A new measure of uncertainty based on knowledge granulation for rough sets

Jiye Liang *, Junhong Wang, Yuhua Qian

Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, Shanxi, China
School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China

ARTICLE INFO

ABSTRACT

In rough set theory, accuracy and roughness are used to characterize uncertainty of a set and approximation accuracy is employed to depict accuracy of a rough classification. Although these measures are effective, they have some limitations when the lower/upper approximation of a set under one knowledge is equal to that under another knowledge. To overcome these limitations, we address in this paper the issues of uncertainty of a set in an information system and approximation accuracy of a rough classification in a decision table. An axiomatic definition of knowledge granulation for an information system is given, under which these three measures are modified. Theoretical studies and experimental results show that the modified measures are effective and suitable for evaluating the roughness and accuracy of a set in an information system and the approximation accuracy of a rough classification in a decision table, respectively, and have a much simpler and more comprehensive form than the existing ones.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Recently, rough set theory developed by Pawlak in [11] has become a popular mathematical framework in areas such as pattern recognition, image processing, feature selection, neural computing, conflict analysis, decision support, data mining and knowledge discovery process from large data sets [1,10,13–16]. As one of the most important issues in rough set theory, uncertainty of a set has been widely studied. As follows, we briefly review some relevant literatures.

To evaluate uncertainty of a system, the concept of entropy was introduced by Shannon in [21]. It is a very useful mechanism for characterizing information contents in various modes and has been applied in diverse fields. The entropy and its variants were adapted for rough set theory in [4,23] and information interpretation of rough set theory was given in [5,17]. Beaubouef et al. [2] addressed information measures of uncertainty of rough sets and rough relation databases. In [6], a new method for evaluating both uncertainty and fuzziness was proposed. Unlike most existing information entropies, Qian and Liang [19] proposed a so-called combination entropy for evaluating uncertainty of a knowledge from an information system. All these studies were dedicated to evaluating uncertainty of a set in terms of the partition ability of a knowledge. As a powerful mechanism, granulation was first introduced by Zadeh in [26]. It presents a more visual and easily understandable description for a partition on the universe. To characterize the granulation, granular computing was introduced in [27,28], which, as a term with many meanings, covers all the research related to granulations. With regard to granular computing, many pieces of nice work were accomplished in [3,9,18,20,25,29]. Especially, closely associated with granular computing, several measures on knowledge in an information system were proposed and the relationships between these

---

* Corresponding author. Address: School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China.
E-mail addresses: ljy@sxu.edu.cn (J. Liang), cswjh@126.com (J. Wang), jinchengqyh@126.com (Y. Qian).

measures were discussed in [7,8]. These measures include granulation measure, information entropy, rough entropy, and knowledge granulation, and have become effective mechanisms for evaluating uncertainty in rough set theory.

To evaluate uncertainty of a set, Pawlak presented several numerical measures in [12], which are accuracy and roughness of a set and approximation accuracy of a rough classification. Although these measures are effective, they have some restrictions. For instances, when the lower/upper approximation of a set under one knowledge is equal to that under another knowledge, the imprecision of a rough set and approximation accuracy of a rough classification cannot be well characterized by these measures. The applications of rough set theory in some fields are hence limited. To solve this problem, an improved accuracy measure for rough sets was given in [24], which calculates the imprecision of a set using an excess entropy. However, this measure has a complex mathematical form. To overcome the shortcomings of the existing measures, we address in this paper the issues of uncertainty of a set in an information system and approximation accuracy of a rough classification in a decision table. An axiomatic definition of knowledge granulation for an information system is given, under which the measures of accuracy, roughness and approximation accuracy are modified. Theoretical studies and experimental results show that the modified measures are effective and suitable for evaluating the roughness and accuracy of a set in an information system and the approximation accuracy of a rough classification in a decision table, respectively, and have a much simpler and more comprehensive form than the existing ones.

The rest of this paper is organized as follows. Some preliminary concepts such as information systems, indiscernibility relation, partition, lower and upper approximations, partial relation of knowledge and decision tables are briefly recalled in Section 2. In Section 3, an axiomatic definition of knowledge granulation for information systems is introduced, and two definitions of knowledge granulation in the literatures are proved to be special forms of the axiomatic definition. In Section 4, through some examples, the limitations of the classical measures are revealed. In Section 5, based on a knowledge granulation, three modified measures (*Accuracy*, *Roughness* and γ) are proposed to an information system, and numerical experiments are performed on practical data sets to show the validity of these three modified measures. Section 6 concludes this paper with some remarks and discussions.

2. Some basic concepts

In this section, several basic concepts are reviewed, which are information systems, indiscernibility relation, partition, lower and upper approximations, partial relation of knowledge and decision tables.

In the following, we first recall the concept of information systems.

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, $a: U \to V_a$ with V_a being 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 $U/IND(P)$ or just U/P.

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

As follows, we give the definitions of a lower approximation and an upper approximation in rough set theory.

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

$$\underline{P}X = \bigcup\{P_i \in U/IND(P) | P_i \subseteq X\}$$

and

$$\overline{P}X = \bigcup\{P_i \in U/IND(P) | P_i \bigcap X \neq \emptyset\},$$

where $\underline{P}X$ 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.

We define a partial relation $\preceq$ on the family $\{U/B | B \subseteq A\}$ as follows [12]: $U/P \preceq U/Q$ (or $U/Q \succeq U/P$) if and only if, for every $P_i \in U/P$, there exists $Q_j \in U/Q$ such that $P_i \subseteq Q_j$, where $U/P = \{P_1, P_2, \ldots, P_m\}$ and $U/Q = \{Q_1, Q_2, \ldots, Q_n\}$ are partitions induced by $P, Q \subseteq A$, respectively. In this case, we say that Q is coarser than P, or P is finer than Q. If $U/P \preceq U/Q$ and $U/P \neq U/Q$, we say Q is strictly coarser than P (or P is strictly finer than Q), denoted by $U/P \prec U/Q$ (or $U/Q \succ U/P$).

