A novel measure of knowledge granularity in rough sets

Qinrong Feng*

Department of Computer Science and Technology,
Room 501, Building of Electronics and Information Engineering,
Jiading Campus,
Tongji University,
4800 Cao'an Highway,
Shanghai 201 804, PR China
and
College of Mathematics and Computer Science,
Shanxi Normal University,
Linfen, Shanxi 041 004, PR China
E-mail: fengqr72@163.com
*Corresponding author

Duoqian Miao, Jie Zhou and Yi Cheng

Department of Computer Science and Technology,
Tongji University,
Shanghai 201 804, PR China
E-mail: miaoduoqian@163.com
E-mail: jie_jpu@163.com
E-mail: chengyimail@sohu.com

Abstract: Knowledge granularity, an average measure of the size of knowledge granules, is a type of uncertainty arises from the indiscernibility relation. Consequently, granularity and indiscernibility are closely connected. In our opinion, knowledge granularity is a measure of uncertainty in an intra-granule. In this paper, a new measure of knowledge granularity for information system is proposed, which is characterised by mathematical expectation of lengths of granules in a partition. Based on the definition of knowledge granularity, relative knowledge granularity for decision table is also defined. The most advantage of relative knowledge granularity in this paper is that it can reveal the fact that granules belong to positive region have no contribution to the value of this measure. With this observation and the monotonicity of positive region, relative knowledge granularity can be computed recursively by adopting the strategy of separate-and-conquer, which is effective, especially for large scale data.

Keywords: knowledge granularity; mathematical expectation; positive region; relative knowledge granularity.

Reference to this paper should be made as follows: Feng, Q., Miao, D., Zhou, J. and Cheng, Y. (2010) 'A novel measure of knowledge granularity in rough sets', *Int. J. Granular Computing, Rough Sets and Intelligent Systems*, Vol. 1, No. 3, pp.233–251.

Biographical notes: Qinrong Feng received her MS in Computer Algebra in 2000 from Shanxi Normal University. Now, she is a PhD Candidate in the Department of Computer Science and Technology at Tongji University, PR China. Her research interests include data mining, rough sets and granular computing.

Duoqian Miao received his Bachelor of Science degree in Mathematics in 1985, MS in Probability and Statistics in 1991 both from Shanxi University and Doctor of Philosophy in Pattern Recognition and Intelligent System at Institute of Automation, Chinese Academy of Sciences in 1997. He is a Professor and Vice Dean at the School of Electronics and Information Engineering of Tongji University, PR China. His current research interests include rough sets, granular computing, principal curve, web intelligence, data mining, etc.

Jie Zhou received his MS in College of Information Science and Engineering from Central South University in 2007. Now, he is a PhD Candidate in the Department of Computer Science and Technology at Tongji University, PR China. His research interests include rough sets and data mining.

Yi Cheng received her MS in Operational Research and Cybernetics in 2002 from Sichan Normal University. Now, she is a PhD Candidate in the Department of Computer Science and Technology at Tongji University, PR China. Her research interests include fuzzy rough sets, data mining and granular computing.

1 Introduction

Granular computing (Yao, 2004), as a way of thinking, has been explored in many fields. It captures and reflects our ability to perceive the world at different granularity and to change granularities in problem solving. Its basic computing and reasoning units are granules, that is, groups, classes or clusters of a universe. The size of a granule is considered as a basic property. Intuitively, the size may be interpreted as the degree of abstraction, concreteness or detail. In the set-theoretic setting, the size of a granule can be measured by the cardinality of the granule.

One can associate quantitative measures to granules and granulations to capture their features. Consider, the case of crisp granulation. The cardinality of a granule, or Hartley information measure, can be used as a measure of the size or uncertainty of a granule (Yao, 2003). We all know that a good quantitative measure of granules may make the task of exploration much easier.

Knowledge granularity, the average measure of knowledge granules, has been researched during the past few years. It was Zadeh (1979) who first introduced the term of information granularity (knowledge granularity). After that several measures of knowledge granularity were discussed. Wierman (1999) introduced the concept of granularity to measure uncertainty of information. This concept has the same form as Shannon's entropy. Miao and Fan (2002) defined the concept of knowledge granularity, introduced the concepts of importance and consistency of an attribute, and a reduction algorithm for information system was designed based on this definition. Liang and Qian (2008) investigated information granulation in complete or incomplete information systems, it has been effectively applied in measuring attribute significance, feature selection, decision rules extracting, etc. Cattaneo et al. (2008) pointed out that entropy is a measure of uncertainty rather than a measure of granularity and the notion of entropy as a measure of uncertainty is distinguish from the notion of co-entropy as a measure of granularity in the context of partition. Liang and Qian (2006) presented an axiom definition of knowledge granularity for information system. Zhao et al. (2008) pointed out that there are some deficiencies about the axiomatic definition of Liang and Qian's, and put forward a new axiomatic definition of knowledge granularity.

In summarize, the measures of knowledge granularity in the above literature can be divided into two classes, one class is defined from the point of view of binary relation (Liang and Qian, 2008; Miao and Fan, 2002), and the other class is defined from the viewpoint of entropy, for example Hartley entropy (Cattaneo et al., 2008).

In fact, as Cattaneo et al. (2008) observed that there exists distinct difference between knowledge granularity and information entropy. We note that the finer a partition, the smaller of its knowledge granularity, but the larger of its information entropy. In other words, knowledge granularity is strictly monotonic decrease with the finer of a partition. On the contrary, the information entropy is strictly monotonic increase. Thus, knowledge granularity and information entropy are different kinds of measures of uncertainty. In our opinion, knowledge granularity is an average measure of the uncertainty of an intra-granule, and information entropy is an average measure of uncertainty of inter-granules. In this paper, we only focus on knowledge granularity, a measure of the uncertainty of an intra-granule. As Pawlak (1998) pointed out that the knowledge granularity is due to the indiscernibility of objects caused by lack of sufficient information about them. Consequently, granularity and indiscerniblity are strictly connected. So it is rational to measure knowledge granularity by the measure of indiscernibility ability, which is in fact an average measure of uncertainty of an intra-granule.

In this paper, we analyse knowledge from an entirely new point of view, that is, we treat every knowledge granule as an object of one-dimension, and the cardinality of a granule can be looked as the length of it. Thus, a new measure of knowledge granularity is defined for information system, which is characterised by mathematical expectation of lengths of granules in a partition. Based on this knowledge granularity, a measure of relative knowledge granularity is defined for decision table. From the definition of relative knowledge granularity, we note a fact that granules belong to positive region have no contribution to the value of relative knowledge granularity. With this observation and the monotonicity of positive region, relative knowledge granularity can be computed recursively follows the strategy of separate-and-conquer, which is very effective especially for large scale data.

The rest of this paper is organised as follows. Knowledge granularity is defined for information system and its properties are discussed in Section 2. In section 3, relative knowledge granularity is defined for decision table and its properties are also discussed. In Section 4, we give the conclusion.

