Granular Structures and Approximations in Rough Sets and Knowledge Spaces

Yiyu Yao[1], Duoqian Miao[2], and Feifei Xu[1,2]

[1] Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2
yyao@cs.uregina.ca
[2] Department of Computer Science and Technology, Tongji University, Shanghai, China 201804
miaoduoqian@163.com, xufeifei1983@hotmail.com

Summary. Multilevel granular structures play a fundamental role in granular computing. In this chapter, we present a general framework of granular spaces. Within the framework, we examine the granular structures and approximations in rough set analysis and knowledge spaces. Although the two theories use different types of granules, they can be unified in the proposed framework.

1 Introduction

Granular computing is an emerging field of study focusing on structured thinking, structured problem solving and structured information processing with multiple levels of granularity [1, 2, 7, 8, 13, 17, 19, 20, 21, 22, 24, 25, 29]. Many theories may be interpreted in terms of granular computing. The main objective of this chapter is to examine the granular structures and approximations used in rough set analysis [10, 11] and knowledge spaces [3, 4].

A primitive notion of granular computing is that of granules. Granules may be considered as parts of a whole. A granule may be understood as a unit that we use for describing and representing a problem or a focal point of our attention at a specific point of time. Granules can be organized based on their inherent properties and interrelationships. The results are a multilevel granular structure. Each level is populated by granules of the similar size or the similar nature. Depending on a particular context, levels of granularity may be interpreted as levels of abstraction, levels of details, levels of processing, levels of understanding, levels of interpretation, levels of control, and many more. An ordering of levels based on granularity provides a hierarchical granular structure.

The formation of granular structures is based on a vertical separation of levels and a horizontal separation of granules in each level. It explores the property of loose coupling and nearly-decomposability [14] and searches for a good approximation [12]. Typically, elements in the same granules interact more than elements in different granules. Granules in the same level are relatively independent and granules in two adjacent levels are closely related.

A. Abraham, R. Falcón, and R. Bello (Eds.): Rough Set Theory, SCI 174, pp. 71–84.
springerlink.com

Rough set analysis and knowledge spaces use set-theoretic formulations of granular structures. A granule is interpreted as a subset of a universal set, satisfying a certain condition. In the case of rough set analysis, a granule must be definable with respect to a logic language [23, 28]. In the case of knowledge spaces, a granule must be defined by a surmise relation or a surmise system [3, 4]. Granules are ordered by the set-inclusion relation. The family of granules forms a subsystem of the power set of the universal set. As a consequence, one needs to study approximations of any subset by granules in the subsystem [16, 26].

The above discussion suggests a unified framework for studying granular structures. It serves as a basis of the current study. The rest of the chapter is organized as follows. Section 2 proposes a framework of granular spaces. Sections 3 and 4 examine the granular structures and approximations in rough sets and knowledge spaces.

2 Granular Spaces

A set-theoretic interpretation of granules and granular structures is presented and a framework of granular spaces is introduced.

2.1 A Set-Theoretic Interpretation of Granules

Categorization or classification is one of the fundamental tasks of human intelligence [12]. In the process of categorization, objects are grouped into categories and a name is given to each category. One obtains high-level knowledge about groups of objects. Such knowledge may be applied later to similar objects of the category, since each object is no longer viewed as a unique entity. In order to obtain a useful categorization, one needs to search for both similarity and dissimilarity between objects. While similarity leads to the integration of individuals into categories, dissimilarity leads to the division of larger categories into smaller subcategories.

The idea of categorization immediately leads to a set-theoretic interpretation of granules. A granule may be simply viewed as the set of objects in a category. The process of categorization covers two important issues of granulation, namely, the construction of granules and the naming of granules. The construction of granules explores both the similarity and dissimilarity of objects. Objects in the same categories must be more similar to each other, and objects in different granules are more dissimilar to each other. Consequently, one may view objects in the same category as being indistinguishable or equivalent from a certain point of view. For those objects, it may be more economic to give a name, so that one can talk about the category by its name instead of many individuals. In addition, we can apply the common properties of the objects in the category to similar objects in order to make meaningful inference in the future [12].

With the set-theoretic interpretation of granules, we can apply a set-inclusion relation on granules to form sub-super relations between granules. We can also apply set-theoretic operations on granules to construct new granules. The resulting family of granules forms a multilevel hierarchical structure.

There is another way to arrange granules. Once names are assigned to granules, we may use these names to form more abstract new granules. In this case, the elements of a granule are names to other granules. That is, a granule is treated as a whole by its name in a higher-level granule. A granule therefore plays dual roles. In the current level, it is a set of individuals; in its adjacent higher-level, it is considered as a whole by its name.

