International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems
Vol. , No. () 1–??

UNCERTAINTY MEASURE OF ROUGH SETS BASED ON A KNOWLEDGE GRANULATION FOR INCOMPLETE INFORMATION SYSTEMS

JUNHONG WANG, JIYE LIANG * AND YUHUA QIAN

Key Laboratory of Ministry of Education for Computational Intelligence and Chinese Information Processing, School of Computer and Information Technology, Shanxi University Taiyuan, 030006, China
∗ljy@sxu.edu.cn

CHUANGYIN DANG

Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong

Received (received date)
Revised (revised date)

Rough set theory is a relatively new mathematical tool for computer applications in circumstances characterized by vagueness and uncertainty. In this paper, we address uncertainty of rough sets for incomplete information systems. An axiom definition of knowledge granulation for incomplete information systems is obtained, under which a measure of uncertainty of a rough set is proposed. This measure has some nice properties such as equivalence, maximum and minimum. Furthermore, we prove that the uncertainty measure is effective and suitable for measuring roughness and accuracy of rough sets for incomplete information systems.

Keywords: Incomplete information system; rough set; knowledge granulation; roughness; accuracy; measure.

1. Introduction

The entropy of a system as defined by Shannon[1] gives a measure of uncertainty about its actual structure. It is a powerful mechanism for characterizing information contents in various modes and applications in many diverse fields.

In rough set theory (see, e.g.[2]), data analysis is based on the conviction that the knowledge about the world is available only up to certain granularity, and that the granularity can be expressed mathematically by partitions and their associated equivalence relations. However, the existing uncertainty measure for rough sets has not taken into consideration the granularity of the partition induced by an equivalence relation. In some cases, the uncertainty of a rough set cannot be well characterized by the existing measure. The applications of rough sets in some domains are hence limited. Several authors (see, e.g.[3,4,5]) used Shannon's entropy and its variants to measure uncertainty in rough set theory. A new definition for information entropy in rough set theory is presented in[6]. Unlike the logarithmic behavior of Shannon entropy, the gain function considered there possesses the complement nature. Wierman[7] presented a well justified measure of uncertainty, the measure of granularity, along with an axiomatic derivation. Its strong connections to the Shannon entropy and the Hartley measure of uncertainty also lend strong support to its correctness and applicability. Furthermore, the relationships among information entropy, rough entropy and knowledge granulation in complete information systems are established in[8,9]. For rough sets in complete information systems, measures of uncertainty and rough relation databases were addressed in[10] and an improved uncertainty measure for rough sets was given in[11], which measures uncertainty of rough sets using excess entropy. However, it is difficult to generalize the results in complete information systems to incomplete information systems.

In order to measure uncertainty of rough sets for incomplete information systems, a axiom definition of knowledge granulation is obtained in this paper, under which the knowledge granulation in[8] becomes a special form. Based on knowledge granulation for rough sets, a measure of uncertainty is proposed in this paper. This measure overcomes the limitation of the existing uncertainty measure and hence can be used to measure roughness and accuracy of rough sets for incomplete information systems.

The rest of this paper is organized as follows. Section 2 gives a brief introduction to an incomplete information system. We discuss limitations of the existing uncertainty measure for rough sets in Section 3. Knowledge granulation is studied in Section 4. A measure of uncertainty of a rough set is proposed and a case study is carried out to illustrate the uncertainty measure in Section 5. Finally, Experimental results are summarized in Section 6.

2. An Incomplete Information System

An information system is a pair $S = (U, A)$, where,

(1) U is a non-empty finite set of objects;

(2) A is a non-empty finite set of attributes; and

(3) for every $a \in A$, there is a mapping a, $a : U \to V_a$ where V_a is called the value set of a.

Each subset of attributes $P \subseteq A$ determines a binary indistinguishable relation $IND(P)$ as follows:

$$IND(P) = \{(u, v) \in U \times U |\ \forall a \in P, a(u) = a(v)\}.$$

It can be easily shown that $IND(P)$ is an equivalence relation on the set U.

For $P \subseteq A$, the relation $IND(P)$ constitutes a partition of U, which is denoted by $U/IND(P)$.