2 Knowledge granularity for information system

A central notion of rough set theory is indiscernibility relation (equivalence relation) introduced by Pawlak (1991). An equivalence relation divides a universal set into a family of pairwise disjoint subsets, called a partition of the universe. A granule in a partition is therefore an equivalence class defined by an equivalence relation. Thus, the knowledge granularity is due to the indiscernibility of objects caused by lack of sufficient information about them.

An information system is a quadruple IS = (U, A, V, f), where U is a non-empty, finite set of objects, called the universe, and A is a non-empty, finite set of attributes, $V = \bigcup_{a \in A} V_a$ where V_a is the value set of a, called the domain of a, f is an information function from $U \times A$ to V, which assigns particular values from domains of attributes to objects such that $f(x_i, a) \in V_a$ for all $x_i \in U$ and $a \in A$. For every subset of attributes $B \subseteq A$ and $B \neq \varnothing, \bigcap B$ (intersection of all equivalence relations belong to B) is also an equivalence relation, and be denoted as IND(B), which is defined by xIND(B) y if and only if $f(x, a) = f(y, a)$, for every $a \in B$.

Every attribute (or knowledge) in IS = (U, A, V, f) determines a partition of U, so it can be looked as an equivalence relation over U. In this paper, we will use the term knowledge and partition exchangeably. Any equivalence class (i.e. elementary set) is interpreted as a granule. A definable set is any subset of U obtained as the set-theoretic union of elementary subsets. The set of all definable set induced from a partition of U plus the empty set $\varnothing$ has the structure of σ-algebra, and every elementary set in this σ-algebra called an event. Thus, every equivalence relation over U can be treated as a random variable defined over the σ-algebra of the set of cardinalities of partition granules. So probability distribution functions and mathematical expectations of these random variables are determined. Mathematical expectation is an important numerical characteristic of a random variable, and it can be used to measure the average behaviour of random variable effectively.

A partition is usually composed of multiple partition granules which are subset of the universe. The cardinality of a granule, or Hartley entropy, can be used as a measure of the size or uncertainty of a granule (Cattaneo et al., 2008; Yao, 2004). In this paper, each granule in a partition can be treated as an object of one-dimension, so the cardinality of a granule can be treated as its length. We will measure the size of a granule by its cardinality, and measure the knowledge granularity by mathematical expectation of the cardinality of granules.

2.1 Basic concepts

In this section, we will first list some basic concepts which are the basis of defining knowledge granularity.

Let IS = (U, A, V, f) be an information system, $P \subseteq A$, $X = \{X_1, X_2, \ldots, X_m\}$ be the partition induced by P. We call $|X_i|$ the length of X_i ($i = 1, 2, \ldots, m$) and the symbol $|X_i|$ means the cardinality of the granule X_i.

Definition 1: *Let IS = (U, A, V, f) be an information system, $P \subseteq A$ and $X = \{X_1, X_2, \ldots, X_m\}$ be the partition induced by P, then the probability distribution of P is defined as*

$$(X; p) = \begin{bmatrix} |X_1| & |X_2| & \cdots & |X_m| \\ p(|X_1|) & p(|X_2|) & \cdots & p(|X_m|) \end{bmatrix},$$

where

$$p(|X_i|) = \frac{|X_i|}{|U|}, \quad i = 1, 2, \ldots, m.$$

Definition 2: *Let $IS = (U, A, V, f)$ be an information system, $P, Q \subseteq A$, $X = \{X_1, X_2, \ldots, X_m\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$ be partitions induced by P and Q, respectively, then the joint probability distribution of P and Q is defined as*

$$(XY; p) = \begin{bmatrix} |X_1 \cap Y_1| & \cdots & |X_i \cap Y_j| & \cdots & |X_m \cap Y_n| \\ p(|X_1 \cap Y_1|) & \cdots & p(|X_i \cap Y_j|) & \cdots & p(|X_m \cap Y_n|) \end{bmatrix},$$

where

$$p(|X_i \cap Y_j|) = \frac{|X_i \cap Y_j|}{|U|}, \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n.$$

Definition 3: (Feng et al., 2009) *Let $IS = (U, A, V, f)$ be an information system, $P \subseteq A$, $X = \{X_1, X_2, \ldots, X_m\}$ be the partition induced by P, then the knowledge granularity of P is defined as*

$$E(P) = \sum_{i=1}^{m} |X_i| p(|X_i|) = \sum_{i=1}^{m} |X_i| \frac{|X_i|}{|U|}.$$

We can see that this measure is mathematical expectation of lengths of granules in a partition. Mathematical expectation can be used to measure the average behaviour of a random variable. On one hand, we can see that this measure is used to measure the average length of partition granules induced by an attribute or a subset of attributes. Thus, it is a measure of knowledge granularity. On the other hand, $|X_i|^2$ can count the number of pairs of objects which are indiscernible in X_i, so this measure can reflect the indiscernible ability. Thus, it is also a measure of uncertainty of an intra-granule.

For a special case, that is, $P = \varnothing$ (empty set), we all know that $\varnothing$ is not an equivalence relation over U, so $E(\varnothing)$ cannot be defined by Definition 3. However, in order to keep the completeness of the definition and facilitate computation, we stipulate $E(\varnothing) = |U| + 1$ in this paper. The rationality of this stipulation will be interpreted later.

We can see that the form of this definition is similar to the definition of knowledge granularity defined in Miao and Fan (2002) Liang and Qian (2008), but the point of departure of these two definitions is entirely different. In Miao and Fan (2002) and Liang and Qian (2008), the authors defined knowledge granularity from the point of view of relation, but the definition above is defined from the viewpoint of expectation (average length of partition granules).

Example 1: *Let $U = \{1, 2, 3, 4\}$ be a finite set, $X = \{\{1, 3\}, \{2, 4\}\}$ and $Y = \{\{2\}, \{1, 3, 4\}\}$ are partitions of U. Compute the knowledge granularities of X and Y, respectively.*

According to Definition 3, we have

$$E(X) = 2 \times \frac{2}{4} + 2 \times \frac{2}{4} = 2 \text{ and } E(Y) = 1 \times \frac{1}{4} + 3 \times \frac{3}{4} = 2.5.$$

This indicates that the average length of granules in partition X is equal to 2, and the average length of granules in partition Y is equal to 2.5. From this example, we can see that the indiscernible ability of X is weaker than that of Y.

Definition 4: *Let $IS = (U, A, V, f)$ be an information system, P, $Q \subseteq A$, $X = \{X_1, X_2, \ldots, X_m\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$ be partitions induced by P and Q, respectively, then the joint knowledge granularity of P and Q is defined as*

$$E(P \cup Q) = \sum_{i,j} |X_i \cap Y_j| p(|X_i \cap Y_j|) = \sum_{i,j} |X_i \cap Y_j| \frac{|X_i \cap Y_j|}{|U|}.$$

Definition 5: *Let U be a set, $X = \{X_1, X_2, \ldots, X_m\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$ are partitions of U induced by P and Q, respectively, then we will call P is finer than Q if and only if for any $X_i \in X$ there always exist $Y_j \in Y$ such that $X_i \subseteq Y_j$, and denoted as $P \preceq Q$. If $X_i \subseteq Y_j$ and $X_i \neq Y_j$, we will call P is strictly finer than Q and denote as $P \prec Q$.*

2.2 Properties of knowledge granularity

In the following, some significant properties about knowledge granularity are presented, and they are valuable in evaluating uncertainty of an information system.