A decision table is an information system $S = (U, C \cup D)$ with $C \cap D = \emptyset$, where an element of C is called a condition attribute, C is called a condition attribute set, an element of D is called a decision attribute, and D is called a decision attribute set. If $U/C \preceq U/D$, then $S = (U, C \cup D)$ is said to be consistent; otherwise it is said to be inconsistent [12]. For example, a decision table about cars is given in Table 1.

Table 1
An information system about cars.

Car	Price	Size	Engine	Max-speed	d
u_1	Low	Compact	Gasoline	Low	Poor
u_2	Low	Full	Diesel	High	Good
u_3	High	Full	Diesel	Medium	Poor
u_4	High	Full	Diesel	Low	Poor
u_5	Low	Full	Gasoline	High	Poor

In Table 1, $U = \{u_1, u_2, u_3, u_4, u_5\}$ is the universe, $C = \{c_1, c_2, c_3, c_4\}$ with $c_1 =$Price, $c_2 =$Size, $c_3 =$Engine, $c_4 =$Max-Speed is a condition attribute set, and $D = \{d\}$ is a decision attribute set.

3. Knowledge granulation

In this section, an axiomatic definition of knowledge granulation in an information system is introduced, and the knowledge granulations given in [7] and [24] are proved to be special forms under this definition.

Definition 1. For any given 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 information system $S = (U, A)$ if G satisfies the following conditions:

(1) $G(P) \geqslant 0$ for any $P \subseteq A$ (Non-negativity);
(2) $G(P) = G(Q)$ for any $P, Q \subseteq A$ if there is a bijective mapping $f : U/P \to U/Q$ such that $|X_i| = |f(X_i)| (\forall i \in \{1, 2, \ldots, m\})$, where $U/P = \{X_1, X_2, \ldots, X_m\}$ and $U/Q = \{Y_1, Y_2, \ldots, Y_m\}$; and
(3) $G(P) < G(Q)$ for any $P, Q \subseteq A$ with $P \prec Q$ (Monotonicity).

In [7], a definition of knowledge granulation was given, which is as follows.

Definition 2 [7]. Let $S = (U, A)$ be an information system and $U/A = \{X_1, X_2, \ldots, X_m\}$. A knowledge granulation of A is given by

$$GK(A) = \frac{1}{|U|^2} \sum_{i=1}^{m} |X_i|^2. \tag{1}$$

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

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

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

(1) Obviously, $GK(R)$ is non-negative.
(2) Let $P, Q \subseteq A$, $U/P = \{X_1, X_2, \ldots, X_m\}$ and $U/Q = \{Y_1, Y_2, \ldots, Y_n\}$. Suppose that there be a bijective mapping $f : U/P \to U/Q$ such that $|X_i| = |f(X_i)| (\forall i \in \{1, 2, \ldots, m\})$. Then, $n = m$ and

$$GK(P) = \frac{1}{|U|^2} \sum_{i=1}^{m} |X_i|^2 = \frac{1}{|U|^2} \sum_{i=1}^{m} |f(X_i)|^2 = \frac{1}{|U|^2} \sum_{i=1}^{n} |Y_i|^2 = GK(Q).$$

(3) Let $P, Q \subseteq A$ satisfying $P \prec Q$, $U/P = \{X_1, X_2, \ldots, X_m\}$ and $U/Q = \{Y_1, Y_2, \ldots, Y_n\}$. Then, $m > n$ and there exists a partition $C = C_1, C_2, \ldots, C_n$ of $1, 2, \ldots, m$ such that

$$Y_j = \bigcup_{i \in C_j} X_i, \quad j = 1, 2, \ldots, n.$$

Hence,

$$GK(Q) = \frac{1}{|U|^2} \sum_{j=1}^{n} |Y_j|^2 = \frac{1}{|U|^2} \sum_{j=1}^{n} \left| \bigcup_{i \in C_j} X_i \right|^2 = \frac{1}{|U|^2} \sum_{j=1}^{n} \left(\sum_{i \in C_j} |X_i| \right)^2 > \frac{1}{|U|^2} \sum_{j=1}^{n} \left(\sum_{i \in C_j} |X_i|^2 \right) = \frac{1}{|U|^2} \sum_{i=1}^{m} |X_i|^2 = GK(P).$$

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

In [24], Xu and Zhou used a graph, called *ERG*, to represent a partition of a given universe. Let R be an equivalence relation defined on U, and the graph *ERG* with respect to R be given by $G(R) = (N(R), E(R))$, where $N(R) = U$ and $E(R) = \{(x, y)|xRy\}$. Then the roughness of R is defined as

$$Rough(R) = \frac{connection(G(R))}{|N(R)|(|N(R)| - 1)\log_2 |N(R)|},$$ (2)

where $N(R) = U$, v is an node in $N(R)$, $connection(G(R)) = \sum_{v \in N(R)} I(G(R)^v) - I(G(R))$ and $I(G(R))$ is the entropy of the information source multiplied by the number of nodes.

Proposition 2. *Rough*(R) *in* [24] *is a knowledge granulation under* Definition 1.

Proof. It is sufficient to show that *GK* meets all the conditions in Definition 1.

(1) Let $R \subseteq A$. Since information theory shows that the entropy of a joint distribution is not over the sum of the entropies of all the components, i.e. $\sum_{v \in N(R)} I(G(R)^v) \geqslant I(G(R))$, hence, $connection(G(R)) = \sum I(G(R)^v) - I(G(R)) \geqslant 0$. And $|N(R)|(|N(R)| - 1)\log_2 |N(R)|$ is a constant greater than zero. Thus, $Rough(R)$ is non-negative.

