
(3) $0 \leq |r_{ij}| \leq 1$.

$|r_{ij}|$ represents the similarity between the ith sensor and the jth sensor. Generally, speaking, the information introduced by different sensors is expected to be diverse and complementary in information fusion, so $|r_{ij}|$ should be a little value.

In [5], Ho introduced an algorithm to calculate the agreement of multiple classifiers in multiple classifier systems. Given a set of n fixed samples and the same weights, the agreement of classifiers can be written as

$$S_{ij} = \frac{1}{n}\sum_{k=1}^{n} f(X_k),$$

where

$$f(X_k) = \begin{cases} 1, & \text{if } C_i(X_k) = C_j(X_k) \\ 0, & \text{otherwise} \end{cases}$$

$C_i(X_k) = C_j(X_k)$ means the output of ith classifier is the same as that of jth classifier with respect to sample X_k. The agreement or similarity degree of two classifiers here is the ratio of the same outputs to all of the outputs. It is easily seen that $0 \leq S_{ij} \leq 1$, and the matrix $(S_{ij})_{n\times n}$ satisfies reflexivity and symmetry for $S_{ii} = 1$ and $S_{ij} = S_{ji}$.

Here, we unify the representation of the similarity of a family of information as a fuzzy similarity matrix, written as

$$S = \begin{bmatrix} 1 & s_{12} & \cdots & s_{1n} \\ s_{21} & 1 & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & 1 \end{bmatrix} = I + \overline{S}.$$

where $0 \leq S_{ij} \leq 1$ and $S_{ij} = S_{ji}$.

Here we do not discuss how we can get the matrix, but study the measure of uncertainty of the matrix and the interpretation of the measure. Matrix I stands for the auto-relation matrix and $\bar{S}$ for the overlap information of the fused system.

Definition 8. Let $W = (w_1, w_2, \ldots, w_n)$ be the weight vector of the sensors or classifiers in the fusing system, such that $0 \le w_i \le 1$ and $\sum_i w_i = 1$. Then, the information quantity is defined as

$$H(\tilde{S}, W) = -\sum_{i=1}^{n} w_i \log_2 \overline{\mathrm{Card}}([x_i]_{\tilde{S}}),$$

where $\overline{\mathrm{Card}}([x_i]_{\tilde{S}}) = \sum_{j=1}^{n} w_j \times s_{ij}$.

Proposition 5 shows that the fused systems will bring no information gain if one of the sensor or classifiers is selected while others is excluded in the fusing system. Proposition 6 shows that there will not be any information increment if the sensors or classifiers are completely identical in the fusing systems. According to Proposition 7 we have $H(\tilde{S}_1, W) \ge H(\tilde{S}_2, W)$ if $\tilde{S}_1 \subseteq \tilde{S}_2$. We can understand that $\tilde{S}_1 \subseteq \tilde{S}_2$ means the sensors or classifiers in the fusing system $\tilde{S}_1$ are more independent than those in $\tilde{S}_2$, therefore, as a whole, the fusing system $\tilde{S}_1$ can provide more information than $\tilde{S}_2$. In practice, a multiple classifier system with little dependence has greater opportunity to achieve a good classification performance [2,7]. Therefore, it is reasonable to select some of the classifiers with great uncertainty quantity in multiple classifier ensembles.

Example 1. Given two classifier systems $C = \{C_1, C_2, C_3, C_4\}$ and $C' = \{C'_1, C'_2, C'_3, C'_4\}$, firstly we consider they take the same weight in the ensemble system and the weight vector $W = (1/4, 1/4, 1/4, 1/4)$. The similarity matrices of two ensembles are

$$M(C) = \begin{bmatrix} 1 & 0.8 & 0.2 & 0.1 \\ 0.8 & 1 & 0.8 & 0.2 \\ 0.2 & 0.8 & 1 & 0.7 \\ 0.1 & 0.2 & 0.7 & 1 \end{bmatrix},$$
$$M(C') = \begin{bmatrix} 1 & 0.7 & 0.3 & 0.1 \\ 0.7 & 1 & 0.6 & 0.2 \\ 0.3 & 0.6 & 1 & 0.9 \\ 0.1 & 0.2 & 0.9 & 1 \end{bmatrix}.$$

By computing we get $H(M(C), W) = 0.7528$, $H(M(C'), W) = 0.7462$. The results show the second ensemble is more diverse than the first one.