2.2 A Formulation of Granules as Concepts

The set-theoretic interpretation of granules can be formally developed based on the notion of concepts, a basic unit of human thought.

The classical view of concepts defines a concept jointly by a set of objects, called the extension of the concept, and a set of intrinsic properties common to the set of objects, called the intension of the concept [15]. The intension reflects the intrinsic properties or attributes shared by all objects (i.e., instances of a concept). Typically, the name of a concept reflects the intension of a concept. The extension of a concept is the set of objects which are concrete examples of a concept. One may introduce a logic language so that the intension of a concept is represented by a formula and the extension is represented by the set of objects satisfying the formula [28].

The language $\mathcal{L}$ is constructed from a finite set of atomic formulas, denoted by $\mathcal{A} = \{p, q, ...\}$. Each atomic formula may be interpreted as basic knowledge. They are the elementary units. In general, an atomic formula corresponds to one particular property of an individual. The construction of atomic formulas is an essential step of knowledge representation. The set of atomic formulas provides a basis on which more complex knowledge can be represented. Compound formulas can be built from atomic formulas by using logic connectives. If ϕ and φ are formulas, then $(\neg\phi)$, $(\phi \wedge \varphi)$, $(\phi \vee \varphi)$, $(\phi \rightarrow \varphi)$, and $(\phi \leftrightarrow \varphi)$ are also formulas.

The semantics of the language $\mathcal{L}$ are defined in the Tarski's style by using the notions of a model and satisfiability [9, 11, 28]. The model is a nonempty domain consisting of a set of individuals, denoted by $U = \{x, y, ...\}$. For an atomic formula p, we assume that an individual $x \in U$ either satisfies p or does not satisfy p, but not both. For an individual $x \in U$, if it satisfies an atomic formula p, we write $x \models p$, otherwise, we write $x \not\models p$. The satisfiability of an atomic formula by individuals of U is viewed as the knowledge describable by the language $\mathcal{L}$. An individual satisfies a formula if the individual has the properties as specified by the formula. To emphasize the roles played by the set of atomic formulas $\mathcal{A}$, the operations $\{\neg, \wedge, \vee\}$ and the set of individuals U, we also rewrite the language $\mathcal{L}$ as $\mathcal{L}(\mathcal{A}, \{\neg, \wedge, \vee\}, U)$.

If ϕ is a formula, the set $m(\phi)$ defined by:

$$m(\phi) = \{x \in U \mid x \models \phi\}, \tag{1}$$

is called the meaning of the formula ϕ. The meaning of a formula ϕ is indeed the set of all objects having the properties expressed by the formula ϕ. In other words, ϕ can be viewed as the description of the set of objects $m(\phi)$. As a result, a concept can be expressed by a pair $(\phi, m(\phi))$, where $\phi \in \mathcal{L}$. ϕ is the intension of a concept while $m(\phi)$ is the extension of a concept. A connection between formulas and subsets of U is established. Similarly, a connection between logic connectives and set-theoretic operations can be stated [11]:

$$\begin{aligned}
&\text{(i)} \quad m(\neg\phi) = (m(\phi))^c,\\
&\text{(ii)} \quad m(\phi \wedge \psi) = m(\phi) \cap m(\psi),\\
&\text{(iii)} \quad m(\phi \vee \psi) = m(\phi) \cup m(\psi),\\
&\text{(iv)} \quad m(\phi \rightarrow \psi) = (m(\phi))^c \cup (m(\psi))^c,\\
&\text{(v)} \quad m(\phi \equiv \psi) = (m(\phi) \cap m(\psi)) \cup ((m(\phi))^c \cap (m(\psi))^c),
\end{aligned}$$

where $(m(\phi))^c = U - m(\phi)$ is the complement of $m(\phi)$. Under this formulation, we can discuss granules in terms of intensions in a logic setting and in terms of extension in a set-theoretic setting.

2.3 Granular Spaces and Granular Structures

Each atomic formula in $\mathcal{A}$ is associated with a subset of U. This subset may be viewed as an elementary granule in U. Each formula is obtained by taking logic operations on atomic formulas. The meaning set of the formula can be obtained from the elementary granules through set-theoretic operations. With the language $\mathcal{L}$, for each formula, we can find its meaning set by equation (1). On the other hand, for an arbitrary subset of the universe U, one may not be able to find a formula to precisely represent it. This leads to the introduction of the definability of granules in a logic language. We say a subset or a granule $X \subseteq U$ is definable if and only if there exists a formula ϕ in the language $\mathcal{L}$ such that,

$$X = m(\phi). \tag{2}$$

Otherwise, it is undefinable [23]. By the formulation of the logic language, it follows that each definable granule can be expressed in terms of elementary granules through set-theoretic operations.