In an information system, it may occur that some of the attribute values for an object are missing. For example, in a medical information systems there may be a group of patients for which it is impossible to perform all the required tests. These missing values can be represented by the set of all possible values for the attribute or equivalence by the domain of the attribute. To indicate such a situation, a distinguished value (the so-called null value) is usually assigned to those attributes.

If V_a contains a null value for at least one of the attribute $a \in A$, then S is called an incomplete information system (see, *e.g.*[12,13]), otherwise it is a complete information system. In the following discussions, we will denote the null value by $*$.

For any given information system $S = (U, A)$ and an attribute subset $P \subseteq A$, We define a binary relation on U as follows:

$$SIM(P) = \{(u, v) \in U \times U |\ \forall a \in P, a(u) = a(v) \ \text{ or } \ a(u) = * \ \text{ or } \ a(v) = *\}.$$

In fact, $SIM(P)$ is a similarity relation on U, i.e., it satisfy reflexive and symmetric relations[12]. The concept of a similarity relation has a wide variety of applications in classifications[12]. It can be easily shown that

$$SIM(P) = \bigcap_{a \in P} SIM(\{a\}).$$

Let $S_P(u)$ denote the object set $\{v \in U | (u, v) \in SIM(P)\}$. Then $S_P(u)$ is the maximal set of objects which are possibly indistinguishable by P with u, called similarity class about u.

Let $U/SIM(P)$ denote a classification, which is the family set $\{S_P(u) | u \in U\}$. The elements of $U/SIM(P)$ constitute a covering of U, i.e., for every $u \in U$, $S_P(u) \neq \emptyset$, and $\bigcup_{u \in U} S_P(u) = U$. $U/SIM(P)$ is called a knowledge in U , and every similarity class $S_P(u)$ is called a knowledge granule. Knowledge granulation is the average measure of knowledge granules (similarity classes) in P.

Example 1. Consider the descriptions of several cars in Table 1.

Table 1. The information system about car[14].

Car	Price	Size	Engine	Max-Speed
u_1	high	full	high	low
u_2	low	full	$*$	low
u_3	$*$	compact	$*$	high
u_4	high	full	$*$	high
u_5	$*$	full	$*$	high
u_6	low	full	high	$*$

This is an incomplete information system, where $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$, and $A = \{a_1, a_2, a_3, a_4\}$ with a_1=Price, a_2=Size, a_3=Engine, a_4=Max-Speed.

It is easy to obtain that

$$U/SIM(A) = \{S_A(u_1), S_A(u_2), S_A(u_3), S_A(u_4), S_A(u_5), S_A(u_6)\},$$

where $S_A(u_1) = \{u_1\}$, $S_A(u_2) = \{u_2, u_6\}$, $S_A(u_3) = \{u_3\}$, $S_A(u_4) = \{u_4, u_5\}$, $S_A(u_5) = \{u_4, u_5, u_6\}$, $S_A(u_6) = \{u_2, u_5, u_6\}$.

In the studies of an incomplete information system, a particular interest is given to the discrete classification $\omega(U) = \{\{x\}|x \in U\}$, and the indiscrete classification $\delta(U) = \{U\}$, or just ω and δ if there is no confusion as to the domain set involved.

For any given incomplete information system $S = (U, A)$, $P \subseteq A$ and $X \subseteq U$, we can define a lower approximation of X in U and a upper approximation of X in U by

$$\underline{P}X = \{x \in U | S_P(x) \subseteq X\} = \{x \in X | S_P(x) \subseteq X\}$$

and

$$\overline{P}X = \{x \in U | S_P(x) \cap X \neq \phi\} = \cup\{S_P(x) | x \in X\}.$$

As what in a complete information system, $\underline{P}X$ is a set of objects that belong to X with certainty, while $\overline{P}X$ is a set of objects that possibly belong to X.

