International Journal of Approximate Reasoning 50 (2009) 174–188

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Knowledge structure, knowledge granulation and knowledge distance in a knowledge base

Yuhua Qian [a,b,c], Jiye Liang [a,b,*], Chuangyin Dang [c]

[a] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China
[b] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
[c] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 15 October 2007
Received in revised form 12 August 2008
Accepted 13 August 2008
Available online 24 August 2008

Keywords:
Rough set theory
Granular computing
Knowledge bases
Knowledge granulation
Knowledge distance

ABSTRACT

One of the strengths of rough set theory is the fact that an unknown target concept can be approximately characterized by existing knowledge structures in a knowledge base. Knowledge structures in knowledge bases have two categories: complete and incomplete. In this paper, through uniformly expressing these two kinds of knowledge structures, we first address four operators on a knowledge base, which are adequate for generating new knowledge structures through using known knowledge structures. Then, an axiom definition of knowledge granulation in knowledge bases is presented, under which some existing knowledge granulations become its special forms. Finally, we introduce the concept of a knowledge distance for calculating the difference between two knowledge structures in the same knowledge base. Noting that the knowledge distance satisfies the three properties of a distance space on all knowledge structures induced by a given universe. These results will be very helpful for knowledge discovery from knowledge bases and significant for establishing a framework of granular computing in knowledge bases.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

In 1982, Pawlak proposed a new set theory, the so-called rough set theory [27,29], which is an effective tool for uncertainty management and uncertainty reasoning, and has a wide variety of applications in artificial intelligence [6,11,12,38,41,56]. In the rough set theory, an attribute set partitions a universe into some knowledge granules or elemental concepts, which is called a knowledge structure. Partition, granulation and approximation are the methods widely used in human's reasoning [53–55]. To date, rough set methodology has been applied in feature selection [44,45], knowledge reduction [16,21,26,39,47,51,57], rule extraction [2,8,34,46,56,58,59], uncertainty reasoning [9,28,32] and granular computing [1,13,22,23,31,50–52,54]. In the past 10 years, some extensions of Pawlak's rough set model have been proposed in terms of various requirements [5,25,37,42,43,49].

Knowledge bases and indiscernibility relations are two basic concepts in the rough set theory and assessing the uncertainty of a knowledge structure in a knowledge base is an important research issue. According to whether or not there missing data (null values), knowledge bases are classified into two categories: complete and incomplete [14–17]. In the rough set theory, there are two main approaches for measuring the uncertainty of a knowledge structure in knowledge bases, which are information entropy [3,4,6,10,18,24,33,40] and knowledge granulation [7,19,20,33,48].

* Corresponding author. Address: School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China. Tel./fax: +86 0351 7018176.
E-mail addresses: jinchengqyh@126.com (Y. Qian), ljy@sxu.edu.cn (J. Liang), mecdang@cityu.edu.hk (C. Dang).

0888-613X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2008.08.004

For a given knowledge base, one of the tasks in data mining and knowledge discovery is to generate new knowledge through using known knowledge. However, in rough set theory, the number of knowledge structures is finite in a given knowledge base, which limits the ability of this knowledge base for approximating an unknown concept. This leads to a task for acquiring more knowledge structures from a given knowledge base. To date, the mechanism that how to generate new knowledge structures based on known knowledge structures in knowledge bases have not been widely researched. Therefore, such a mechanism is desirable and very helpful for rule extraction and knowledge discovery from knowledge bases. In addition, knowledge granulation can be used to characterize the degree of the coarseness of a knowledge structure. The finer the knowledge structure is, the smaller the knowledge granulation is. In recent years, several various forms of knowledge granulations have been given in [19,20,33]. From these existing knowledge granulations, we find that they all satisfy several same constraints. In other words, there may exist an uniform description for the existing knowledge granulations. In the view of granular computing, an axiom definition of knowledge granulation may be needed in order to measure the uncertainty of knowledge structures in a knowledge base. It is deserved to point out that when the knowledge granulation (or information entropy) of one knowledge structure is equal to that of the other knowledge structure, these two knowledge structures have the same uncertainty. Nevertheless, it does not mean that these two knowledge structures are equivalent each other. That is to say, information entropy and knowledge granulation cannot characterize the difference between any two knowledge structures in a given knowledge base. In fact, we often need to distinguish two knowledge structures for uncertain data processing in some practical applications.

Based on the above these analyses, main objective of this study has three hands, which are establishing a mathematical framework of granular computing in the context of knowledge bases for acquiring more knowledge structures, constructing an axiom definition of knowledge granulation and giving a knowledge distance among knowledge structures for characterizing the difference among knowledge structures from a knowledge base, respectively.

The rest of the paper is organized as follows. Some basic concepts in rough set theory are briefly reviewed in Section 2. In Section 3, we establish four operators ($\bigcap, \bigcup, \wr$ and $-$) on a knowledge base and investigate their operation properties. Noting that $(K, \bigcap, \bigcup)$ is an assignment lattice and $(K, \bigcap, \bigcup, \wr)$ is a complemented lattice. In Section 4, an axiom definition of knowledge granulation is constructed, under which several existing forms of knowledge granulations become its special instances. In Section 5, to characterize the difference among knowledge structures in a knowledge base, the notion of a knowledge distance is defined and some of its major properties are obtained. Finally, Section 6 concludes this paper with some remarks and discussions.

2. Preliminaries

In this section, we will review several basic concepts in rough set theory and knowledge bases. Throughout this paper, we suppose that the universe U is a finite non-empty set.

Let U be a finite and non-empty set (called a universe) and $R \subseteq U \times U$ an equivalence relation on U, then $K = (U, R)$ is called an knowledge structure (also called an approximation space) [27,30]. The equivalence relation R partitions the set U into disjoint subsets. This partition of the universe is called a quotient set induced by R, denoted by U/R. It represents a very special type of similarity between elements of the universe. If two elements $x, y \in U (x \neq y)$ belong to the same equivalence class, we say that x and y are indistinguishable under the equivalence relation R, i.e., they are equal in R. We denote the equivalence class including x by $E_R(x)$. For our further development, we denote a knowledge structure induced by U/R on U by $K(R)$. In fact, the knowledge structure is formally defined as $K(R) = \{E_R(x) | x \in U\}$. Each equivalence class $E_R(x) (x \in U)$ may be viewed as a knowledge granule consisting of indistinguishable elements [35,50]. The granulation structure induced by an equivalence relation is a partition of the universe.

We say $K = (U, \mathbf{R})$ is a knowledge base, where U is a finite and non-empty set and $\mathbf{R}$ is a family of equivalence relations. Through using a given knowledge structure, one can construct a rough set of any subset on the universe in the following definition.

Definition 1 [27]. Let $K = (U, \mathbf{R})$ be a knowledge base, X a subset of U and $R \in \mathbf{R}$ an equivalence relation, two sets are defined as

$$\underline{R}X = \bigcup \{E_R(x) \in U/R | E_R(x) \subseteq X\}, \tag{1}$$

$$\overline{R}X = \bigcup \{E_R(x) \in U/R | E_R(x) \cap X \neq \emptyset\}, \tag{2}$$

where $\underline{R}X$ and $\overline{R}X$ are called R-lower approximation and R-upper approximation with respect to R, respectively. The order pair $\langle \underline{R}X, \overline{R}X \rangle$ is called a rough set of X with respect to the equivalence relation R.

Let $K = (U, \mathbf{R})$ be a knowledge base, if $R (R \in \mathbf{R})$ is an equivalence relation, then one can get a cover of U by $U/R = \{E_R(x) | x \in U\}$, i.e., for $\forall x \in U$, one has that $E_R(x) \neq \emptyset$ and $\bigcup_{x \in U} E_R(x) = U$. Obviously, $\forall x, y \in U (x \neq y)$, if x, y are partitioned into the same equivalence class, then $E_R(x) = E_R(y)$, otherwise $E_R(x) \cap E_R(y) = \emptyset$. One can define a partial relation $\preceq$ as follows: $P \preceq Q$ $(P, Q \in \mathbf{R})$ if and only if, one has $E_P(x_i) \subseteq E_Q(x_i)$ for any $i \in \{1, 2, \ldots, |U|\}$ [20,36,50]. Here, we denote that P is finer than Q by $P \preceq Q$. Obviously, $(\mathbf{R}, \preceq)$ is a poset [50].

Another important binary relation is tolerance relation, which satisfies reflexivity and symmetry. For example, in an incomplete information system $S = (U, A)$, we define a binary relation on U by $SIM(A) = \{(x, y) \in U \times U | \ \forall a \in A, a(x) = a(y)$ or $a(x) = *$ or $a(y) = *\}$, where $*$ is a missing value. Clearly, $SIM(A)$ is a tolerance relation on U. Similarly, let $R \subseteq U \times U$ denote a tolerance relation on U, the tolerance relation R classifies the universe U into some subsets, i.e., a cover of U [14,15]. This cover of the universe is called a knowledge structure induced by R, denoted by U/R or $K(R)$. If y belongs to the tolerance class determined by x with respect to R, we say two elements x and y are indistinguishable under the tolerance relation R, i.e., they are similar in R [14–17]. We denote the tolerance class of x by $S_R(x)$ [14,15,21]. Each tolerance class $S_R(x)(x \in R)$ is viewed as a knowledge granule [20,33,36]. The granulation structure induced by a tolerance relation is a cover of the universe. Conveniently, we say $K = (U, \mathbf{R})$ is also a knowledge base, where U is a finite and non-empty set and $\mathbf{R}$ is a family of tolerance relations. The following definition gives a rough set of a subset of the universe based on a tolerance relation.