Property 1: *Let $IS = (U, A, V, f)$ be an information system, $P \subseteq A$ and $X = \{X_1, X_2, \ldots, X_m\}$ be the partition induced by P, then*

$$1 \le \frac{|U|}{m} \le E(P) \le |U|$$

where m is the cardinality of X.

Proof: Set $|U| = n$, and $|X_i| = k_i, i = 1, 2, \ldots, m$, clearly, $m \le n$. From Definition 3, we have

$$E(P) = \sum_i |X_i| \frac{|X_i|}{|U|} = \sum_i \frac{k_i^2}{n}.$$

Since $k_1 + k_2 + \ldots + k_m = n$, so we have

$$n^2 = (k_1 + k_2 + \cdots + k_m)^2 = k_1^2 + k_2^2 + \cdots + k_m^2 + 2\sum_{s<t} k_s k_t \le m(k_1^2 + k_2^2 + \cdots + k_m^2) \quad (1)$$

Hence, we have

$$\sum_i k_i^2 \ge \frac{n^2}{m}, \text{ So } E(P) = \sum_i \frac{k_i^2}{n} \ge \frac{n}{m} = \frac{|U|}{m} \ge 1.$$

By the Formula (1), we have that $E(p) = n/m$ if and only if $k_i = k_j (i, j = 1, 2, \cdots, m)$, that is, every granule include the same number of elements. Especially, $E(p) = 1$ if and only if $m = n$, that is, every granule include only one element.

On the other hand, since $k_1^2 + k_2^2 + \cdots + k_m^2 \le (k_1 + k_2 + \cdots + k_m)^2 = n^2$, so we have

$$E(P) = \sum_i \frac{k_i^2}{n} \le n = |U|,$$

and $E(P) = |U|$ holds if and only if all of elements are included in one granule. □

This property indicates that the knowledge granularity is greater or equal to 1 and lesser or equal to $|U|$ for any partition of a finite set U. In other words, the knowledge granularity is bounded.

Especially, when $m = |U|$, that is, each partition granule includes only one element, then its knowledge granularity is equal to 1, and when $m = 1$, that is, all of elements in U are included in one granule, then the knowledge granularity is equal to $|U|$.

Next, we will explain the rationality of the stipulation mentioned above, namely, $E(\varnothing) = |U| + 1$. From Property 1, we can see that the knowledge granularity is increase with the decrease of the number of granules. The knowledge granularity increase monotonously from 1 to $|U|$ when the number of granules decrease from $|U|$ to 1, so we can regard $\varnothing$ as a partition include zero granule, hence we will have $E(\varnothing) > |U|$ with the trend of $E(P)$. Thus, the stipulation $E(\varnothing) = |U| + 1$ is rational.

Lemma 1: *Let IS = (U, A, V, f) be an information system, P, Q ⊆ A,* $X = \{X_1, X_2, ..., X_m\}$ *and* $Y = \{X_1, X_2, \ldots, X_{m-1}, X'_m, X'_{m+1}\}$ *be partitions induced by P and Q, respectively, then* $E(P) = E(Q) + \frac{2|X'_m||X'_{m+1}|}{|U|}$, *where* $X_m = X'_m \cup X'_{m+1}$ *and* $X'_m \cap X'_{m+1} = \varnothing$.

Proof: This result can be easily obtained from Definition 3.

From this lemma, we can find that if $Y \prec X$, that is $Q \prec P$, we have $E(Q) < E(P)$. Moreover, based on this lemma, we can compute knowledge granularity incrementally.

Property 2: *Let IS = (U, A, V, f) be an information system, P, Q ⊆ A, if partition induced by P is finer than that of Q, denoted as,* $P \prec Q$ *then E(P) < E(Q).*

Proof: Let $X = \{X_1, X_2, \ldots, X_m\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$ be partitions induced by P and Q. Since $P \prec Q$, so for any $X_i \in X$, there always exist $Y_j \in Y$ such that $X_i \subset Y_j$. As $X = \{X_1, X_2, \ldots, X_m\}$ is a partition, so for any $i \neq j$, we have $X_i \cap X_j = \varnothing$, thus there exists a set N such that $Y_j = \bigcup_{i \in N} X_i$ so we have $|Y_j| = \sum_{i \in N} |X_i|$, thus

$$\sum_{i \in N} |X_i|^2 < \left(\sum_{i \in N} |X_i| \right)^2 = |Y_j|^2,$$

hence we have $E(P) < E(Q)$. □

This property indicates the monotonicity of knowledge granularity, that is, knowledge granularity is strictly monotonic decrease with the finer of a partition.

Corollary 1: *Let IS = (U, A, V, f) be an information system, P, Q ⊆ A, then we have* $E(P \cup Q) \le min(E(P), E(Q))$.

This result can be easily obtained from Property 2.

This Corollary shows that the more knowledge you own, the smaller its knowledge granularity. Or we can say that the more knowledge you own, the weaker the indiscernible ability.

Definition 6**:** *Let IS = (U, A, V, f) be an information system, P, $Q \subseteq A$, $X = \{X_1, X_2, \ldots, X_m\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$ be partitions induced by P and Q, respectively, we call two partitions X and Y are isomorphic or P and Q are isomorphic if and only if there exists a bijection $g: X \to Y$, such that $|X| = |Y|$ and for any $X_i \in X$, $|X_i| = |g(X_i)|$.*

Property 3: *Let IS = (U, A, V, f) be an information system, P, $Q \subseteq A$, if two partitions induced by P and Q are isomorphic, then we have E(P) = E(Q). Obviously, if IND(P) = IND(Q), then E(P) = E(Q).*

The proof can be easily obtained from Definition 3.

This property indicates that if two partitions are isomorphic, then they have the same knowledge granularity. But the inverse is not necessarily hold.

Property 4: *Let IS = (U, A, V, f) be an information system, P, $Q \subseteq A$ and $P \subseteq Q$. If E(P) = E(Q), then IND(P) = IND(Q).*

Proof: Let $U/\mathrm{IND}(P) = \{X_1, X_2, \ldots, X_m\}$ and $U/\mathrm{IND}(Q-P) = \{Y_1, Y_2, \ldots, Y_n\}$ are partitions induced from P and $Q-P$, where $Q-P$ denotes the set difference of Q and P, it is a set of elements of Q not in P.

By definition 3, we have

$$E(P) = \sum_i \frac{|X_i|^2}{|U|},$$

and

$$E(Q) = E(P \cup (Q-P)) = \sum_{i,j} \frac{|X_i \cap Y_j|^2}{|U|} = \sum_i \sum_j \frac{|X_i \cap Y_j|^2}{|U|}.$$

For every i, since $|X_i| = \sum_j |X_i \cap Y_j|$, so we have

$$|X_i|^2 = \left(\sum_j |X_i \cap Y_j| \right)^2 = \sum_j |X_i \cap Y_j|^2 + 2\sum_{j<k} |X_i \cap Y_j||X_i \cap Y_k|,$$

and since $2\sum_{j<k} |X_i \cap Y_j| \times |X_i \cap Y_k| \geq 0$, we therefore get $|X_i|^2 \geq \sum_j |X_i \cap Y_j|^2$, Hence, we have $\sum_i |X_i|^2 \geq \sum_i \sum_j |X_i \cap Y_j|^2$, that is, $E(P) \geq E(Q)$.