We consider a new weight vector $W_1 = (1/8, 1/8, 3/8, 3/8)$. Then, $H(M(C), W_1) = 0.6473$, $H(M(C'), W_1) = 0.5633$. In this case, $H(M(C), W_1) > H(M(C'), W_1)$, so we should choice the first ensemble.

Note that we do not consider the influence of the quality of the individual classifier here.

Comparing the proposed measure with the correlation entropy introduced in [20], we can find the correlation entropy does not refer to the weight vector. However, the weighted fusions or ensembles are widely used [2].

5. Measuring granularity in granular computing

Granulation plays a key role in human cognition and reasoning. Granulation of an object set X leads to a collection of granules of X, which are drawn together by indistinguishability, similarity, proximity or functionality [32,33]. In classical rough set model, information granules are the crisp equivalence classes induced by an indiscernibility relation [28]. Zadeh suggested although models of information granulation in which the granules are crisp play important roles in a wide variety of methods, the fuzziness of granules, their attributes and their values is characteristic of the ways in which human concepts are formed, organized and manipulated. Fuzzy information granulation underlies the remarkable human ability to make rational decisions in an unknown or uncertain environment. Zhang [35] presented a theoretical framework of fuzzy reasoning model under quotient space structure. Here, we will introduce this work as the basic model of granular computing.

A problem space is described as a triple $\Psi = (X, \Im, f)$, where X is the universe, $\Im$ is the structure of universe X and f denotes the attributes or knowledge of the universe X. f induces a family of equivalence classes of the universe, correspondingly builds a structure of the universe. If f is crisp knowledge, the equivalence classes are crisp; otherwise, the equivalence classes are fuzzy.

Definition 9. Given two problem spaces $\Psi_1 = (X, \Im_1, f_1)$ and $\Psi_2 = (X, \Im_2, f_2)$, $\tilde{E}_1$ and $\tilde{E}_2$ are two fuzzy equivalence relations generated by the knowledge f_1 and f_2, respectively. We say Ψ_1 is finer than Ψ_2 if $\tilde{E}_1 \subseteq \tilde{E}_2$, denoted by $\Psi_1 \prec \Psi_2$.

This relation is called a ‘‘coarse–fine’’ relation on the problem space. Its easily shown that this relation satisfies transitivity, namely, If $\Psi_1 \prec \Psi_2$, $\Psi_2 \prec \Psi_3$, we have $\Psi_1 \prec \Psi_3$. Define two basic granular spaces: the F-space and C-space as follows.

Definition 10. Given a problem space $(X, \Im, f)$, the knowledge f generates the relation matrix $(r_{ij})_{n\times n}$ of the elements in the space, where n is the cardinality of X. If the matrix satisfies $r_{ij} = 1$ for $i = j$, otherwise $r_{ij} = 0$, we say the space is the finest problem space, denoted by F-space or Ψ_F.

In F-space, all elements are discernible. There is no uncertainty in this case, which can be understood as a white box.

Definition 11. Given a problem space $(X, \Im, f)$, the knowledge f generates the relation matrix $(r_{ij})_{n\times n}$ of the elements. If the matrix satisfies that $\forall i, j$, $r_{ij} = 1$, we say the space is the coarsest problem space, denoted by C-space or Ψ_C.

Obviously, we have $\Psi_F \prec \Psi_C$. F-space and C-space are two limits of the problem space. If we say F-space is a white box of the problem, C-space can be considered as a black box. Other spaces between Ψ_F and Ψ_C are grey boxes.

Definition 12. Given a problem space $\Psi = (X, \Im, f)$, the knowledge f generates a fuzzy equivalence relation on the universe, and the relation matrix $\tilde{E} = (r_{ij})_{n\times n}$, the granularity of the problem space is defined as

$$G(\Psi) = H(\tilde{E}) = -\sum_{i=1}^{n}\frac{1}{n}\log_2 \overline{\text{Card}}([x_i]_{\tilde{R}}),$$

where $||[x_i]_{\tilde{E}}||$ is the relative cardinality of the fuzzy information granule $[x_i]_{\tilde{E}}$, and n is the cardinality of the universe. Given a universe X with n elements, according to Propositions 2 and 3, we have