Now we define a partial order on the set of all the classifications of U. Let $P, Q \subseteq A$. We say that Q is coarser than P (or P is finer that Q), denoted by $P \preceq Q$, if and only if $S_P(u_i) \subseteq S_Q(u_i)$ for $\forall i \in \{1, 2, \cdots, |U|\}$. If $P \preceq Q$ and $P \neq Q$, and $\exists j \in \{1, 2, \cdots, |U|\}$ such that $S_P(u_j) \subset S_Q(u_j)$, then we say that Q is strictly coarser than P (or P is strictly finer than Q) and denoted by $P \prec Q$.

3. An Existing Measure for Rough Sets and Its Limitation

Pawlak[2] discusses two numerical characterizations of uncertainty of a rough set: accuracy and roughness. The accuracy measures the degree of completeness of knowledge about the given rough set X, and is defined by the ratio of the cardinalities of the lower and upper approximation sets of X as follows:

$$\alpha_R(A) = \frac{|\underline{R}X|}{|\overline{R}X|}. \tag{1}$$

The roughness represents the degree of incompleteness of knowledge about the rough set, and is calculated by subtracting the accuracy from 1:

$$\rho_R(A) = 1 - \alpha_R(A). \tag{2}$$

These measures require knowledge of the number of elements in each of the approximation sets and are good metrics for uncertainty as it arises from the boundary region, implicitly taking into account equivalence classes as they belong entirely or partially to the set. However, accuracy and roughness measures do not necessarily provide us with information on the uncertainty related to the granularity of the indiscernibility relation.

Example 2. (Continued from Example 1). Let $P = \{a_2, a_4\}$, then,

$$U/SIM(P) = \{S_P(u_1), S_P(u_2), S_P(u_3), S_P(u_4), S_P(u_5), S_P(u_6)\},$$

where $S_P(u_1) = \{u_1, u_2, u_6\}$, $S_P(u_2) = \{u_1, u_2, u_6\}$, $S_P(u_3) = \{u_3\}$, $S_P(u_4) = \{u_4, u_5, u_6\}$, $S_P(u_5) = \{u_4, u_5, u_6\}$, $S_P(u_6) = \{u_1, u_2, u_4, u_5, u_6\}$.

It can be easily observed that

$$U/SIM(A) \subset U/SIM(P).$$

Considering the set $X = \{u_1, u_2, u_4, u_6\}$, we have that

$$\underline{A}X = \underline{P}X = \{u_1, u_2\} \text{ and } \overline{A}X = \overline{P}X = \{u_1, u_2, u_4, u_5, u_6\}.$$

Thus,

$$\alpha_A(X) = \alpha_P(X) = 0.4 \text{ and } \rho_A(X) = \rho_P(X) = 0.6.$$

Notice that, in Example 2, an inclusion relation exists in two knowledge representation systems $U/SIM(A)$ and $U/SIM(P)$, but the same accuracy or roughness can be obtained for the rough set X. Therefore, it is necessary for us to introduce a more accurate measure for rough sets. The problem of measuring uncertainty of rough sets has been solved for a complete information system. However, it remains unsolved for an incomplete information system.

4. Knowledge Granulation

In the view of granular computing, a granule is a dump of objects in the universe of discourse, drawn together by indistinguishability, similarity, proximity, or functionality. In fact, knowledge granulation is the average measure of knowledge granules on the universe and also represents the ability of classifications.

In fact, the granulation gives an easily understandable description for partitioning of the universe. In 1979, the problem of fuzzy information granule was introduced by Zadeh[15]. Especially, several measures in an information system closely associated with granular computing such as granulation measure, information entropy, rough entropy and knowledge granulation and their relationships were discussed in[8]. However, there exists no unified description for knowledge granulation. In the following, an axiom definition of knowledge granulation is given. We prove that the above granular measure is a special form of the axiom definition.

Definition 1. For any given incomplete information system $S = (U, A)$, let G be a mapping from the power set of A to the set of real numbers. We say that G is a knowledge granulation in an incomplete information system $S = (U, A)$ if G satisfies the following conditions:

(1) $G(P) \geq 0$ for any $P \subseteq A$ (Non-negativity);

(2) $G(P) = G(Q)$ for any $P, Q \subseteq A$ if there is a bijective mapping $f : U/SIM(P) \longrightarrow U/SIM(Q)$ such that $|S_P(u_i)| = |f(S_P(u_i))|$

($\forall i \in \{1,2,\cdots,|U|\}$), where $U/SIM(P) = \{S_P(u_1), S_P(u_2), \cdots, S_P(u_{|U|})\}$ and $U/SIM(Q) = \{S_Q(u_1), S_Q(u_2), \cdots, S_Q(u_{|U|})\}$; and

(3) $G(P) \leq G(Q)$ for any $P, Q \subseteq A$ with $P \preceq Q$ (Monotonicity).