Definition 2 ([14,15]). Let $K = (U, \mathbf{R})$ be a knowledge base, X a subset of U and $R \in \mathbf{R}$ a tolerance relation, two sets are defined as

$$\underline{R}X = \bigcup\{x \in X | S_R(x) \subseteq X\}, \tag{3}$$
$$\overline{R}X = \bigcup\{S_R(x) | x \in X\}, \tag{4}$$

where $\underline{R}X$ and $\overline{R}X$ are called R-lower approximation and R-upper approximation with respect to R, respectively. The order pair $\langle \underline{R}X, \overline{R}X \rangle$ is called a rough set of X with respect to the tolerance relation R.

Let $K = (U, \mathbf{R})$ be a knowledge base, if $R(R \in \mathbf{R})$ is a tolerance relation, then we denote a cover of U by $U/R = \{S_R(x) | x \in U\}$, i.e., $\forall x \in U$, one has $S_R(x) \neq \emptyset$ and $\bigcup_{x \in U} S_R(x) = U$. In [20], Liang et al. defined a partial relation $\preceq$ as follows: $P \preceq Q (P, Q \in \mathbf{R})$ if and only if, for every $i \in \{1, 2, \ldots, |U|\}$, one has that $S_P(x_i) \subseteq S_Q(x_i)$. Here, we also denote that P is finer than Q by $P \preceq Q$. It is easy to see that $(\mathbf{R}, \preceq)$ is also a poset.

3. Operators on a knowledge base

In this section, by uniformly representing a complete knowledge structure and an incomplete knowledge structure, we will propose four operators on a knowledge base and discuss their fundamental algebra properties.

In [20], Liang et al. established the relationship between a complete knowledge structure and an incomplete knowledge structure in the same knowledge base. Let $K = (U, R)$ be a knowledge structure, R a equivalence relation, $U/R = \{X_1, X_2, \ldots, X_m\}$, $U/R = \{S_R(x_1), S_R(x_2), \ldots, S_R(x_{|U|})\}$ and $X_i = \{x_{i1}, x_{i2}, \ldots, x_{is_i}\}$, where $|X_i| = s_i$ and $\sum_{i=1}^{m} s_i = |U|$, then

$$X_i = S_R(x_{i1}) = S_R(x_{i2}) = \ldots = S_R(x_{is_i}). \tag{5}$$

Through this mechanism, one can denote $U/R = \{E_R(x) | x \in U\}$ by using $U/R = \{S_R(x) | x \in U\}$. The mechanism gives uniform representations of knowledge structures in a knowledge base. It is illustrated by the following example.

Example 1. Let $U = \{x_1, x_2, \ldots, x_6\}$, R a equivalence relation and $U/R = \{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6\}\}$. Then, $U/R = \{S_R(x) | x \in U\}$ can be represented equivalently as

$$U/R = \{S_R(x_1), S_R(x_2), S_R(x_3), S_R(x_4), S_R(x_5), S_R(x_6)\} = \{\{x_1, x_2\}, \{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_3, x_4, x_5\}, \{x_3, x_4, x_5\}, \{x_6\}\}.$$

For convenience, we denote the knowledge structure induced by R on U as $K(R)$ in the rest of this paper, where R is an equivalence relation or a tolerance relation.

There are two types of operators to be considered in granular computing based on rough set theory. One is operations among knowledge granules, the other is operations among knowledge structures in a knowledge base. As operations among knowledge granules is based on classical sets, we still operate on them by $\cap$, $\cup$, $-$ and $\sim$, i.e., a new knowledge granule can be generated by $\cap$, $\cup$, $-$ and $\sim$ on known knowledge granules. However, operations among knowledge structures are performed through composing and decomposing known knowledge structures in knowledge bases in essence. Therefore, the operators on a knowledge base to generate new knowledge structures are very desirable. In the following, we introduce four operators among knowledge structures in a knowledge base.

Definition 3. Let $K = (U, \mathbf{R})$ be a knowledge base and $K(P), K(Q) \in K$ two knowledge structures. Four operators $\bigcap$, $\bigcup$, $-$ and $\wr$ on K are defined as

$$K(P) \bigcap K(Q) = \{S_{P \cap Q}(x) | S_{P \cap Q}(x) = S_P(x) \cap S_Q(x), x \in U\}, \tag{6}$$
$$K(P) \bigcup K(Q) = \{S_{P \cup Q}(x) | S_{P \cup Q}(x) = S_P(x) \cup S_Q(x), x \in U\}, \tag{7}$$
$$K(P) - K(Q) = \{S_{P-Q}(x) | S_{P-Q}(x) = x \cup (S_P(x) - S_Q(x)), x \in U\}, \tag{8}$$
$$\wr K(P) = \{\wr S_P(x) | \wr S_P(x) = x \cup \sim S_P(x), x \in U\}, \tag{9}$$

where $\sim S_P(x) = U - S_P(x)$.

Here, we regard $\bigcap$, $\bigcup$, $-$ and $\wr$ as four atomic formulas and finite connection on them are all formulas. Through using these operators, one can obtain a new knowledge structure via some known knowledge structures on U. Let $\mathbf{K}(U)$ denote the set of all knowledge structures on U, then these four operators $\bigcap$, $\bigcup$, $-$ and $\wr$ on $\mathbf{K}(U)$ are close. As follows, we investigate several fundamental algebra properties of these four operators.

Theorem 1. *Let $\bigcap$, $\bigcup$ be two operators on K, then*

(1) $K(P)\bigcap K(P) = K(P)$, $K(P)\bigcup K(P) = K(P)$;
(2) $K(P)\bigcap K(Q) = K(Q)\bigcap K(P)$, $K(P)\bigcup K(Q) = K(Q)\bigcup K(P)$;
(3) $K(P)\bigcap(K(P)\bigcup K(Q)) = K(P)$, $K(P)\bigcup(K(P)\bigcap K(Q)) = K(P)$; *and*
(4) $(K(P)\bigcap K(Q))\bigcap K(R) = K(P)\bigcap(K(Q)\bigcap K(R))$, $(K(P)\bigcup K(Q))\bigcup K(R) = K(P)\bigcup(K(Q)\bigcup K(R))$.

Proof. They are straightforward from Definition 3. □

Theorem 2. *Let $\bigcap$, $\bigcup$ and $\wr$ be three operators on K, then*

(1) $\wr(\wr K(P)) = K(P)$,
(2) $K(P)\bigcap \wr K(P) = \{x_i | x_i \in U\}$,
(3) $\wr(K(P)\bigcap K(Q)) = \wr K(P)\bigcup \wr K(Q)$, *and*
(4) $\wr(K(P)\bigcup K(Q)) = \wr K(P)\bigcap \wr K(Q)$.

Proof. For any $x_i \in U$, $K(P), K(Q) \in K$, $S_P(x_i)$ is the tolerance class induced by x_i in $K(P)$.

(1) From Definition 3, one can easily see that $\wr(S_P(x_i)) = x_i \cup \sim S_P(x_i)$ and $\wr(\wr(S_P(x_i))) = x_i \cup (x_i \cup S_P(x_i)) = S_P(x_i)$. Therefore, $\wr(\wr K(P)) = K(P)$.
(2) From Definition 3, it follows that $S_P(x_i) \cap \wr(S_P(x_i)) = x_i$, $\forall x_i \in U$. Then, $K(P)\bigcap \sim K(P) = \{x_i | x_i \in U\}$.
(3) According to Definition 3, for $\forall x_i \in U$, it follows that

$$\wr(S_P(x_i) \cap S_Q(x_i)) = x_i \cup \sim (S_P(x_i) \cap S_Q(x_i)) = x_i \cup (\sim S_P(x_i) \cup \sim S_Q(x_i)) = (x_i \cup \sim S_P(x_i)) \cup (x_i \cup \sim S_Q(x_i)) = \wr S_P(x_i) \cup \wr S_Q(x_i).$$

Therefore, one can get that $\wr(K(P)\bigcap K(Q)) = \wr K(P)\bigcup \wr K(Q)$.
(4) According to Definition 3, for $\forall x_i \in U$, one has that

$$\wr(S_P(x_i) \cup S_Q(x_i)) = x_i \cup \sim (S_P(x_i) \cup S_Q(x_i)) = x_i \cup (\sim S_P(x_i) \cap \sim S_Q(x_i)) = (x_i \cup \sim S_P(x_i)) \cap (x_i \cup \sim S_Q(x_i)) = \wr S_P(x_i) \cap \wr S_Q(x_i).$$

Hence, one can obtain that $\wr(K(P)\bigcup K(Q)) = \wr K(P)\bigcap \wr K(Q)$. □

Theorem 2 shows that (1) is reflexive, (2) is complementary, and (3) and (4) are two dual principles.

Theorem 3. *Let $\bigcap$, $\bigcup$, $-$ and $\wr$ be operators on K, then*

(1) $K(P) - K(Q) = K(P)\bigcap \wr K(Q)$,
(2) $K(P) - K(Q) = K(P) - (K(P)\bigcap K(Q))$,
(3) $K(P)\bigcap(K(Q) - K(R)) = (K(P)\bigcap K(Q)) - (K(P)\bigcap K(R))$, *and*
(4) $(K(P) - K(Q))\bigcup K(Q) = K(P)$.

Proof. They are straightforward from Definition 3. □

The above three theorems are illustrated by the following example.

Example 2. Let $U = \{x_1, x_2, x_3, x_4\}$, $K(P) = \{\{x_1, x_2\}, \{x_1, x_2\}, \{x_3, x_4\}, \{x_3, x_4\}\}$ and $K(Q) = \{\{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_4\}\}$, one can acquire some new knowledge structures through using $K(P)$ and $K(Q)$.