(2) Let $P, Q \subseteq A$, $U/P = \{X_1, X_2, \ldots, X_m\}$ and $U/Q = \{Y_1, Y_2, \ldots, Y_n\}$. Suppose that there be a bijective mapping $f : U/P \to U/Q$ such that $|X_i| = |f(X_i)| (\forall i \in \{1, 2, \ldots, m\})$. Then, $n = m$, $I(G(P)) = I(G(Q))$, $I(G(P)^v) = I(G(Q)^w)$, and for $v, w \in N(R))$, $v \in X_i, w \in Y_i$ and $|Y_i| = |X_i|$. Thus, $Rough(P) = Rough(Q)$.

(3) Let $P, Q \subseteq A$ satisfying $Q \prec P$, $U/P = \{X_1, X_2, \ldots, X_m\}$ and $U/Q = \{Y_1, Y_2, \ldots, Y_n\}$. Let $G(P) = (N(P), E(P))$ and $G(Q) = (N(Q), E(Q))$ be the *ERG* graphs of P and Q, respectively. Since P and Q are two equivalence relations on U, hence, $N(P) = N(Q) = U$. It follows from $P \prec Q$ that if $(x, y) \in IND(Q)$, then $(x, y) \in IND(P)$ for any $x, y \in U$, i.e. $E(Q) \subset E(P)$. Therefore, $connection(G(Q)) < connection(G(P))$. Due to that P and Q are equivalence relations defined on U, $|N(P)|(|N(P)| - 1)\log_2 |N(P)| = |N(Q)|(|N(Q)| - 1)\log_2 |N(Q)| = |U|(|U| - 1)\log_2^{|U|}$. It follows immediately that $Rough(Q) < Rough(P)$.

These results together implies that $Rough(R)$ in [24] is a knowledge granulation under Definition 1. $\square$

Remark. In general, a knowledge granulation provides an important approach to measuring the discernibility ability of a knowledge in rough set theory. The smaller the knowledge granulation is, the stronger its discernibility ability is.

4. Limitations of classical measures in rough set theory

In this section, through two illustrative examples, we reveal the limitations of existing classical measures for evaluating uncertainty of a set and approximation accuracy of a rough classification.

In [12], Pawlak proposed two numerical measures for evaluating uncertainty of a set: accuracy and roughness. The accuracy measure is equal to the degree of completeness of a knowledge about the given object set X, and is defined by the ratio of the cardinalities of the lower and upper approximation sets of X as follows:

$$\alpha_R(X) = \frac{|\underline{R}X|}{|\overline{R}X|}.$$ (3)

The roughness measure represents the degree of incompleteness of a knowledge about the set, and is calculated by subtracting the accuracy from one:

$$\rho_R(X) = 1 - \alpha_R(A).$$ (4)

These measures take into account the number of elements in each of the approximation sets and are good metrics for evaluating uncertainty that arises from the boundary region. However, the accuracy and roughness do not provide the information that is caused by the uncertainty related to the granularity of the indiscernibility relation. Their limitations are revealed by the following example.

Example 1. In Table 1, suppose that $R_1, R_2, R_3 \subseteq C$, where $R_1 = \{c_1\}$, $R_2 = \{c_1, c_2\}$ and $R_3 = \{c_1, c_2, c_3\}$.
By calculating, one can have

$$U/R_1 = \{\{u_1, u_2, u_5\}, \{u_3, u_4\}\},$$
$$U/R_2 = \{\{u_1\}, \{u_2, u_5\}, \{u_3, u_4\}\}$$

and

$$U/R_3 = \{\{u_1\}, \{u_2\}, \{u_3, u_4\}, \{u_5\}\}.$$

It is easy to see that

$$R_3 \prec R_2 \prec R_1.$$

Suppose $X = \{u_1, u_2, u_3, u_5\}$. Then, it is obvious that

$$\underline{R_1}X = \underline{R_2}X = \underline{R_3}X = \{u_1, u_2, u_5\}$$

and

$$\overline{R_1}X = \overline{R_2}X = \overline{R_3}X = \{u_1, u_2, u_3, u_4, u_5\}.$$

Thus,

$$\alpha_{R_1}(X) = \alpha_{R_2}(X) = \alpha_{R_3}(X) = 0.6$$

and

$$\rho_{R_1}(X) = \rho_{R_2}(X) = \rho_{R_3}(X) = 0.4.$$

Note that, in Example 1, there is a partial relation in three knowledge representation systems U/R_1, U/R_2 and U/R_3, but the same accuracy or roughness is obtained for each of the three rough sets of X induced by the three knowledge systems, respectively. Therefore, it is necessary to introduce more effective measures.

In [12], approximation accuracy of a rough classification was introduced by Pawlak. Let $F = \{Y_1, Y_2, \ldots, Y_n\}$ be a classification of the universe U, and C an attribute set in an information system. Then, C-lower and C-upper approximations of F are given by $\underline{C}F = \{\underline{C}Y_1, \underline{C}Y_2, \ldots, \underline{C}Y_n\}$ and $\overline{C}F = \{\overline{C}Y_1, \overline{C}Y_2, \ldots, \overline{C}Y_n\}$, respectively, where

$$\underline{C}Y_i = \bigcup \{x \in U | [x]_C \subseteq Y_i \in F\}, \quad 1 \leqslant i \leqslant n$$

and

$$\overline{C}Y_i = \bigcup \{x \in U | [x]_C \cap Y_i \neq \emptyset, Y_i \in F\}, \quad 1 \leqslant i \leqslant n.$$

The approximation accuracy of F by C is defined as

$$\alpha_C(F) = \frac{\sum_{Y_i \in U/D} |\underline{C}Y_i|}{\sum_{Y_i \in U/D} |\overline{C}Y_i|}. \tag{5}$$

The approximation accuracy provides the percentage of possible correct decisions when classifying objects by employing the attribute set C. Although, in some situations, $\alpha_C(F)$ can be used to evaluate the approximation accuracy of a rough classification in an information system, its limitations are revealed by the following example.