In (8), a different kind of knowledge granulation was given, which is as follows.

Definition 2^8**.** Granulation of knowledge of A is given by

$$GK(A) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |S_A(u_i)|, \tag{3}$$

where $U/SIM(A) = \{S_A(u_1), S_A(u_2), \cdots, S_A(u_{|U|})\}$.

Obviously, when $S = (U, A)$ is an incomplete information system, $1/|U| \leq GK(R) \leq 1$ for any subset R of A.

Proposition 1. *GK in Definition 2 is a knowledge granulation under definition 1.*

Proof. It suffices to show that GK meets all the conditions in Definition 1.

(1) Obviously, $GK(R)$ is non-negative.

(2) Let $P, Q \subseteq A$, $U/SIM(P) = \{S_P(u_1), S_P(u_2), \cdots, S_P(u_{|U|})\}$ and $U/SIM(Q) = \{S_Q(u_1), S_Q(u_2), \cdots, S_Q(u_{|U|})\}$. Suppose that there be a bijiective mapping function $f : U/SIM(P) \longrightarrow U/SIM(Q)$ such that $|S_P(u_i)| = |f(S_P(u_i))|$ ($\forall i \in \{1,2,\cdots,|U|\}$). Let $f(S_P(u_i) = S^c_Q(u_i)$, where $S^c_Q(u_i) \in \{S_Q(u_1), S_Q(u_2), \cdots, S_Q(u_{|U|})\}$. Then, we have that

$$\begin{aligned} GK(P) &= \tfrac{1}{|U|^2} \sum_{i=1}^{|U|} |S_P(u_i)| \\ &= \tfrac{1}{|U|^2} \sum_{i=1}^{|U|} |S^c_Q(u_i)| \\ &= \tfrac{1}{|U|^2} \sum_{i=1}^{|U|} |S_Q(u_i)| \\ &= GK(Q). \end{aligned}$$

(3) Let $P, Q \subseteq A$ satisfying $P \preceq Q$. Then, for $\forall i \in \{1,2,\cdots,|U|\}$, $S_P(u_i) \subseteq S_Q(u_i)$, i.e., $|S_P(u_i)| \leq |S_Q(u_i)|$.

Hence,

$$\begin{aligned} GK(P) &= \tfrac{1}{|U|^2} \sum_{i=1}^{|U|} |S_P(u_i)| \\ &\leq \tfrac{1}{|U|^2} \sum_{i=1}^{|U|} |S_Q(u_i)| \\ &= GK(Q). \end{aligned}$$

Thus, GK in Definition 2 is the knowledge granulation under Definition 1.
This completes the proof. □

5. Measure of Rough Sets Based on Knowledge Granulation

In this section, based on knowledge granulation in incomplete information systems, a measure of uncertainty of rough sets is proposed.

Definition 3. Let $S = (U, A)$ be an incomplete information system, $\emptyset \subset X \subseteq U$, and $R \subseteq A$. The roughness of X with respect to R is defined as follows:

$$Roughness_R(X) = \rho_R(X)GK(R), \tag{4}$$

where $\rho_R(X) = 1 - \frac{|\underline{R}X|}{|\overline{R}X|}$.

Apparently, the granularity of the partition induced by the equivalence relation R has been considered in the new definition. In the following, we show that the redefined roughness measure has some meaningful properties and is valid in measuring uncertainty of rough sets.