By computing, some new knowledge structures constructed are listed as follows:

$$\wr K(P) = \{\{x_1, x_3, x_4\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3\}, \{x_1, x_2, x_4\}\},$$
$$\wr K(Q) = \{\{x_1, x_2, x_3\}, \{x_1, x_2, x_4\}, \{x_1, x_3\}, \{x_1, x_2, x_3, x_4\}\},$$
$$K(P)\bigcap K(Q) = \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \{x_4\}\},$$
$$K(P)\bigcup K(Q) = \{\{x_1, x_2, x_4\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_4\}\},$$
$$\wr K(P)\bigcap \wr K(Q) = \{\{x_1, x_3\}, \{x_2, x_4\}, \{x_1, x_3\}, \{x_1, x_2, x_4\}\},$$
$$\wr K(P)\bigcup \wr K(Q) = \{\{x_1, x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}, \{x_1, x_2, x_3\}, \{x_1, x_2, x_3, x_4\}\},$$
$$K(P) - K(Q) = \{\{x_1, x_2\}, \{x_1, x_2\}, \{x_3\}, \{x_3, x_4\}\} \text{ and}$$
$$\wr K(P) - \wr K(Q) = \{\{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_3\}, \{x_4\}\}.$$

Suppose $K=(U,R)$ be a knowledge base, $P,Q\in R$, and $K(P)$, $K(Q)\in K$ be two knowledge structures induced by P, Q, respectively. To investigate properties of the operations among knowledge structures on a knowledge base, we will denote $K(P)\preceq K(Q)$ iff $P\preceq Q$.

Theorem 4. *Let $\bigcap$, $\bigcup$ and $\wr$ be three operators on K, the following properties hold:*

(1) *If $K(P)\preceq K(Q)$, then $\wr K(Q)\preceq \wr K(P)$;*
(2) *$K(P)\bigcap K(Q)\preceq K(P)$, $K(P)\bigcap K(Q)\preceq K(Q)$; and*
(3) *$K(P)\preceq K(P)\bigcup K(Q)$, $K(Q)\preceq K(P)\bigcup K(Q)$.*

Proof. The terms (2) and (3) can be easily proved from (6) and (7) in Definition 3, respectively.

From Definition 3, one can obtain that

$$K(P)\preceq K(Q)\Rightarrow \text{for } \forall x_i\in U, S_P(x_i)\subseteq S_Q(x_i)\Rightarrow \text{for } \forall x_i\in U, \sim S_Q(x_i)\subseteq \sim S_P(x_i)\Rightarrow \text{for } \forall x_i\in U, x_i\cup\sim S_Q(x_i)\subseteq x_i\cup\sim S_P(x_i)$$
$$\Rightarrow \wr K(Q)\preceq \wr K(P).$$

Hence, the term (1) in this theorem holds. □

Definition 4 [57]. Let $(L,\leqslant)$ be a poset, if there exist two operators $\wedge,\vee$ on $L:L^2\to L$ such that

(1) $a\wedge b=b\wedge a$, $a\vee b=b\vee a$,
(2) $(a\wedge b)\wedge c=a\wedge(b\wedge c)$, $(a\vee b)\vee c=a\vee(b\vee c)$, and
(3) $a\wedge b=b\iff b\leqslant a$, $a\vee b=b\iff a\leqslant b$.Then we call L is a lattice.
Furthermore, if
(4) $a\wedge(b\vee c)=(a\wedge b)\vee(a\wedge c)$, $a\vee(b\wedge c)=(a\vee b)\wedge(a\vee c)$.Then we call L is an assignment lattice.

We call L a complemented lattice, if for any $a\in L$, there exists a' such that $(a')'=a$ and $a\leqslant b\iff b'\leqslant a'$. If there exist $0,1\in L$ such that $0\leqslant a\leqslant 1$ for any $a\in L$, then we call 0 and 1 its minimal element and maximal element, respectively.

Theorem 5. *$(K,\bigcup,\bigcap)$ is an assignment lattice.*

Proof. At first, we prove $(K,\preceq)$ is a lattice.

From (2) and (4) in Theorem 1, the terms (1) and (2) in Definition 4 are obvious.

Let $K(P),K(Q),K(R)\in K$ be three knowledge structures, where $K(P)=\{S_P(x)|x\in U\}$, $K(Q)=\{S_Q(x)|x\in U\}$ and $K(R)=\{S_R(x)|x\in U\}$. One can obtain that

$$K(P)\bigcap K(Q)=K(P)\iff \text{for } \forall x_i\in U,\quad S_{P\cap Q}(x_i)=S_P(x_i),\quad i\leqslant|U|\iff S_P(x_i)\cap S_Q(x_i)$$
$$=S_P(x_i)\iff S_P(x_i)\subseteq S_Q(x_i)\quad \text{for } \forall x_i\in U\iff K(P)\preceq K(Q).$$

According to the dual principle a in lattice, one can easily get that $K(P)\bigcup K(Q)=K(P)\iff K(Q)\preceq K(P)$. Thus, the term (3) in Definition 4 holds.

In addition, for $K(P),K(Q),K(R)\in K$, we know that

$$S_P(x_i)\cap(S_Q(x_i)\cup S_R(x_i))=(S_P(x_i)\cap(S_Q(x_i))\cup(S_P(x_i)\cap S_R(x_i))\quad \forall x_i(i\leqslant|U|).$$

Hence, $K(P)\bigcap(K(Q)\bigcup K(R))=(K(P)\bigcap K(Q))\bigcup(K(P)\bigcap K(R))$. From the dual principle in a lattice, one can get that

$$K(P)\bigcup(K(Q)\bigcap K(R))=(K(P)\bigcup K(Q))\bigcap(K(P)\bigcup K(R)).$$

Therefore, $(K,\bigcup,\bigcap)$ is an assignment lattice. □

Theorem 6. *Let $\mathbf{K}(U)$ be the set of all knowledge structures on U, then $(\mathbf{K}(U),\bigcup,\bigcap,\wr)$ is a complemented lattice.*

Proof. From Theorem 5, it is obvious that $(\mathbf{K}(U),\bigcup,\bigcap,\wr)$ is an assignment lattice. Furthermore, from (1) in Theorem 2, one can get that $\wr(\wr K(P))=K(P)$. In addition, from (3) in Definition 3, one has that

$$K(P)\preceq K(Q)\iff \quad \text{for } \forall x_i\in U,\quad S_P(x_i)\subseteq S_Q(x_i)\iff \quad \text{for } \forall x_i\in U,\sim S_P(x_i)\supseteq\sim S_Q(x_i)\iff \text{for } \forall x_i\in U, x_i\cup\sim S_P(x_i)$$
$$\supseteq x_i\cup\sim S_Q(x_i)\iff \text{for } \forall x_i\in U, \wr S_P(x_i)\supseteq \wr S_Q(x_i)\iff K(Q)\preceq K(P).$$

Hence, $(\mathbf{K}(U),\bigcup,\bigcap,\wr)$ is a complemented lattice. □

In a complemented lattice $(\mathbf{K}(U),\bigcup,\bigcap,\wr)$, the knowledge structure $K(\omega)=\{x_i|x_i\in U\}$ and the knowledge structure $K(\delta)=\{S_P(x_i)|S_P(x_i)=U,x_i\in U\}$ are two special knowledge structures, where $K(\omega)$ is the discrete classification and $K(\delta)$ is the indiscrete classification. For any $K(P) \in \mathbf{K}(U)$, one has that $K(\omega) \preceq K(P) \preceq K(\delta)$. Then, we can call $K(\omega)$ and $K(\delta)$ the minimal element and the maximal element on the lattice $(\mathbf{K}(U), \bigcup, \bigcap, \wr)$, respectively.

Remark. One of the strengths of rough set theory is the fact that an unknown target concept can be characterized approximately by existing knowledge structures in a knowledge base. From the above analyses, it is shown that these four operators ($\bigcup$, $\bigcap$, $\wr$ and $-$) can be applied to generate new knowledge structures on a knowledge base. That is to say, one can use these new knowledge structures to approximate an unknown target. Therefore, this mechanism may be used to rule extraction and knowledge discovery from knowledge bases.

4. Knowledge granulation

As we know, knowledge granulation, in a broad sense, is the average measure of knowledge granules of a knowledge structure in a given knowledge base. It can be used to characterize the classification ability of a given knowledge structure [19,20,33,50,55].

In recent years, some researchers have discussed and investigated that how to measure the classification ability of a knowledge structure and what is the essence of knowledge granulation in knowledge bases. Generally, the partial relation "$\preceq$" are concerned for investigating various definitions of knowledge granulation. However, the partial relation "$\preceq$" may be not strict in terms of characterizing the properties of knowledge granulation in knowledge bases. In order to discover the essence of knowledge granulation, we introduce a new binary relation "$\underline{\ll}$" on $\mathbf{K}(U)$ in the following.

Let $K = (U, \mathbf{R})$ be a knowledge base, $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x) | x \in U\}$ and $K(Q) = \{S_Q(x) | x \in U\}$. We define a binary relation $\underline{\ll}$ as follows: $K(P) \underline{\ll} K(Q)$ $(P, Q \in \mathbf{R})$ if and only if, for $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$, there exists a sequence $K'(Q)$ of $K(Q)$, where $K'(Q) = \{S_Q(x'_1), S_Q(x'_2), \ldots, S_Q(x'_{|U|})\}$, such that $|S_P(x_i)| \leqslant |S_Q(x'_i)|$, $i \leqslant |U|$.

The following theorem gives some properties of the binary relation $\underline{\ll}$.

Theorem 7. *The following properties hold*

(1) $\underline{\ll}$ *is reflexive,*
(2) $\underline{\ll}$ *is transitive,*
(3) $K(P) \bigcap K(Q) \underline{\ll} K(P)$, $K(P) \bigcap K(Q) \underline{\ll} K(Q)$, *and*
(4) $K(P) \underline{\ll} K(P) \bigcup K(Q)$, $K(Q) \underline{\ll} K(P) \bigcup K(Q)$.

Proof. They are straightforward.

