http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 11 (12): 1714-1721, 2012
ISSN 1812-5638 / DOI: 10.3923/itj.2012.1714.1721

Granular Computing-based Granular Structure Model and its Application in Knowledge Retrieval

Lin Sun, Jiucheng Xu, Chuan Wang, Tianhe Xu and Jinyu Ren
College of Computer and Information Technology, Henan Normal University, Henan 453007, China

Abstract: This study, from the viewpoint of granularity, investigates the extended formulas and the formulation representation of granules and then introduces some operations of granules in rough sets. Within the framework of granular spaces presented, we examine their granular structure model. Moreover, some of their important propositions and properties are derived, the performances of which are shown through two illustrative examples. Furthermore, from the viewpoints of user interests and granular information processing, we develop a conceptual framework of knowledge retrieval based on the granular structure model which enlarges the application areas of granular computing.

Key words: Granular computing, granule, granular space, granular structure, knowledge retrieval

INTRODUCTION

It has been argued, both from philosophical and theoretical points of views, that information granularity is essential to human problem solving and hence has very significant impact on the design and implementation of intelligent system (Pal *et al*., 2005). A granule is a clump of objects (points), in the universe of discourse, drawn together by indistinguishability, similarity, proximity, or functionality. Granulation leads to information compression/summarization. In many situations, when a problem involves incomplete, uncertain, or vague information, it may be difficult to differentiate distinct elements and one is forced to consider granules. Granular computing may be regarded as a unified framework for theories, methodologies, techniques and tools that make use of granules, i.e., groups, classes, or clusters of a universe, in the process of problem solving (Wu *et al*., 2009; Sun *et al*., 2011; Zeng *et al*., 2011). By examining existing studies in a unified framework of granular computing and extracting their commonalities, one may be able to develop a general theory for problem solving. Furthermore, granular computing is a way of thinking that relies on our ability to perceive the real world under various grain sizes, to abstract and consider only those things that serve our present interest and to switch among different granularities (Yao, 2004a).

Currently, many models and methods of granular computing have been proposed and studied. A general framework of granular computing based on fuzzy set theory was proposed (Zadeh, 1997). In 1982, Pawlak put forward the theory of rough sets (Miao *et al*., 2007), namely a concrete example of granular computing. Hobbs (1985) set up a theory of granularity which is similar to the theory of rough sets in terms of formulation. A quotient space theory of problem solving based on hierarchical description and representation of a problem were developed (Zhang and Zhang, 2003). Lin (2006) proposed to use neighborhood systems for the formulation of granular computing. A unified framework of granular computing was proposed by Yao and Yao (2002). The new framework extends results obtained in the set-theoretic setting and extracts high-level common principles from a wide range of scientific disciplines. Based on a simple granulation structure, namely a partition of the universe of discourse, Yao (2004a) proposed a model of granular computing. The model is constructed by granulating a finite universe of discourse through a family of pairwise disjoint subsets under an equivalence relation. And many other researchers also proposed some available models and methods of granular computing (Pei, 2007; Liu *et al*., 2007). These models of granular computing have been proposed over the years. As a concrete theory of granular computing, rough set models enable us to precisely define and analyze many notions of granular computing. Though one can gain a better understanding of granular computing within the rough set framework, there are many fundamental issues which have not been thoroughly investigated. Granulation of the universe of discourse, description of granules, relationship between granules and computing with granules, to name but a few examples, are issues that

Corresponding Author: Lin Sun, College of Computer and Information Technology, Henan Normal University, Henan 453007, China

need further scrutiny (Ma *et al*., 2007). Granulation of a universe of discourse is one of the important aspects whose solution has significant bearing on granular computing.

In an information system, a_v which can be also denoted by (a, v), is defined as a descriptor, denoted by a (x) = v, where v is the value of attribute a with respect to individual variable x∈U (U is a given universe). Thus a_v is considered as a proposition in rough logic (Liu *et al*., 2001). The granule is defined via propositional formula a_v in rough logic, so it is called elementary granular logical formula (also called atomic formula). In the practical application, knowledge discovery in databases has become a research hot spot nowadays (Chen *et al*., 2010; Qi and Wei, 2008) and then a attribute value *v* can denote tens, hundreds and thousands of values. That is, only one attribute has tens, hundreds and thousands of corresponding atomic formulas. However, what we investigate is only a case that the values of the attributes lie in some region, so that we have to use a great number of atomic formulas to represent them. For example, when searching in database and describing granules like $a_v>x$ and $a_v<y$, a mass of atomic formulas are used to describe the granules. Thus, it shows that the current methods have great limitations in real-value information systems. But until now, few literatures have been reported to overcome the limitations by constructing different granular spaces and granular structures in rough sets. Therefore, it is very desirable to present some elementary operations of granules and establish efficient granular spaces and granular structure model in information systems. This paper focuses on creating such a solution.