Obviously, $E(P) = E(Q)$ if and only if for every i, we have $\left|X_i\right|^2 = \sum_j \left|X_i \cap Y_j\right|^2$, and if and only if for every i, we have

$$\sum_{j<k} \left|X_i \cap Y_j\right| \cdot \left|X_i \cap Y_k\right| = 0. \tag{2}$$

Clearly, the Formula (2) holds if and only if for every pair of $j < k$, we have $|X_i \cap Y_j| \cdot |X_i \cap Y_k| = 0$. Since $\bigcup_{j=1}^{n} Y_j = U$ and $X_i \subseteq U$, so it is impossible for any pair $j < k$, there always have $|X_i \cap Y_j| = |X_i \cap Y_k| = 0$, hence there will exist a pair of j, k, such that $|X_i \cap Y_j| = 0$ and $|X_i \cap Y_k| \neq 0$. For the case of $|X_i \cap Y_k| \neq 0$, if $X_i \not\subset Y_k$, then there will exist $l \neq k$, such that $|X_i \cap Y_l| \neq \varnothing$. However, this is impossible. Otherwise, if this is the case, then we will have $|X_i \cap Y_l| \times |X_i \cap Y_k| \neq 0$, hence we have $\sum_{j<k} |X_i \cap Y_j| \times |X_i \cap Y_k| \neq 0$. Obviously, this is contrary to $E(P) = E(Q)$. So for every X_i, there always exist Y_k, such that $X_i \subset Y_k$.

Hence, we have $\text{IND}(P) \subseteq \text{IND}(Q-P)$, namely, $P \Rightarrow Q-P$, so $\text{IND}(P \cup (Q-P)) = \text{IND}(P)$, that is, $\text{IND}(Q) = \text{IND}(P)$. □

This property indicates that if P and Q have the same knowledge granularities, and $P \subseteq Q$, then they will induce the same partition. In other words, we have $E(P) = E(Q)$ if and only if $\text{IND}(P) = \text{IND}(Q)$ under the condition of $P \subseteq Q$.

Property 5: *Let IS = (U, A, V, f) be an information system, $P \subseteq A$, a relation $R \in P$ is indispensable in P if and only if $E(P-\{R\}) = E(P)$.*

The proof can be easily obtained from the definition of indispensable and the result of Property 3.

From this property, we can see that a relation $R \in P$ is indispensable in P if and only if the knowledge granularity is unchanged whether we remove R from P or not.

Corollary 2: *Let IS = (U, A, V, f) be an information system, $P \subseteq A$, a relation $R \in P$ is dispensable in P if and only if $E(P-\{R\}) > E(P)$.*

Property 6: *Let IS = (U, A, V, f) be an information system, $P \subseteq A$, then P is independent if and only if for any $R \in P$, there always have $E(P-\{R\}) > E(P)$.*

The proof can be easily obtained from the definition of independent and Corollary 2.

This property shows that every attribute in an independent subset of attributes has contribution to the knowledge granularity of this subset.

Property 7: *Let IS = (U, A, V, f) be an information system, then $Q \subseteq A$ be a reduct of A if and only if Q satisfies the following conditions:*

1 $E(Q) = E(A)$.

2 For any $R \in Q$, we have $E(Q-\{R\}) > E(Q)$.

The proof can be easily obtained from the definition of reduct and Property 6.

This property indicates that a reduct of IS = (U, A, V, f) is an independent subset of attributes, which has the same knowledge granularity with it.

Definition 7: *Let IS = (U, A, V, f) be an information system and* $R \subset A$*, then for any* $a \in A-R$*, its significance Sig(a, R) is defined as*

$$Sig(a, R) = E(R) - E(R \cup \{a\}).$$

Especially, if R = Ø, then Sig(a, Ø) = E(Ø)–E({a}).

Based on this definition, we can design a heuristic reduction algorithm for information system. The detail is refer to (Feng et al., 2007a).

3 Relative knowledge granularity for decision table

In existing literature, there have some different forms of relative knowledge granularity, such as conditional Hartley entropy (Cattaneo et al., 2008), conditional rough entropy (Liang and Shi, 2004) and relative attribute dependency (Han et al., 2004; Hu et al., 2003). However, in this paper, we will define it as the difference of two knowledge granularities.

Based on the definition of knowledge granularity in this paper, we will define relative knowledge granularity for decision table to measure the knowledge granularity of conditional attributes relative to decision attribute. From this definition, we note a fact that granules belong to positive region have no contribution to relative knowledge granularity, this has not been mentioned in existing literature. Based on this observation and the monotonicity of positive region, we adopt the strategy of separate-and-conquer to compute relative knowledge granularity recursively. This is very effective, especially for large scale data.

3.1 Relative knowledge granularity

Let $\mathrm{DT} = (U, C \cup D, V, f)$ be a decision table, where U is the universe, C is the conditional attributes and D is the decision attribute. $V = \bigcup_{a \in C \cup D} V_a$, where V_a is the value set of a, called the domain of a, f is an information function from $U \times (C \cup D)$ to V, which assigns particular values from domains of attributes to objects such that $f(x_i, a) \in V_a$ for all $x_i \in U$ and $a \in C \cup D$.

Let $A \subseteq C$, we call $\mathrm{DT} = (U, A \cup D, V, f)$ the decision table induced by A, and the decision table $\mathrm{DT} = (U, A \cup D, V, f)$ can be decomposed into two pairwise disjoint parts, one part is $\mathrm{DT} = (\mathrm{POS}_A(D), A \cup D, V, f)$ which is totally consistent, and the other part is $\mathrm{DT} = (U - \mathrm{POS}_A(D), A \cup D, V, f)$ which is entirely inconsistent.

Since the decision table $\mathrm{DT} = (\mathrm{POS}_A(D), A \cup D, V, f)$ is consistent, we will say that decision attribute D totally depends on A over $\mathrm{POS}_A(D)$, in other words, the partition $U/\mathrm{IND}(A)$ is finer than $U/\mathrm{IND}(D)$ over $\mathrm{POS}_A(D)$, thus we have

$$\left[U / IND(A) = U / IND(A \bigcup D)\right]\Big|_{POS_A(D)} \tag{3}$$

where the symbol $A|_B$ means restrict A to B. The Formula (3) indicates that the partition induced by A is the same as that induced by $A \cup D$ over $\mathrm{POS}_A(D)$, so according to Property 3, A and $A \cup D$ have the same knowledge granularity over $\mathrm{POS}_A(D)$.

