Teng *et al. SpringerPlus (2016) 5:117*
DOI 10.1186/s40064-016-1670-2

SpringerPlus
a SpringerOpen Journal

RESEARCH Open Access

Integrated measures for rough sets based on general binary relations

Shuhua Teng[1], Fan Liao[2], Mi He[3], Min Lu[1] and Yongjian Nian[3*]

*Correspondence:
yjnian@126.com
[3] School of Biomedical Engineering, Third Military Medical University, Chongqing 400038, China
Full list of author information is available at the end of the article

Abstract

Uncertainty measures are important for knowledge discovery and data mining. Rough set theory (RST) is an important tool for measuring and processing uncertain information. Although many RST-based methods for measuring system uncertainty have been investigated, the existing measures cannot adequately characterise the imprecision of a rough set. Moreover, these methods are suitable only for complete information systems, and it is difficult to generalise methods for complete information systems to incomplete information systems. To overcome these shortcomings, we present new uncertainty measures, integrated accuracy and integrated roughness, that are based on general binary relations, and we study important properties of these measures. A theoretical analysis and examples show that the proposed integrated measures are more precise than existing uncertainty measures, they are suitable for both complete and incomplete information systems, and they are logically consistent. Therefore, integrated accuracy and integrated roughness overcome the limitations of existing measures. This research not only develops the theory of uncertainty, it also expands the application domain of uncertainty measures and provides a theoretical basis for knowledge acquisition in information systems based on general binary relations.

Keywords: Rough set, Uncertainty measure, General binary relation, Information system

Background

Uncertainty is an important topic in research on artificial intelligence (Li and Du 2005). Rough set theory (RST) is a mathematical tool for handling imprecise, incomplete and uncertain data (Pawlak 1991), and it is an effective method to deal with uncertainty problems. In classical RST, the uncertainty of rough sets depends on two factors, knowledge uncertainty (the size of information granularities) and set uncertainty (the size of the rough set boundary) (Pawlak 1991). Set uncertainty in RST is measured with two quantities, accuracy and roughness, but they do not adequately reflect the uncertainty of a rough set. In some cases, the accuracy measure reflects only the size of the boundary region but not the size of the information granularities formed by the attribute sets, which limits the applicability of classical rough sets (Pawlak 1991). To solve this problem, researchers have proposed a number of integrated uncertainty measures based on certain binary relations (Teng et al. 2016; Wang et al. 2008a; Liang et al. 2009) that consider both the knowledge uncertainty and the set uncertainty. Although these measures are

effective, they have certain restrictions. These measures change with information granularities which are unrelated to rough set X, i.e., information granularities in the negative region of X; this is inconsistent with human cognition in uncertainty problems (Wang and Zhang 2008). Intuitively, a rough measure that reflects two types of uncertainty should have a higher value than that of a measure which reflects only one type of uncertainty, but this property is not satisfied by the existing integrated uncertainty measures. In addition, the existing integrated uncertainty measures do not sufficiently characterise the uncertainty in certain cases. Wang and Zhang (2008) proposed a fuzziness measure for rough sets based on information entropy, which overcomes the problem of existing uncertainty measures for rough sets. However, a fuzziness measure based on the equivalence relation is not suitable for the incomplete information system and ordered information system. In practice, knowledge acquisition usually involves information that is incomplete for various reasons such as data measurement errors, a limited understanding and the conditions under which the data were acquired (Kryszkiewicz et al. 1998). Incompleteness in an information system is one of the main causes of uncertainty. RST, which is based on the traditional equivalence relation (i.e., reflexivity, symmetry, and transitivity) cannot directly deal with incomplete information systems, which greatly constrains the use of RST in practical applications (Gantayat et al. 2014; Sun et al. 2014). Hence, several extended models and methods for RST such as the tolerance relation (i.e., reflexivity, symmetry) (Wang and Zhang 2008), the asymmetric similarity relation (i.e., reflexivity, transitivity) (Stefanowski and Tsoukias 1999), the limited tolerance relation (i.e., reflexivity, symmetry) (Wang 2002), the dominance relation (reflexivity, transitivity) (Greco et al. 2002; Hu et al. 2012), and the general binary relation (i.e., reflexivity) (Yao 1998; Teng et al. 2009; Zhu 2007) which can directly process an incomplete information system, have been proposed. Based on these relations, directly measuring the uncertainty of incomplete data has caused considerable concern (Huang et al. 2004; Qian et al. 2009; Xu and Li 2011; Dai and Xu 2012; Sun et al. 2012; Dai et al. 2014; Chen et al. 2014; Dai et al. 2013).

The various uncertainty measures mentioned above are mostly aimed at one special binary relation without universality, and do not adequately reflect the uncertainty of rough sets in certain cases. Little attention has been paid to uncertainty measures based on general binary relations (Huang et al. 2004; Wang et al. 2008b). To overcome the limitations of the existing uncertainty measures and to analyse data more efficiently, it is necessary to find an uncertainty measure that is universal and more accurate.

This paper begins with an analysis of the limitations of the existing uncertainty measures for rough sets. Next, a knowledge uncertainty measure based on general binary relations is presented, which is applicable in classical systems as well, i.e., it is an effective technique to deal with complex data sets. Novel integrated measures based on general binary relations are proposed, and the properties of these integrated measures are analysed. At last, Examples are used to verify the validity of the proposed uncertainty measures.

Preliminary concepts of RST

Information system is a pair $S = (U, A)$, where $U = \{u_1, u_2, \ldots, u_{|U|}\}$ is a non-empty finite set of objects ($|\cdot|$ denotes the cardinality of the set), $A = \{a_1, a_2, \ldots, a_{|A|}\}$ is a non-empty finite set of attributes such that a_j: $a_j \to V_{a_j}$ for every $a_j \in A$. The set V_{a_j} is called the value set of a_j.

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

$$\mathrm{IND}(P) = \left\{ \left(u_i, u_j\right) \in U \times U \middle| \forall a \in P, f(u_i, a) = f\left(u_j, a\right) \right\} \tag{1}$$

Obviously, IND(P) is an equivalence relation. If $\left(u_i, u_j\right) \in \mathrm{IND}(P)$, then u_i and u_j are indiscernible with respect to attribute set P. The partition generated by IND(P) is denoted by $U/\mathrm{IND}(P)$, which can be abbreviated as U/P. The partition $U/P = \{P_1, P_2, \ldots, P_m\}$ denotes knowledge associated with the equivalence relation IND(P), where P_i is an equivalence class, $1 \le i \le m$, and $1 \le m \le |U|$. Each equivalence class is an information granularity. Thus, the attribute set P will also be called the knowledge. The equivalence class determined by u_i with respect to the attribute set P is denoted by $[u_i]_P = \left\{u_j \in U | (u_i, u_j) \in \mathrm{IND}(P)\right\}$. Obviously, if $u_i \in P_k$, then $[u_i]_P = P_k$. For any set $X \subseteq U$, the P-lower and P-upper approximations of X are $\underline{P}X = \{u_i \in U | [u_i]_P \subseteq X\}$ and $\overline{P}X = \{u_i \in U | [u_i]_P \cap X \neq \emptyset\}$, respectively. The boundary region of X is represented by $BN_P(X) = \overline{P}X - \underline{P}X$.

An information system S (= (U, A)) is an incomplete information system if the attribute values include an empty value "*"; otherwise, S is a complete information system.