A definable granule is represented by a pair $(\phi, m(\phi))$ or simply $m(\phi)$. The family of all definable granules is given by:

$$Def(\mathcal{L}(\mathcal{A}, \{\neg, \wedge, \vee\}, U)) = \{m(\phi) \mid \phi \in \mathcal{L}(\mathcal{A}, \{\neg, \wedge, \vee\}, U)\}, \tag{3}$$

which is a subsystem of the power set 2^U, closed under set complement, intersection and union. Based on these notions, we formally define a granular space by the triplet:

$$(U, \mathcal{S}_0, \mathcal{S}), \tag{4}$$

where

U is the universe,
$\mathcal{S}_0 \subseteq 2^U$ is a family of elementary granules, i.e., $\mathcal{S}_0 = \{m(p) \mid p \in \mathcal{A}\}$,
$\mathcal{S} \subseteq 2^U$ is a family of definable granules, i.e., $\mathcal{S} = Def(\mathcal{L}(\mathcal{A}, \{\neg, \wedge, \vee\}, U))$.

Note that $\mathcal{S}$ can be generated from $\mathcal{S}_0$ through set-theoretic operations.

The family $\mathcal{S}$ of definable granules is called a granular structure. Since the logic language uses logic operations $\neg$, $\wedge$, and $\vee$, the family of definable granules $\mathcal{S}$ is closed under set complement, intersection and union. That is, $\mathcal{S}$ is an σ-algebra. Note that $\mathcal{S}_0$ is not necessarily the basis of the σ-algebra $\mathcal{S}$.

There are additional requirements to make the granular space more practical. For example, the family of elementary granules normally can not be all singleton subsets of U, as a singleton subset is equivalent to its unique object. The set of all granules constructed from the family of elementary granules is normally a superset of the family of elementary granules. Furthermore, it is typically a subset of the power set of U. Otherwise, we do not have the benefits of granulation. It also requires that the union of all the elementary granules covers the universe U. That is, an object satisfies at least one basic formula in $\mathcal{A}$.

From the above discussion, the granular structure can be described by a logic language $\mathcal{L}(\mathcal{A}, \{\neg, \wedge, \vee\}, U)$. To make the granular structure more practical, one may consider a logic language using a subset of logic connectives. As a special case, suppose that the granular structure is closed under set intersection and union but not complement. Since the logic operators $\wedge$ and $\vee$ correspond to set intersection and union, we can use a logic language $\mathcal{L}(\mathcal{A}, \{\wedge, \vee\}, U)$ to describe this type of granular structures. Two more special cases of granular structures are defined by the languages $\mathcal{L}(A, \{\wedge\}, U)$ and $\mathcal{L}(A, \{\vee\}, U)$, respectively [28]. While the former is closed under set intersection, the latter is closed under set union. Note that a granular structure containing U and being closed under set intersection is called a closure system.

3 Rough Set Analysis

The granular space of rough set analysis is a quotient space induced by an equivalence relation.

3.1 Granular Spaces and Granular Structures

Rough set analysis studies relationships between objects and their attribute values in an information table [10, 11]. An information table provides a convenient way to describe a finite set of objects by a finite set of attributes. Formally, an information table can be expressed as:

$$M = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}), \tag{5}$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, V_a is a nonempty set of values for an attribute $a \in At$, and $I_a : U \to V_a$ is an information function.

For a set of attributes $P \subseteq At$, one can define an equivalence relation on the set of objects:

$$xE_P y \Longleftrightarrow \forall a \in P(I_a(x) = I_a(y)). \tag{6}$$

Two objects are equivalent if they have the same values on all attributes in P. Rough set theory is developed based on such equivalence relations.

Let $E \subseteq U \times U$ denote an equivalence relation on U. The pair $apr = (U, E)$ is called an approximation space [10]. The equivalence relation E partitions the set U into disjoint subsets called equivalence classes. This partition of the universe is denoted by U/E. The partition U/E may be considered as a granulated view of the universe. For an object $x \in U$, the equivalence class containing x is given by:

$$[x]_E = \{y \mid yEx\}. \tag{7}$$

Equivalence classes are referred to as elementary granules.

By taking the union of a family of equivalence classes, we can obtain a composite granule. The family of all such granules contains the entire set U and the empty set $\emptyset$, and is closed under set complement, intersection and union. More specifically, the family is an σ-algebra, denoted by $\sigma(U/E)$, with the basis U/E.