In this study, we introduce the extended formulas and investigate the formulation representation of granules and their operations. Then, we present general framework of granular spaces and within the framework, the granular structure model in rough sets is examined. Moreover, some of their important propositions and properties are derived. Finally, we describe a conceptual framework of knowledge retrieval system based on the granular structure model and give its main process.

FORMULATION OF GRANULES AND OPERATIONS

As we know, rough sets analysis studies relationships between objects and their attribute values in an information system (IS). An information system provides a convenient way to describe a finite set of objects by a finite set of attributes (Pawlak, 1991). Formally, an information system is usually expressed in the following form: IS = (U, A, V, f), where U is a finite non-empty set of objects indicating a given universe; A is a finite non-empty set of attributes; V is the union of attribute domains such that $V = \cup_{a\in A} V_a$ and V_a denotes the value domain of attribute a∈A; f: U×A→V is an information function which associates a unique value of each attribute with every object belonging to U, so that for any a∈A and u∈U, $f(u, a)\in V_a$. Also, IS = (U, A, V, f) can be written more simply as (U, A), if V and f are understood.

Definition 1: Let IS = (U, A) be an information system, the formulas on IS can be inductively described as follows:

1: Any atomic formula (a, v) on IS is a formula on IS
2: If φ is a formulas on IS, then (¬φ) is a formula on IS
3: If φ and Ψ are formulas on IS, then (φ∧Ψ) (φ∨Ψ) and (φ→Ψ) are formulas on IS
4: Formulas are generated by using rule (1) (2) or (3) in finite steps on IS

It should be noticed that the formulas above have been presented by Liu and Liu (2004) and Yan and Liu (2007). 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 (Yao *et al*., 2009). The intension reflects the intrinsic properties or attributes shared by all objects and the extension of a concept is the set of objects which are concrete examples of a concept. Then, we further introduce a logic language to extend atomic formulas, 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.

Definition 2: Given an information system IS = (U, A), with a∈A and $M\subseteq V_a$, the formulas on IS are inductively defined as follows:

1: Any atomic formula (a, M) or in short a_M on IS is a formula on IS
2: If $M = V_a$, then a_M is a dispensable atomic formula on IS which is not involved in the following operations on IS
3: If φ and Ψ are two atomic formulas on IS, then (φ∧Ψ) and (φ∨Ψ) are formulas on IS
4: Formulas are generated by using rule (1) (2) or (3) in finite steps on IS

It is known that a primitive notion of granular computing is that of granules. A granule is interpreted as a subset of a universal set, satisfying a certain condition.

The family of granules forms a subsystem of the power set of the universal set. Thus, based on the extended formulas above, we investigate the formulation representation of granules as follows:

Definition 3: Given an information system IS = (U, A), let a function h(a, M) denote a object set whose attribute value is equal to M on IS, where $M \subseteq V_a$ and $a \in A$, i.e., $h(a, M) = \{x | a(x) \in M, x \in U\}$. Then a granule of the IS is defined as follows: Gr = ((a, M), h(a, M)), where (a, M) refers to the intension of granule Gr and h(a, M) represents the extension of granule Gr. Here, if $M = V_a$, then Gr is called an elementary granule of attribute a on IS.

Definition 4: Given an information system IS = (U, A), let $B = \{a_1, a_2, \ldots, a_k\} \subseteq A$ be a subset of attributes and $\{(a_1, M_1)\ (a_2, M_2), \ldots\ (a_k, M_k)\}$ such that $M_i \in V_{ai}$ is a set of attribute values corresponding to B. Then, the intension of a granule can be defined as $\omega = \{(a_1, M_1)\ (a_2, M_2), \ldots\ (a_k, M_k)\}$ and the extension can be defined as $h(\omega) = \{u | f(u, a_1) = M_1 \wedge f(u, a_2) = M_2 \wedge \ldots \wedge f(u, a_k) = M_k, u \in U, a_i \in B, i \in \{1, 2, \ldots, k\}\}$. So that Gr = (ω, h(ω)) is called a combination granule. Here, h(ω) describes the internal structure of the granule. The collection of the extensions of all granules is denoted by GK. The map: U→GK is called the granulation of the IS.

Definition 5: Given an information system IS = (U, A) with a combination granule Gr = (ω, h(ω)), the decomposition operation of the Gr is denoted by $Dec(Gr) = (Gr_1, Gr_2, \ldots, Gr_n)$, where $Gr_i = ((a_i, M_i), h(a_i, M_i))$, $i \in \{1, 2, \ldots, n\}$.