In an information system, a relation derived from the attribute sets is generally not an equivalence relation but a general binary relation. In this paper, we use R^P to represent a general binary relation derived from the knowledge P. In an information system S, $P \subseteq A$. We define the function R_S^P as follows:

1. The set-valued function $R_S^P : U \to P(U)$ is defined as $R_S^P(u_i) = \{u_j \in U | (u_i, u_j) \in R^P\}$, where $R_S^P(u_i)$ is the subsequent neighbour of u_i under the binary relation R^P. The relation R^P and the corresponding subsequent neighbour $R_S^P(u_i)$ can be uniquely determined from each other, i.e., $u_i R^P u_j \Leftrightarrow u_j \in R_S^P(u_i)$. Let $U / R^P = \{R_S^P(u_i) | u_i \in U\}$ represent the classification of U divided by the knowledge P, where $R_S^P(u_i)$ is called a classification granularity under the general binary relation. The classification granularity $R_S^P(u_i)$ can be understood as the largest set of objects that cannot be distinguished from object u_i given the knowledge P; i.e., objects in $R_S^P(u_i)$ should belong to the same class as u_i given the knowledge P. Obviously, $R_S^P(u_i)$ will be an equivalence class, a dominance class, a tolerance class, a limited tolerance class, or an asymmetric similarity class of an object u_i if R^P is an equivalence relation, a dominance relation, a tolerance relation, a limited tolerance relation or an asymmetric similarity relation, respectively. Note that classification granularities in U / R^P do not always constitute partitions or covers of U (Wang et al. 2008b). The lower and upper approximation sets of $X \subseteq U$ with respect to a general binary relation R^P are defined as $\underline{R^P}(X) = \{u_i \in U | R_S^P(u_i) \subseteq X\}$ and $\overline{R^P}(X) = \{u_i \in U | R_S^P(u_i) \cap X \neq \emptyset\}$, respectively.
2. If Q and $P \subseteq A$, we define a partial relation $\preceq$ as follows: $P \preceq Q \Leftrightarrow R_S^P(u_i) \subseteq R_S^Q(u_i)$ for $\forall u_i \in U$, which means that the knowledge P is finer (i.e., has finer classification

granularities) than the knowledge Q. If $R_S^P(u_i) \subseteq R_S^Q(u_i)$ for $\forall u_i \in U$ and $\exists u_j \in U$ satisfies $R_S^P(u_j) \subset R_S^Q(u_j)$, then we say that the knowledge P is strictly finer than the knowledge Q, or the knowledge Q entirely depends on the knowledge P, which is denoted by $P \prec Q$. The notation $P \approx Q$ represents $R_S^P(u_i) = R_S^Q(u_i)$ for $\forall u_i \in U$.

Limitations of existing uncertainty measures

In classical RST, there are two main causes of uncertainty: the information granularity derived from the binary relation in the universe, which is knowledge uncertainty, and the boundary of the rough set in the given approximation space, which is set uncertainty (Pawlak 1991). Beaubouef et al. (1998) proposed a new integrated uncertainty measure for complete information systems, which they called rough entropy.

Given an information system $S = (U, A)$, $P, Q \subseteq A$, and $U/P = \{P_1, P_1, \ldots, P_m\}$. The rough entropy of $X \subseteq U$ with respect to P is defined as (Beaubouef et al. 1998)

$$H(X, P) = \rho_P(X) H^G(P) \tag{2}$$

where $H^G(P) = -\left|\sum_{i=1}^{m} \frac{|P_i|}{|U|} \log_2 \frac{1}{|P_i|}\right|$ is called the granularity measure of the knowledge P. In Eq. (2), $H^G(P)$ measures knowledge uncertainty, and the roughness $\rho_P(X) = 1 - \frac{|\underline{P}X|}{|\overline{P}X|}$ measures set uncertainty. Rough entropy considers two types of uncertainty and is therefore an integrated uncertainty measure.

Yang and John (2008) noted that existing uncertainty measures cannot correctly measure the uncertainty of boundary rough sets, whose lower approximation is an empty set. Thus, Yang and John (2008) defined the measures global accuracy $\sigma_P(X)$ and global roughness $G_P(X)$ under the equivalence relation to measure the uncertainty of rough sets:

$$\sigma_p(X) = \frac{|U - BN_P(X)|}{|U|} \tag{3}$$

$$G_P(X) = 1 - \sigma_p(X) \tag{4}$$

where $BN_P(X) = \left|\overline{P}X\right| - |\underline{P}X|$. The global accuracy and the global roughness reveal the global uncertainty with respect to the universe of discourse, which addresses the shortcomings of classical measures for boundary rough sets. However, similar to classical measures, global accuracy and global roughness cannot measure the knowledge uncertainty.

If the boundary region of $X \subseteq U$ with respect to the knowledge A is an empty set, the rough set X can be precisely described by the knowledge A. In this case, the rough set X becomes a precise set; i.e., the uncertainty of X is 0. Thus, the uncertainty of a rough set X is related only to the size of the boundary region and the information granularity of the boundary region and not to the information granularity in the positive and negative regions (Wang and Zhang 2008). Although the rough entropy in Eq. (2) can measure two types of uncertainty, it is not always effective in certain cases. In the following, two examples reveal the limitations of the existing uncertainty measures for both complete and incomplete information systems.

Example 1 In a complete information system $S = (U, A)$, $U = \{u_1, u_2, \ldots, u_{3600}\}$, $X \subseteq U$ and $P \subseteq A$. Figure 1 presents the lower and upper approximations and the boundary region of X as the information granularity induced by the knowledge P changes, where in subfigures (1)–(7) the information granularity is progressively finer. In subfigure (1), the lower approximation set is an empty set and the boundary region is the entire universe. Parts of the universe in Fig. 1 (2) are finer than those in Fig. 1 (1), i.e., 6 units in Fig. 1 (1) are equally divided into 24 smaller units. The lower approximation set remains empty, and the boundary region comprises 22 smaller units. Similarly, Fig. 1 (3) shows the results as parts of the universe [i.e., two of the large units in Fig. 1 (2)] are further divided evenly. Figure 1 (4) presents the results when the largest unit in Fig. 1 (3) is further divided evenly. Figure 1 (5) shows the results when all of the smaller units in Fig. 1 (4) are further divided evenly, and Fig. 1 (6) presents the results when the negative region of Fig. 1 (5) is divided evenly. Figure 1 (7) presents the results when the positive domain of Fig. 1 (6) is further divided evenly.

The values of various uncertainty measures for the rough set X in each subfigure of Fig. 1 are shown in Table 1, where *Num_L*, *Num_U* and *Num_B* represent the number of objects in the lower approximation, the upper approximation, and the boundary region,

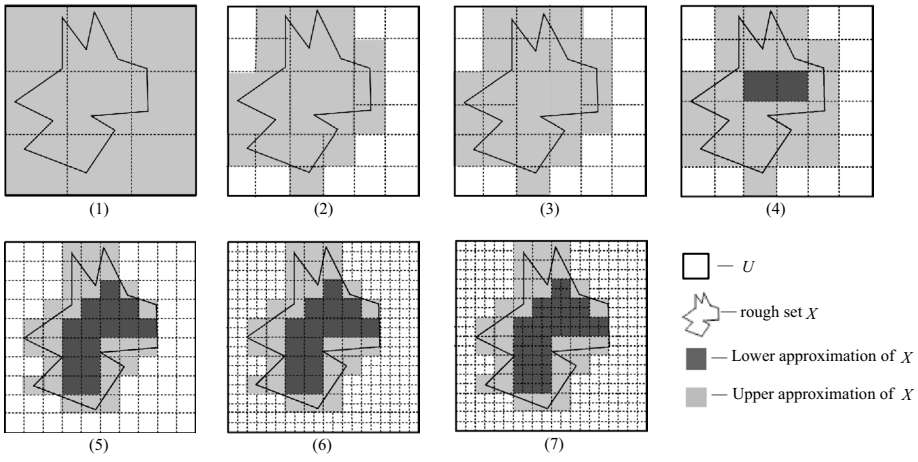

Fig. 1 Lower and upper approximations of a rough set for various levels of information granularities