The above formulation can be expressed in terms of a logic language. For an attribute-value pair (a, v), where $a \in At$ and $v \in V_a$, we have an atomic formula $a = v$. The meaning of $a = v$ is the following set of objects:

$$m(a = v) = \{x \in U \mid I_a(x) = v\}. \tag{8}$$

It immediately follows that the equivalence class $[x]$ is defined by the formula $\bigwedge_{a \in At} a = I_a(x)$. That is, $[x]$ is a definable granule. One can easily see that the union of a family of equivalence classes is also a definable granule. Thus, the set of all definable granules is $\sigma(U/E)$.

Based on the above discussion, we can conclude that a granular space used in rough set theory is given by:

$$(U, U/E, \sigma(U/E)),$$

where

U is the universe,
$U/E \subseteq 2^U$ is the family of equivalence classes,
$\sigma(U/E) \subseteq 2^U$ is the σ-algebra generated from U/E.

That is, the granular space and granular structure of rough set analysis are a special case of the ones introduced in the last sect.

3.2 Rough Set Approximations

An arbitrary set $A \subseteq U$ may not necessarily be the union of some equivalence classes. This implies that one may not be able to describe A precisely using the logic language $\mathcal{L}$. In order to infer information about such undefinable granules, it is necessary to approximate them by definable granules.

For a subset of objects $A \subseteq U$, it may be approximated by a pair of lower and upper approximations:

$$\begin{aligned} \underline{apr}(A) &= \bigcup \{X \in \sigma(U/E) \mid X \subseteq A\}, \\ \overline{apr}(A) &= \bigcap \{X \in \sigma(U/E) \mid A \subseteq X\}. \end{aligned} \tag{9}$$

The lower approximation $\underline{apr}(A)$ is the union of all the granules in $\sigma(U/E)$ that are subsets of A. The upper approximation $\overline{apr}(A)$ is the intersection of all the granules in $\sigma(U/E)$ that contain A. This is referred to as the subsystem based definition [16, 18, 26]. Since $\sigma(U/E)$ is closed under $\cap$ and $\cup$, the definition is well defined. In addition, $\underline{apr}(A) \in \sigma(U/E)$ is the largest granule in $\sigma(U/E)$ that is contained by A, and $\overline{apr}(A) \in \sigma(U/E)$ is the smallest granule that contains A. That is, the pair $(\underline{apr}(A), \overline{apr}(A))$ is the tightest approximation.

Lower and upper approximations are dual to each other in the sense:

$$\text{(Ia)} \quad \underline{apr}(A) = (\overline{apr}(A^c))^c,$$
$$\text{(Ib)} \quad \overline{apr}(A) = (\underline{apr}(A^c))^c.$$

The set A lies within its lower and upper approximations:

$$\text{(II)} \quad \underline{apr}(A) \subseteq A \subseteq \overline{apr}(A).$$

One can also verify the following properties:

$$\text{(IIIa)} \quad \underline{apr}(A \cap B) = \underline{apr}(A) \cap \underline{apr}(B),$$
$$\text{(IIIb)} \quad \overline{apr}(A \cup B) = \overline{apr}(A) \cup \overline{apr}(B).$$

The lower (upper) approximation of the intersection (union) of a finite number of sets can be obtained from their lower (upper) approximations. However, we only have:

$$\text{(IVa)} \quad \underline{apr}(A \cup B) \supseteq \underline{apr}(A) \cup \underline{apr}(B),$$
$$\text{(IVb)} \quad \overline{apr}(A \cap B) \subseteq \overline{apr}(A) \cap \overline{apr}(B).$$

It is impossible to obtain the lower (upper) approximation of the union (intersection) of some sets from their lower (upper) approximations. Additional properties of rough set approximations can be found in [10] and [27].

3.3 An Example

Suppose $U = \{a, b, c, d, e\}$. Given an equivalence relation:

$$aEa, \quad aEb, \quad bEa, \quad bEb, \quad cEc, \quad cEe, \quad dEd, \quad eEc, \quad eEe,$$

it induces the partition $U/E = \{\{a, b\}, \{c, e\}, \{d\}\}$. We can construct an σ-algebra by taking union of any family of equivalence classes:

$$\sigma(U/E) = \{\emptyset, \{a, b\}, \{c, e\}, \{d\}, \{a, b, c, e\}, \{a, b, d\}, \{c, d, e\}, U\}.$$

The corresponding granular space is:

$$(U, U/E, \sigma(U/E)),$$

where

$$U = \{a, b, c, d, e\},$$

$$U/E = \{\{a,b\},\{c,e\},\{d\}\},$$
$$\sigma(U/E) = \{\emptyset,\{a,b\},\{c,e\},\{d\},\{a,b,c,e\},\{a,b,d\},\{c,d,e\},U\}.$$