Example 2. In Table 1, $U/D = \{\{u_1, u_3, u_4, u_5\}, \{u_2\}\}$ gives a classification of the universe U.
For $C_1 = \{c_3\}$ and $C_2 = \{c_2, c_3\}$, one has that

$$\alpha_{C_1}(U/D) = \alpha_{C_2}(U/D) = 0.4.$$

In Example 2, although there is a partial relation between the two knowledge representation systems U/C_1 and U/C_2, they achieve the same approximation accuracy for the rough classification.

5. Uncertainty measures based on a knowledge granulation

In this section, based on a knowledge granulation in an information system, the three existing measures of accuracy, roughness and approximation accuracy are modified. Furthermore, theoretical studies and experimental results are provided to show that the modified measures are effective and suitable for evaluating the roughness and accuracy of a set in an information system and the approximation accuracy of a rough classification in a decision table, respectively.

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

$$Roughness_R(X) = \rho_R(X)GK(R), \tag{6}$$

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

Clearly, the granularity of the partition induced by the equivalence relation R is taken into account in the new definition. In the following, we show that the modified roughness measure has some significant properties and is valuable in evaluating uncertainty of a set.

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

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

Property 3 (Minimum). *Let $S = (U, A)$ be an information system, $R \subseteq A$ and $\emptyset \subset X \subseteq U$. The minimum roughness of X with respect to R is zero. This value is achieved only if $U/R = \omega$. Obviously, when $S = (U, A)$ is an information system, one has that $0 \leqslant Roughness_R(X) \leqslant 1$ for any subset R of A.*

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

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

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

$$Roughness_P(X) = \rho_P(X)GK(P) \leqslant \rho_Q(X)GK(Q) = Roughness_Q(X).$$

This completes the proof. □

Proposition 3 states that the roughness of X with respect to R decreases with R becoming finer.

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

$$Accuracy_R(X) = 1 - \rho_R(X)GK(R), \tag{7}$$

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

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

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

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

Obviously, when $S = (U, A)$ is an information system, one has that $0 \leqslant Accuracy_R(X) \leqslant 1$ for any subset R of A.

Proposition 4. *Let $S = (U, A)$ be an information system, $P, Q \subseteq A$ and $\emptyset \subset X \subseteq U$. If $P \preceq Q$, then*

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

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

Thus

$$Accuracy_P(X) = 1 - \rho_P(X)GK(P) = 1 - Roughness_P(X) \geqslant 1 - Roughness_Q(X) = Accuracy_Q(X).$$

This completes the proof. □

Proposition 4 states that the accuracy of X with respect to R increases with R becoming finer.

As follows, we analyze the complexity for computing uncertainty of a set. Let $S = (U, A)$ be an information system, where $U = \{u_1, u_2, \ldots, u_{|U|}\}$ is a universe and A is a set of attributes, and let $R \subseteq A$ and X a non-empty subset of $U : \emptyset \subset X \subseteq U$. By performing the following steps, one can compute the roughness of X with respect to R.

(1) Compute $U/R = \{X_1, X_2, \ldots, X_n\}$ and $GK(R) = \frac{1}{|U|^2} \sum_{i=1}^{n} |X_i|^2$;
(2) compute $\underline{R}X = \bigcup\{X_i \in U/R | X_i \subseteq X\}$, $\overline{R}X = \bigcup\{X_i \in U/R | X_i \bigcap X \neq \emptyset\}$, $\alpha_R(X) = \frac{|\underline{R}X|}{|\overline{R}X|}$ and $\rho_R(X) = 1 - \alpha_R(X)$; and
(3) finally, compute $Roughness_R(X) = \rho_R(X)GK(R)$.

In Step (1), to compute all $X_i(i = 1, 2, \ldots, n)$, it needs to decide whether $(u_i, u_j) \in IND(R)(i, j = 1, 2, \ldots, |U|)$ or not. Thus, its time complexity is $O(|A| |U|^2)$. The time complexity of (2) is $O(|U|)$ and the time complexity of the step (3) is a constant. Therefore, the overall time complexity is $O(|A| |U|^2)$.

Example 3 (*Continued from Example 1*). By computing, we have that

$$GK(R_1) = \frac{1}{25}\left(3^2 + 2^2\right) = 0.52,$$

$$GK(R_2) = \frac{1}{25}\left(1 + 2^2 + 2^2\right) = 0.36$$

and

$$GK(R_3) = \frac{1}{25}\left(1 + 1 + 2^2 + 1\right) = 0.28.$$

Hence, the roughnesses of X with respect to R_1, R_2 and R_3 are given by

$$Roughness_{R_1}(X) = 0.4 \times 0.52 = 0.208,$$
$$Roughness_{R_2}(X) = 0.4 \times 0.36 = 0.144$$

and

$$Roughness_{R_3}(X) = 0.4 \times 0.28 = 0.112,$$

respectively. Obviously, $Roughness_{R_1}(X) > Roughness_{R_2}(X) > Roughness_{R_3}(X)$. It is clear that the roughness of X with respect to R decreases with R becoming finer.

The accuracies of X with respect to R_1, R_2 and R_3 are given by

$$Accuracy_{R_1}(X) = 0.792,$$
$$Accuracy_{R_2}(X) = 0.856$$

and

$$Accuracy_{R_3}(X) = 0.888,$$

respectively. Thus, $Accuracy_{R_1}(X) < Accuracy_{R_2}(X) < Accuracy_{R_3}(X)$. Therefore, the accuracy of X with respect to R increases as R becoming finer.