Table 1 Uncertainty measures of the rough set *X* with various information granularities

Subfig. no.	Uncertainty measures						
	Num_L	*Num_U*	*Num_B*	$\rho_P(X)$	$G_P(X)$	$H^G(P)$	$H(X,P)$
(1)	0	3600	3600	1	1	8.64	8.64
(2)	0	2200	2200	1	0.61	7.31	7.31
(3)	0	2200	2200	1	0.61	6.87	6.87
(4)	200	2200	2000	0.91	0.56	6.64	6.04
(5)	5400	1548	1008	0.65	0.28	5.17	3.36
(6)	5400	1548	1008	0.65	0.28	4.03	2.62
(7)	5400	1548	1008	0.65	0.28	3.25	2.11

respectively. From Table 1, we can observe that the number of objects in the boundary of X decreases as the information granularity becomes finer, i.e., the number of objects surely belonging or not belonging to X increases. The uncertainty measures decrease monotonically as the information granularities become smaller through finer classification. However, the existing uncertainty measures are not always effective in certain cases; their limitations are revealed by the following five observations:

1. Rough set X is a boundary rough set (i.e., the lower approximation of X is an empty set) in Figs. 1 and 2. From the differences between partitions (1) and (2), we can observe that the boundary region becomes smaller and the information granularities in the boundary region become finer. Obviously, the uncertainty of the rough set X should become smaller. However, $\rho_P(X)$ in Table 1 does not change; although $H^G(P)$ decreases, it reflects only the variation in the information granularity and not the uncertainty of the set. Thus, $\rho_P(X)$ and $H^G(P)$ cannot adequately describe the uncertainty of a boundary rough set. The measure $H(X,P)$ reflects only the set uncertainty of the boundary rough set and not the knowledge uncertainty.
2. It can be observed that from partitions (2) and (3) that the boundary region does not change, but the information granularity in the boundary region becomes finer, which shows that the set uncertainty remains the same while the knowledge uncertainty decreases. In Table 1, $\rho_P(X)$ and $G_P(X)$ do not change whereas $H(X,P)$ decreases, which illustrates that $\rho_P(X)$ and $G_P(X)$ do not reflect the uncertainty of the knowledge whereas rough entropy $H(X,P)$ does.
3. Comparing partitions (3) with (4) and (4) with (5), it can be observed that the boundary region becomes smaller and the information granularity in the boundary region becomes finer. Therefore, the uncertainty of the rough set X decreases. In Table 1, $\rho_P(X)$, $G_P(X)$, $H^G(P)$ and $H(X,P)$ all decrease. However, $\rho_P(X)$ and $G_P(X)$ reflect only the set uncertainty, $H^G(P)$ reflects only the knowledge uncertainty, and $H(X,P)$ reflects both types of uncertainty.
4. Comparing partitions (5) with (6) and (6) with (7), we can observe that the boundary region and the information granularity in the boundary region remain the same. Accordingly, the uncertainty of X should not change (Wang and Zhang 2008). Although the information granularity becomes finer in the negative region from

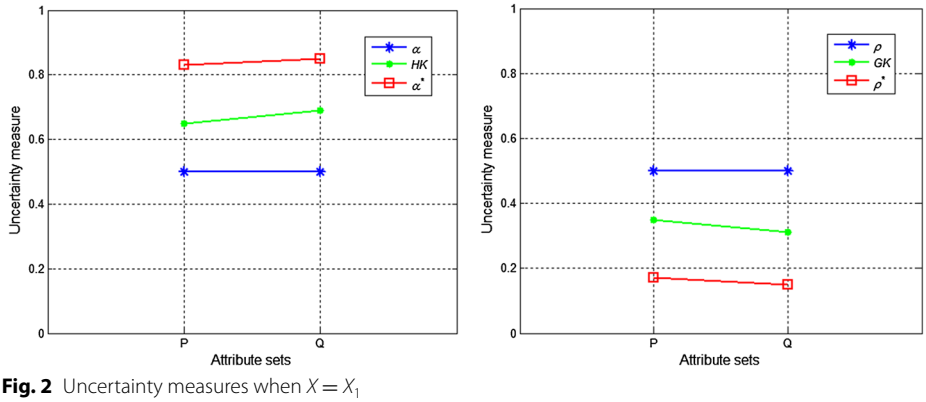

Fig. 2 Uncertainty measures when $X = X_1$

(5) to (6) and in the positive region from (6) to (7), the uncertainty of rough set X should remain unaffected (Wang and Zhang 2008). In Table 1, $\rho_P(X)$ and $G_P(X)$ are constant, which is consistent with human cognition, but $H(X,P)$ decreases, which shows that $H(X,P)$ does not accurately reflect the uncertainty of a rough set in this case.

5. An integrated measure of uncertainty in RST includes both types of uncertainty. Intuitively, the value of an integrated roughness measure that includes both types of uncertainty should be larger than that of a measure that considers only one type of uncertainty. However, rough entropy does not satisfy this requirement: although rough entropy includes both types of uncertainty, the numerical values can be smaller than those of the knowledge uncertainty measure, as shown in Table 1.

From the preceding analysis, it may be concluded that the existing uncertainty measures for a complete information system do not accurately reflect the uncertainty of rough sets. Next, the characteristics of uncertainty measures for an incomplete information system will be analysed.

In an incomplete information system, the equivalence relation of classical measures is extended to a tolerance relation R_T^P, which is expressed as:

$$\alpha_{R_T^P}(X) = \frac{\left|\underline{R_T^P}(X)\right|}{\left|\overline{R_T^P}(X)\right|} \tag{5}$$

$$\rho_{R_T^P}(X) = 1 - \alpha_{R_T^P}(X) \tag{6}$$

In Eqs. (5) and (6), $\alpha_{R_T^P}(X)$ and $\rho_{R_T^P}(X)$ are the accuracy and the roughness, respectively. Obviously, $0 \le \alpha_{R_T^P}(X), \rho_{R_T^P}(X) \le 1$. The larger the uncertainty of a rough set, the smaller $\alpha_{R_T^P}(X)$ is and the larger $\rho_{R_T^P}(X)$ is. Therefore, the accuracy and the roughness can be used to measure the set uncertainty. As was the case for a complete information system, Eqs. (5) and (6) measure only set uncertainty and not knowledge uncertainty for an incomplete information system (Wang et al. 2008a). Wang et al. (2008a) proposed new definitions of accuracy and roughness based on the tolerance relation:

$$\alpha^*_{R_T^P}(X) = 1 - \rho_{R_T^P}(X) \times GK(R_T^P) \tag{7}$$

$$\rho^*_{R_T^P}(X) = \rho_{R_T^P}(X) \times GK(R_T^P) \tag{8}$$

Knowledge granularity, defined as $GK(R_T^P) = \sum_{i=1}^{|U|} \left|R_T^P(u_i)\right| / |U|^2$, was employed to measure the roughness of knowledge. In contrast to knowledge granularity, $HK(R_T^P) = 1 - GK(R_T^P)$ was used to characterise the precision of knowledge. Obviously, Eqs. (7) and (8) consider both set uncertainty and knowledge uncertainty, which corrects the problems with the classical definitions of accuracy and roughness to some extent. However, certain limitations remain for an incomplete information system, and these are revealed by the following example.