(1) $\Psi_1 \prec \Psi_2 \Rightarrow G(\Psi_1) > G(\Psi_2)$;
(2) $G(\Psi_C) = 0; \quad G(\Psi_F) = \log_2 n$;
(3) $\forall \Psi_i: \quad G(\Psi_F) \geq G(\Psi_i) \geq G(\Psi_C)$

Proposition 9. *Given a problem space $\Psi = (X, \Im, f)$ with N objects, f generates a crisp equivalence relation $E = (r_{ij})_{N\times N}$ on X. The n distinct equivalence classes are written as $X_1, X_2, \ldots, X_n$. We define a probability distribution on X_1, where $p(X_i) = |X_i|/|X|$. Shannon's information entropy is denoted as $H(P)$, then we have $H(P) = H(E)$.*

Proposition 9 shows Shannon's entropy can be used as an information granularity of crisp problem spaces, and it also shows the relationship between the proposed uncertainty measure and Shannon's entropy.

Example 2. $\Psi = (X, \Im, f)$ is a problem space with three objects, f induces a crisp equivalence relation

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

Two equivalence classes $X_1 = \{x_1, x_2\}$ and $X_2 = \{x_3\}$ are induced by the relation, $P(X_1)=2/3$, $P(X_2)=1/3$.

$$H(P) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.9183,$$

$$H(E) = -\frac{1}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.9183.$$

We have $H(P) = H(E)$.

Definition 13. $\Psi = (X, \Im, f, P)$ is a problem space with n objects, P is a probability distribution. f generates a fuzzy equivalence relation $\tilde{E}$ on X. The information granularity of the problem space is defined as

$$G(\Psi) = H(\tilde{E}, P) = -\sum_{i=1}^{n} p_i \log_2 \sum_{j=1}^{n} p_j \times r_{ij}.$$

Definition 13 presents an algorithm to measure the uncertainty quantity of a fuzzy granular world in probability spaces. Given two spaces $\Psi_1 = (X, \Im_1, f_1, P)$ and $\Psi_2 = (X, \Im_2, f_2, P)$, if $\tilde{E}_1 \subset \tilde{E}_2$ we have $G(\Psi_1) > G(\Psi_2)$. Its worth noting if $G(\Psi_1) > G(\Psi_2)$ we do not have the conclusion $\tilde{E}_1 \subset \tilde{E}_2$.

Proposition 10. *Given $\Psi = (X, \Im, f, P)$, f generates a fuzzy equivalence relation $\tilde{E}$ on X. For any $0 \leq \alpha_1 \leq \alpha_2 \leq 1$, we have $\tilde{E}_{\alpha_1} \subseteq \tilde{E}_{\alpha_2}$. $\Psi_1 = (X, \Im_1, f_1, P)$ and $\Psi_2 = (X, \Im_2, f_2, P)$ are two problem spaces induced by crisp equivalence relations $\tilde{E}_{\alpha_1}, \tilde{E}_{\alpha_2}$, then we have*

$$\Psi_2 \prec \Psi_1, \qquad G(\Psi_1) \leq G(\Psi_2).$$

Example 3. X is a universe with five elements, and f_1 and f_2 are two knowledge sets about the universe X. They generate two problem spaces $\Psi_1 = (X, \Im_1, f_1, P)$ and $\Psi_2 = (X, \Im_2, f_2, P)$, the probability distribution and two fuzzy equivalence relations generated by f_1 and f_2 are as follows:

$$P = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.15 & 0.05 & 0.1 & 0.6 & 0.1 \end{pmatrix},$$

$$\tilde{E}_1 = \begin{bmatrix} 1 & 0.4 & 0.8 & 0.5 & 0.5 \\ 0.4 & 1 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.4 & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{bmatrix},$$

$$\tilde{E}_2 = \begin{bmatrix} 1 & 0.8 & 0.8 & 0.2 & 0.8 \\ 0.8 & 1 & 0.85 & 0.2 & 0.85 \\ 0.8 & 0.85 & 1 & 0.2 & 0.9 \\ 0.2 & 0.2 & 0.2 & 1 & 0.2 \\ 0.8 & 0.85 & 0.9 & 0.2 & 1 \end{bmatrix}.$$

The fuzzy equivalence classes of x_1 with respect to two spaces are

$$[x_1]_{\tilde{E}_1} = \left\{\frac{1}{x_1} + \frac{0.4}{x_2} + \frac{0.8}{x_3} + \frac{0.5}{x_4} + \frac{0.5}{x_5}\right\},$$

and

$$[x_1]_{\tilde{E}_2} = \left\{\frac{1}{x_1} + \frac{0.8}{x_2} + \frac{0.8}{x_3} + \frac{0.2}{x_4} + \frac{0.8}{x_5}\right\},$$

respectively.