Definition 6: Given an information system IS = (U, A) with $B = \{a_1, a_2, \ldots, a_n\} \subseteq A$, let $Gr_i = (\omega_i, h(\omega_i))$ and $Gr_j = (\omega_j, h(\omega_j))$ be two arbitrary combination granules on IS, where $i \neq j$, $\omega_i = \{(a_1, M_1)\ (a_2, M_2), \ldots\ (a_n, M_n)\}$ and $\omega_j = \{(a_1, N_1)\ (a_2, N_2), \ldots\ (a_n, N_n)\}$. Then, the intersection operation of Gr_i and Gr_j is denoted by $Gr_i \wedge Gr_j = (\omega_k, h(\omega_k))$, where $\omega_k = \{(a_1, M_1 \wedge N_1)\ (a_2, M_2 \wedge N_2), \ldots\ (a_n, M_n \wedge N_n)\}$.

Definition 7: Given an information system IS = (U, A) with $B = \{a_1, a_2, \ldots, a_n\} \subseteq A$, let $Gr_i = (\omega_i, h(\omega_i))$ and $Gr_j = (\omega_j, h(\omega_j))$ be two arbitrary combination granules on IS, where $i \neq j$, $\omega_i = \{(a_1, M_1)\ (a_2, M_2), \ldots\ (a_n, M_n)\}$ and $\omega_j = \{(a_1, N_1)\ (a_2, N_2), \ldots\ (a_n, N_n)\}$. Then, the union operation of Gr_i and Gr_j is denoted by $Gr_i \vee Gr_j = (\omega_k, h(\omega_k))$, where $\omega_k = \{(a_1, M_1 \vee N_1)\ (a_2, M_2 \vee N_2), \ldots\ (a_n, M_n \vee N_n)\}$.

Theorem 1: Let IS = (U, A) be an information system with $B = \{a_1, a_2, \ldots, a_n\} \subseteq A$. If $Gr_i = (\omega_i, h(\omega_i))$ and $Gr_j = (\omega_j, h(\omega_j))$ are two arbitrary combination granules on IS, where $i \neq j$, $\omega_i = \{(a_1, M_1)\ (a_2, M_2), \ldots\ (a_n, M_n)\}$ and $\omega_j = \{(a_1, N_1)\ (a_2, N_2), \ldots\ (a_n, N_n)\}$. Then, one has that $Gr_i \wedge Gr_j = \wedge(Dec(Gr_i) \vee Dec(Gr_j))$ and $Gr_i \vee Gr_j = \vee(Dec(Gr_i) \wedge Dec(Gr_j))$.

Proof: From Definition 5, one has that $Dec(Gr_i) = \{Gr_{i1}, Gr_{i2}, \ldots, Gr_{in}\}$, where $Gr_{i1} = ((a_1, M_1), h(a_1, M_1))$, $Gr_{i2} = ((a_2, M_2), h(a_2, M_2))$, …, $Gr_{in} = ((a_n, M_n), h(a_n, M_n))$. Similarly, $Dec(Gr_j) = \{Gr_{j1}, Gr_{j2}, \ldots, Gr_{jn}\}$, where $Gr_{j1} = ((a_1, N_1), h(a_1, N_1))$, $Gr_{j2} = ((a_2, N_2), h(a_2, N_2))$, …, $Gr_{jn} = ((a_n, N_n), h(a_n, N_n))$. Then, we can obtain that $\wedge(Dec(Gr_i) \vee Dec(Gr_j)) = \wedge\{Gr_{i1}, Gr_{i2}, \ldots, Gr_{in}, Gr_{j1}, Gr_{j2}, \ldots, Gr_{jn}\} = (Gr_{i1} \wedge Gr_{j1}) \wedge (Gr_{i2} \wedge Gr_{j2}) \wedge \ldots \wedge (Gr_{in} \wedge Gr_{jn}) = ((a_1, M_1 \wedge N_1), h(a_1, M_1 \wedge N_1)) \wedge ((a_2, M_2 \wedge N_2), h(a_2, M_2 \wedge N_2)) \wedge \ldots \wedge ((a_n, M_n \wedge N_n), h(a_n, M_n \wedge N_n))$. It follows from Definition 6 that $Gr_i \wedge Gr_j = (\omega_k, h(\omega_k))$, where $\omega_k = \{(a_1, M_1 \wedge N_1)\ (a_2, M_2 \wedge N_2), \ldots\ (a_n, M_n \wedge N_n)\}$. Hence, $\wedge(Dec(Gr_i) \vee Dec(Gr_j)) = Gr_i \wedge Gr_j$, i.e., $Gr_i \wedge Gr_j = \wedge(Dec(Gr_i) \vee Dec(Gr_j))$ holds. Like this proof above, the equation $Gr_i \vee Gr_j = \vee(Dec(Gr_i) \wedge Dec(Gr_j))$ can be proved.