Here, we say that $K(P)$ is granulation finer than $K(Q)$ if $K(P) \underline{\ll} K(Q)$. If $K(P) \underline{\ll} K(Q)$ and there exists $S_Q(x'_i) \in K'(Q)$ such that $|S_P(x_i)| < |S_Q(x'_i)|$, we say that $K(Q)$ is strictly granulation coarser than $K(P)$ (or $K(P)$ is strictly granulation finer than $K(Q)$), denoted by $K(P) \ll K(Q)$. □

Theorem 8. *The partial relation $\preceq$ is a special instance of the relation $\underline{\ll}$.*

Proof. Let $K = (U, \mathbf{R})$ be a knowledge base, $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x) | x \in U\}$ and $K(Q) = \{S_Q(x) | x \in U\}$. If $K(P) \preceq K(Q)$, one can obtain that $S_P(x_i) \subseteq S_Q(x_i)$ for any $x_i \in U$, i.e., $|S_P(x_i)| \leqslant |S_Q(x_i)|$. That is to say, one can find an array of all tolerance classes in $K(Q)$ such that $K(P) \underline{\ll} K(Q)$. Therefore, the partial relation $\preceq$ is a special instance of the relation $\underline{\ll}$. □

Definition 5. Let $K = (U, \mathbf{R})$ be a knowledge base, if for $\forall P \in \mathbf{R}$, there is a real number $G(P)$ with the following properties:

(1) $G(P) \geqslant 0$ (non-negative);
(2) for $\forall P, Q \in \mathbf{R}$, let $K(P) = \{S_P(x_i) | x_i \in U\}$ and $K(Q) = \{S_Q(x_i) | x_i \in U\}$, if there is a bijective mapping function $f : K(P) \to K(Q)$ such that $|S_P(x_i)| = |f(S_P(x_i))|$, then $G(P) = G(Q)$ (invariability);
(3) if $\forall P, Q \in \mathbf{R}$ and $K(P) \ll K(Q)$, then $G(P) < G(Q)$ (monotonicity).

Then G is called a knowledge granulation on K.

As a result of the above discussions, we come to the following three theorems.

Theorem 9 (Extremum). *Let $K = (U, \mathbf{R})$ be a knowledge base and $\forall K(P) \in K$, then $G(P)$ achieves its minimum value if $U/P = \omega$ and $G(P)$ achieves its maximum value if $U/P = \delta$, where ω denotes the identity relation and δ denotes the universal relation.*

Proof. Let $K(\omega) = \{S_\omega(x_i) | S_\omega(x_i) = x_i, x_i \in U\}$ and $K(\delta) = \{S_\delta(x_i) | S_\delta(x_i) = U, x_i \in U\}$. Hence, for $\forall R \in \mathbf{R}$, $K(R) = \{S_R(x_i) | x_i \in U\}$, one has that $x_i \subseteq S_R(x_i)$ and $1 \leqslant |S_R(x_i)|$, i.e., $K(\omega) \underline{\ll} K(R)$, and $S_R(x_i) \subseteq U$ and $|S_R(x_i)| \leqslant |U|$, i.e., $K(R) \underline{\ll} K(\delta)$. Therefore, $K(\omega) \ll K(R) \ll K(\delta)$ $\forall R \in \mathbf{R}$. From (3) in Definition 5, one can get that $G(\omega) \leqslant G(R) \leqslant G(\delta)$, i.e., $G(P)$ achieves its minimum value if $U/P = \omega$ (identity relation) and $G(P)$ achieves its maximum value if $U/P = \delta$ (universal relation). □

From Definition 5 and Theorem 9, it is easy to see that the size of $G(P)$ only depends on the cardinality of every class in the knowledge structure $K(P)$. The minimum value of $G(P)$ can be obtained when $P = \omega$ and the maximum value of $G(P)$ can be approached when $P = \delta$.

Theorem 10. *Let $K = (U, \mathbf{R})$ be a knowledge base and $K(P), K(Q)$ two knowledge structures on K, then $G(P) \leqslant G(Q)$ if $K(P) \preceq K(Q)$.*

Proof. From the definition of $\preceq$, one can see that $K(P) \preceq K(Q)$ $(P, Q \in \mathbf{R})$ if and only if $S_P(x_i) \subseteq S_Q(x_i)$ for every $i \in \{1, 2, \ldots, |U|\}$. Hence, for every $S_P(x_i) \in K(P)$, there exists $S_Q(x_i) \in K(Q)$ such that $|S_P(x_i)| \leqslant |S_Q(x_i)|$, i.e., $K(P) \lll K(Q)$. Therefore, one can easily obtain that $G(P) \leqslant G(Q)$ from (3) in Definition 5. □

Theorem 11. *The following properties hold:*

(1) $G(P) = G(\wr\wr P)$;
(2) $G(P \bigcap Q) \leqslant G(P)$, $G(P \bigcap Q) \leqslant G(Q)$;
(3) $G(P) \leqslant G(P \bigcup Q)$, $G(Q) \leqslant G(P \bigcup Q)$; *and*
(4) $G(P \bigcap \wr P) = G(\omega)$, $G(P \bigcup \wr P) = G(\delta)$.

Proof. They are straightforward. □

In [19,20,33], several different kinds of knowledge granulations have been given, in the following, we prove that these knowledge granulations are all special forms under Definition 6.

Definition 6 [19]. Let $K = (U, \mathbf{R})$ be a knowledge base, $R \in \mathbf{R}$ and $U/R = \{X_1, X_2, \ldots, X_m\}$. Knowledge granulation of the knowledge structure $K(R)$ is defined as

$$GK(R) = \frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2, \quad \frac{1}{|U|} \leqslant GK(R) \leqslant 1, \tag{10}$$

where $\sum_{i=1}^{m}|X_i|^2$ is the cardinality of the equivalence relation $\bigcup_{i=1}^{m}(X_i \times X_i)$ determined by R.

Theorem 12. *GK in Definition 6 is a knowledge granulation under Definition 5.*

Proof

(1) Obviously, it is non-negative.
(2) Let $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$. Supposing that there be a bijective mapping function $f : K(P) \to K(Q)$ such that $|S_P(x_i)| = |f(S_Q(x_i))|$ and $f(S_P(x_i)) = S_Q(x_{j_i})$, $j_i \in \{1, 2, \ldots, |U|\}$, then one has that

$$GK(P) = \frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2 = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(x_i)| = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x_{j_i})| = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x_i)| = GK(Q).$$

(3) Let $P, Q \in \mathbf{R}$ with $K(P) \ll K(Q)$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$, then there exists a sequence $K'(Q)$ of $K(Q)$, where $K'(Q) = \{S_Q(x'_1), S_Q(x'_2), \ldots, S_Q(x'_{|U|})\}$, such that $|S_P(x_i)| < |S_Q(x'_i)|$. Hence, one has that

$$GK(P) = \frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2 = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(x_i)| < \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x'_i)| = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x_i)| = GK(Q).$$

Therefore, GK in Definition 6 is a knowledge granulation under Definition 5. □

Definition 7 [20]. Let $K = (U, \mathbf{R})$ be a knowledge base, $P \in \mathbf{R}$ and $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$. Knowledge granulation of the knowledge structure $K(P)$ is defined as

$$G(P) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i)|}{|U|}, \tag{11}$$

where $\frac{|S_P(x_i)|}{|U|}$ is the probability of tolerance class $S_P(x_i)$ within the universe U.

If $K(P) = K(\omega)$, $G(P)$ achieves its minimum value $G(P) = \frac{1}{|U|}$; if $K(P) = K(\delta)$, $G(P)$ achieves its maximum value $G(P) = 1$. It is obvious that $\frac{1}{|U|} \leqslant G(P) \leqslant 1$.

Theorem 13. *G in Definition 7 is a knowledge granulation under Definition 5.*

Proof

(1) Obviously, it is non-negative.
(2) Let $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$. Supposing that there be a bijective mapping function $f : K(P) \to K(Q)$ such that $|S_P(x_i)| = |f(S_Q(x_i))|$ and $f(S_P(x_i)) = S_Q(x_{j_i})$, $j_i \in \{1, 2, \ldots, |U|\}$, then one has that

$$G(P) = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(x_i)| = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x_{j_i})| = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x_i)| = G(Q).$$

(3) Let $P, Q \in \mathbf{R}$ with $K(P) \ll K(Q)$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$, then there exists a sequence $K'(Q)$ of $K(Q)$, where $K'(Q) = \{S_Q(x'_1), S_Q(x'_2), \ldots, S_Q(x'_{|U|})\}$, such that $|S_P(x_i)| < |S_Q(x'_i)|$. Hence, one has that

$$G(P) = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(x_i)| < \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x'_i)| = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(x_i)| = G(Q).$$

Thus, G in Definition 7 is a knowledge granulation under Definition 5. □

Theorem 14. *Let $K = (U, \mathbf{R})$ be a knowledge base, $P \in \mathbf{R}$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $\wr P$ the relation induced by $\wr K(P)$, then $G(P) + G(\wr P) = 1 + \frac{1}{|U|}$.*

Proof. From Definition 7, it follows that

$$G(P) + G(\wr P) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i)|}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|\wr S_P(x_i)|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i)|}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|x_i \cup \sim S_P(x_i)|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|U|+1}{|U|} = 1 + \frac{1}{|U|}.$$

That is $G(P) + G(\wr P) = 1 + \frac{1}{|U|}$. □

Theorem 15. *Let $\mathbf{K}(U)$ be the set of all knowledge structures on U and $K(P), K(Q) \in \mathbf{K}(U)$ two knowledge structures, then $G(P) - G(Q) = G(\wr(Q) - G(\wr P)$.*

Proof. Obviously, we have that

$$G(\wr P) - G(\wr Q) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{1 + |U| - |S_P(x_i)|}{|U|} - \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{1 + |U| - |S_Q(x_i)|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_Q(x_i)| - |S_P(x_i)|}{|U|} = -(G(P) - G(Q)),$$

i.e., $G(P) - G(Q) = G(\wr Q) - G(\wr P)$. □

Definition 8 [33]. Let $K = (U, \mathbf{R})$ be a knowledge base, $R \in \mathbf{R}$ and $U/R = \{X_1, X_2, \ldots, X_m\}$. Combination granulation of knowledge structure $K(R)$ is defined as

$$CG(R) = \sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}}, \tag{12}$$

where $0 \leqslant CG(R) \leqslant 1$, $\frac{|X_i|}{|U|}$ represents the probability of equivalence class X_i within the universe U, and $\frac{C^2_{|X_i|}}{C^2_{|U|}}$ denotes the probability of pairs of elements on equivalence class X_i within the whole pairs of elements on the universe U.