In [24], the accuracy of X with respect to R is defined as

$$D_R(X) = 1 - \rho_R(X) \times rough(R),\tag{8}$$

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

For Example 3, one can also compute the accuracies of X with respect to R_1, R_2 and R_3 with the formula (8), which are as follows:

$$D_{R_1}(X) = 1 - \rho_R(X) \times rough(R) = 1 - 0.4 \times \left[1 - \frac{3 \times 2 \times \log_2 3 + 2 + 2 \times 3 \times \log_2 3}{5 \times 4 \times \log_2 5} \right] = 0.781,$$

$$D_{R_2}(X) = 1 - \rho_R(X) \times rough(R) = 1 - 0.4 \times \left[1 - \frac{2 \times (2 + 2 \times 3 \times \log_2 3) - \log_2 \frac{1}{5^5}}{5 \times 4 \times \log_2 5} \right] = 0.898$$

and

$$D_{R_3}(X) = 1 - \rho_R(X) \times rough(R) = 1 - 0.4 \times \left[1 - \frac{2 + 2 \times 3 \times \log_2 3 - 3 \times \log_2 \frac{3}{5^5}}{5 \times 4 \times \log_2 5} \right] = 0.958.$$

Clearly, $D_{R_1}(X) < D_{R_2}(X) < D_{R_3}(X)$. Thus, the results are consistent with those in Example 3.

The time complexity of the method in [24] is equivalent to that of computing an equivalence relation, i.e. $O(|A||U|^2)$, which is the same as the method proposed in this paper.

Remark. One can see that the measure in the literature [24] and the measure in this paper improve the Pawlak's accuracy measure. It has been shown that $Rough(R)$ in [24] is a kind of knowledge granulation. So, the measures in the literature [24] and this paper both come from the knowledge granulation defined in this paper. But the definition of knowledge granulation given in this paper has a much simpler form. Thus, the measure proposed in this paper is much simpler and more comprehensive than that in [24].

In the following, the measure proposed in this paper is exemplified through an information system about planning tennis ball (see Table 2 in [22]), where $U = \{u_1, u_2, \ldots, u_{24}\}$ 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 $R_1 = \{a_1, a_2\}$ and $R_2 = \{a_1, a_2, a_3\}$. By computing, we have $GK(R_1) = 0.132$ and $GK(R_2) = 0.090$. Clearly, it is different between the knowledge granulation of R_1 and that of R_2. In fact, $U/R_1 \subset U/R_2$, i.e., R_2 is finer than R_1.

Let

$$X_1 = \{u_4, u_5, u_6, u_{17}, u_{23}\},$$
$$X_2 = \{u_1, u_2, u_3, u_8, u_{11}\},$$
$$X_3 = \{u_8, u_{11}, u_{12}, u_{15}\},$$
$$X_4 = \{u_1, u_4, u_5, u_9, u_{14}, u_{19}, u_{21}, u_{22}, u_{23}\}$$

and

$X_5 = \{u_6, u_7, u_{11}, u_{12}, u_{15}, u_{17}, u_{18}, u_{21}, u_{24}\}$. By testing on these five sets, X_1, X_2, X_3, X_4 and X_5, one can obtain the same lower approximation and upper approximation for these sets according to R_1 and R_2 (see Table 3), which are the same as those obtained by using Pawlak's roughness measure. However, these sets have different $Roughness(X)$ (see Table 4). This is caused

Table 2
An information system about planning tennis ball.

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

Table 3
The lower approximations and upper approximations of sets.

X	$\underline{R_1}X = \underline{R_2}X$	$\overline{R_1}X = \overline{R_2}X$
X_1	$\{u_4, u_5, u_{23}\}$	$\{u_4, u_5, u_6, u_7, u_{17}, u_{18}, u_{23}, u_{24}\}$
X_2	$\{u_1, u_2, u_3\}$	$\{u_1, u_2, u_3, u_8, u_{10}, u_{11}, u_{12}\}$
X_3	$\{u_{11}, u_{12}\}$	$\{u_8, u_{10}, u_{11}, u_{12}, u_{15}, u_{16}\}$
X_4	$\{u_4, u_5, u_9, u_{21}, u_{22}, u_{23}\}$	$\{u_1, u_2, u_3, u_4, u_5, u_9, u_{13}, u_{14}, u_{19}, ..., u_{23}\}$
X_5	$\{u_6, u_7, u_{11}, u_{12}, u_{17}, u_{18}, u_{24}\}$	$\{u_6, u_7, u_{11}, u_{12}, u_{15}, u_{16}, u_{17}, u_{18}, u_{21}, u_{24}\}$

Table 4
Comparison of classical roughness measure and the measure in this paper.

X	$\rho_{R_1}(X) = \rho_{R_2}(X)$	$Roughness_{R_1}(X)$	$Roughness_{R_2}(X)$
X_1	0.625	0.082	0.056
X_2	0.571	0.075	0.052
X_3	0.667	0.088	0.060
X_4	0.534	0.071	0.049
X_5	0.364	0.048	0.033

by that R_1 and R_2 have different knowledge granulations. Thus, under the measure proposed in this paper, the uncertainties of X with respect to different equivalence relations are well characterized.

In order to demonstrate the advantage of the measure proposed in this paper, we have downloaded three public data sets from UCI Repository of Machine Learning Databases [30], which are described in Table 5.

In Table 5, the data set Servo is from a simulation of a servo system involving a servo amplifier, a motor, a lead screw/nut, and a sliding carriage of some sort; the data set Tic-tac-toe is the encoding of the complete set of possible board configurations at the end of tie-tac-toe games, which is used to obtain possible ways to create a "three-in-a-row"; and the Nursery data set is derived from a hierarchical decision model originally developed to rank applications for nursery schools.

Table 5
Data sets description.

Data sets	Samples	Number of attributes
Servo	167	4
Tic-tac-toe	958	9
Nursery	12960	8

Table 6
Pawlak's accuracy measure α and the accuracy measure *Accuracy* proposed in this paper with different numbers of attributes for X_1 in the data set Servo.