First, we do not take the probability into account. The cardinalities and relative cardinalities of the two fuzzy classes are

$$|[x_1]_{\tilde{E}_1}| = 3.4; \qquad |[x_1]_{\tilde{E}_2}| = 3.6; \qquad ||[x_1]_{\tilde{E}_1}|| = \frac{3.4}{5} = 0.68;$$
$$||[x_1]_{\tilde{E}_2}|| = 0.72.$$

The information granularities of two spaces are

$$G(\Psi_1) = H(\tilde{E}_1) = -\sum_{i=1}^{5}\frac{1}{5}\log_2 ||[x_i]_{\tilde{E}_1}|| = 0.7235;$$

$$G(\Psi_2) = H(\tilde{E}_2) = -\sum_{i=1}^{5}\frac{1}{5}\log_2 ||[x_i]_{\tilde{E}_2}|| = 0.6245.$$

Now, lets take the probability distribution into consideration. The fuzzy probabilities of two fuzzy equivalence classes

$[x_1]_{\tilde{E}_1}$ and $[x_1]_{\tilde{E}_2}$ are

$$P([x_1]_{\tilde{E}_1}) = 0.15 \times 1 + 0.05 \times 0.4 + 0.1 \times 0.8 + 0.6 \times 0.5 + 0.1 \times 0.5 = 0.60;$$

$$P([x_1]_{\tilde{E}_2}) = 0.15 \times 1 + 0.05 \times 0.8 + 0.1 \times 0.8 + 0.6 \times 0.2 + 0.1 \times 0.8 = 0.47.$$

The information granularities of the two spaces are

$$G(\Psi_1) = H(\tilde{E}_1, P) = -\sum_{i=1}^{5} p_i \log_2 \sum_{j=1}^{5} p_j \times r_{ij} = 0.5078;$$

$$G(\Psi_2) = H(\tilde{E}_2, P) = -\sum_{i=1}^{5} p_i \log_2 \sum_{j=1}^{5} p_j \times r_{ij} = 0.7770.$$

We can find that the probability distributions change the order of granularities of two problem spaces. In the second space, element x_4 is with little similarity with other elements, however, the probability of x_4 is the greatest one. Therefore, the granularity of the second space is greater than that of the first one.

6. Applications and experimental analysis

Shannon's entropy was widely used in feature selection algorithms for categorical data [18,19], where categorical features generate a partition of the sample space, then entropy is used to measure the consistency between the partitions by condition and decision, respectively [19]. However, it cannot be used to measure numerical and fuzzy features. Based on numerical features, Hu et al. presented a way directly generating fuzzy similarity relations on the data set, and then transformed the similarity relations into fuzzy equivalence ones [8]. However, it is time-consuming to conduct the transform, especially when data sets are of great size. Here, we show a numerical feature selection technique directly with similarity relations.

Definition 14 (*attribute significance*).
Give a problem space $\langle U, C, d\rangle$, where U is the set of samples; C the set of numerical condition attributes; d is the decision variable. $\forall a \in C$, a induces a fuzzy similarity relation over samples R_a. R_d is the relation generated by d. Then, we define the significance of a relative to d as the condition entropy

$$H(d|a) = H(R_d \cap R_a) - H(R_a).$$

Accordingly, the significance of a subset of attributes B is defined as

$$H(d|B) = H(R_d \cap R_B) - H(R_B),$$

where $R_B = \cap_{a_i \in B} R_{a_i}$. Obviously, $H(d|B) \geq 0$ and $H(d|B) \geq H(d|C)$. If $R_B = R_d$, we have $H(d|B) = 0$.