Example 2 Let $S = (U, A)$ be an incomplete information system with $U = \{u_1, u_2, \ldots, u_7\}$, $P, Q \subseteq A$. Assume that

$$\begin{cases} U/R_T^P = \{\{u_1, u_2\}, \{u_2, u_1\}, \{u_3, u_4, u_5\}, \{u_4, u_3\}, \{u_5, u_3, u_6\}, \{u_6, u_5, u_7\}, \{u_7, u_6\}\} \\ U/R_T^Q = \{\{u_1, u_2\}, \{u_2, u_1\}, \{u_3, u_4\}, \{u_4, u_3\}, \{u_5, u_6\}, \{u_6, u_5, u_7\}, \{u_7, u_6\}\} \end{cases} \tag{9}$$

Obviously, $U/R_T^Q \subset U/R_T^P$. Table 2 shows the upper and lower approximations, and the boundary region of the rough set X, while Table 3 shows the values of the uncertainty measures of the rough sets X for the knowledge P and Q. Figures 2 and 3 present the uncertainty measures of X_1 and X_2, respectively. The subscripts of the uncertainty measures in Figs. 2 and 3 are omitted, e.g., $\alpha_{R_T^P}(X)$ is abbreviated as α and $GK(R_T^P)$ is abbreviated as GK.

From Tables 2 and 3, Figs. 2, and 3, we can make the following observations:

1. When $X = X_1$, the lower and upper approximations of X_1 with respect to the knowledge P and Q are identical, and the classification granularities in the upper approximations $\{u_1, u_2, u_6, u_7\}$ induced by the knowledge P and Q are also identical. Therefore, the roughness and the accuracy of the knowledge P and Q are equal, which is logically consistent. However, $\alpha^*_{R_T^P}(X) < \alpha^*_{R_T^Q}(X)$ and $\rho^*_{R_T^Q}(X) < \rho^*_{R_T^P}(X)$. These results are caused by the subdivision of the classification granularities $R_T^P(u_3)$ and $R_T^P(u_5)$ in the negative region of set X_1 with the knowledge Q. Obviously, $R_T^P(u_3)$ and

Table 2 Upper and lower approximations and the boundary region of the rough set *X*

X	$\underline{R_T^P}(X)$	$\underline{R_T^Q}(X)$	$\overline{R_T^P}(X)$	$\overline{R_T^Q}(X)$	$BN_{R_T^P}(X)$	$BN_{R_T^Q}(X)$
X_1	$\{u_1, u_2\}$	$\{u_1, u_2\}$	$\{u_1, u_2, u_6, u_7\}$	$\{u_1, u_2, u_6, u_7\}$	$\{u_6, u_7\}$	$\{u_6, u_7\}$
X_2	$\emptyset$	$\emptyset$	$\{u_1, u_2, u_3, u_4, u_5\}$	$\{u_1, u_2, u_3, u_4\}$	$\{u_1, u_2, u_3, u_4, u_5\}$	$\{u_1, u_2, u_3, u_4\}$

Table 3 Uncertainty measures of the rough set *X*

X	$\rho_{R_T^P}(X)$	$\rho_{R_T^Q}(X)$	$GK(R_T^P)$	$GK(R_T^Q)$	$\rho^*_{R_T^P}(X)$	$\rho^*_{R_T^Q}(X)$
X_1	0.50	0.50	0.35	0.31	0.17	0.15
X_2	1	1	0.35	0.31	0.35	0.31

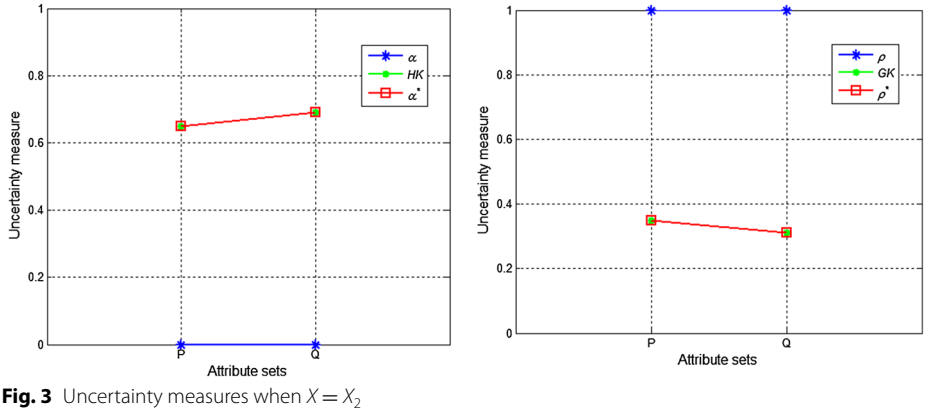

Fig. 3 Uncertainty measures when $X = X_2$

$R_T^P(u_5)$ are unrelated to X, and thus $\alpha^*_{R_T^P}(X)$ and $\rho^*_{R_T^P}(X)$ are inconsistent with human cognition.

2. When $X = X_2$, the lower approximation of set X_2 is an empty set, and as a result, X_2 is a boundary rough set. The boundary regions of X_2 with respect to the knowledge P and Q are different. In this case, the larger the boundary region is, the coarser the knowledge (Yang and John 2008). However, $\rho_{R_T^P}(X) = \rho_{R_T^Q}(X)$ and $\alpha_{R_T^P}(X) = \alpha_{R_T^Q}(X)$, so from Tables 2 and 3 we obtain $HK(R_T^P) = \alpha^*_{R_T^P}(X) < \alpha^*_{R_T^Q}(X) = HK(R_T^Q)$ and $GK(R_T^Q) = \rho^*_{R_T^Q}(X) < \rho^*_{R_T^P}(X) = GK(R_T^P)$, which shows that $\rho_{R_T^P}(X)$ and $\alpha_{R_T^P}(X)$ do not accurately reflect the uncertainty of the rough set when $BN_{R_T^P}(X) = \emptyset$; $\alpha^*_{R_T^P}(X)$ and $\rho^*_{R_T^P}(X)$ can measure the knowledge uncertainty but not the set uncertainty.
3. From Fig. 2 and Fig. 3, it can be observed that $\alpha_{R_T^P}(X) \le \alpha^*_{R_T^P}(X)$, $\alpha_{R_T^Q}(X) \le \alpha^*_{R_T^Q}(X)$, $HK(R_T^P) \le \alpha^*_{R_T^P}(X)$, and $HK(R_T^Q) \le \alpha^*_{R_T^Q}(X)$; therefore, $\rho^*_{R_T^P}(X) < \rho_{R_T^P}(X)$, $\rho^*_{R_T^Q}(X) < \rho_{R_T^Q}(X)$, $\rho^*_{R_T^P}(X) \le GK(R_T^P)$ and $\rho^*_{R_T^Q}(X) \le GK(R_T^Q)$ when $X = X_1$ or $X = X_2$. That is, the value of the roughness measure that includes two types of uncertainty is smaller than that of the measure reflecting only one type of uncertainty, whereas the value of the accuracy measure that includes two types of uncertainty is greater than that of the measure reflecting only one type of uncertainty. Obviously, these results are logically inconsistent.

Example 2 shows that, similar to the results for a complete information system, uncertainty measures for an incomplete information system have certain limitations. Xu et al. (Xu et al. 2009) presented a new integrated uncertainty measure for ordered information systems with properties similar to those of $\alpha^*_{R_T^P}(X)$ and $\rho^*_{R_T^P}(X)$. Therefore, this uncertainty measure has the same limitations.

From Examples 1 and 2, we can conclude that the imprecision of rough sets is not well characterised by existing measures for both complete and incomplete information systems. Therefore, it is necessary to find a more comprehensive and effective uncertainty measure based on general binary relations.