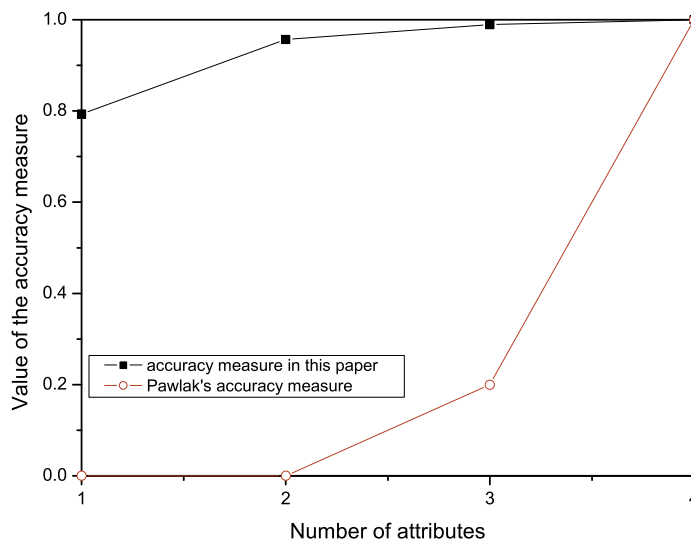| Measure	Attributes			
	1	2	3	4
α	0.000000	0.000000	0.200000	1.000000
Accuracy	0.793288	0.957008	0.989587	1.000000

Table 7
Pawlak's accuracy measure α and the accuracy measure *Accuracy* proposed in this paper with different numbers of attributes for X_2 in the data set Tie-tac-toe.

Measure	Attributes								
	1	2	3	4	5	6	7	8	9
α	0.000000	0.000000	0.100257	0.111406	0.311080	0.653716	0.897533	1.000000	1.000000
Accuracy	0.641548	0.876951	0.958570	0.984367	0.994799	0.998832	0.999828	1.000000	1.000000

Table 8
Pawlak's accuracy measure α and the accuracy measure *Accuracy* proposed in this paper with different numbers of attributes for X_3 in the data set Nursery.

Measure	Attributes							
	1	2	3	4	5	6	7	8
α	0.000000	0.500000	0.800000	0.947368	0.982143	0.991071	0.997006	1.000000
Accuracy	0.666667	0.966667	0.996667	0.999781	0.999975	0.999994	0.999999	1.000000

Here, we compare the accuracy measure proposed in this paper with Pawlak's accuracy measure for a set on these three practical data sets. Let X_1 be the set of first 10 objects in the data set Servo, and X_2, X_3 be the set of first 500 and 1000 objects in the data set Tie-tac-toe and Nursery, respectively. The comparisons of values of two measures for the sets X_1, X_2, and X_3 with the numbers of attributes in these three data sets are shown in Tables 6–8 and Figs. 1–3.

It can be seen from Tables 6–8 that the values of the accuracy measure in this paper and Pawlak's accuracy measure increase with the number of selected attributes becoming bigger. However, from Fig. 1, it is easy to see that the values of α is equal to zero when the number of features equals 1 or 2. In this situation, the lower approximation of the object set equals an empty set in the data set. Hence, Pawlak's accuracy measure cannot be used to effectively characterize the accuracy of the object set. But, for the same situation as that the number of attributes equals 1 and 2, the values of the accuracy measure in this paper equal 0.793288 and 0.957008, respectively. It shows that unlike Pawlak's accuracy measure, the accuracy measure in this paper induced by two attributes is higher than that induced by only one attribute. Thus, the measure in this paper may be much better than Pawlak's accuracy measure for this situation. One can draw the same conclusion from Figs. 2

Fig. 1. Variation of Pawlak's accuracy measure α and the accuracy measure *Accuracy* in this paper with the number of attributes for X_1 (data set Servo).
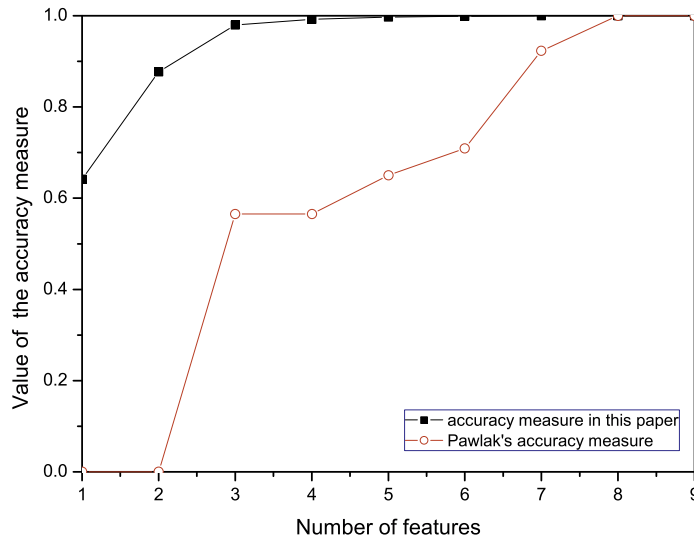
Fig. 2. Variation of Pawlak's accuracy measure α and the accuracy measure *Accuracy* in this paper with the number of attributes for X_2 (data set Tie-tac-toe).
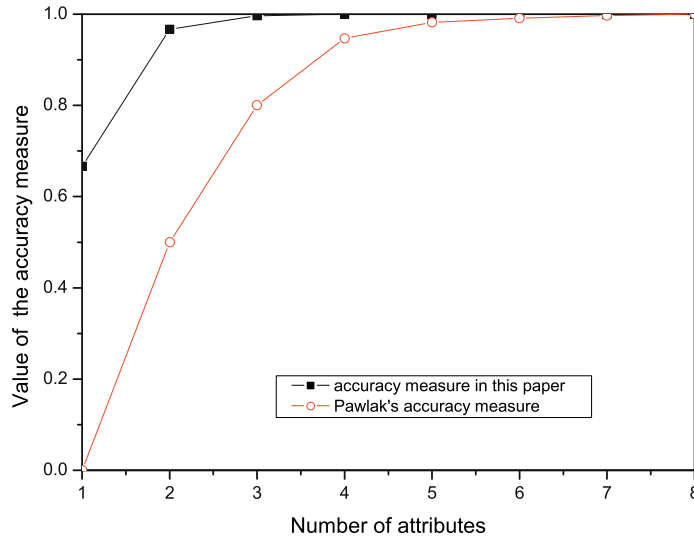

Fig. 3. Variation of Pawlak's accuracy measure α and the accuracy measure *Accuracy* in this paper with the number of attributes for X_3 (data set Nursery).

and 3. In other words, when α equals zero or has the same value except one for different numbers of attributes in Figs. 1–3, the measure in this paper is still valid for evaluating the accuracy of the set by using the selected attributes. Therefore, the measure proposed in this paper may be better than Pawlak's measure for evaluating uncertainty of a set.