Consider a subset of objects $A = \{a,b,c,d\}$. It can not be obtained by taking union of some elementary granules. That is, it is an undefinable granule. We approximate it by a pair of subsets from below and above in the σ-algebra $\sigma(U/E)$. From equation (9), we have:

$$\underline{apr}(A) = \bigcup\{X \in \sigma(U/E) \mid X \subseteq A\} = \{a,b\} \cup \{d\} \cup \{a,b,d\} = \{a,b,d\},$$
$$\overline{apr}(A) = \bigcap\{X \in \sigma(U/E) \mid A \subseteq X\} = \{a,b,c,d,e\}.$$

It follows that $\underline{apr}(A) = \{a,b,d\} \subseteq A \subseteq \overline{apr}(A) = \{a,b,c,d,e\}$.

4 Knowledge Space Theory

Knowledge spaces [3, 4, 5, 6] represent a new paradigm in mathematical psychology. It is a systematic approach to the assessment of knowledge by constructing sequences of questions to be asked.

In knowledge spaces, we consider a pair $(Q,\mathcal{K})$, where Q is a finite set of questions and $\mathcal{K} \subseteq 2^Q$ is a collection of subsets of Q. Each element $K \in \mathcal{K}$ is called a knowledge state and $\mathcal{K}$ is the set of all possible knowledge states. Intuitively, the knowledge state of an individual is represented by the set of questions that he is capable of answering. Each knowledge state can be considered as a granule. The collection of all the knowledge states together with the empty set $\emptyset$ and the whole set Q is called a knowledge structure, and may be viewed as a granular knowledge structure in the terminology of granular computing.

There are two types of knowledge structures. One is closed under set union and intersection, and the other is closed only under set union, and the latter is called a knowledge space.

4.1 Granular Spaces Associated to Surmise Relations

One intuitive way to study a knowledge structure is through a surmise relation. A surmise relation on the set Q of questions is a reflexive and transitive relation S. By aSb, we can surmise the mastery of a if a student can correctly answer question b. A surmise relation imposes conditions on the corresponding knowledge structure. For example, aSb means that if a knowledge state contains b, it must also contain a.

Formally, for a surmise relation S on the (finite) set Q of questions, the associated knowledge structure $\mathcal{K}$ is defined by:

$$\mathcal{K} = \{K \mid \forall q,q' \in Q((qSq', q' \in K) \Longrightarrow q \in K)\}. \tag{10}$$

The knowledge structure associated to a surmise relation contains the empty set $\emptyset$, the entire set Q, and is closed under set intersection and union.

For each question q in Q, under a surmise relation, we can find one unique prerequisite question set $R_p(q) = \{q' \mid q'Sq\}$. The family of the prerequisite question sets for all the questions is denoted by $\mathcal{B}$, which is a covering of Q. Each prerequisite set for a question is called an elementary granule. By taking the union of prerequisite sets for a family of questions, we can obtain a knowledge structure $\mathcal{K}$ associated to the surmise relation S. It defines a granular space $(Q, \mathcal{B}, \mathcal{K})$. All knowledge states are called granules in $(Q, \mathcal{B}, \mathcal{K})$.

As a result, we have a granular space based on a surmise relation:

$$(Q, \mathcal{B}, \mathcal{K}),$$

where

Q is a set of question set,
$\mathcal{B} \subseteq 2^Q$ is a family of prerequisite sets, i.e., $\mathcal{B} = \{R_p(q) \mid q \in Q\}$,
$\mathcal{K} \subseteq 2^Q$ is a family of knowledge states.

Note that $\mathcal{B} \subseteq \mathcal{K}$ and each knowledge state can be expressed as a union of some elements of $\mathcal{B}$.

4.2 An Example

Suppose $Q = \{a, b, c, d, e\}$. Given a surmise relation:

$$aSa, \quad aSd, \quad bSb, \quad bSc, \quad bSd, \quad bSe, \quad cSc, \quad cSd, \quad cSe, \quad dSd, \quad eSe,$$

we have a knowledge structure:

$$\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, \{b, c, e\}, \{a, b, c, e\}, \{a, b, c, d\}, Q\}.$$

It can be easily seen that the knowledge structure is closed under set union and intersection. It is a knowledge structure associated to a surmise relation. As a result, we can find the prerequisite set for each question:

$$\begin{aligned} R_p(a) &= \{a\}, \\ R_p(b) &= \{b\}, \\ R_p(c) &= \{b, c\}, \\ R_p(d) &= \{a, b, c, d\}, \\ R_p(e) &= \{b, c, e\}. \end{aligned}$$

As a result, we have a granular space based on a surmise relation:

$$(Q, \mathcal{B}, \mathcal{K}),$$

where

$$\begin{aligned} &Q = \{a, b, c, d, e\}, \\ &\mathcal{B} = \{\{a\}, \{b\}, \{b, c\}, \{a, b, c, d\}, \{b, c, e\}\}, \\ &\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, \{b, c, e\}, \{a, b, c, e\}, \{a, b, c, d\}, Q\}. \end{aligned}$$

A granule in $\mathcal{K}$ can be attained by taking union of some elementary granules in $\mathcal{B}$. For example, $\{a, b\} = \{a\} \cup \{b\}$ and $\{a, b, c, e\} = \{a\} \cup \{b, c, e\}$.