Integrated measures based on general binary relations

In classical RST (Pawlak 1991), uncertainty includes knowledge uncertainty and set uncertainty. Various integrated uncertainty measures have been proposed that are based on a given binary relation and include both types of uncertainty (Wang et al. 2008a; Liang et al. 2009; Xu et al. 2009). The values of these measures depend on the classification granularity, which is unassociated with the set $X \subseteq U$, specifically the classification granularity in the negative region of X. This behaviour is inconsistent with human cognition (Wang and Zhang 2008). Intuitively, the value of an integrated roughness measure (i.e., the roughness of a rough set) that evaluates two types of uncertainty should be greater than that of a measure which evaluates only one type of uncertainty, but this property is not satisfied by almost all the existing integrated measures. In addition, the existing integrated uncertainty measures cannot be used to effectively characterise the roughness of rough sets in certain cases. In this section, the limitations of existing integrated uncertainty measures are addressed. First, a knowledge uncertainty measure that

is based on general binary relations is presented. Based on this uncertainty measure, novel and logically consistent integrated uncertainty measures are presented.

Information entropy measure based on general binary relations

Classical RST starts from an equivalence relation. Knowledge is based on the ability to partition a "universe" using the equivalence relation. The finer the partitioning, the more precise the knowledge will be. In an incomplete information system, overlaps may occur among several similar classes defined by the tolerance relation, the similarity relation, or the limited tolerance relation. Moreover, a covering is substituted for the partition of the universe. Thus, the equivalence relation cannot be satisfied for an incomplete information system. The same problems appear for general binary relations. However, research on uncertainty measures based on general binary relations is lacking (Huang et al. 2004). This lack of research motivates the investigation of an effective uncertainty measure based on general binary relations. In the following, an uncertainty measure based on general binary relations will be discussed.

Let $R^P \subseteq U \times U$ be a general binary relation on U, $P \subseteq A$. For two elements $u_i, u_j \in U$, if u_j has the same properties as u_i with respect to R^P, i.e., $u_i R^P u_j$, we say that u_j is R^P -related to u_i. A general binary relation may be more conveniently represented using successor neighbourhoods or a classification granularity:

$$R_S^P(u_i) = \{u_j \Big| u_j \in U, u_i R_S^P u_j\} \tag{10}$$

The classification granularity $R_S^P(u_i)$ consists of all R^P-related elements of u_i. If $R_S^P(u_i)$ contains more elements, more objects will belong to the same class as u_i, the classification granularities will be larger, and the capability of the knowledge P to classify the object u_i will be weaker. Given these characteristics, a definition of an uncertainty measure based on general binary relations is given as follows.

Definition 1 Given an information system $S = (U, A)$, $u_i \in U$ and $1 \le i \le |U|$, the information entropy of the knowledge $P \subseteq A$ is defined as

$$H'(P) = 1 - G'(P) \tag{11}$$

where

$$G'(P) = \sum_{u_i \in BN_{R^P}(X)} \frac{\left|R_S^P(u_i)\right| - 1}{|U|(|U| - 1)} \tag{12}$$

Theorem 1 (Monotonicity) *Given an information system $S = (U, A)$, $P, Q \subseteq A$ and $P \preceq Q$, the information entropy satisfies $H'(Q) \le H'(P)$, where equality holds if and only if $P \approx Q$.*

The proof of this theorem follows from the definition of the partial relation and Definition 1.

Corollary 1 *Given an information system $S = (U, A)$, $P \subseteq A$, $H'(P)$ reaches a minimum value of 0 if and only if $R_S^P(u_i) = U$ for $\forall u_i \in U$, and $H'(P)$ reaches a maximum value of 1 if and only if $R_S^P(u_i) = u_i$ for $\forall u_i \in U$.*

Theorem 1 and Corollary 1 indicate that the information entropy monotonically increases as the classification granularity becomes smaller through finer classification. If the knowledge P cannot distinguish between any two objects in the universe U, the information entropy is at the minimum and the knowledge P has the weakest classification capability and the greatest roughness. If the knowledge P can distinguish all objects in the universe U, the information entropy is at the maximum and the knowledge P has the strongest classification capability and accuracy. Therefore, information entropy describes the roughness of knowledge in the context of granularity.

Integrated measures of rough sets

To measure the uncertainty of rough sets more precisely, Yang and John (2008) proposed two complementary uncertainty measures for a complete information system, global accuracy and global roughness. These two complementary uncertainty measures can measure the set uncertainty more comprehensively than other uncertainty measures. However, these two complementary uncertainty measures are based on the equivalence relation and are not suitable for an incomplete information system. However, global accuracy and global roughness can be extended to incomplete systems using a general binary relation. The new definition for global accuracy is

$$\sigma'_P(X) = 1 - \frac{\left|BN'_P(X)\right|}{2|U|} \tag{13}$$

where $BN'_P(X) = \overline{R^P}(X) - \underline{R^P}(X)$. Global roughness is then defined as $\omega'_P(X) = 1 - \sigma'_P(X)$. Based on these definitions, we propose two novel integrated measures.

Definition 2 Given an information system $S = (U, A)$, $P \subseteq A$, $X \subseteq U$ and the general binary relation R^P, the integrated roughness and the integrated accuracy of X are defined as:

$$\rho'_P(X) = 1 - \alpha'_P(X) \tag{14}$$

$$\alpha'_P(X) = \sigma'_P(X) \times H'(P) \tag{15}$$

$H'(P)$ is used to measure knowledge uncertainty, and $\sigma'_P(X)$ is used to measure set uncertainty. Obviously, Definition 2 considers not only the size of the boundary region of a rough set but also the classification granularity of the boundary region. Therefore, integrated roughness and integrated accuracy measure two types of uncertainty.

Theorem 2 (Monotonicity) *Given an information system $S = (U, A)$, $P, Q \subseteq A$, $P \prec Q$, and $X \subseteq U$, the following relations hold:*

(1) $\sigma'_Q(X) \leq \sigma'_P(X)$; (2) $\rho'_P(X) \leq \rho'_Q(X)$.

Proof (1) Because $P \prec Q$, we have that $R_S^P(u_i) \subseteq R_S^Q(u_i)$ for $\forall u_i \in U$, and $\exists u_k \in U$ satisfies $R_S^P(u_k) \subset R_S^Q(u_k)$. For $\forall u_i \in \underline{R^Q}(X)$, $R_S^Q(u_i) \subseteq X$, we obtain $R_S^P(u_i) \subseteq X$, i.e., $u_i \in \underline{R^P}(X)$. Thus, $\underline{R^Q}(X) \subseteq \underline{R^P}(X)$. Similarly, $R_S^P(u_i) \cap X \neq \emptyset$ for $\forall u_i \in \overline{R^P}(X)$. Because $R_S^P(u_i) \subseteq R_S^Q(u_i)$, we have $R_S^Q(u_i) \cap X \neq \emptyset$, i.e., $u_i \in \overline{R^Q}(X)$. Therefore, $\overline{R^P}(X) \subseteq \overline{R^Q}(X)$ and $BN_P'(X) \subseteq BN_Q'(X)$. According to Eq. (13), we have $\sigma_Q'(X) \leq \sigma_P'(X)$, where equality holds if and only if $BN_P'(X) = BN_Q'(X)$.

(2) Because $P \prec Q$, we have $R_S^P(u_i) \subseteq R_S^Q(u_i)$ for any $u_i \in U$, and $\exists u_k \in U$ satisfies $R_S^P(u_k) \subset R_S^Q(u_k)$. To simplify the proof, we assume that only one object $u_k \in U$ satisfies $R_S^P(u_k) \subset R_S^Q(u_k)$, so we have $R_S^P(u_i) = R_S^Q(u_i)$ for any other $u_i \neq u_k$ (the proof for many objects is similar). Three cases are discussed:

① $R_S^Q(u_k) \subseteq X$: Because $R_S^P(u_k) \subset R_S^Q(u_k)$, it follows that $R_S^P(u_k) \subseteq X$ and $u_k \notin BN_P'(X) = BN_Q'(X)$. From the proof of (1), we have $\sigma_Q'(X) = \sigma_P'(X)$. Because $R_S^P(u_i) = R_S^Q(u_i)$ for $\forall u_i \neq u_k$, from Eq. (8) we obtain $H'(P) = H'(Q)$. According to Definition 2, we have $\alpha_Q'(X) = \alpha_P'(X)$ and $\rho_P'(X) = \rho_Q'(X)$.