Denote the partition induced by *A* as

$$U / IND(A) = \{X_1, X_2, \cdots, X_k, X_{k+1}, \cdots, X_m\}$$

We note that there has close relation between the partitions $U / \mathrm{IND}(A \cup D)$ and $U/\mathrm{IND}(A)$. Namely, each partition granule in $U/\mathrm{IND}(A)$ either be kept unchanged or be divided into multiple finer partition granules due to adding the decision attribute *D* into *A*. So we might as well denote the partition induced from $A \cup D$ by

$$U / IND(A \bigcup D) = \{X_1, X_2, \cdots, X_k, Y_{k+1}, \cdots, Y_n\}$$

Apparently, those partition granules kept unchanged are all included in *A* – positive region of *D*, that is, $\mathrm{POS}_A(D) = X_1 \cup X_2 \cup \ldots \cup X_k$. According to Definition 3, we have

$$E(A) = \sum_{X_i \subseteq \mathrm{POS}_A(D)} |X_i| \frac{|X_i|}{|U|} + \sum_{X_i \subseteq U - \mathrm{POS}_A(D)} |X_i| \frac{|X_i|}{|U|},$$

and

$$E(A \cup D) = \sum_{X_i \subseteq \mathrm{POS}_A(D)} |X_i| \frac{|X_i|}{|U|} + \sum_{Y_i \subseteq U - \mathrm{POS}_A(D)} |Y_j| \frac{|Y_j|}{|U|}.$$

Thus, we have

$$E(A) - E(A \cup D) = \sum_{X_i \subseteq U - \mathrm{POS}_A(D)} \frac{|X_i|^2}{|U|} - \sum_{Y_j \subseteq U - \mathrm{POS}_A(D)} \frac{|Y_j|^2}{|U|}.$$

This formula indicates that the difference of *E*(*A*) and $E(A \cup D)$ is only determined by the entirely inconsistent part of *A* relative to *D*, and has nothing to do with the totally consistent part. Thus, the value of $E(A) - E(A \cup D)$ can be used to measure the uncertainty of $(U, A \cup D, V, f)$ or uncertainty *A* relative to *D*.

Chen et al. (2000) pointed out that the uncertainty coming from the granularity of the partition includes inconsistency and randomness. Düntsch and Gediga (1998) constructed an information entropy based uncertainty measure $H^{\det}$ which can deal with the two aspects of the uncertainty. However, the measure of $H^{\det}$ is strictly monotonic increase with the finer of a partition. Next, we will give a new measure of relative knowledge granularity which can also deal with the two aspects of the uncertainty, and it is strictly monotonic decrease with the finer of a partition. This is accord with our cognition.

Definition 8 (Feng et al., 2008b): *Let* $DT = (U, C \cup D, V, f)$ *be a decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute,* $A \subseteq C$*, then the relative knowledge granularity of A relative to D is defined as* $E(A; D) = E(A) - E(A \cup D)$.

The relative knowledge granularity is defined as the difference of two knowledge granularities. From this definition and the above analysis, we find that granules belong to positive region have no contribution to relative knowledge granularity.

Table 1 Weather

U	*Outlook(a_1)*	*Temperature(a_2)*	*Humidity(a_3)*	*Windy(a_4)*	*Class(D)*
1	Sunny	Hot	High	False	N
2	Sunny	Hot	High	True	N
3	Overcast	Hot	High	False	P
4	Rain	Mild	High	False	P
5	Rain	Cool	Normal	False	P
6	Rain	Cool	Normal	True	N
7	Overcast	Cool	Normal	True	P
8	Sunny	Mild	High	False	N
9	Sunny	Cool	Normal	False	P
10	Rain	Mild	Normal	False	P
11	Sunny	Mild	Normal	True	P
12	Overcast	Mild	High	True	P
13	Overcast	Hot	Normal	False	P
14	Rain	Mild	High	True	N

Example 2: *Given a decision table* $DT = (U, C \cup D, V, f)$ *described as Table 1, where U = {1, 2, …, 14} is the universe, C = {a_1, a_2, a_3, a_4} is the conditional attributes, D is the decision attribute. Compute the E(a_1;D).*

We can easily get partitions from Table 1 as follows,

$$U/IND(a_1) = \{\{1,2,8,9,11\},\{3,7,12,13\},\{4,5,6,10,14\}\},$$
$$U/IND(D) = \{\{1,2,6,8,14\},\{3,4,5,7,9,10,11,12,13\}\},$$
$$U/IND(\{a_1\} \cup D) = \{\{1,2,8\},\{6,14\},\{9,11\},\{3,7,12,13\},\{4,5,10\}\}.$$

By Definition 3, we have

$$E(a_1) = 5 \times \frac{5}{14} + 4 \times \frac{4}{14} + 5 \times \frac{5}{14} = \frac{66}{14} \approx 4.7$$

and

$$E(a_1 \cup D) = 3 \times \frac{3}{14} + 2 \times \frac{2}{14} + 2 \times \frac{2}{14} + 4 \times \frac{4}{14} + 3 \times \frac{3}{14} = 3,$$

so by Definition 8, we have $E(a_1; D) \approx 4.7 - 3 = 1.7.$

On the other hand, we can easily get $\mathrm{POS}_{a_1}(D) = \{3,7,12,13\}$. From the above analysis, we know that granules belong to positive region have no contribution to the relative knowledge granularity. Thus, we only need to consider granules not belong to positive region, namely, we will compute relative knowledge granularity $E(a_1; D)$ over set {1,2,4,5,6,8,9,10,11,14}, so we also have

$$E(a_1; D) = \left(5 \times \frac{5}{14} + 5 \times \frac{5}{14}\right) - \left(3 \times \frac{3}{14} + 2 \times \frac{2}{14} + 3 \times \frac{3}{14} + 2 \times \frac{2}{14}\right) \approx 1.7.$$

3.2 Properties of relative partition granularity

In this section, we will show that relative knowledge granularity has many significant properties, and it is valuable in evaluating uncertainty of a decision table.

Property 8: *Let $DT = (U, C \cup D, V, f)$ be a decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute, then we have $0 \le E(C; D) < |U|$.*

The proof can be easily obtained from Definition 8 and Property 1.

This property indicates that relative knowledge granularity is bounded.

Property 9: *Let $DT = (U, C \cup D, V, f)$ be a decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute, then decision table $DT = (U, C \cup D, V, f)$ is consistent if and only if $E(C; D) = 0$.*

Proof: Let $U/\mathrm{IND}(C) = \{X_1, X_2, \ldots, X_m\}$ and $U/\mathrm{IND}(D) = \{Y_1, Y_2, \ldots, Y_n\}$ if $\mathrm{DT} = (U, C \cup D, V, f)$ is a consistent decision table, then we have $POS_C(D) = U$. That is, $U/\mathrm{IND}(C)$ *is* finer than $U/IND(D)$, so we have $U/\mathrm{IND}(C \cup D) = U/\mathrm{IND}(C)$, hence we have $E(C; D) = E(C) - E(C \cup D) = 0.$

Conversely, if $E(C; D) = E(C) - E(C \cup D) = 0$ then we have $E(C) = E(C \cup D),$ obviously, since $C \subseteq C \cup D,$ so according to Property 4, we have $U/\mathrm{IND}(C \cup D) = U/\mathrm{IND}(C),$ that is, $U/\mathrm{IND}(C)$ is finer than $U/\mathrm{IND}(D)$, so $\mathrm{DT} = (U, C \cup D, V, f)$ is a consistent decision table. □

This property gives the sufficient and necessary condition of $E(C; D)$ getting to its minimum. In other words, the relative knowledge granularity of C relative to D is equal to zero for a consistent decision table.