From the definitions and the theorem above, the following property can be obtained. Given an information system IS = (U, A), the following property holds.

Property 1: Let IS = (U, A) be an information system with $B = \{a_1, a_2, \ldots, a_n\} \subseteq A$. If $Gr_i = (\omega_i, h(\omega_i))$ and $Gr_j = (\omega_j, h(\omega_j))$ are two arbitrary combination granules on IS. Then $Gr_i = Gr_j$ if and only if $Gr_i \wedge Gr_j = Gr_i \vee Gr_j$.

Definition 8: Given an information system IS = (U, A) with $B = \{a_1, a_2, \ldots, a_n\} \subseteq A$, let $Gr_i = (\omega_i, h(\omega_i))$ and $Gr_j = (\omega_j, h(\omega_j))$ be two arbitrary combination granules on IS. If $Gr_i \wedge Gr_j = Gr_j$ and $Gr_i \vee Gr_j = Gr_i$, then the Gr_i is a parent granule of the Gr_j, or the Gr_j is a child granule of the Gr_i, denoted by $Gr_j \prec Gr_i$.

Property 2: Let IS = (U, A) be an information system with $B = \{a_1, a_2, \ldots, a_n\} \subseteq A$ and Gr = (ω, h(ω)) be a combination granule on IS. Then, one has that $Gr \prec K$ for any $K \in Dec(Gr)$.

Proof: Given a combination granule Gr = (ω, h(ω)) on IS, suppose any $K \in Dec(Gr)$, it follows from Definition 5 that $K = ((a_i, M_i), h(a_i, M_i))$ $(i \in \{1, 2, \ldots, n\})$. Since it is obvious that $\omega \wedge (a_i, M_i) = \omega$, then we have that $Gr \wedge K = Gr$. And since $\omega \vee (a_i, M_i) = (a_i, M_i)$, then we have that $Gr \vee K = K$. Hence, it shows from Definition 8 that K is a parent granule of Gr, i.e., $Gr \prec K$.

In the following, we show the performances of granules which are proposed above in an information system through an illustrative example.

Table 1: Information system

U	a	b	c
x_1	1	3	0
x_2	-1	0.5	1
x_3	3	1	2
x_4	0	0	4
x_5	5	0	1
x_6	2	1	1

Example 1: Let IS = (U, A) be an information system, shown in Table 1, where $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, A = {a, b, c}.

From Table 1, it can be observed that the atomic formula of attribute a with the value greater than 0 can be expressed as (a, {a(x)| a(x)>0}) which can be written more simply as {a(x)>0}. Similarly, the atomic formula of attribute b with the value not equal to 0 can be expressed as {b(x)≠0} and that of attribute c with the value equal to 1 can be expressed as {c(x) = 1}. Thus, the granule Gr_1 can be expressed as $(\omega_1, h(\omega_1))$, where $\omega_1 = \{(a, a(x)>0)\ (b, b(x)\neq 0)\ (c, c(x) = 1)\}$. Similarly, the granule Gr_2 can be expressed as $(\omega_2, h(\omega_2))$, where $\omega_2 = \{(a, a(x)<3), (c, c(x)>1)\}$. Therefore, one has that $Gr_1 \wedge Gr_2 = (\omega_3, h(\omega_3))$, where $\omega_3 = \{(a, 0<a(x)<3)\ (b, b(x)\neq 0)\ (c, c(x) = 1)\}$. Similarly, $Gr_1 \vee Gr_2 = (\omega_4, h(\omega_4))$, where $\omega_4 = \{(c, c(x)>0)\}$.

GRANULAR SPACES AND GRANULE STRUCTURE MODEL

It is known that knowledge structures can be built based on concepts and a concept is considered to be the basic unit of human thought and knowledge. Then a concept can be conveniently interpreted as a granule. The representation, interpretation, connection and organization of concepts lead to granular structures (Yao, 2004b). 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.

Let IS = (U, A) be an information system. Suppose that G is a finite set of elementary granules on IS and an elementary granule ((a, M), h(a, M))∈G, then we establish two maps between the object set U and the elementary granule set G as f: G→U and g: U→G. The main objective is to construct a new granular space. Based on these notions above, we formally define a granular space as follows.