Theorem 16. *CG in Definition 8 is a knowledge granulation under Definition 5.*

Proof

(1) Obviously, it is non-negative.
(2) Let $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$. Supposing that there be a bijective mapping function $f : K(P) \to K(Q)$ such that $|S_P(x_i)| = |f(S_Q(x_i))|$ and $f(S_P(x_i)) = S_Q(x_{j_i})$, $j_i \in \{1, 2, \ldots, |U|\}$. Then,

$$CG(P) = \sum_{i=1}^{m}\frac{|P_i|}{|U|}\frac{C^2_{|P_i|}}{C^2_{|U|}} = \sum_{i=1}^{|U|}\frac{|S_P(u_i)|}{|U|}\frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}} = \sum_{i=1}^{|U|}\frac{|S_Q(u_{j_i})|}{|U|}\frac{C^2_{|S_Q(u_{j_i})|}}{C^2_{|U|}} = \sum_{i=1}^{|U|}\frac{|S_Q(u_i)|}{|U|}\frac{C^2_{|S_Q(u_i)|}}{C^2_{|U|}} = \sum_{j=1}^{n}\frac{|Q_j|}{|U|}\frac{C^2_{|Q_j|}}{C^2_{|U|}} = CG(Q).$$

(3) Let $P, Q \in \mathbf{R}$ with $P \ll Q$, $K(P) = \{S_P(u_1), S_P(u_2), \ldots, S_P(u_{|U|})\}$ and $K(Q) = \{S_Q(u_1), S_Q(u_2), \ldots, S_Q(u_{|U|})\}$, then for arbitrary $S_P(u_i)(i \leqslant |U|)$, there exists a sequence $\{S_Q(u'_1), S_Q(u'_2), \ldots, S_Q(u'_{|U|})\}$ such that $|S_P(u_i)| < |S_Q(u'_i)|$ $(i = 1, 2, \ldots, |U|)$. Therefore,

$$CG(P) = \sum_{i=1}^{m} \frac{|P_i|}{|U|} \frac{C^2_{|P_i|}}{C^2_{|U|}} = \sum_{i=1}^{|U|} \frac{|S_P(u_i)|}{|U|} \frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}} < \sum_{i=1}^{|U|} \frac{S_Q(u'_i)}{|U|} \frac{C^2_{|S_Q(u'_i)|}}{C^2_{|U|}} = \sum_{i=1}^{|U|} \frac{S_Q(u_i)}{|U|} \frac{C^2_{|S_Q(u_i)|}}{C^2_{|U|}} = \sum_{j=1}^{n} \frac{|Q_j|}{|U|} \frac{C^2_{|Q_j|}}{C^2_{|U|}} = CG(Q).$$

Hence, CG in Definition 8 is a knowledge granulation under Definition 5. □

Definition 9 [33]. Let $K = (U, \mathbf{R})$ be a knowledge base, $P \in \mathbf{R}$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$. Combination granulation of knowledge structure $K(P)$ is defined by

$$CG(P) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C^2_{|S_P(x_i)|}}{C^2_{|U|}}, \tag{13}$$

where $\frac{C^2_{|S_P(x_i)|}}{C^2_{|U|}}$ denotes the probability of pairs of elements on tolerance class $S_P(x_i)$ within the whole pairs of elements on the universe U.

Theorem 17. *CG in Definition 9 is a knowledge granulation under Definition 5.*

Proof

(1) Obviously, it is non-negative.
(2) Let $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$. Supposing that there be a bijective mapping function $f : K(P) \to K(Q)$ such that $|S_P(x_i)| = |f(S_Q(x_i))|$ and $f(S_P(x_i)) = S_Q(x_{j_i}), j_i \in \{1, 2, \ldots, |U|\}$, one has that

$$CG(P) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}} = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C^2_{|S_Q(u_{j_i})|}}{C^2_{|U|}} = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C^2_{|S_Q(x_i)|}}{C^2_{|U|}} = CG(Q).$$

(3) Let $P, Q \in \mathbf{R}$ with $P \ll Q$, $K(P) = \{S_P(u_1), S_P(u_2), \ldots, S_P(u_{|U|})\}$ and $K(Q) = \{S_Q(u_1), S_Q(u_2), \ldots, S_Q(u_{|U|})\}$, then for any $S_P(u_i)(i \leqslant |U|)$, there exists a sequence $\{S_Q(u'_1), S_Q(u'_2), \ldots, S_Q(u'_{|U|})\}$ such that $|S_P(u_i)| < |S_Q(u'_i)|$ $(i = 1, 2, \ldots, |U|)$. Therefore,

$$CG(P) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}} < \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C^2_{|S_Q(u'_i)|}}{C^2_{|U|}} = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C^2_{|S_Q(u_i)|}}{C^2_{|U|}} = CG(Q).$$

Thus, CG in Definition 9 is a knowledge granulation under Definition 5. □

Through using the axiom definition of knowledge granulation, one can construct some new knowledge granulations according to various opinions. In the following, we show the significance of the axiom definition of knowledge granulation by constructing a new form of knowledge granulation in a given knowledge base.

Definition 10. Let $K = (U, \mathbf{R})$ be a knowledge base, $P \in \mathbf{R}$ and $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$. Knowledge granulation of the knowledge structure $K(P)$ is defined as

$$GE(P) = \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 |S_P(x_i)|. \tag{14}$$

If $K(P) = K(\omega)$, $G(P)$ achieves its minimum value $G(P) = 0$; if $K(P) = K(\delta)$, $G(P)$ achieves its maximum value $G(P) = \log_2 |U|$. It is obvious that $0 \leqslant GE(P) \leqslant \log_2 |U|$.

Theorem 18. *GE in Definition 10 is a knowledge granulation under Definition 5.*

Proof

(1) Obviously, it is non-negative.
(2) Let $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$. Supposing that there be a bijective mapping function $f : K(P) \to K(Q)$ such that $|S_P(x_i)| = |f(S_Q(x_i))|$ and $f(S_P(x_i)) = S_Q(x_{j_i})$, $j_i \in \{1, 2, \ldots, |U|\}$. Hence,

$$GE(P) = \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 |S_P(x_i)| = \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 |S_Q(x_{j_i})| = \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 |S_Q(x_i)| = GE(Q).$$

(3) Let $P, Q \in \mathbf{R}$ with $K(P) \ll K(Q)$, $K(P) = \{S_P(x_1), S_P(x_2), \ldots, S_P(x_{|U|})\}$ and $K(Q) = \{S_Q(x_1), S_Q(x_2), \ldots, S_Q(x_{|U|})\}$, then there exists a sequence $K'(Q)$ of $K(Q)$, where $K'(Q) = \{S_Q(x'_1), S_Q(x'_2), \ldots, S_Q(x'_{|U|})\}$, such that $|S_P(x_i)| < |S_Q(x'_i)|$. Hence,

$$GE(P) = \frac{1}{|U|}\sum_{i=1}^{|U|}\log_2|S_P(x_i)| \leqslant \frac{1}{|U|}\sum_{i=1}^{|U|}\log_2|S_Q(x'_i)| = \frac{1}{|U|}\sum_{i=1}^{|U|}\log_2|S_Q(x_i)| = GE(Q).$$

Therefore, GE in Definition 10 is a knowledge granulation under Definition 5. □

5. Knowledge distance

In rough set theory, information entropy and knowledge granulation are two main approaches to measuring the uncertainty of a knowledge structure in knowledge bases. If the knowledge granulation (or information entropy) of one knowledge structure is equal to that of the other knowledge structure, we say that these two knowledge structures have the same uncertainty. However, it does not mean that these two knowledge structures are equivalent each other. In other words, information entropy and knowledge granulation cannot characterize the difference between any two knowledge structures in a knowledge base. In this section, we introduce a notion of knowledge distance to differentiate two given knowledge structures and investigate some of its important properties.

In [50], Yao presented the concept of set closeness between two classical sets to measure the degree of the sameness between sets. Let A and B be two finite sets, the measure is defined by $H(A,B) = \frac{|A\cap B|}{|A\cup B|}$ $(A \cup B \neq \emptyset)$ [50]. Obviously, the formula $1 - H(A,B) = 1 - \frac{|A\cap B|}{|A\cup B|}$ $(A \cup B \neq \emptyset)$ can characterize the difference between two finite classical sets. To characterize the relationship among knowledge structures, based on the view of set closeness, we introduce an approach called knowledge distance for measuring the difference between two knowledge structures on the same knowledge base in the following.

Definition 11. Let $K = (U, \mathbf{R})$ be a knowledge base, $P, Q \in \mathbf{R}$, $K(P) = \{S_P(x_i)|x_i \in U\}$ and $K(Q) = \{S_Q(x_i)|x_i \in U\}$. Knowledge distance between $K(P)$ and $K(Q)$ is defined as

$$D(K(P), K(Q)) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_Q(x_i)|}{|U|}, \tag{15}$$

where $|S_P(x_i) \oplus S_Q(x_i)| = |S_P(x_i) \cup S_Q(x_i)| - |S_P(x_i) \cap S_Q(x_i)|$.

The knowledge distance represents the measure of difference between two knowledge structures in the same knowledge base. Obviously, $0 \leqslant D(K(P), K(Q)) \leqslant 1 - \frac{1}{|U|}$.