Lemma 2: *Let $DT = (U, C \cup D, V, f)$ be a decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute, P, Q $\subseteq$ C and partitions induced by P, Q and D over U are $X = \{X_1, \ldots, X_{m-1}, X'_m, X'_{m+1}\}$, $Y = \{X_1, \ldots, X_{m-1}, X_m\}$* and $Z = \{Z_1, Z_2, \ldots, Z_k\}$*, respectively, where $X_m = X'_m \cup X'_{m+1}$ and $X'_m \cap X'_{m+1} = \varnothing$, then we have $E(Q; D) \ge E(P; D)$.*

Proof: By Definition 8, we have

$$E(Q; D) - E(P; D) = \big(E(Q) - E(P)\big) - \big(E(Q \cup D) - E(P \cup D)\big)$$

while $E(Q) - E(P) = \dfrac{|X_m|^2}{|U|} - \dfrac{|X'_m|^2}{|U|} - \dfrac{|X'_{m+1}|^2}{|U|},$

since $X_m = X'_m \cup X'_{m+1}$ and $X'_m \cap X'_{m+1} = \varnothing,$ so we have $|X_m| = |X'_m| + |X'_{m+1}|$ and hence, we have

$$E(Q) - E(P) = \frac{2|X'_m| \cdot |X'_{m+1}|}{|U|}.$$

Likewise, in the same way, we have

$$E(Q \cup D) - E(P \cup D) = 2\sum_i \frac{|X'_m \cap Z_i| \cdot |X'_{m+1} \cap Z_i|}{|U|},$$

hence we have

$$E(Q;D)-E(P;D) = 2\frac{|X'_m|.|X'_{m+1}|}{|U|} - 2\sum_i \frac{|X'_m \cap Z_i|.|X'_{m+1}\cap Z_i|}{|U|}$$
$$= \frac{2}{|U|}\left(|X'_m|.|X'_{m+1}| - \sum_i |X'_m \cap Z_i|.|X'_{m+1}\cap Z_i|\right).$$

Since $X'_m = (X'_m \cap Z_1)\cup\cdots\cup(X'_m \cap Z_k)$ and for any pair of $i \neq j$, $(X'_m \cap Z_i)\cap(X'_m \cap Z_j)=\varnothing$, so $|X'_m| = |X'_m \cap Z_1| + \cdots + |X'_m \cap Z_k|$, similarly, we have $|X'_{m+1}| = |X'_{m+1}\cap Z_1| + \cdots + |X'_{m+1}\cap Z_k|$. Clearly, we can see that

$$|X'_m|\cdot|X'_{m+1}| \geq \sum_t |X'_m \cap Z_t|\cdot|X'_{m+1}\cap Z_t|,$$

so we have $E(Q)-E(P) \geq E(Q\cup D)-E(P\cup D)$, hence $E(Q; D) \geq E(P; D)$.

Especially, $E(Q;D)-E(P;D)=\sum_{i\neq j} 2\cdot|X'_m \cap Z_i|\cdot|X'_{m+1}\cap Z_j| = 0$ if and only if there exists $1 \leq t \leq k$, such that $X'_m \subseteq Z_t$ and $X'_{m+1} \subseteq Z_t$, that is, X'_m and X'_{m+1} are included in the same granule Z_t. Thus, there exists a granule Z_t, such that X_m is entirely included in it. □

This lemma shows that $E(P; D) \leq E(Q; D)$ under the condition of $P \prec Q$, and if $E(Q; D) = E(P; D)$, then we have $POS_P(D) = POS_Q(D)$, but the inverse is not necessarily hold.

Property 10: *Let $DT=(U, C\cup D, V, f)$ be a decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute, $P \subseteq C$, then for all $r \in P$ we have $E(P-\{r\};D) \geq E(P; D)$.*

The proof can be easily obtained from Lemma 2.

This property indicates that the more the knowledge you own, the less the relative knowledge granularity.

Corollary 3: *Let $DT=(U, C\cup D, V, f)$ be a decision table, where U is the universe, $C = \{C_1, C_2, \ldots, C_m\}$ is the set of conditional attributes and D is the decision attribute, then we have*

$$E(C_1;D) \geq E(C_1 \cup C_2;D) \geq \cdots \geq E(C_1\cup\cdots\cup C_m;D).$$

The proof can be easily obtained from Property 10.

This property indicates that the value of relative knowledge granularity will decrease monotonously with the conditional attributes increase one by one.

In fact, Lemma 2, Property 10 and Corollary 3 are all show the monotonicity of relative knowledge granularity.

We also note the monotonicity of positive regions in traditional rough sets which are as followings.

Property 11: *Let $DT=(U, C\cup D, V, f)$ be a decision table, where U is the universe, $C = \{C_1, C_2, \ldots, C_m\}$ is the set of conditional attributes and D is the decision attribute, then $POS_{\{C_1\}}(D) \subseteq POS_{\{C_1,C_2\}}(D) \subseteq \cdots \subseteq POS_{\{C_1,C_2,\ldots,C_n\}}(D)$.*

We will obtain the following theorem by combining the monotonicity of positive region and the properties of positive region, which can used to compute positive region recursively.

Theorem 1 (Feng et al., 2007b): *Let $DT = (U, C \cup D, V, f)$ be a decision table, where U is the universe, C = {C_1, C_2, …, C_m} is the set of conditional attributes and D is the decision attribute, $P \subseteq C$, then $POS_C(D) = POS_P(D) \bigcup POS_C(D)|_{U-POS_P(D)}$, where the symbol $POS_C(D)|_{U-POS_P(D)}$ means restrict $POS_C(D)$ to $U–POS_P(D)$.*

This theorem shows that it is easily to compute positive region $POS_C(D)$ recursively by adopting the strategy of separate-and-conquer (Fürnkranz, 1999) which has been coined by Pagallo and Haussler(1990). That is, we compute $\text{POS}_{C_1}(D)$ firstly, remove $\text{POS}_{C_1}(D)$ from the universe *U,* then compute $\text{POS}_{\{C_1,C_2\}}(D)$ over $U - \text{POS}_{C_1}(D)$, remove $\text{POS}_{\{C_1,C_2\}}(D)$ from the remaining universe $U - \text{POS}_{C_1}(D)$, by analogy we can get the positive region $POS_C(D)$. This method of computing positive region is effective, especially for large scale data.

Based on Theorem 1 and Corollary 3, we can compute relative knowledge granularity recursively, which is illustrated through the following example. Due to our observation that granules belong to positive region have no contribution to relative knowledge granularity, we only need to focus on granules not included in positive region.