Definition 15 (*relative attribute significance*).
Give a problem space $\langle U, C, d\rangle$, $B \subseteq C$, $\forall a \in C - B$, we define the relative significance of a to B as

$$\mathrm{SIG}(a, B, d) = H(d|B) - H(d|B \cup a).$$

Table 1
Data description

Data set	Abbreviation	Samples	Numerical features	Categorical features	Classes
1. Ionosphere	Iono	351	34	0	2
2. Sonar, mines vs. rocks	Sonar	208	60	0	2
3. Small soybean	Soy	47	35	0	4
4. Wisconsin diagnostic breast cancer	WDBC	569	31	0	2
5. Wisconsin prognostic breast cancer	WPBC	198	33	0	2
6. Wine recognition	Wine	178	13	0	3

Here, we have $\mathrm{SIG}(a, B, d) \geq 0$.

Based on the proposed measure, we can construct a forward greedy search algorithm for numerical feature selection.

Algorithm. numerical feature subset selection
Input: sample data set $\langle U, A = C \cup d\rangle$.
Output: feature subset *red*
Step 1: $\forall a \in A$: compute the fuzzy relationR_a;
Step 2: $\emptyset \rightarrow red$;
Step 3: For each $a_i \in C\text{-}red$
Compute $H_i = \mathrm{SIG}(a_i, red, d)$
End
Step 4: select the attribute satisfying:
$\mathrm{SIG}(a_i, red, d) = \max_i(H_i)$
Step 5: if $\mathrm{SIG}(a, red, d) > 0$, then
$red \cup a \rightarrow red$ go to Step 3
Else return, *red*

Several data sets are downloaded from the machine learning data repository, University of California at Irvine. They are described in Table 1. There are only numerical attributes in these data sets. We compare the classification accuracies and numbers of selected features with the classical rough method, where the numerical features are discretized with FCM. Two popular leaning algorithms (CART and RBF-SVM) are introduced to evaluate the quality of selected features. Table 2 shows the results with classical rough method [36], where N1 and N2 mean number of features in original data and reduced data, respectively. Table 3 gives the results based on fuzzy information entropy reduction.

Here, the fuzzy similarity relation is computed with the triangle membership function (Fig. 1). In fact, we also can employ other symmetric membership functions, such as Gaussian function.

Give a sample set $U = \{x_1, x_2, \ldots, x_n\}$ described with some numerical attributes $C = \{a_1, a_2, \ldots, a_N\}$ and a decision attribute d. $\forall a_i \in C$, we denote the maximal value and minimal value by $\max(a_i)$ and $\min(a_i)$, respectively. Then, the variable a_i is standardized with

$$\overline{f(x, a_i)} = \frac{f(x, a_i) - \min(a_i)}{\max(a_i)}.$$

Table 2
Classical rough set-based feature selection where numerical attributes are discretized

Data	Feature		CART		RBF-SVM	
	N1	N2	Raw data	Reduced data	Raw data	Reduced data
Iono	34	10	0.8755 ± 0.0693	0.9089 ± 0.0481	0.9379 ± 0.0507	0.9348 ± 0.0479
Sonar	60	6	0.7207 ± 0.1394	0.6926 ± 0.0863	0.8510 ± 0.0948	0.7074 ± 0.1004
Soy	35	2	0.9750 ± 0.0791	1.0000 ± 0.0000	0.9300 ± 0.1135	1.0000 ± 0.0000
Wdbc	31	8	0.9050 ± 0.0455	0.9351 ± 0.0339	0.9808 ± 0.0225	0.9649 ± 0.0183
Wpbc	33	7	0.6963 ± 0.0826	0.6955 ± 0.1018	0.7779 ± 0.0420	0.7837 ± 0.0506
Wine	13	4	0.8986 ± 0.0635	0.8972 ± 0.0741	0.9889 ± 0.0234	0.9486 ± 0.0507
Average	34.33	6.17	0.8452	0.8549	0.9111	0.8899

Table 3
Fuzzy information entropy-based numerical feature selection where fuzzy similarity relations are constructed with numerical data

Data	Feature		CART		RBF-SVM	
	N1	N2	Accuracy 1	Accuracy 2	Accuracy 1	Accuracy 2
Iono	34	13	0.8755 ± 0.0693	0.9068 ± 0.0564	0.9379 ± 0.0507	0.9462 ± 0.0365
Sonar	60	12	0.7207 ± 0.1394	0.7160 ± 0.0857	0.8510 ± 0.0948	0.8271 ± 0.0902
Soy	35	2	0.9750 ± 0.0791	1.0000 ± 0.0000	0.9300 ± 0.1135	1.0000 ± 0.0000
Wdbc	31	17	0.9050 ± 0.0455	0.9193 ± 0.0318	0.9808 ± 0.0225	0.9702 ± 0.0248
Wpbc	33	17	0.6963 ± 0.0826	0.7103 ± 0.1092	0.7779 ± 0.0420	0.8087 ± 0.0601
Wine	13	9	0.8986 ± 0.0635	0.9097 ± 0.0605	0.9889 ± 0.0234	0.9833 ± 0.0268
Average	34.33	11.67	0.8452	0.8603	0.9111	0.9226