4.3 Granular Spaces Associated to Surmise Systems

Modeling a knowledge structure with a surmise relation is sometimes too restrictive. That is, a question can only have one prerequisite set. In real-life situations, a question may have several prerequisite sets. This leads to the concept of surmise systems.

A surmise system on a (finite) set Q is a mapping σ that associates any element $q \in Q$ to a nonempty collection $\sigma(q)$ of subsets of Q satisfying the following three conditions [4]:

$$\begin{aligned}
&(1) \quad C \in \sigma(q) \Longrightarrow q \in C, \\
&(2) \quad (C \in \sigma(q), q' \in C) \Longrightarrow \exists C' \in \sigma(q')(C' \subseteq C), \\
&(3) \quad C \in \sigma(q) \Longrightarrow \forall C' \in \sigma(q)(C' \not\subseteq C),
\end{aligned}$$

where C is a subset in $\sigma(q)$ called a clause for question q. A surmise system may be interpreted as a neighborhood system. A clause for question q is actually a prerequisite set for q. Each question may have several prerequisite sets, namely, the clauses for question q are not always unique. Condition (1) generalizes the reflexivity condition for a relation, while the second condition extends the notion of transitivity. Condition (3) requires that the clauses for question q are the maximal sets.

Formally, for a surmise system (Q, σ), the associated knowledge structure is given by:

$$\mathcal{K} = \{K \mid \forall q \in Q(q \in K \Longrightarrow \exists C \in \sigma(q)(C \subseteq K))\}, \tag{11}$$

which is closed under set union. Any knowledge structure that is closed under union is called a knowledge space. There is a one-to-one correspondence between surmise systems on Q and knowledge spaces on Q.

There always exists exactly one minimal sub-collection $\mathcal{B}$ of $\mathcal{K}$. For a minimal sub-collection $\mathcal{B}$, any knowledge state in the sub-collection can not be the union of any other knowledge states in $\mathcal{B}$. All the knowledge states in a knowledge space can be obtained by the union of some subsets in the minimal sub-collection. We call a minimal sub-collection a basis of $\mathcal{K}$. It can be easily shown that the basis is a covering of Q. The subsets in the basis are called elementary granules. The corresponding knowledge space defines a granular space $(Q, \mathcal{B}, \mathcal{K})$. The knowledge states in the knowledge space are also called granules.

We have a granular space in a knowledge space:

$$(Q, \mathcal{B}, \mathcal{K}),$$

where

Q *is the question set,*
$\mathcal{B} \subseteq 2^Q$ *is the basis of a knowledge space called elementary granules,*
$\mathcal{K} \subseteq 2^Q$ *is a family of knowledge states in a knowledge space called granules.*

It has been proven that the basis of a knowledge space is the family of all clauses [6]. Each of the clauses is an element of the basis $\mathcal{B}$. Conversely, each element of the basis is a clause for some question. Let $\mathcal{B}_q$ represent the set of all minimal states containing question q in a knowledge space $\mathcal{K}$ with a basis $\mathcal{B}$. Thus, $\mathcal{B}_q \subseteq \mathcal{K}_q$, where $\mathcal{K}_q$ is the set of all states containing q. We also have $\mathcal{B}_q \subseteq \mathcal{B}$ for any question q, and the set $\sigma(q)$ of all the clauses for q is obtained by setting:

$$\sigma(q) = \mathcal{B}_q. \tag{12}$$

Equation (12) specifies how the clauses can be constructed from the basis of a knowledge space.

4.4 An Example

Suppose $Q = \{a, b, c, d, e\}$. Given a surmise system:

$$\begin{aligned}
\sigma(a) &= \{\{a\}\},\\
\sigma(b) &= \{\{b, d\}, \{a, b, c\}, \{b, c, e\}\},\\
\sigma(c) &= \{\{a, b, c\}, \{b, c, e\}\},\\
\sigma(d) &= \{\{b, d\}\},\\
\sigma(e) &= \{\{b, c, e\}\},
\end{aligned}$$

we can obtain a knowledge structure $\mathcal{K} = \{\emptyset,\ \{a\},\ \{b, d\},\ \{a, b, c\},\ \{b, c, e\},$ $\{a, b, d\}, \{a, b, c, d\}, \{a, b, c, e\}, \{b, c, d, e\}, Q\}$. It can be easily verified that it is a knowledge space closed under set union.