Property 1. (Equivalence) *Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$ and $\emptyset \subset X \subseteq U$. If $U/SIM(P) = U/SIM(Q)$, then $Roughness_P(X) = Roughness_Q(X)$.*

Property 2. (Maximum) *Let $S = (U, A)$ be an incomplete information system, $R \subseteq A$ and $\emptyset \subset X \subseteq U$. The maximum roughness of X with respect to R is 1. This value is achieved only when $U/SIM(R) = \delta$.*

Property 3. (Minimum) *Let $S = (U, A)$ be an incomplete information system, $R \subseteq A$ and $\emptyset \subset X \subseteq U$. The minimum roughness of X with respect to R is 0. This value is achieved only when $U/SIM(R) = \omega$.*

Obciously, when $S = (U, A)$ is an incomplete information system, $0 \leq Roughness_R(X) \leq 1$ for any subset R of A.

Proposition 2. *Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$ and $\emptyset \subset X \subseteq U$. If $P \preceq Q$, then*

$$Roughness_P(X) \leq Roughness_Q(X).$$

.

Proof. Let $P \preceq Q$. It is easy to obtain that $GK(P) \leq GK(Q)$ and $0 \leq \rho_R(X) \leq \rho_Q(X)$. Then

$$\begin{aligned} Roughness_P(X) &= \rho_P(X)GK(P) \\ &\leq \rho_Q(X)GK(Q) \\ &= Roughness_Q(X). \end{aligned}$$

This completes the proof. □

Proposition 2 states that the roughness of X with respect to R decreases as R become finer.

Definition 4. Let $S = (U, A)$ be an incomplete information system, $\emptyset \subset X \subseteq U$, and $R \subseteq A$. The accuracy of X with respect to R is defined as follows:

$$Accuracy_R(X) = 1 - \rho_R(X)GK(R), \tag{5}$$

where $\rho_R(X) = 1 - \frac{|\underline{R}X|}{|\overline{R}X|}$.

Property 4. (Equivalence) *Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$ and $\emptyset \subset X \subseteq U$. If $U/SIM(P) = U/SIM(Q)$, then $Accuracy_P(X) = Accuracy_Q(X)$.*

Property 5. (Maximum) *Let $S = (U, A)$ be an incomplete information system, $R \subseteq A$ and $\emptyset \subset X \subseteq U$. The maximum accuracy of X with respect to R is 1. This value is achieved only when $U/SIM(R) = \omega$.*

Property 6. (Minimum) *Let $S = (U, A)$ be an incomplete information system, $R \subseteq A$ and $\emptyset \subset X \subseteq U$. The minimum accuracy of X with respect to R is 0. This value is achieved only when $U/SIM(R) = \delta$.*

Obciously, when $S = (U, A)$ is an incomplete information system, $0 \leq Accuracy_R(X) \leq 1$ for any subset R of A.

Proposition 3. *Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$ and $\emptyset \subset X \subseteq U$. If $P \preceq Q$, then*

$$Accuracy_P(X) \geq Accuracy_Q(X).$$

.

Proof. Let $P \preceq Q$. It is easy to obtain that $Roughness_P(X) \leq Roughness_Q(X)$. Thus

$$\begin{aligned} Accuracy_P(X) &= 1 - \rho_P(X)GK(P) \\ &= 1 - Roughness_P(X) \\ &\geq 1 - Roughness_Q(X) \\ &= Accuracy_Q(X). \end{aligned}$$

This completes the proof. □

Proposition 3 states that the Accuracy of X with respect to R increases as R become finer.

Example 3. (Continued from Example 2). Let $\rho_A(X) = \rho_P(X) = 0.6$. Then,

$$GK(A) = \frac{1}{36}(1 + 2 + 1 + 2 + 3 + 3) = 0.333$$

and

$$GK(P) = \frac{1}{36}(3+3+1+3+3+5) = 0.5.$$

Hence, the roughness of X with respect to A and P is given by

$$Roughness_A(X) = 0.200$$

and

$$Roughness_P(X) = 0.300,$$

respectively. It is clear that the roughness of X with respect to R decreases as R becomes finer.

The accuracy of X with respect to A and P is given by

$$Accuracy_A(X) = 0.800$$

and

$$Accuracy_P(X) = 0.700,$$

The accuracy of X with respect to R increases as R becomes finer.

6. Experiments

In this section, we will describe our main experiments results.