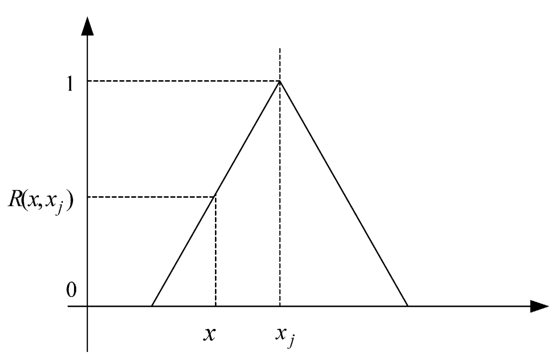

Fig. 1. Compute the similarity relation with triangle membership function.

$\forall x_j \in U$, the fuzzy similarity relation of x_j in terms of a_i is computed as

$$R_a(x,x_j) = 1 - 4 \times |\overline{f(x,a_i)} - \overline{f(x_j,a_i)}| \text{ if } |\overline{f(x,a_i)} - \overline{f(x_j,a_i)}| \leq 0.25;$$

$$R_a(x,x_j) = 0| \text{ if } |\overline{f(x,a_i)} - \overline{f(x_j,a_i)}| > 0.25.$$

Comparing Tables 2 and 3, we can find that more features are selected as to the fuzzy information entropy-based method than the discretization-based algorithm; however, the classification performance is improved with the entropy-based algorithm. Comparing with the original data, although two-thirds features are deleted in the reduced data sets, the average classification accuracies increase.

7. Conclusion

Relations and fuzzy relations are important concepts in mathematics and artificial intelligence. In particular, crisp similarity and equivalence relations, as well as fuzzy similarity and equivalence relations are the foundations of information fusion, fuzzy set, rough set and granular computing. In this paper, we generalize Yager's work and give an uncertainty measure for a general fuzzy binary relation.

The proposed measure is used to compute the diversity quantity of an information fusion system and the granularity of a granulated problem space. It is shown that the fusion system whose classifiers are of little similarity gets a great uncertainty quantity. It means that much of the complementary information is achieved with a diverse multiple classifier system. What's more the proposed diversity measure can compute the influence of the weight vector assigned to the classifiers.

As a granularity measure, the proposed algorithm calculates the uncertainty quantity of a granulated problem space with fuzzy equivalence relations. A "coarse–fine" relation is introduced for a family of problem spaces. The fine granulated problem space will produce a great uncertainty quantity relative to the coarse problem space.

Based on the observation, we employ the proposed measure to evaluate the significance of numerical attributes. Each numerical attribute generates a fuzzy similarity relation over the sample space. We compute the condition entropy of a numerical attribute or a set of numerical attribute relative to the decision, where the greater the condition entropy is, the less important the attribute subset is. A forward greedy search algorithm for numerical feature selection is constructed with the proposed measure. Experimental results show that the proposed method presents an efficient and effective solution for numerical feature analysis.

References

[1] C. Bertoluzza, V. Doldi, G. Naval, Uncertainty measure on fuzzy partitions, Fuzzy Sets Syst. 142 (2004) 105–116.

[2] P.M. Granitto, P.F. Verdes, H.A. Ceccatto, Neural network ensembles: evaluation of aggregation algorithms, Artific. Intell. 163 (2) (2005) 139–162.

[3] S. Greco, B. Matarazzo, R. Slowinski, Rough sets theory for multicriteria decision analysis, Eur. J. Operat. Res. 129 (1) (2001) 1–47.

[4] E. Hernandez, J. Recasens, A reformulation of entropy in the presence of indistinguishability operators, Fuzzy Sets Syst. 128 (2002) 185–196.

[5] T.K. Ho, Random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 832–844.