We have a granular space in the knowledge space:

$$(Q, \mathcal{B}, \mathcal{K}),$$

where

$$\begin{aligned}
Q &= \{a, b, c, d, e\},\\
\mathcal{B} &= \{\{a\}, \{b, d\}, \{a, b, c\}, \{b, c, e\}\},\\
\mathcal{K} &= \{\emptyset, \{a\}, \{b, d\}, \{a, b, c\}, \{b, c, e\}, \{a, b, d\}, \{a, b, c, d\}, \{a, b, c, e\}, \{b, c, d, e\},\\
&\quad\ Q\}.
\end{aligned}$$

$\mathcal{K}$ can be constructed by taking union of any clauses in the basis $\mathcal{B}$.

4.5 Rough Set Approximations in Knowledge Spaces

The two types of knowledge structures defined by a surmise relation and a surmise system satisfy different properties. They produce different rough-set-like approximations.

Knowledge Structure Defined by a Surmise Relation

Each knowledge state is considered as a granule. For an arbitrary subset of questions $A \subseteq Q$, it may not be a knowledge state. We can use a pair of states from below and above to approximate A. Since the knowledge structure associated to a surmise relation is closed under set intersection and union, we can extend the subsystem based definition [16, 18, 26]. The lower and upper approximations are unique. They are defined by:

$$\underline{apr}(A) = \bigcup\{K \in \mathcal{K} \mid K \subseteq A\},$$
$$\overline{apr}(A) = \bigcap\{K \in \mathcal{K} \mid A \subseteq K\}. \quad (13)$$

If A is a knowledge state, we have $\underline{apr}(A) = A = \overline{apr}(A)$. The lower approximation $\underline{apr}(A)$ is the union of all the knowledge states which are subsets of A. The upper approximation $\overline{apr}(A)$ is the intersection of all the knowledge states which contain A. The knowledge structure associated to a surmise relation is not closed under complement. It follows that the lower and upper approximations are no longer dual to each other.

The physical interpretation of approximations in knowledge spaces may be given as follows. Suppose A is a set of questions that can be answered correctly. Since it is not a knowledge state, we approximate it by two knowledge states. The lower approximation represents the affirmatory mastery of a subset of questions, while the upper approximation describes the possible mastery of a subset of questions.

Knowledge Structure Defined by a Surmise System

A knowledge structure associated to a surmise system is only closed under set union, but not necessarily closed under set intersection. In this case, we can still use a subsystem-based definition [16, 18, 26] to define the lower approximation. However, we must introduce a new definition for upper approximation.

By keeping the interpretation of lower and upper approximations as the greatest knowledge states that are contained in A and the least knowledge states that contain A, we have the following definition:

$$\underline{apr}(A) = \{\cup\{K \in \mathcal{K} \mid K \subseteq A\}\},$$
$$\overline{apr}(A) = \{K \in \mathcal{K} \mid A \subseteq K, \forall K' \in \mathcal{K}(K' \subset K \Longrightarrow A \nsubseteq K')\}. \quad (14)$$

The lower approximation in knowledge spaces is unique while the upper approximation is a family of sets [16]. If A is a knowledge state, we have $\underline{apr}(A) = \{A\} = \overline{apr}(A)$.

5 Conclusion

Granules and granular structures are two fundamental notions of granular computing. A set-theoretic framework of granular spaces is proposed. The framework considers three levels of characterization of a universal set U, namely, the triplet $(U, \mathcal{S}_0, \mathcal{S})$. They are the ground level which is U, the elementary granule level which is a subsystem $\mathcal{S}_0$ of 2^U, and the granular structure level which is a system $\mathcal{S}$ of 2^U. Typically, $\mathcal{S}_0$ is a covering of U and $\mathcal{S}$ is a hierarchical structure generated from $\mathcal{S}_0$. The framework enables us to study rough set analysis and knowledge spaces in a common setting. Moreover, results from the two theories can be applied to each other, which brings more insights into the research of the two theories.

With the proposed framework, we demonstrate that rough set analysis and knowledge spaces both consider a similar type of granular space. Their main differences lie in the construction and interpretation of elementary granules and granular structures. This observation immediately opens up new avenues of research. As an illustration, we introduce the notion of approximations from rough set analysis to knowledge spaces theory. The results are a new type of approximations not considered in rough set analysis.