Theorem 19 (Extremum). *Let $K(U, \mathbf{R})$ be a knowledge base, $K(P)$, $K(Q)$ two knowledge structures on K. Then, $D(K(P), K(Q))$ achieves its minimum value $D(K(P), K(Q)) = 0$ iff $K(P) = K(Q)$ and $D(K(P), K(Q))$ achieves its maximum value $D(K(P), K(Q)) = 1 - \frac{1}{|U|}$ iff $K(P) = \wr K(Q)(\iff K(Q) = \wr K(P))$.*

Proof. For $\forall P, Q \in \mathbf{R}$, one has that $1 \leqslant |S_P(x_i) \cap S_P(x_i)| \leqslant |U|$ and $1 \leqslant |S_P(x_i) \cup S_P(x_i)| \leqslant |U|$. Therefore, $\forall P, Q \in \mathbf{R}$, $0 \leqslant |S_P(x_i) \oplus S_Q(x_i)| \leqslant |U| - 1$, i.e., $0 \leqslant \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i)\oplus S_Q(x_i)|}{|U|} \leqslant 1 - \frac{1}{|U|}$.

If $K(P) = K(Q)$, then $K(P) \bigcap K(Q) = K(P)$ and $K(P) \bigcup K(Q) = K(P)$. Hence,

$$D(K(P), K(Q)) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_Q(x_i)|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{0}{|U|} = 0,$$

i.e., $D(K(P), K(Q))$ achieves its minimum value 0 if and only if $K(P) = K(Q)$.

If $K(P) = \wr K(Q)$, then $K(P) \bigcap K(Q) = K(\omega)$ and $K(P) \bigcup K(Q) = K(\delta)$. Hence,

$$D(K(P), K(Q)) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_Q(x_i)|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|U| - |x_i|}{|U|} = 1 - \frac{1}{|U|},$$

i.e., $D(K(P), K(Q))$ achieves its maximum value $1 - \frac{1}{|U|}$ if and only if $K(P) = \wr K(Q)$ or $K(Q) = \wr K(P)$. □

In particular, one has that $D(K(\omega), K(\delta)) = 1 - \frac{1}{|U|}$. In fact, we have that $K(\omega) = \wr K(\delta)$ and $K(\delta) = \wr K(\omega)$.

Let $K(P) = \{S_P(x_i)|x_i \in U\}$, $K(Q) = \{S_Q(x_i)|x_i \in U\}$ and $K(R) = \{S_R(x_i)|x_i \in U\}$ be three knowledge structures on U. For $S_P(x_i) \in K(P)$, $S_Q(x_i) \in K(Q)$ and $S_R(x_i) \in K(Q)$, $x_i \in U$, we note $S_{P\cup Q\cup R}(x_i) = S_P(x_i) \cup S_Q(x_i) \cup S_R(x_i)$. One can give a certain array of all elements in $S_{P\cup Q\cup R}(x_i)$ and denote the array by $Array = (x_{i_1}, x_{i_2}, \ldots, x_{i_{|S_{P\cup Q\cup R}(x_i)|}})$. Therefore, one can represent $S_P(x_i)$ by the following array

$$x_{i_k} = \begin{cases} 1 & \text{if } x_{i_k} \in S_P(x_i), \\ 0 & \text{else}, \end{cases}$$

for $x_{i_k} \in Array$, $k \leqslant |S_{P\cup Q\cup R}(x_i)|$.

Similarly, the expressions of $S_Q(x_i)$ and $S_R(x_i)$ can also be obtained. In fact, the expression of *Array* is various, so the expression of $S_P(x_i)$, $S_Q(x_i)$ and $S_R(x_i)$ should also be changed according to *Array*, respectively. This kind of representations about the tolerance classes is illustrated by the following example.

Example 3. Consider three tolerance classes $S_P(x_i) = \{1,2,3\}$, $S_Q(x_i) = \{2,3,4\}$ and $S_R(x_i) = \{3,4,5\}$. Compute the expressions of $S_P(x_i)$, $S_Q(x_i)$ and $S_R(x_i)$ through using the above method.

By computing, one has that $S_{P\cup Q\cup R}(x_i) = S_P(x_i) \cup S_Q(x_i) \cup S_R(x_i) = \{1,2,3,4,5\}$. Assume that $Array = (1,2,3,4,5)$.

For $S_P(x_i)$, one can obtain that $S_P(x_i) = (1,1,1,0,0)$. Similarly, it follows that $S_Q(x_i) = (0,1,1,1,0)$ and $S_R(x_i) = (0,0,1,1,1)$.

Let A, B, C be three classical sets, $Array = (t_1, t_2, \ldots, t_{|A\cup B\cup C|})$, $t_i \cap t_j = \emptyset$, $t_i, t_j \in A \cup B \cup C$. Hence, from the above expression method, one can get the array expressions of A, B and C as follows:

$$A' = \{a_1, a_2, \ldots, a_{|A\cup B\cup C|}\},$$
$$B' = \{b_1, b_2, \ldots, b_{|A\cup B\cup C|}\}, \quad \text{and}$$
$$C' = \{c_1, c_2, \ldots, c_{|A\cup B\cup C|}\}.$$

Based on these denotations, we then measure the distance between two classical sets by the following formula

$$d(A,B) = \sum_{i=1}^{|A\cup B\cup C|} (a_i + b_i), \quad a_i \in A', \quad b_i \in B'. \tag{16}$$

Analogously, one has that $d(B,C) = \sum_{i=1}^{|A\cup B\cup C|}(b_i + c_i)$ and $d(A,C) = \sum_{i=1}^{|A\cup B\cup C|}(a_i + c_i)$.

From these denotations, we come to the following lemma.

Lemma 1. *Let A, B, C be three classical sets, then* $d(A,B) + d(B,C) \geqslant d(A,C)$, $d(A,B) + d(A,C) \geqslant d(B,C)$ *and* $d(A,C) + d(B,C) \geqslant d(A,B)$.

Proof. Suppose that $A' = \{a_1, a_2, \ldots, a_{|A\cup B\cup C|}\}$, $B' = \{b_1, b_2, \ldots, b_{|A\cup B\cup C|}\}$ and $C' = \{c_1, c_2, \ldots, c_{|A\cup B\cup C|}\}$. From $(a_i + b_i) + (b_i + c_i) \geqslant (a_i + c_i)$, it follows that

$$d(A,B) + d(B,C) = \sum_{i=1}^{|A\cup B\cup C|}(a_i + b_i) + \sum_{i=1}^{|A\cup B\cup C|}(b_i + c_i) = \sum_{i=1}^{|A\cup B\cup C|}((a_i + b_i) + (b_i + c_i)) \geqslant \sum_{i=1}^{|A\cup B\cup C|}(a_i + c_i) = d(B,C).$$

Similarly, $d(A,B) + d(A,C) \geqslant d(B,C)$ and $d(A,C) + d(B,C) \geqslant d(A,B)$. □

Theorem 20. *Let* $\mathbf{K}(U)$ *be the set of all knowledge structures induced by U, then* $(\mathbf{K}(U), D)$ *is a distance space.*

Proof

(1) One can obtain easily that $D(K(P), K(Q)) \geqslant 0$ from Definition 7.
(2) It is obvious that $D(K(P), K(Q)) = D(K(Q), K(P))$.
(3) For the proof of the triangle inequality principle, one only need to prove that $D(K(P), K(Q)) + D(K(P), K(R)) \geqslant D(K(Q), K(R))$, $D(K(R), K(Q)) + D(K(P), K(R)) \geqslant D(K(Q), K(P))$ and $D(K(R), K(Q)) + D(K(P), K(Q)) \geqslant D(K(P), K(R))$ for any $K(P)$, $K(Q)$, $K(R) \in \mathbf{K}(U)$.From Lemma 1, we know that for $x_i \in U$, $D(S_P(x_i), S_Q(x_i)) + D(S_P(x_i), S_R(x_i)) \geqslant D(S_Q(x_i), S_R(x_i))$, $D(S_P(x_i), S_Q(x_i)) + D(S_Q(x_i), S_R(x_i)) \geqslant D(S_P(x_i), S_R(x_i))$ and $D(S_P(x_i), S_R(x_i)) + D(S_Q(x_i), S_R(x_i)) \geqslant D(S_P(x_i), S_Q(x_i))$. Hence,

$$\begin{aligned} D(K(P), K(Q)) + D(K(P), K(R)) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_Q(x_i)|}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_R(x_i)|}{|U|} \\ &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{d(S_P(x_i), S_Q(x_i))}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{d(S_P(x_i), S_R(x_i))}{|U|} \\ &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{1}{|U|}(d(S_P(x_i), S_Q(x_i)) + d(S_P(x_i), S_R(x_i))) \geqslant \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{d(S_Q(x_i), S_R(x_i))}{|U|} \\ &= \frac{1}{|U|}\sum_{i=1}^{|U|} D(K(Q), K(R)). \end{aligned}$$

Similarly, one can obtain that $D(K(R), K(Q)) + D(K(P), K(R)) \geqslant D(K(Q), K(P))$, $D(K(R), K(Q)) + D(K(P), K(Q)) \geqslant D(K(R), K(P))$.

Therefore, $(\mathbf{K}(U), D)$ is a distance space. □

The above theorem is explained by the following example.

Example 4. Assume that $K = \{x_1, x_2, x_3, x_4, x_5\}$ and $K(P), K(Q), K(R)$ be three knowledge structures induced by equivalence relations P, Q, R on K, where $K(P) = \{\{x_1, x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_4, x_5\}, \{x_4, x_5\}\}$, $K(Q) = \{\{x_1,\ x_2\}, \{x_1, x_2\}, \{x_3\}, \{x_4, x_5\}, \{x_4, x_5\}\}$ and $K(R) = \{\{x_1, x_2\}, \{x_1, x_3\}, \{x_3\}, \{x_4\}, \{x_5\}\}$.