The idea in the measure definition of this paper can be generalized to defining approximation accuracy of a rough classification. Based on the knowledge granulation, a new approximation accuracy of a rough classification in a decision table can be proposed, which is as follows.

Definition 5. Let $S = (U, C \cup D)$ be a decision table, $U/D = \{Y_1, Y_2, \ldots, Y_n\}$ and $R \subseteq C$. An approximation accuracy of U/D with respect to R is defined as:

$$\gamma_R(U/D) = 1 - (1 - \alpha_R(U/D))GK(R). \tag{9}$$

Property 7 (Equivalence). *Let $S = (U, C \cup D)$ be a decision table, $U/D = \{Y_1, Y_2, \ldots, Y_n\}$ and $P, Q \subseteq C$. If $U/P = U/Q$, then $\gamma_P(U/D) = \gamma_Q(U/D)$.*

Property 8 (Maximum). *Let $S = (U, C \cup D)$ be a decision table, $U/D = \{Y_1, Y_2, \ldots, Y_n\}$ and $R \subseteq C$. The maximum approximation accuracy of U/D with respect to R is one only if $\sum_{Y_i \in U/D} |\underline{C}Y_i| = \sum_{Y_i \in U/D} |\overline{C}Y_i|$.*

Property 9 (Minimum). *Let $S = (U, C \cup D)$ be a decision table, $U/D = \{Y_1, Y_2, \ldots, Y_n\}$ and $R \subseteq C$. The minimum approximation accuracy of U/D with respect to R is zero only if $U/R = \delta$.*

Obviously, when $S = (U, C \cup D)$ is a decision table, one has that $0 \leqslant \gamma_R(U/D) \leqslant 1$ for any subset R of C.

Proposition 5. *Let $S = (U, C \cup D)$ be a decision table, $U/D = \{Y_1, Y_2, \ldots, Y_n\}$ and $P, Q \subseteq C$. If $P \preceq Q$, then*

$$\gamma_P(U/D) \geqslant \gamma_Q(U/D).$$

Proof. Let $P \preceq Q$. It is easy to obtain that $a_Q(U/D) \leqslant a_P(U/D)$ and $GK(Q) \geqslant GK(P)$.

Thus

$$\gamma_P(U/D) = 1 - (1 - a_P(U/D))GK(P) \geqslant \gamma_Q(U/D).$$

This completes the proof. □

Proposition 5 states that the approximation accuracy of U/D with respect to R increases with R becoming finer. The time complexity of computing $\gamma_C(U/D)$ is $O(|C||U|^2)$, which is the same as $\alpha_C(U/D)$.

Example 4 (*Continued from Example 2*). By computing, one has that $GK(C_1) = 0.52$ and $GK(C_2) = 0.44$. Thus,

$$\gamma_{C_1}(U/D) = 0.688$$

and

$$\gamma_{C_2}(U/D) = 0.736.$$

Therefore, when $C_2 \preceq C_1$, $\gamma_{C_2}(U/D) > \gamma_{C_1}(U/D)$.

For a general decision table, to illustrate the difference between $\alpha_C(U/D)$ and $\gamma_C(U/D)$, the practical data set Tic-tac-toe in Table 5 is used again. The data set has two decision classes. The comparisons of values of the two measures with the numbers of attributes in the data set is shown in Table 9 and Fig. 4.

It can be seen from Table 9 that the values of the measure γ and the approximation accuracy $\alpha_C(U/D)$ increase as the number of selected attributes becoming bigger in the data set. The measure γ and the approximation accuracy will achieve the

Table 9
$a_C(U/D)$ and $\gamma_C(U/D)$ with different numbers of attributes in the data set Tie-tac-toe.

Measure	Features								
	1	2	3	4	5	6	7	8	9
$a_C(U/D)$	0.0000	0.0000	0.0640	0.0988	0.2788	0.6539	0.8943	1.0000	1.0000
$\gamma_C(U/D)$	0.6415	0.8769	0.9569	0.9841	0.9945	0.9988	0.9998	1.0000	1.0000
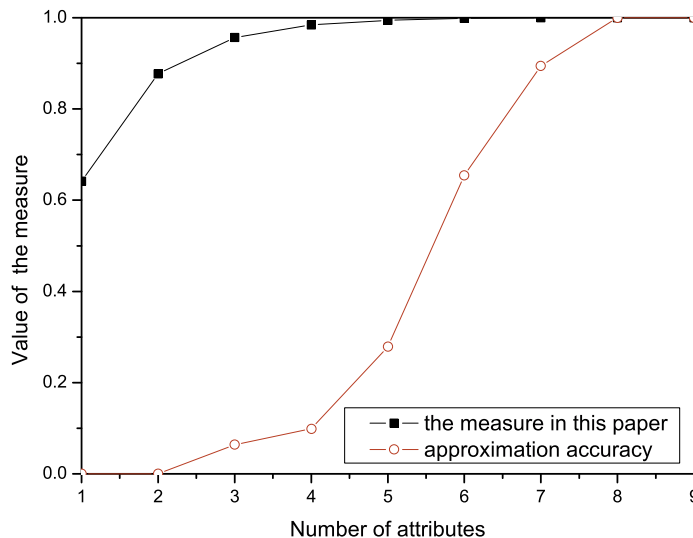
Fig. 4. Variation of the measure γ in this paper and the approximation accuracy with the number of attributes (data set Tic-tac-toe).

same value of one if the decision table becomes consistent by adding selected attributes. However, from Fig. 4, it is easy to see that the values of approximation accuracy is equal to zero when the number of attributes equals 1 or 2. In this situation, the lower approximation of the rough classification equals an empty set in the decision table. But, for the same situation as that the number of attributes equals 1 and 2, the values of the measure γ equal 0.6415 and 0.8769, respectively. It shows that unlike the approximation accuracy, the measure in this paper with two attributes is higher than that with only one attribute. In other words, when $\alpha_C(U/D) = 0$ in Fig. 4, the measure γ is still valid for evaluating the approximation accuracy of a classification by using the selected attributes. Hence, the measure γ may be better than the approximation accuracy for evaluating the approximation accuracy of a rough classification.