Example 3: *Given a decision table $DT = (U, C \cup D, V, f)$ described as Table 1, where U = {1, 2, …, 14} is the universe, C = {a_1, a_2, a_3, a_4} is the conditional attributes, D is the decision attribute.*

Compute the *E*(*C*; *D*).

1 Compute E({a_1}; D): we can easily get $POS_{a_1}(D) = \{3,7,12,13\}$, then remove $POS_{a_1}(D)$ from the universe and compute relative knowledge granularity *E*({a_1}; *D*) over {1, 2, 4, 5, 6, 8, 9, 10, 11, 14},

$U / IND(\{a_1\}) = \{\{1,2,8,9,11\},\{4,5,6,10,14\}\}$,

$U / IND(D) = \{\{1,2,6,8,14\},\{4,5,9,10,11\}\}$,

$U / IND(\{a_1\} \bigcup D) = \{\{1,2,8\},\{6,14\},\{9,11\},\{4,5,10\}\}$ so we have

$$E(\{a_1\}; D) = \left(5 \times \frac{5}{14} + 5 \times \frac{5}{14}\right) - \left(3 \times \frac{3}{14} + 2 \times \frac{2}{14} + 3 \times \frac{3}{14} + 2 \times \frac{2}{14}\right) \approx 1.7 .$$

2 Compute *E*({a_1, a_2}; D): we can easily get $POS_{\{a_1,a_2\}}(D) = \{1,2,9\}$ over {1, 2, 4, 5, 6, 8, 9, 10, 11, 14}, remove $POS_{\{a_1,a_2\}}(D)$ from {1, 2, 4, 5, 6, 8, 9, 10, 11, 14} and we only need to compute *E*({a_1, a_2}; D) over {4, 5, 6, 8, 10, 11, 14}.

$U / IND(\{a_1,a_2\}) = \{\{4,10,14\},\{5,6\},\{8,11\}\}$,

$U / IND(D) = \{\{6,8,14\},\{4,5,10,11\}\}$,

$U / IND(\{a_1, a_2\} \cup D) = \{\{4,10\}, \{14\}, \{5\}, \{6\}, \{8\}, \{11\}\}$ so

$$E(\{a_1, a_2\}; D) = 3 \times \frac{3}{14} + 2 \times \frac{2}{14} + 2 \times \frac{2}{14} - 2 \times \frac{2}{14} - 1 \times \frac{1}{14} - 1 \times \frac{1}{14} - 1 \times \frac{1}{14} - 1 \times \frac{1}{14} - 1 \times \frac{1}{14} \approx 0.57.$$

3 Compute $E(\{a_1, a_2, a_3\}; \mathrm{D})$: we can get $POS_{\{a_1,a_2,a_3\}}(D) = \{8,10,11\}$ over {4, 5, 6, 8, 10, 11, 14}, remove $POS_{\{a_1,a_2,a_3\}}(D)$ from {4, 5, 6, 8, 10, 11, 14} and we only need to compute $E(\{a_1, a_2, a_3\}; \mathrm{D})$ over {4, 5, 6, 14}.

$U / IND(\{a_1, a_2, a_3\}) = \{\{4,14\}, \{5,6\}\}$,

$U / IND(D) = \{\{6,14\}, \{4,5\}\}$,

$U / IND(\{a_1, a_2, a_3\} \cup D) = \{\{4\}, \{5\}, \{6\}, \{14\}\}$, so we have

$E(\{a_1, a_2, a_3\}; \mathrm{D}) \approx 0.29$.

4 Compute $E(C; D)$: we can get $POS_C(D)$ = {4, 5, 6, 14} over {4, 5, 6, 14}, remove $POS_C(D)$ from the remaining universe, we then get an empty set, that is, this decision table is a consistent one, so $E(C; D) = 0$.

This example shows that we can compute relative knowledge granularity recursively, and the cardinality of universe is decrease greatly with the increase of conditional attributes.

Property 12: *Let $DT = (U, C \cup D, V, f)$ be a consistent decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute, then we will say that $r \in C$ is D-dispensable in C if and only if $E(C; D) = E(C - \{r\}; D)$.*

Proof: Let $U/\mathrm{IND}(C) = \{X_1, X_2, ..., X_m\}$ and $U/\mathrm{IND}(D) = \{Y_1, Y_2, ..., Y_n\}$, since $\mathrm{DT} = (U, C \cup D, V, f)$ is a consistent decision table, that is, $POS_C(D) = U$, so $U/\mathrm{IND}(C)$ is finer than $U/\mathrm{IND}(D)$, hence we have $U/\mathrm{IND}(C \cup D) = U/\mathrm{IND}(C)$, so $E(C; D) = E(C) - E(C \cup D) = 0$. If r is D-dispensable in C, then

$$POS_{C-\{r\}}(D) = POS_C(D) = U,$$

so $U/\mathrm{IND}(C-\{r\})$ is finer than $U/\mathrm{IND}(D)$, that is,

$$U / IND((C - \{r\}) \cup D) = U / IND(C - \{r\})$$

so $E(C - \{r\}; D) = E(C - \{r\}) - E((C - \{r\} \cup D) = 0$, hence $E(C; D) = E(C-\{r\}; D)$.

Conversely, we assume that $\mathrm{POS}_{C-\{r\}}(D) \neq POS_C(D) = U$, Let $U/\mathrm{IND}(C-\{r\}) = \{Z_1, Z_2, \ldots, Z_k\}$, then for any Y_s, $Y_t \in U/\mathrm{IND}(D)$, there will at least exist a Z_i, such that $Z_i \cap Y_s \neq \varnothing$ and $Z_i \cap Y_t \neq \varnothing$, so according to Property 7, we have $E(C - \{r\}; D) = E(C - \{r\}) - E((C - \{r\}) \cup D) > 0$. Obviously, this is contrary to the equality of $E(C; D) = E(C-r\}; D) = 0$, so the hypothesis is not hold, that is, $\mathrm{POS}_{C-\{r\}}(D) \neq U$ is not hold, hence $\mathrm{POS}_{C-\{r\}}(D) = POS_C(D) = U$. So r is D-dispensable in C, thus we have $E(C; D) = E(C-\{r\}; D) = 0$. □

This property shows that a conditional attribute is D-dispensable in C if and only if it has no contribution to the relative knowledge granularity.

Corollary 4: *Let* $DT = (U, C \cup D, V, f)$ *be a consistent decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute, then we will say that* $r \in C$ *is D-indispensable in C if and only if* $E(C; D) \neq E(C-\{r\}; D)$.

Property 13: *Let* $DT = (U, C \cup D, V, f)$ *be a consistent decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute, then we will say that* $r \in C$ *is D-independent in C if and only if for* $\forall r \in C$, $E(C; D) \neq E(C-\{r\}; D)$.

The proof can be easily obtained from Corollary 4.

This property shows that every *D*-independent attribute in *C* has its contribution to relative knowledge granularity.

Corollary 5: *Let* $DT = (U, C \cup D, V, f)$ *be a consistent decision table,* $a \in C$, *then* $a \in CORE_C(D)$ *if and only if* $E(C; D) \neq E(C-\{a\}; D)$.