② $R_S^Q(u_k) \cap X = \emptyset$: Because $R_S^P(u_k) \subset R_S^Q(u_k)$, we have $R_S^P(u_k) \cap X = \emptyset$ and $u_k \notin BN_P'(X) = BN_Q'(X)$. From the proof of (1), we have $\sigma_Q'(X) = \sigma_P'(X)$. Because $R_S^P(u_i) = R_S^Q(u_i)$ for $\forall u_i \neq u_k$, from Eq. (11) we obtain $H'(Q) = H'(P)$. According to Definition 2, we have that $\alpha_Q'(X) = \alpha_P'(X)$ and $\rho_P'(X) = \rho_Q'(X)$.

③ $R_S^Q(u_k) \cap X \neq \emptyset$ and $R_S^Q(u_k) \cap X \neq R_S^Q(u_k)$. We have $u_k \in BN_Q'(X)$. Three cases must be considered:

(1) If $R_S^P(u_k) \cap X \neq \emptyset$ and $R_S^P(u_k) \cap X \neq R_S^P(u_k)$, then $u_k \in BN_P'(X) = BN_Q'(X)$. From the proof of (1), we obtain $0 < \sigma_Q'(X) = \sigma_P'(X)$. Because $R_S^P(u_i) = R_S^Q(u_i)$ for $\forall u_i \neq u_k$, $R_S^P(u_k) \subset R_S^Q(u_k)$, from Eq. (8) and Definition 2 we have that $H'(Q) < H'(P)$, $\alpha_Q'(X) < \alpha_P'(X)$ and $\rho_P'(X) < \rho_Q'(X)$.

(2) If $R_S^P(u_k) \subseteq X$, then $u_k \notin BN_P'(X)$. Thus, $BN_P'(X) \subset BN_Q'(X) \neq \emptyset$. From the proof of (1), we have $\sigma_Q'(X) < \sigma_P'(X)$. Because $R_S^P(u_i) = R_S^Q(u_i)$ and $R_S^P(u_k) \subset R_S^Q(u_k)$ for $\forall u_i \neq u_k$, according to Eq. (8) we have that $H'(Q) < H'(P)$. From Definition 2, we have that $\alpha_Q'(X) < \alpha_P'(X)$ and $\rho_P'(X) < \rho_Q'(X)$.

(3) If $R_S^P(u_k) \cap X = \emptyset$, then $u_k \notin BN_P'(X)$. Therefore, $BN_P'(X) \subset BN_Q'(X) \neq \emptyset$. From the proof of (1), we have $\sigma_Q'(X) < \sigma_P'(X)$. Because $R_S^P(u_i) = R_S^Q(u_i)$ and $R_S^P(u_k) \subset R_S^Q(u_k)$ for $\forall u_i \neq u_k$, according to Eq. (13) we obtain $H'(Q) < H'(P)$. From Definition 2, we have that $\alpha_Q'(X) < \alpha_P'(X)$ and $\rho_P'(X) < \rho_Q'(X)$.

This concludes the proof of Theorem 2.

Corollary 2 *Given an information system* $S = (U, A)$, $P, Q \subseteq A$, $P \prec Q$ *and* $X \subseteq U$, *where* $U' = \{u_k \in U \,|\, R_S^P(u_k) \subset R_S^Q(u_k)\}$, *then* $\rho_P^\alpha(X) = \rho_Q^\alpha(X)$ *if and only if* $u_k \notin BN_Q'(X)$ *for* $\forall u_k \in U'$.

The proof of this corollary follows from Theorem 2. From Theorem 2 and Corollary 2, we can observe that the integrated accuracy does not strictly monotonically increase, and the integrated roughness does not strictly monotonically decrease as the classification granularity becomes smaller through finer classification. That is, the integrated

accuracy and the integrated roughness are unrelated to the classification granularity $R_S^Q(u_i)$, where $u_i \in \{U - BN'_Q(X)\}$. If the classification granularity $R_S^Q(u_k)$ defined by the knowledge P satisfies $u_k \in BN'_Q(X)$, the integrated accuracy (integrated roughness) strictly monotonically increases (decreases), which is accords to human cognition.

Corollary 3 *Given an information system $S = (U, A)$, $P \subseteq A$ and $X \subseteq U$, the integrated roughness satisfies $0 \leq \rho'_P(X) \leq 1$. Equality holds on the right side if and only if $R_S^P(u_i) = U$ for $\forall u_i \in U$, and equality holds on the left side if and only if $BN'_P(X) = \emptyset$.*

The proof of this corollary follows from Eqs. (11), (13), (14) and (15).

Theorem 3 *Given an information system $S = (U, A)$, $P \subseteq A$ and $X \subseteq U$, the integrated accuracy and the integrated roughness satisfy the relations $\alpha'_P(X) \leq \sigma'_P(X)$ and $\omega'_P(X) \leq \rho'_P(X)$.*

It can be concluded from Theorem 3 that the value of the integrated accuracy $\alpha'_P(X)$, which measures two types of uncertainty, will be less than that of $\sigma'_P(X)$, which measures only one type of uncertainty, and the value of the integrated roughness $\rho'_P(X)$, which measures two types of uncertainty, will be greater than that of $\omega'_P(X)$, which measures only one type of uncertainty. Therefore, the new integrated measures $\alpha'_P(X)$ and $\rho'_P(X)$ are logically consistent.

Corollary 4 *Given an information system $S = (U, A)$, $P, Q \subseteq A$, $P \preceq Q$ and $X \subseteq U$,*

(1) *If X is a boundary rough set (i.e., $\underline{R^P}(X) = \underline{R^Q}(X) = \emptyset$) and $\overline{R^Q}(X) = \overline{R^P}(X)$, then $\rho_Q(X) = \rho_P(X)$ and $\omega'_Q(X) = \omega'_P(X)$, but $\rho'_P(X) \leq \rho'_Q(X)$;*
(2) *If $\rho'_P(X) = \rho'_Q(X)$, then $\rho_Q(X) = \rho_P(X)$ and $\omega'_Q(X) = \omega'_P(X)$;*
(3) *If $\rho_P(X) < \rho_Q(X)$ or $\omega'_P(X) < \omega'_Q(X)$, then $\rho'_P(X) \leq \rho'_Q(X)$;Property (1) in Corollary 4 indicates that the integrated roughness $\rho'_P(X)$ measures both set uncertainty and knowledge uncertainty for X; however, $\rho_P(X)$ and $\omega'_P(X)$ measure only set uncertainty. Property (2) in Corollary 4 shows that $\rho_P(X)$ and $\omega'_P(X)$ are invariant if the integrated roughness $\rho'_P(X)$ remains unchanged, although the classification granularity is smaller through finer classification. However, $\rho_P(X)$ and $\omega'_P(X)$ may not decrease if the integrated roughness $\rho'_P(X)$ decreases. Property (3) in Corollary 4 shows that the integrated roughness $\rho'_P(X)$ decreases when $\rho_P(X)$ and $\omega'_P(X)$ decrease. The converses of properties (2) and (3) are not always true. Corollary 4 implies that the integrated roughness is more sensitive than $\rho_P(X)$ and $\omega'_P(X)$ for a general binary relation.*

The preceding properties characterise the variation of the integrated roughness with the classification granularity. The effectiveness of the proposed measure is verified in the following example.