Using the method in the paper, we have performed on some real data. As an example, we consider the information system about planning tennis ball (see table 2). Where $U = \{u_1, u_2, ..., u_{24}\}$, and $A = \{a_1, a_2, a_3, a_4\}$ with a_1=Outlook, a_2=Temperature, a_3=Humidity and a_4=Windy.

Let $P = \{a_2, a_4\}$. By computing, we have $GK(A) = 0.257$ and $GK(P) = 0.486$. It show that Granulation of knowledge of A and Granulation of knowledge of P is different, because $U/SIM(A) \subset U/SIM(P)$ or A is finer than P.

By performing our experiment on some sets of objects in table 2, we can get a number of especial data showed in table 3 and 4. These sets have same lower approximation and upper approximation according to A and P, also they have same $\rho(X)$, but they have different $Roughness(X)$. The reason is A and P have different knowledge granulation. Thus, the uncertainty of X with respect to a different equivalance relation is well characterized. The experiments data illuminated that the method in the paper is effective.

where

$X_1 = \{u_1, u_4, u_6, u_8, u_9, u_{15}, u_{17}, u_{21}, u_{24}\}$,

$X_2 = \{u_3, u_4, u_5, u_6, u_{10}, u_{11}, u_{12}, u_{15}, u_{16}, u_{19}, u_{21}, u_{22}, u_{23}\}$,

Table 2. The information system about planning tennis ball.

Events	Outlook	Temperature	Humidity	Windy
u_1	Overcast	Hot	High	*
u_2	Overcast	Hot	High	Very
u_3	*	Hot	*	Medium
u_4	*	Hot	High	Not
u_5	Sunny	*	High	*
u_6	Rain	Mild	*	Not
u_7	Rain	*	High	Medium
u_8	Rain	Hot	*	Not
u_9	Rain	Cool	Normal	*
u_{10}	*	Hot	*	Very
u_{11}	*	*	Normal	very
u_{12}	Sunny	Cool	Normal	*
u_{13}	Overcast	*	High	*
u_{14}	Overcast	Mild	*	*
u_{15}	*	Cool	Normal	Not
u_{16}	*	Cool	Normal	Medium
u_{17}	Rain	*	Normal	Not
u_{18}	Rain	*	Normal	Medium
u_{19}	*	Mild	*	Medium
u_{20}	Overcast	Mild	Normal	Very
u_{21}	*	Mild	*	Very
u_{22}	Sunny	Mild	High	*
u_{23}	Sunny	*	Normal	*
u_{24}	Rain	*	High	*

Table 3. The lower approximation and upper approximation of some rough sets.

X	$\underline{A}X = \underline{P}X$	$\overline{A}X = \overline{P}X$
X_1	$\{u_6, u_8, u_{17}\}$	$\{u_1, u_2, ..., u_{24}\}$
X_2	$\{u_5, u_{12}, u_{22}, u_{23}\}$	$\{u_1, u_2, ..., u_{24}\}$
X_3	$\{u_1, u_2, u_{13}, u_{14}, u_{20}\}$	$\{u_1, u_2, ..., u_{24}\}$
X_4	$\{u_7, u_{18}\}$	$\{u_1, u_3, ..., u_{19}, u_{21}, ..., u_{24}\}$
X_5	$\{u_2, u_{20}\}$	$\{u_1, ..., u_5, u_9, ..., u_{16}, u_{19}, ..., u_{24}\}$
X_6	$\{u_3, u_4, u_6, u_7, u_8, u_{15}, ..., u_{19}\}$	$\{u_1, u_2, ..., u_{24}\}$

$$X_3 = \{u_1, u_2, u_3, u_4, u_6, u_{10}, u_{11}, u_{13}, u_{14}, u_{15}, u_{16}, u_{19}, u_{20}, u_{21}\},$$

$$X_4 = \{u_3, u_7, u_9, u_{16}, u_{18}, u_{19}, u_{24}\},$$

$$X_5 = \{u_1, u_2, u_{10}, u_{11}, u_{12}, u_{13}, u_{14}, u_{20}, u_{21}\},$$

$$X_6 = \{u_1, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{12}, u_{13}, u_{14}, u_{15}, u_{16}, u_{17}, u_{18}, u_{19}, u_{22}, u_{23}, u_{24}\}.$$

Table 4. The comparison of classical roughness measure and the measure in the paper.