Property 14: *Let* $DT = (U, C \cup D, V, f)$ *be a consistent decision table, where U is the universe, C is the set of conditional attributes and D is the decision attribute,* $P \subseteq C$, *then P is a D-reduct of C if and only if P satisfies the following two conditions.*

1 $E(P; D) = E(C; D) = 0$.

2 $\forall r \in P$, $E(P; D) \neq E(P-\{r\}; D)$.

The proof can be easily obtained from the definition of relative reduct and Property 13.

This property indicates that a *D*-reduct of *C* is an independent subset of *C*, which has the same relative knowledge granularity with *C*.

Definition 9: *Let* $DT = (U, C \cup D, V, f)$ *be a decision table and* $P \subset C$, *for every attribute* $a \in C-P$, *its significance is denoted as Sig(a ,P, D) and is defined by*

$$Sig(a, P, D) = E(P; D) - E(P \cup \{a\}; D).$$

Especially, if $P = \varnothing$, *then*

$$Sig(a, \varnothing, D) = E(\varnothing; D) - E(\{a\}; D).$$

We have $E(\varnothing; D) = |U| + 1 - E(D)$ *according to Definition 8 and Definition 3.*

Based on this definition, we can design a heuristic reduction algorithm for decision table. The algorithm and analysis about computation complexity is referred to (Feng et al., 2008a).

4 Conclusions

Knowledge granularity, an average measure of the size of the knowledge granule, is also a type of uncertainty arises from the indiscernibility relation. It is caused by lack of sufficient information about them. Consequently, granularity and indiscerniblity are closely connected. In our opinion, knowledge granularity is a measure of uncertainty of an intra-granule. As Chen et al. (2000) pointed out that the uncertainty coming from the granularity of the partition includes inconsistency and randomness. Düntsch and Gediga (1998) constructed an information entropy based uncertainty measure $H^{\det}$ which can deal with the two aspects of the uncertainty. However, the measure of $H^{\det}$ is strictly monotonic increase with the finer of a partition.

In this paper, a new measure of knowledge granularity is defined in the context of rough set theory, which is characterised by mathematical expectation of lengths of granules in a partition. Based on the definition of knowledge granularity, relative knowledge granularity for decision table is defined, which can also deal with the two aspects of the uncertainty, namely, inconsistency and randomness. Moreover, it is strictly monotonic decrease with the finer of a partition. The most advantage of relative knowledge granularity in this paper is that it can reveal the fact that granules belong to positive region have no contribution to the value of the measure. With this observation and the monotonicity of positive region, we adopt the strategy of separate-and-conquer to compute relative knowledge granularity recursively, which is effective, especially for large scale data.

References

Cattaneo, G., Ciucci, D. and Bianucci, D. (2008) 'Entropy and co-entropy of partitions and coverings with applications to roughness theory', in R. Bello et al. (Eds), *Granular Computing*, *STUDFUZZ*, Vol. 224, pp.55–77.

Chen, X., Zhu, S. and Ji, Y. (2000) 'Entropy based uncertainty measures for classification rules with inconsistency tolerance', *2000 IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 4, pp.2816–2821.

Düntsch, I. and Gediga, G. (1998) 'Uncertainty measures of rough set prediction', *Artificial Intelligence*, Vol. 106, No. 1, pp.109–137.

Feng, Q., Miao, D. and Cheng, Y. (2007a) 'Reduction algorithm for information systems based on knowledge partition granularity', *Computer Engineering and Applications*, Vol. 43, No. 34, pp.19–22 (in Chinese).

Feng, Q., Miao, D. and Cheng, Y. (2007b) 'A matrix-based algorithm for computing positive region', *Proceedings of 2007 National Conference on Artificial Intelligence*, pp.203–209 (in Chinese).

Feng, Q. Miao, D. and Cheng, Y. (2008a) 'An approach to knowledge reduction based on relative partition granularity', *Proceedings of the IEEE International Conference of GrC*, pp.226–231.

Feng, Q., Miao, D. and Cheng, Y. (2008b) 'The presentation of relative partition granularity of attributes reduction for decision table', *Journal of Chinese Computer Systems*, Vol. 29, No. 12, pp.2305–2308.

Feng, Q. et al., (2009) 'A partition granularity representation for knowledge', *Pattern Recognition and Artificial Intelligence*, Vol. 22, No. 1, pp.64–69.

Furnkranz, J. (1999) 'Separate-and-conquer rule learning', *Artificial Intelligence Review*, Vol. 13, No. 1, pp.3–54.

Han, J., Hu, X. and Lin, T.Y. (2004) 'Feature subset selection based on relative dependency between attributes', in S. Tsumoto et al. (Eds), *RSCTC 2004, LNAI*, Vol. 3066, pp.176–185.

Hu, X.T., Lin, T.Y. and Han, J. (2003) 'A new rough sets model based on database systems', in G. Wang et al. (Eds), *RSFDGrC 2003, LNAI 2639*, pp.114–121.

Liang, J. and Qian, Y. (2006) 'Axiomatic approach of knowledge granulation in information system', in A. Sattar and B.H. Kang (Eds), *AI 2006*, *Lecture Notes in Artificial Intelligence*, Vol. 4304, pp.1074–1078.

Liang, J. and Qian, Y. (2008) 'Information granules and entropy theory in information systems', *Science in China Series F: Information Sciences*, Vol. 51, No. 9, pp.1–18.

Liang, J.Y. and Shi, Z.Z. (2004) 'The information entropy, rough entropy and knowledge granulation in rough set theory', *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 12, No. 1, pp.37–46.

Miao, D. and Fan, S. (2002) 'The calculation of knowledge granulation and its application', *Systems Engineering Theory and Practice*, Vol. 22, No. 1, pp.48–56 (in Chinese).

Pagallo, G. and Haussler, D. (1990) 'Boolean feature discovery in empirical learning', *Machine Learning*, Vol. 5, No. 1, pp.71–99.

Pawlak, Z. (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Pawlak, Z. (1998) 'Granularity of knowledge, indiscernibility and rough sets', *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp.106–110.

Wierman, M.J. (1999) 'Measuring uncertainty in rough set theory', *Int. J. General Systems*, Vol. 28, No. 4, pp.283–297.

Yao, Y.Y. (2003) 'Probabilistic approaches to rough sets', *Expert Systems*, Vol. 20, No. 5, pp.287–297.

Yao, Y.Y. (2004) 'Granular computing', *Computer Science* (*Ji Suan Ji Ke Xue*), Vol.31, pp.1–5.

Zadeh, L.A. (1979) 'Fuzzy sets and information granularity', in M. Gupta, R. Ragade, R.R. Yager (Eds), *Advances in Fuzzy Set Theory and Applications*, Amsterdam, pp.3–18.

Zhao, M., Yang, Q. and Gao, D. (2008) 'Axiomatic definition of knowledge granularity and its constructive method', *Proceedings of Rough Sets and Knowledge Technology, Lecture Notes in Computer Science*, Berlin; Heidelberg: Springer, Vol. 5009, pp.348–354.