Example 3 (Continued from Example 1) Results for the uncertainty measures based on an equivalence relation were obtained from Eqs. (11), (13), (14) and (15), and these results are listed in Table 4.

Table 4 New uncertainty measures of a rough set *X* with various classification granularities

	Num_L	*Num_U*	*Num_B*	$\sigma'_P(X)$	$\omega'_P(X)$	$H'(P)$	$G'(P)$	$\alpha'_P(X)$	$\rho'_P(X)$
(1)	0	3600	3600	0.5	0.5	0.889	0.111	0.445	0.555
(2)	0	2200	2200	0.694	0.306	0.955	0.045	0.663	0.337
(3)	0	2200	2200	0.694	0.306	0.974	0.026	0.676	0.324
(4)	200	2200	2000	0.722	0.278	0.985	0.015	0.711	0.289
(5)	5400	1548	1008	0.860	0.140	0.997	0.003	0.857	0.143
(6)	5400	1548	1008	0.860	0.140	0.997	0.003	0.857	0.143
(7)	5400	1548	1008	0.860	0.140	0.997	0.003	0.857	0.143

From Table 4, we can make the following observations:

1. Comparing partitions (1) with (2), (3) with (4) and (4) with (5), we can observe that the boundary region becomes smaller, and thus $\sigma'_P(X)$ becomes smaller and $\omega'_P(X)$ becomes larger. In addition, the classification granularity in the boundary region becomes finer, which leads to an increase in the discernibility of objects in the boundary region, and thus $\rho'_P(X)$ becomes smaller and $\alpha'_P(X)$ becomes larger. Obviously, the new integrated measures $\alpha'_P(X)$ and $\rho'_P(X)$ reflect not only the set uncertainty but also the knowledge uncertainty in the boundary region.
2. Comparing partition (2) with (3), it can be observed that the boundary region, the global accuracy $\sigma'_P(X)$ and the global roughness $\omega'_P(X)$ do not change. However, the classification granularity in the boundary region becomes finer, i.e., the discernibility of objects in the boundary region increases, and thus $H'(P)$ becomes larger. Obviously, an increase in $\alpha'_P(X)$ and a decrease in $\rho'_P(X)$ in this case reflect the decrease of the knowledge uncertainty in the boundary region, whereas the set uncertainty does not change.
3. Comparing partitions (5) with (6) and (6) with (7), it can be observed that the boundary region and the classification granularity in the boundary region remain the same, and thus the uncertainty of the rough set X does not change. Accordingly, $\sigma'_P(X)$, $\omega'_P(X)$, $H'(P)$, $G'(P)$, $\rho'_P(X)$ and $\alpha'_P(X)$ all do not change, which shows that the new integrated measures are unassociated with subdivision of classification granularities unrelated to rough set X. Therefore, the new integrated measures are consistent with human cognition.
4. The integrated accuracy $\alpha'_P(X)$ and the integrated roughness $\rho'_P(X)$ reflect two types of uncertainty. Therefore, the value of the integrated accuracy is smaller than those of $\sigma'_P(X)$ and $H'(P)$, and the value of the integrated roughness $\rho'_P(X)$ is larger than those of $\omega'_P(X)$ and $G'(P)$. These results are logically consistent.

Example 3 illustrates that the new integrated measures $\alpha'_P(X)$ and $\rho'_P(X)$ for a complete information system overcome the limitations of the existing uncertainty measures, better characterise the imprecision of rough sets and are consistent with human cognition.

Example 4 (Continued from Example 2) We calculate the new uncertainty measures for the tolerance relation using Eqs. (11), (13), (14) and (15). The results are shown in

Table 5. Figures 4 and 5 present the new uncertainty measures for X_1 and X_2, respectively. The subscripts of the uncertainty measures in Figs. 4 and 5 are omitted (as in Figs. 2 and 3).

We can draw the following conclusions from Table 5, Fig. 4 and Fig. 5:

1. When $X = X_1$, the upper and lower approximations of set X_1 are equal, and the classification granularities of objects in the boundary region are also the same with respect to the knowledge P and Q. Thus, subdividing the classification granularities $R_S^P(u_3)$ and $R_S^P(u_5)$ (which are unrelated to X) in the negative region of set X does not alter the values of $\alpha'_P(X)$ and $\rho'_P(X)$, which shows that $\alpha'_P(X)$ and $\rho'_P(X)$ are consistent with human cognition.
2. When $X = X_2$, X is a boundary rough set. The boundary regions of X with respect to the knowledge P and Q are different. Consequently, $\sigma'_P(X) < \sigma'_Q(X)$ and $\omega'_Q(X) < \omega'_P(X)$. In addition, the classification granularities of objects in the boundary region with respect to the knowledge P and Q are different. Further-

Table 5 The proposed uncertainty measures for an incomplete information system

X	$\sigma'_P(X)$	$\sigma'_Q(X)$	$H'(P)$	$H'(Q)$	$\alpha'_P(X)$	$\alpha'_Q(X)$
X_1	0.857	0.857	0.929	0.929	0.796	0.796
X_2	0.643	0.714	0.833	0.905	0.536	0.646

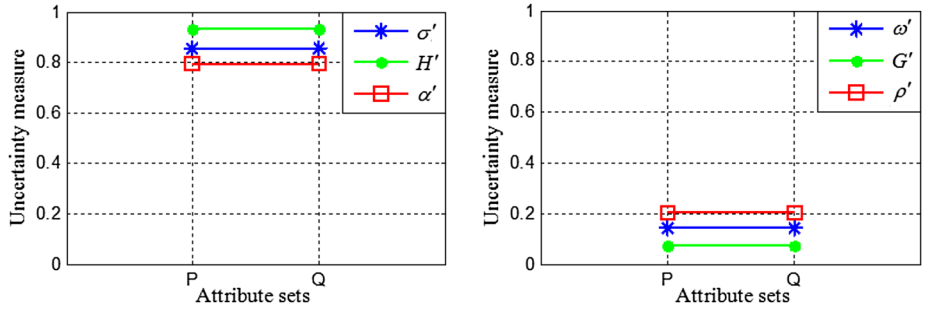

Fig. 4 The proposed uncertainty measures when $X = X_1$

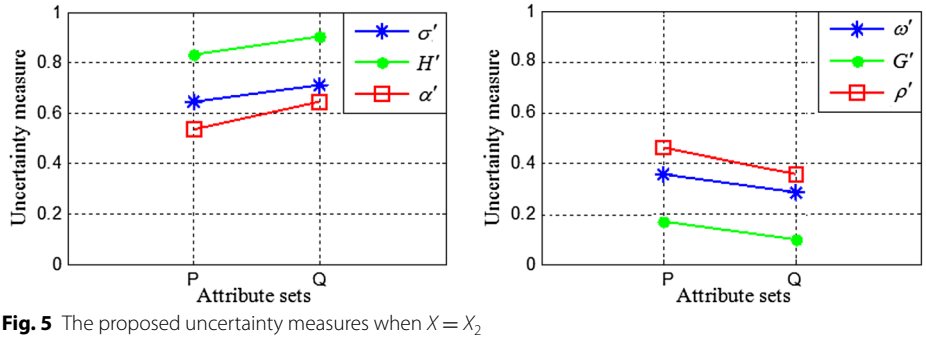

Fig. 5 The proposed uncertainty measures when $X = X_2$

more, $H'(P) < H'(Q)$ and $G'(Q) < G'(P)$. Finally, the integrated measures satisfy $\rho'_Q(X) < \rho'_P(X)$ and $\alpha'_P(X) < \alpha'_Q(X)$. Obviously, the proposed integrated accuracy and integrated roughness can not only correctly reflect set uncertainty but also correctly measure knowledge uncertainty for a boundary rough set. Therefore, $\alpha'_P(X)$ and $\rho'_P(X)$ can adequately characterise the uncertainty of rough sets.