6. Conclusions and discussion

In this paper, we have addressed the issues of uncertainty of a set in an information system and approximation accuracy of a rough classification in a decision table. Two examples have been employed to reveal the limitations of three existing classical measures for evaluating uncertainty of a set and approximation accuracy of a rough classification. An axiomatic definition of knowledge granulation for an information system have been given. Under this definition, it has been proved that the knowledge granulations given in [7,24] are two special forms of the definition. Based on the knowledge granulation defined in this paper, the three existing measures have been modified. Theoretical studies and numerical experiments have been carried out to show that the modified measures are effective and suitable for evaluating the roughness and accuracy of a set in an information system and the approximation accuracy of a rough classification in a decision table, respectively, and have a much simpler and more comprehensive form than the existing ones.These new measures may be helpful for rule evaluation and knowledge discovery.

Acknowledgements

References

[1] J. Bazan, J.F. Peters, A. Skowron, H.S. Nguyen, M. Szczuka, Rough set approach to pattern extraction from classifiers, Electronic Notes in Theoretical Computer Science 82 (4) (2003) 1–10.
[2] T. Beaubouef, F.E. Petry, G. Arora, Information-theoretic measures of uncertainty for rough sets and rough relational databases, Information Sciences 109 (1998) 185–195.
[3] G.J. Klir, Basic issues of computing with granular computing, in: Proceedings of IEEE International Conference on Fuzzy Systems, 1998, pp. 101–105.
[4] J.Y. Liang, K.S. Qu, Information measures of roughness of knowledge and rough sets for information systems, Journal of Systems Science and Systems Engineering 10 (1) (2002) 95–103.
[5] J.Y. Liang, D.Y. Li, Uncertainty and Knowledge Acquisition in Information Systems, Science Press, Beijing, 2005.
[6] J.Y. Liang, K.S. Chin, C.Y. Dang, C.M. YAM, A new method for measuring uncertainty and fuzziness in rough set theory, International Journal of General Systems 31 (4) (2002) 331–342.
[7] J.Y. Liang, Z.Z. Shi, The information entropy rough entropy knowledge granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (1) (2004) 37–46.
[8] J.Y. Liang, Z.Z. Shi, D.Y. Li, Information entropy, rough entropy and knowledge granulation in incomplete information systems, International Journal of General Systems 35 (6) (2006) 641–654.
[9] T.Y. Lin, Granular computing on binary relations I: Data mining and neighborhood systems, II: Rough sets representations and belief functions, in: L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery 1, Physica-Verlag, Heidelberg, 1998, pp. 107–140.
[10] S.K. Pal, W. Pedrycz, A. Skowron, R. Swiniarski, Presenting the special issue on rough-neuro computing, Neurocomputing 36 (2001) 1–3.
[11] Z. Pawlak, Rough sets, International Journal of Computer and Information Science 11 (1982) 341–356.
[12] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.
[13] Z. Pawlak, Some remarks on conflict analysis, Europe Journal of Operational Research 166 (3) (2005) 649–654.
[14] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.
[15] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (2007) 28–40.
[16] Z. Pawlak, A. Skowron, Rough sets and boolean reasoning, Information Sciences 177 (2007) 41–73.
[17] Z. Pawlak, S.K.M. Wong, W. Ziarko, Rough sets: probabilistic versus deterministic approach, International Journal of Man–Machine Studies 29 (1988) 81–95.
[18] Z. Pawlak, Granularity of knowledge, indiscernibility and rough sets, in: Proceedings of IEEE International Conference on Computing and Information, vol. I, 1988, pp. 186–189.
[19] Y.H. Qian, J.Y. Liang, Combination entropy and combination granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based systems 16 (2) (2008) 179–193.
[20] Y.H. Qian, J.Y. Liang, D.Y. Li, H.Y. Zhang, C.Y. Dang, Measures for evaluating the decision performance of a decision table in rough set theory, Information Sciences 178 (1) (2008) 181–202.
[21] C.E. Shannon, The mathematical theory of communication, The Bell System Technical Journal 27 (3 and 4) (1948) 373–423 (see also 623–656).
[22] Z.Z. Shi, Knowledge Discovery, Tsinghua University Press, Beijing, 2002.
[23] M.J. Wierman, Measuring uncertainty in rough set theory, International Journal of General Systems 28 (1999) 283–297.
[24] B.W. Xu, Y.M. Zhou, H.M. Lu, An improved accuracy measure for rough sets, Journal of Computer and System Sciences 71 (2005) 163–173.
[25] Y.Y. Yao, Granular computing on basic issues and possible solutions, in: Proceedings of the Fifth International Conference on Computing and Information, vol. I, 2000, pp. 186–189.

[26] L.A. Zadeh, Fuzzy sets and information granularity, in: M. Gupta, R. Ragade, R. Yager (Eds.), Advances in Fuzzy Set Theory and Application, North-Holland, Amsterdam, 1979, pp. 3–18.
[27] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Transactions on Fuzzy Systems 4 (1) (1996) 103–111.
[28] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (1997) 111–127.
[29] L. Zhang, B. Zhang, Fuzzy reasoning model under quotient space structure, Information Sciences 173 (4) (2005) 353–364.
[30] The UCI machine learning repository, <http://mlearn.ics.uci.edu/MLRepository.html>.