By computing their knowledge distances, one can obtain that

$$D(K(P), K(Q)) = \frac{1}{5}\left[\frac{1}{5}(1+1+2+0+0)\right] = \frac{4}{25},$$
$$D(K(P), K(R)) = \frac{1}{5}\left[\frac{1}{5}(1+1+2+1+1)\right] = \frac{6}{25}, \quad \text{and}$$
$$D(K(Q), K(R)) = \frac{1}{5}\left[\frac{1}{5}(0+2+0+1+1)\right] = \frac{4}{25}.$$

Hence, one has that $\frac{4}{25} + \frac{6}{25} = \frac{10}{25} > \frac{4}{25}$, $\frac{4}{25} + \frac{4}{25} = \frac{8}{25} > \frac{6}{25}$. It is easy to see that $D(K(P), K(Q)) + D(K(P), K(R)) \geqslant D(K(Q), K(R))$, $D(K(P), K(Q)) + D(K(Q), K(R)) \geqslant D(K(P), K(R))$ and $D(K(Q), K(R)) + D(K(P), K(R)) \geqslant D(K(P), K(Q))$. □

For further development, we give the following Lemma 2.

Lemma 2. *Let A, B, C be three classical sets with $A \subseteq B \subseteq C$ or $A \supseteq B \supseteq C$, then $d(A, B) + d(B, C) = d(A, C)$.*

Proof. Suppose that $A' = \{a_1, a_2, \ldots, a_{|A\cup B\cup C|}\}$, $B' = \{b_1, b_2, \ldots, b_{|A\cup B\cup C|}\}$ and $C' = \{c_1, c_2, \ldots, c_{|A\cup B\cup C|}\}$. Let $A \subseteq B \subseteq C$, thus $A \cup B \cup C = A$ and $B \cup C = B$. Therefore,

$$d(A,B) + d(B,C) = \sum\nolimits_{i=1}^{|A\cup B\cup C|}(a_i + b_i) + \sum_{i=1}^{|A\cup B\cup C|}(b_i + c_i) = (|A\cup B| - |A\cap B|) + (|B\cup C| - |B\cap C|) = (|A| - |B|) + (|B| - |C|)$$
$$= |A| - |C| = \sum\nolimits_{i=1}^{|A\cup B\cup C|}(a_i + c_i) = d(A, C).$$

For $A \supseteq B \supseteq C$, similarly, one can draw the same conclusion. □

By Definition 11 and Lemma 2, one can obtain the following theorem.

Theorem 21. *Let $K = (U, \mathbf{R})$ be a knowledge base, P, Q, $R \in \mathbf{R}$ and $K(P) \preceq K(Q) \preceq K(R)$ or $K(R) \preceq K(Q) \preceq K(P)$. Then, $D(K(P), K(R)) = D(K(P), K(Q)) + D(K(Q), K(R))$.*

Proof. For $K(P)$, $K(Q)$, $K(R) \in K$ and $K(P) \preceq K(Q) \preceq K(R)$, one can easily get that $S_P(x_i) \subseteq S_Q(x_i) \subseteq S_R(xi)$, $x_i \in U$. Hence, it follows from Lemma 2 that

$$\begin{aligned}D(K(P), K(Q)) + D(K(Q), K(R)) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_Q(x_i)|}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_Q(x_i) \oplus S_R(x_i)|}{|U|}\\
&= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{d(S_P(x_i), S_Q(x_i))}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{d(S_Q(x_i), S_R(x_i))}{|U|}\\
&= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{1}{|U|}(d(S_P(x_i), S_Q(x_i)) + d(S_Q(x_i), S_R(x_i))) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{d(S_P(x_i), S_R(x_i))}{|U|}\\
&= \frac{1}{|U|}\sum_{i=1}^{|U|}D(K(P), K(R)).\end{aligned}$$

For $K(R) \preceq K(Q) \preceq K(P)$, similarly, one can draw the same conclusion. □

Example 5. Let $U = \{x_1, x_2, x_3, x_4, x_5\}$ and $K(P), K(Q), K(R)$ be three knowledge structures induced by equivalence relations P, Q, R on K, where $K(P) = \{\{x_1, x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_4, x_5\}, \{x_4, x_5\}\}$, $K(Q) = \{\{x_1, x_2\}, \{x_1, x_2\}, \{x_3\}, \{x_4, x_5\}, \{x_4, x_5\}\}$ and $K(R) = \{\{x_1, x_2\}, \{x_1, x_2\}, \{x_3\}, \{x_4\}, \{x_5\}\}$.

It is obvious that $K(R) \preceq K(Q) \preceq K(P)$. By computing the knowledge distances among them, one can obtain that

$$D(K(P), K(Q)) = \frac{1}{5}\left[\frac{1}{5}(1+1+2+0+0)\right] = \frac{4}{25},$$
$$D(K(Q), K(R)) = \frac{1}{5}\left[\frac{1}{5}(0+0+0+1+1)\right] = \frac{2}{25}, \quad \text{and}$$
$$D(K(P), K(R)) = \frac{1}{5}\left[\frac{1}{5}(1+1+2+1+1)\right] = \frac{6}{25}.$$

It is clear that $D(K(P), K(Q)) + D(K(Q), K(R)) = \frac{4}{25} + \frac{2}{25} = \frac{6}{25} = D(K(P), K(R))$.

Theorem 22. *Let $\mathbf{K}(U)$ be the set of all knowledge structures induced by U and $K(P), K(Q) \in \mathbf{K}(U)$ two knowledge structures, then $D(K(P), K(Q)) = D(\wr K(P), \wr K(Q))$.*

Proof. It follows from the definition of D that

$$\begin{aligned}
D(\wr K(P), \wr K(Q)) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|(x_i \cup \sim S_P(x_i)) \oplus (x_i \cup \sim S_Q(x_i))|}{|U|} \\
&= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|(x_i \cup \sim S_P(x_i)) \cup (x_i \cup \sim S_Q(x_i))| - |(x_i \cup \sim S_P(x_i)) \cup (x_i \cap \sim S_Q(x_i))|}{|U|} \\
&= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|x_i \cup (\sim S_P(x_i) \cup \sim S_Q(x_i))| - |(x_i \cup (\sim S_P(x_i) \cap \sim S_Q(x_i))|}{|U|} \\
&= (1 + |U| - |S_P(x_i) \cap S_Q(x_i)|) - (1 + |U| - |S_P(x_i) \cup S_Q(x_i)|) = |S_P(x_i) \cup S_Q(x_i)| - |S_P(x_i) \cap S_Q(x_i)| \\
&= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus S_Q(x_i)|}{|U|} = D(K(P), K(Q)).
\end{aligned}$$

That is $D(K(P), K(Q)) = D(\wr K(P), \wr K(Q))$. □

As a result of the above discussions and analyses, we come to the following three corollaries.

Corollary 1. *Let $\mathbf{K}(U)$ be the set of all knowledge structures induced by U and $K(P), K(Q) \in \mathbf{K}(U)$ two knowledge structures. If $K(P) \preceq K(Q)$, then $D(K(P), K(\omega)) \leqslant D(K(Q), K(\omega))$.*

Proof. From the knowledge structure $K(\omega) = \{x_i | x_i \in U\}$ and $K(P) \preceq K(Q)$, one has that $\{x_i\} \subseteq S_P(x_i) \subseteq S_Q(x_i)$, $x_i \in U$. Therefore,

$$\begin{aligned}
D(K(P), K(\omega)) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus \{x_i\}|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i)| - 1}{|U|} \leqslant \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_Q(x_i)| - 1}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_Q(x_i) \oplus \{x_i\}|}{|U|} \\
&= D(K(Q), K(\omega)),
\end{aligned}$$

i.e., $D(K(P), K(\omega)) \leqslant D(K(Q), K(\omega))$. □

Corollary 2. *Let $\mathbf{K}(U)$ be the set of all knowledge structures induced by U and $K(P), K(Q) \in \mathbf{K}(U)$ two knowledge structures. If $K(P) \prec K(Q)$, then $D(K(P), K(\delta)) > D(K(Q), K(\delta))$.*

Proof. Since the knowledge structure $K(\delta) = \{S_P(x_i) | S_P(x_i) = U, x_i \in U\}$ and $K(P) \preceq K(Q)$, so $S_P(x_i) \subseteq S_Q(x_i) \subseteq U$, $x_i \in U$. Hence,

$$\begin{aligned}
D(K(P), K(\delta)) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i) \oplus U|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|U| - |S_P(x_i)|}{|U|} \geqslant \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|U| - |S_Q(x_i)|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_Q(x_i) \oplus U|}{|U|} \\
&= D(K(Q), K(\delta)),
\end{aligned}$$

That is $D(K(P), K(\delta)) \geqslant D(K(Q), K(\delta))$. □

Corollary 3. *Let $\mathbf{K}(U)$ be the set of all knowledge structures induced by U and $K(P)$ a knowledge structure on $\mathbf{K}(U)$, then $D(K(P), K(\delta)) + D(K(P), K(\omega)) = 1 - \frac{1}{|U|}$.*

Proof. Since $K(\omega) \preceq K(P) \preceq K(\delta)$, one can obtain that

$$\begin{aligned}
D(K(P), K(\delta)) + D(K(P), K(\omega)) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i)| - 1}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|U| - |S_P(x_i)|}{|U|} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_P(x_i)| - 1 + |U| - |S_P(x_i)|}{|U|} \\
&= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|U| - 1}{|U|} = 1 - \frac{1}{|U|}.
\end{aligned}$$

Obviously, $D(K(P), K(\delta)) + D(K(P), K(\omega)) = 1 - \frac{1}{|U|}$. □

Remark. Unlike information entropy and knowledge granulation, the knowledge distance can characterize the difference between two knowledge structures in knowledge bases. From the definition of knowledge distance, it is easy to see that it is valid for complete knowledge structures and incomplete knowledge structures. It has some potential applications. For example, based on the knowledge distance between the knowledge structure induced by each condition attribute and the knowledge structure induced by the decision attribute, one can construct a heuristic function to extract decision rules with much higher certainty from a complete/incomplete decision table. Further experimental analysis may be desirable, but it is beyond the scope of this paper.