The framework of granular spaces can also be used to interpret other theories. For example, it can be used to study formal concept analysis [26]. As future research, the potential of the proposed framework will be further explored.

References

1. Bargiela, A., Pedrycz, W.: Granular Computing: An Introduction. Kluwer Academic Publishers, Boston (2002)
2. Bargiela, A., Pedrycz, W.: Toward a theory of granular computing for human-centred information processing. IEEE Transactions On Fuzzy Systems (to appear)
3. Doignon, J.P., Falmagne, J.C.: Spaces for the assessment of knowledge. International Journal of Man-Machine Studies 23, 175–196 (1985)
4. Doignon, J.P., Falmagne, J.C.: Knowledge spaces. Springer, Heidelberg (1999)
5. Duntsch, I., Gediga, G.: A note on the correspondences among entail relations, rough set dependencies, and logical consequence. Mathematical Psychology 43, 393–401 (2001)
6. Falmagne, J.C., Koppen, M., Villano, M., Doignon, J.P., Johanessen, L.: Introduction to knowledge spaces: how to test and search them. Psychological Review 97, 201–224 (1990)
7. Lin, T.Y., Yao, Y.Y., Zadeh, L.A.: Data Mining, Rough Sets and Granular Computing. Physica-Verlag, Heidelberg (2002)
8. Nguyen, S.H., Skowron, A., Stepaniuk, J.: Granular computing: a rough set approach. Computational Intelligence 17, 514–544 (2001)
9. Orlowska, E.: Logical aspects of learning concepts. International Journal of Approximate Reasoning 2, 349–364 (1988)
10. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
11. Pawlak, Z.: Rough Sets. In: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
12. Pinker, S.: How the mind works. Norton, New York (1997)
13. Polkowski, L., Semeniuk-Polkowska, M.: On foundations and applications of the paradigm of granular rough computing. International Journal of Cognitive Informatics and Natural Intelligence 2, 80–94 (2008)
14. Simon, H.A.: The Sciences of the Artificial. MIT Press, Cambridge (1996)
15. Solso, R.L., Maclin, M.K., Maclin, O.H.: Cognitive psychology. Allyn and Bacon, Boston (2005)

16. Xu, F.F., Yao, Y.Y., Miao, D.Q.: Rough set approximations in formal concept analysis and knowledge spaces. In: An, A., Matwin, S., Raś, Z.W., Ślęzak, D. (eds.) Foundations of Intelligent Systems. LNCS, vol. 4994, pp. 319–328. Springer, Heidelberg (2008)
17. Yao, J.T.: A ten-year review of granular computing. In: Proceedings of 2007 IEEE International Conference on Granular Computing, pp. 734–739 (2007)
18. Yao, Y.Y.: On generalizing rough set theory. In: Wang, G., Liu, Q., Yao, Y., Skowron, A. (eds.) RSFDGrC 2003. LNCS, vol. 2639, pp. 44–51. Springer, Heidelberg (2003)
19. Yao, Y.Y.: A partition model of granular computing. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Świniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 232–253. Springer, Heidelberg (2004)
20. Yao, Y.Y.: Granular computing. Computer Science (Ji Suan Ji Ke Xue). 31, 1–5 (2004)
21. Yao, Y.Y.: Perspectives of granular computing. In: Proceedings of 2005 IEEE International Conference on Granular Computing, pp. 85–90 (2005)
22. Yao, Y.Y.: Three perspectives of granular computing. Journal of Nanchang Institute of Technology 25(2), 16–21 (2006)
23. Yao, Y.Y.: A note on definability and approximations. In: Peters, J.F., Skowron, A., Marek, V.W., Orłowska, E., Słowiński, R., Ziarko, W.P. (eds.) Transactions on Rough Sets VII. LNCS, vol. 4400, pp. 274–282. Springer, Heidelberg (2007)
24. Yao, Y.Y.: The art of granular computing. In: Proceeding of the International Conference on Rough Sets and Emerging Intelligent Systems Paradigms, pp. 101–112 (2007)
25. Yao, Y.Y.: The rise of granular computing. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition) 20(3), 1–10 (2008)
26. Yao, Y.Y., Chen, Y.H.: Rough set approximations in formal concept analysis. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets V. LNCS, vol. 4100, pp. 285–305. Springer, Heidelberg (2006)
27. Yao, Y.Y., Lin, T.Y.: Generalization of rough sets using modal logic. Intelligent Automation and Soft Computing 2, 103–120 (1996)
28. Yao, Y.Y., Zhou, B.: A logic language of granular computing. In: Proceedings 6th IEEE International Conference on Cognitive Informatics, pp. 178–185 (2006)
29. Zadeh, L.A.: Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90, 111–127 (1997)