3. From Figs. 4 and 5, it can be observed that $\alpha'_P(X) \leq \sigma'_P(X)$, $\alpha'_P(X) \leq H'(P)$, $\alpha'_Q(X) \leq \sigma'_Q(X)$ and $\alpha'_Q(X) \leq H'(Q)$ when $X = X_1$ or $X = X_2$. That is to say, the value of the integrated accuracy, which is based on two types of uncertainty, is smaller than that of the measure based on only one type of uncertainty. In addition, $\omega'_P(X) \leq \rho'_P(X)$, $G'(P) \leq \rho'_P(X)$, $\omega'_Q(X) \leq \rho'_Q(X)$ and $G'(Q) \leq \rho'_Q(X)$, which indicates that the value of the integrated roughness, which reflects two types of uncertainty, is greater than that of the measure reflecting only one type of uncertainty. Obviously, these results are logically consistent.

Comparing Examples 3 and 4 with Examples 1 and 2, we can conclude that the new integrated measures $\alpha'_P(X)$ and $\rho'_P(X)$ under general binary relations are suitable for both complete and incomplete information systems. These new measures overcome the limitations of existing uncertainty measures and can satisfactorily characterise the imprecision of rough sets. Therefore, the proposed integrated measures are more comprehensive and effective uncertainty measures for both complete and incomplete information systems.

Conclusion

The extension of RST to incomplete information systems is important for making RST practical. Uncertainty measures are the basis for information processing and knowledge acquisition in an incomplete information system. At present, direct processing of an incomplete information system lacks a theoretical basis. By considering the nature of the roughness of sets, we developed new integrated measures based on general binary relations. Several desirable properties of the proposed measures have been shown. We have demonstrated that the new measures overcome the limitations of existing uncertainty measures and can be used to measure with a simple and comprehensive form the roughness and the accuracy of a rough set, and the results are logically consistent. Research on the application of our proposed integrated measures for rule acquisition is planned.

Authors' contributions

SHT and YJN carried out the studies of RST-based methods for measuring system uncertainty, presented the new uncertainty measures, and drafted the manuscript. ML participated in the provement of the theorems and the design of the examples. MH participated in the analysis of the proposed algorithm and helped to polish the manuscript. FL participated in the revision of this paper. All authors read and approved the final manuscript.

Author details

[1] Science and Technology on Automatic Target Recognition Laboratory, National University of Defense Technology, Changsha 410073, China. [2] PLA Units: 66295, Baoding 072750, China. [3] School of Biomedical Engineering, Third Military Medical University, Chongqing 400038, China.

Acknowledgements

This work has been sponsored by a grant from the National Natural Science Foundation of China (Nos. 41201363 and 61471371) and the Natural Science Foundation of Hunan Province of China (No. 2015jj3022). Moreover, the authors would like to thank the anonymous reviewers for their insightful comments in improving the quality of this paper.

Competing interests

The authors declare that they have no competing interests.

Received: 4 August 2015 Accepted: 5 January 2016
Published online: 09 February 2016

References

Beaubouef T, Petry FE, Arora G (1998) Information-theoretic measures of uncertainty for rough sets and rough relational databases. Inf Sci 109(1–4):185–195

Chen YM, Wu KS, Chen XH, Tang CH, Zhu QX (2014) An entropy-based uncertainty measurement approach in neighborhood systems. Inf Sci 279:239–250

Dai JH, Xu Q (2012) Approximations and uncertainty measures in incomplete information systems. Inf Sci 198:62–80

Dai JH, Wang WT, Xu Q (2013) An uncertainty measure for incomplete decision tables and its applications. IEEE Trans Cybern 43(4):1277–1289

Dai JH, Huang DB, Su HS, Tian HW (2014) Uncertainty measurement for covering rough sets. Int J Uncertainty Fuzziness Knowl Based Syst 22(2):217–233

Gantayat SS, Misra A, Panda BS (2014) A study of incomplete data—a review. In: Satapathy SC, Udgata SK, Biswal BN (eds) Proceedings of the international conference on frontiers of intelligent computing: theory and applications. Springer, Berlin, pp 401–408

Greco S, Matarazzo B, Slowinski R (2002) Rough approximation by dominance relations. Int J Intell Syst 17(2):153–171

Hu QH, Che XJ, Zhang L, Zhang D, Guo MZ, Yu DR (2012) Rank entropy-based decision trees for monotonic classification. IEEE Trans Knowl Data Eng 24(11):2052–2064

Huang B, Zhou XZ, Shi YC (2004) Entropy of knowledge and rough set based on general binary relation. Syst Eng Theory Pract 24(1):93–96

Kryszkiewicz M (1998) Rough set approach to incomplete information systems. Inf Sci 112(1–4):39–49

Li DY, Du Y (2005) Artificial intelligence with uncertainty. National Defense Industry Press, Beijing

Liang JY, Wang JH, Qian YH (2009) A new measure of uncertainty based on knowledge granulation for rough sets. Inf Sci 179(4):458–470

Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publisher, London

Qian YH, Liang JY, Wang F (2009) A new method for measuring the uncertainty in incomplete information systems. Int J Uncertain Fuzziness Knowl Based Syst 17(6):855–880

Stefanowski J, Tsoukias A (1999) On the extension of rough sets under incomplete information. In: New directions in rough sets, data mining, and granular-soft computing. 7th International Workshop, Yamaguchi, Japan

Sun L, Xu JC, Tian Y (2012) Feature selection using rough entropy-based uncertainty measures in incomplete decision systems. Knowl Based Syst 36:206–216

Sun L, Xu JC, Xu TH (2014) Information entropy and information granulation-based uncertainty measures in incomplete information systems. Appl Math Inf Sci 8(4):2073–2083

Teng SH, Sun JX, Li ZY, Zou G (2009) A new heuristic reduction algorithm based on general binary relations. In: Sixth international symposium on multispectral image processing and pattern recognition. SPIE: International Society for Optics and Photonics, Yichang, China

Teng SH, Lu M, Yang AF, Zhang J, Zhuang ZW, He M (2016) Efficient attribute reduction from the viewpoint of discernibility. Inf Sci 326:297–314

Wang GY (2002) Extension of rough set under incomplete information systems. J Comput Res Dev 39(10):1238–1243

Wang GY, Zhang QH (2008) Uncertainty of rough sets in different knowledge granularities. Chin J Comput 31(9):1588–1598

Wang JH, Liang JY, Qian YH, Dang CY (2008a) Uncertainty measure of rough sets based on a knowledge granulation of incomplete information systems. Int J Uncertain Fuzziness Knowl Based Syst 16(2):233–244

Wang CZ, Wu CX, Chen DG (2008b) A systematic study on attribute reduction with rough sets based on general binary relations. Inf Sci 178(9):2237–2261

Xu Y, Li LS (2011) Variable precision rough set model based on (α, λ) connection degree tolerance relation. Acta Autom Sin 37(3):303–308

Xu WH, Zhang XY, Zhang WX (2009) Knowledge granulation, knowledge entropy and knowledge uncertainty measure in ordered information systems. Appl Soft Comput 9(4):1244–1251

Yang YJ, John R (2008) Global roughness of approximation and boundary rough sets. In: IEEE world congress on computational intelligence. IEEE, Hong Kong

Yao YY (1998) Relational interpretations of neighborhood operators and rough set approximation operators. Inf Sci 111(1–4):239–259

Zhu W (2007) Generalized rough sets based on relations. Inf Sci 177(22):4997–5011