X	$\rho_A(X) = \rho_P(X)$	$Roughness_A(X)$	$Roughness_P(X)$
X_1	0.875	0.225	0.425
X_2	0.833	0.214	0.405
X_3	0.792	0.204	0.385
x_4	0.909	0.234	0.442
X_5	0.895	0.230	0.435
X_6	0.583	0.150	0.283

7. Conclusions

In this paper, an axiom definition of knowledge granulation has been given. We have proved that the knowledge granulation in[8] is a special form of the axiom definition. Based on knowledge granulation, a measure of uncertainty of a rough set has been proposed. Several nice properties of this measure have been derived. We have demonstrated that the new measure overcomes the limitation of the existing uncertainty measure and can be used to measure with a simple and comprehensive form the roughness and accuracy of a rough set for an incomplete information system.

Acknowledgements

This work was supported by SRG: 7001977 of City University of Hong Kong, Hi-tech Research and Development Program of China (No. 2007AA01Z165), the National Natural Science Foundation of China (No. 60773133, No. 70471003, No. 60275019), the Foundation of Doctoral Program Research of Ministry of Education of China (No. 20050108004), the Top Scholar Foundation of Shanxi, China, Key Project of Science and Technology Research of the Ministry of Education of China (No. 206017), the Key Laboratory Open Foundation of Shanxi (No. 200603023).

References

1. C. E. Shannon, “The mathematical theory of communication”, *The Bell System Technical Journal* **27** (3 and 4) (1948) 373-423 (see also 623-656).
2. Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data* (Kluwer Academic Publishers, Dordecht, 1991).
3. F. E. Chakik, A. Shahine, J. Jaam and A. Hasnah, “An approach for constructing complex discriminating surfaces based on bayesian interference of the maximum entropy”, *Information Sciences* **163** (2004) 275-291.
4. I. Düntsch, and G. Gediga, “Uncertainty measures of rough set prediction” , *Artificial Intelligence* **106** (1998) 109-137.
5. J. Klir and M. J. Wierman, *Uncertainty-Based Information: Elements of Generalished Information Theory* (Physica-Verlag, New York, 1999).
6. J. Y. Liang, K. S. Chin, C. Y. Dang and C. M. YAM, “A new method for measure-

ing uncertainty and fuzziness in rough set theory", *International Journal of General Systems* **31** (4) (2002) 331-342.
7. M. J. Wierman, "Measuring uncertainty in rough set theory", *International Journal of General Systems* **28** (1999) 283-297.
8. J. Y. Liang and Z. Z. Shi, "The information entropy, rough entropy and knowledge granulation in rough set theory", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **12** (1) (2004) 37-46.
9. J. Y. Liang and D. Y. Li, *Uncertainty and Knowledge Acquisition in Information System, Science Press (in Chinese)* (Beijing, 2005).
10. T. Beaubouef, F. E. Petry and G. Arora, "Information-theoretic measures of uncertainty for rough sets and rough relational databases", *Information Sciences* **109** (1998) 185-195.
11. B. W. Xu, Y. M. Zhou and H. M. Lu, "An improved accuracy measure for rough sets", *Journal of Computer and System Sciences* **71** (2005) 163-173.
12. M. Kryszkiewicz, "Rough set approach to incomplete information systems", *Inforamtion Sciences* **112** (1998) 39-49
13. M. Kryszkiewicz, "Rules in incomplete information systems", *Inforamtion Sciences* **113** (1999) 271-292
14. J. Y. Liang and Z. B. Xu, "The algorithm on knowledge reduction in incomplete information systems", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10** (1) (2002) 95-103.
15. L. A. Zadeh, *Fuzzy Sets and information granularity, in: M. Gupta, R. Ragade, and R. Yager, (Eds), Advances in Fuzzy Set Theory and Application* (North-Holland, Amsterdam 1979) pp.3-18.