[6] Q. Hu, D. Yu, Entropies of fuzzy indiscernibility relation and its operations, Int. J. Uncert. Fuzz. Knowl. Based Syst. 12 (2004) 575–589.

[7] Q. Hu, D. Yu, M. Wang, Constructing rough decision forests, RSFDGrC (2), Lect. Notes Comp. Sci. 3642 (2005) 147–156.

[8] Q. Hu, D. Yu, Z. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recogn. Lett. 27 (2006) 414–423.

[9] Q. Hu, D. Yu, Z. Xie, J. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 14 (2) (2006) 191–201.

[10] G.J. Klir, R.M. Smith, On measuring uncertainty and uncertainty-based information: recent developments, Ann. Math. Artif. Intell. 32 (2001) 5–33.

[11] J.S. Mi, Y. Leung, W.Z. Wu, An uncertainty measure in partition-based fuzzy rough sets, Int. J. Gen. Syst. 34 (1) (2005) 77–90.

[12] N.R. Pal, J.C. Bezdek, Measuring fuzzy uncertainty, IEEE Trans. Fuzzy Syst. 2 (2) (1994) 107–118.

[13] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, 1991.

[14] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1986) 81–106.

[15] C.R. Rao, Statistics and Truth: Putting Chance to Work, 2nd ed., World Scientific, 1997.

[16] C. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Champaign, IL, 1964.

[17] D. Slezak, Association reducts: a framework for mining multi-attribute dependencies, in: Proceedings of the 15th International Symposium on Methodologies for Intelligent Systems, LNCS 3488, 2005, pp. 354–363.

[18] D. Slezak, Approximate entropy reducts, Fund. Informat. 53 (3/4) (2002) 365–390.

[19] G. Wang, Rough set reduction: in algebra view and in information view, Int. J. Intell. Syst. 18 (2003) 679–688.

[20] Q. Wang, Y. Shen, Y. Zhang, et al., A quantitative method for evaluating the performances of hyperspectral image fusion, IEEE Trans. Instrum. Measure. 52 (2003) 1041–1047.

[21] Q. Wang, Y. Shen, J.Q. Zhang, A nonlinear correlation measure for multivariable data set, Phys. D: Non-lin. Phenom. 200 (3/4) (2005) 287–295.

[22] Q. Wang, Y. Shen, Y. Zhang, et al., Fast quantitative correlation analysis and information deviation analysis for evaluating the performances of image fusion techniques, IEEE Trans. Instrum. Measure. 53 (2004) 1441–1447.

[23] X. Wang, Y. Wang, L. Wang, Improving fuzzy c-means clustering based on feature-weight learning, Pattern Recogn. Lett. 25 (2004) 1123–1132.

[24] W.Z. Wu, W.X. Zhang, Neighborhood operator systems and approximations, Inform. Sci. 144 (2002) 201–217.

[25] R.R. Yager, Entropy measures under similarity relations, Int. J. Gen. Syst. 20 (1992) 341–358.

[26] R.R. Yager, On the entropy of fuzzy measures, IEEE Trans. Fuzzy Syst. 8 (4) (2000) 453–461.

[27] R.R. Yager, Uncertainty representation using fuzzy measures, IEEE Trans. Syst. Man Cybernet. Part B: Cybernet. 32 (1) (2002) 13–20.

[28] Y.Y. Yao, Information granulation and rough set approximation, Int. J. Intell. Syst. 16 (2001) 87–104.

[29] L. Zadeh, Fuzzy sets, Inform. Contr. 8 (1965) 338–353.

[30] L. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 10 (1968) 421–427.

[31] L. Zadeh, Similarity relations and fuzzy ordering, Inform. Sci. 3 (1971) 177–200.

[32] L. Zadeh, Fuzzy logic equals Computing with words, IEEE Trans. Fuzzy Syst. 4 (2) (1996) 103–111.

[33] L. Zadeh, Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets Syst. 19 (1997) 111–127.

[34] L. Zhang, B. Zhang, Fuzzy reasoning model under quotient space structure, Informat. Sci. 173 (2005) 353–364.

[35] Y. Zhang, Constructive granular systems with universal approximation and fast knowledge discovery, IEEE Trans. Fuzzy Syst. 13 (1) (2005) 48–57.

[36] N. Zhong, J. Dong, S. Ohsuga, Using rough sets with heuristics for feature selection, J. Intell. Inform. Syst. 16 (2001) 199–214.