6. Conclusions and discussion

The contributions of this paper have three hands. In this paper, by uniformly representing a complete knowledge structure and an incomplete knowledge structure, firstly, we have proposed four operators ($\bigcup$, $\bigcap$, $\wr$ and $-$) on a knowledge base, which can be applied to generate new knowledge structures. For a decision problem in the context of decision tables, these four operators can be used to extract decision rules with much higher certainty from a given decision table. Then, in this framework of knowledge representation proposed in this study, we have established an axiom definition of knowledge granulation and have proved that some existing knowledge granulations are all its special forms. The analysis shows that one can apply this axiom definition to construct a new knowledge granulation, which can be used to restrict a new definition of knowledge granulation according to practical demands and user requirements. Finally, we have introduced the definition of a knowledge distance for calculating the difference between two knowledge structures in the same knowledge base. Noting that the knowledge distance satisfies the three properties of a distance space on all knowledge structures induced by a given universe and $(\mathbf{K}(U), D)$ is a distance space. The knowledge distance can be used to distinguish the difference between two knowledge structures with the same knowledge granulation and to characterize the essence of uncertainty of knowledge structures in knowledge bases. These results have been shown to be very helpful for knowledge discovery from knowledge bases and significant for establishing a framework of granular computing in knowledge bases.

Acknowledgements

This work was partially supported by the national natural science foundation of China (Nos. 60773133, 70471003, and 60573074), the high technology research and development program (No. 2007AA01Z165), the National Key Basic Research and Development Program of China (973) (2007CB311002), the foundation of doctoral program research of the ministry of education of China (No. 20050108004) and key project of science and technology research of the ministry of education of China.

References

[1] A. Bargiela, W. Pedrycz, Granular mappings, IEEE Transactions on Systems, Man and Cybernetics Part A 35 (2) (2005) 292–297.
[2] 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.
[3] T. Beaubouef, F.E. Petry, Fuzzy rough set techniques for uncertainty processing in a relational database, International Journal of Intelligent Systems 15 (5) (2000) 389–424.
[4] 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.
[5] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, European Journal of Operational Research 134 (3) (2001) 592–605.
[6] I. Düntsch, G. Gediga, Uncertainty measures of rough set prediction, Artificial Intelligence 106 (1998) 109–137.
[7] R.V.L. Hartley, Transmission of information, The Bell Systems Technical Journal 7 (1928) 535–563.
[8] T.P. Hong, L. Tseng, S. Wang, Learning rules from incomplete training examples by rough sets, Expert Systems with Applications 22 (4) (2002) 285–293.
[9] Q. Hu, D. Yu, Entropies of fuzzy indiscernibility relation and its operations, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (5) (2004) 575–589.
[10] Q. Hu, D. Yu, Z. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognition Letters 27 (5) (2006) 414–423.
[11] R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15 (1) (2007) 73–89.
[12] G. Jeon, D. Kim, J. Jeong, Rough sets attributes reduction based expert system in interlaced video sequences, IEEE Transactions Consumer Electronics 52 (4) (2006) 1348–1355.
[13] G.J. Klir, Basic issues of computing with granular probabilities, in: Proceedings of 1998 IEEE International Conference on Fuzzy Systems, 1998, pp. 101–105.
[14] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences 112 (1998) 39–49.
[15] M. Kryszkiewicz, Rules in incomplete information systems, Information Sciences 113 (1999) 271–292.
[16] Y. Leung, D.Y. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, Information Sciences 153 (2003) 85–106.
[17] Y. Leung, W.Z. Wu, W.X. Zhang, Knowledge acquisition in incomplete information systems: a rough set approach, European Journal of Operational Research 168 (1) (2006) 164–180.
[18] J.Y. Liang, C.Y. Dang, K.S. Chin, C.M. Yam Richard, A new method for measuring uncertainty and fuzziness in rough set theory, International Journal of General Systems 31 (4) (2002) 331–342.
[19] J.Y. Liang, Z.Z. Shi, The information entropy, rough entropy and knowledge granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-based Systems 12 (1) (2004) 37–46.
[20] J.Y. Liang, Z.Z. Shi, D.Y. Li, M.J. Wierman, The information entropy, rough entropy and knowledge granulation in incomplete information systems, International Journal of General Systems 35 (6) (2006) 641–654.
[21] J.Y. Liang, Z.B. Xu, The algorithm on knowledge reduction in incomplete information systems, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 24 (1) (2002) 95–103.
[22] T.Y. Lin, From rough sets and neighborhood systems to information granulation and computing in words, in: Proceedings of European Congress Intelligent Techniques and Soft Computing, 1997, pp. 1602–1606.
[23] T.Y. Lin, Introduction to special issues on data mining and granular computing, International Journal of Approximate Reasoning 40 (2005) 1–2.
[24] J.S. Mi, Y. Leung, W.Z. Wu, An uncertainty measure in partition-based fuzzy rough sets, International Journal General Systems 34 (2005) 77–90.
[25] J.S. Mi, W.X. Zhang, An axiomatic characterization of a fuzzy generalization of rough sets, Information Sciences 160 (2004) 235–249.
[26] H.S. Nguyen, D. Slezak, Approximation reducts and association rules correspondence and complexity results, Lecture Notes in Artificial Intelligence 1711 (1999) 137–145.
[27] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, 1991.
[28] Z. Pawlak, Rough sets, decision algorithms and Bayes's theorem, European Journal of Operational Research 136 (1) (2002) 181–189.
[29] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.
[30] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (2007) 28–40.

[31] W. Pedrycz, A. Bargiela, Granular clustering: a granular signature of data, IEEE Transactions on Systems, Man and Cybernetics Part B 32 (2) (2002) 212–224.
[32] L. Polkowski, A. Skowron, Rough mereology: A new paradigm for approximate reasoning, International Journal of Approximate Reasoning 15 (4) (1996) 333–365.
[33] Y.H. Qian, J.Y. Liang, Combination entropy and combination granulation in incomplete information systems, Lecture Notes in Artificial Intelligence 4062 (2006) 184–190.
[34] Y.H. Qian, J.Y. Liang, C.Y. Dang, Converse approximation and rule extracting from decision tables in rough set theory, Computers and Mathematics with Applications 55 (2008) 1754–1765.
[35] 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 (2008) 181–202.
[36] Y.H. Qian, J.Y. Liang, C.Y. Dang, H.Y. Zhang, J.M. Ma, On the evaluation of the decision performance of an incomplete decision table, Data & Knowledge Engineering 65 (3) (2008) 373–400.
[37] M. Quafatou, α-RST: a generalization of rough set theory, Information Sciences 124 (2000) 301–316.
[38] M. Rebolledo, Rough intervals-enhancing intervals f or qualitative modeling of technical systems, Artificial Intelligence 170 (2006) 667–685.
[39] A. Skwwron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic, Dordrecht, 1992, pp. 331–362.
[40] C.E. Shannon, The mathematical theory of communication, The Bell System Technical Journal 27 (3–4) (1948) 373–423.
[41] Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, Pattern Recognition 35 (11) (2002) 2425–2438;
. recognitionD. Ślezak, W. Ziarko, The investigation of the Bayesian rough set model, International Journal of Approximate Reasoning 40 (2005) 81–91.
[42] D. Ślezak, Searching for dynamic reducts in inconsistent decision tables, in: Proceedings of IPMU'98, vol. 2, 1998, pp. 1362–1369.
[43] D. Ślezak, Approximation reducts in decision tables, in: Proceedings of IPMU'96, vol. 3, 1996, pp. 1159–1164.
[44] W. Swiniarski, Roman, L. Hargis, Rough sets as a front end of neural-networks texture classifiers, Neurocomputing 36 (1–4) (2001) 85–102.
[45] W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letters 24 (6) (2003) 849–883.
[46] S. Tsumoto, Automated extraction of hierarchical decision rules from clinical databases using rough set model, Expert Systems with Applications 24 (2) (2003) 189–197.
[47] G. Wang, H. Yu, D. Yang, Decision table reduction based on conditional information entropy, Chinese Journal of Computers 25 (7) (2002) 1–9.
[48] M.J. Wierman, Measuring uncertainty in rough set theory, International Journal of General Systems 28 (4) (1999) 283–297.
[49] W.Z. Wu, J.S. Mi, W.X. Zhang, Generalized fuzzy rough sets, Information Sciences 151 (2003) 263–282.
[50] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (1) (2001) 87–104.
[51] Y.Y. Yao, A partition model of granular computing, LNCS Transactions on Rough Sets I (2004) 232–253.
[52] Y.Y. Yao, Neighborhood systems and approximate retrieval, Information Sciences 176 (2006) 3431–3452.
[53] L. Zadeh, Fuzzy logic equals computing with words, IEEE Transactions on Fuzzy Systems 4 (2) (1996) 103–111.
[54] L. Zadeh, A new direction in AI-Toward a computational theory of perceptions, AI Magazine 22 (1) (2001) 73–84.
[55] L. 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.
[56] A. Zeng, D. Pan, Q.L. Zheng, H. Peng, Knowledge acquisition based on rough set theory and principal component analysis, IEEE Intelligent Systems (2006) 78–85.
[57] W.X. Zhang, J.S. Mi, W.Z. Wu, Knowledge reductions in inconsistent information systems, International Journal of Intelligent Systems 18 (2003) 989–1000.
[58] N. Zhong, J. Dong, S. Ohsuga, Rule discovery by soft induction techniques, Neurocomputing 36 (1–4) (2001) 171–204.
[59] W. Ziarko, Acquisition of hierarchy-structured probabilistic decision tables and rules from data, Expert Systems 20 (5) (2003) 305–310.