Definition 9: Given an information system IS = (U, A), let $G \subseteq 2^U$ be a family of elementary granules, i.e., $G = \{((a, M), h(a, M)) | a \in A, M \subseteq V_a\}$. If f and g satisfy the following properties: $f(Gr_1 \cup Gr_2) = f(Gr_1) \cap f(Gr_2)$ for any $Gr_1, Gr_2 \subseteq G$ and $g(X_1 \cup X_2) = g(X_1) \cap g(X_2)$ for any $X_1, X_2 \subseteq U$, then a triplet GS = (U, G, I) is called a granular space, where $I = U \times G$ is a binary relation between objects and granules.

Definition 10: Given an information system IS = (U, A), let GS = (U, G, I) be a granular space. For any elementary granule Gr∈G and x∈U, if (x, Gr)∈I, then the object x satisfies Gr and if (x, Gr)∉I, then the object x does not satisfy Gr.

It is known that we present a general framework of granular space above. Within the framework, we examine the granular structure model in rough sets. The formation of granular structures is based on a vertical separation of levels and a horizontal separation of granules in each level (Yao *et al*., 2009). 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. From the definitions above, the following property can be obtained.

Property 3: Let IS = (U, A) be an information system with a granular space GS = (U, G, I). For any $X, X_1, X_2 \subseteq U$ and $Gr, Gr_1, Gr_2 \subseteq G$, one has that:

1: $Gr_1 \subseteq Gr_2 \Rightarrow f(Gr_2) \subseteq f(Gr_1)$
2: $X_1 \subseteq X_2 \Rightarrow g(X_2) \subseteq g(X_1)$
3: $Gr \subseteq g(f(Gr))$
4: $X \subseteq f(g(X))$
5: $f(Gr_1) \cup f(Gr_2) \subseteq f(Gr_1 \cap Gr_2)$

Proof:

1: Suppose $Gr_1 \subseteq Gr_2$, one has that $Gr_1 \cup Gr_2 = Gr_2$. Then we have $f(Gr_1 \cup Gr_2) = f(Gr_2)$. It follows from Definition 9 that $f(Gr_1 \cup Gr_2) = f(Gr_1) \cap f(Gr_2)$. Since $f(Gr_1 \cup Gr_2) = f(Gr_2)$, then $f(Gr_1) \cap f(Gr_2) = f(Gr_2)$. And since $f(Gr_1) \cap f(Gr_2) \subseteq f(Gr_1)$, then one has that $f(Gr_2) \subseteq f(Gr_1)$
2: It is similar to the proof of (1)
3: It is straightforward
4: It is straightforward
5: Suppose $Gr_1 \subseteq Gr_2$, one has that $Gr_1 \cap Gr_2 \subseteq Gr_1$ and $Gr_1 \cap Gr_2 \subseteq Gr_2$. It follows from (1) that

$f(Gr_1) \subseteq f(Gr_1 \cap Gr_2)$ and $f(Gr_2) \subseteq f(Gr_1 \cap Gr_2)$. Then, we have that $f(Gr_1) \cup f(Gr_2) \subseteq f(Gr_1 \cap Gr_2)$

Definition 11: Given an information system IS = (U, A) with $B = \{a_1, a_2, \ldots, a_n\} \subseteq A$, let $Gr = (\omega, h(\omega))$ be a combination granule on IS. If $Dec(Gr) = \{Gr_1, Gr_2, \ldots, Gr_n\}$, then the granular space corresponding to Gr on IS is expressed as (U', G', I), where $U' = U - \{x|(x, Gr_i) \notin I, Gr_i \in Dec(Gr)\}$, $G' = Dec(Gr)$.

Definition 12: Given an information system IS = (U, A), let (U', G', I) be a granular space corresponding to a combination granule Gr on IS. For any elementary granule $Gr \subseteq G'$, we define (Gr, f(Gr)) as a binary pair. If Gr = g(f(Gr)) holds, then the binary pair is regarded as a child granule node of the Gr.

Theorem 2: Let IS = (U, A) be an information system with a granular space (U', G', I) corresponding to a combination granule Gr on IS. For any $Gr_{n-1} \subseteq G$ and $X_{n-1} \subseteq U$ (n>1), we construct the corresponding binary pair (Gr_{n-1}, X_{n-1}). In this case, the finite iterative operations for each binary pair can be stated by: $X_n = X_{n-1} \cup f(Gr_{n-1})$, $Gr_n = g(X_n)$, $Gr_{n+1} = Gr_n \cup g(X_n)$ and $X_{n+1} = f(Gr_{n+1})$, then there exists a child granule node $(Gr_{n'}, f(Gr_{n'}))$ corresponding to each binary pair.

Proof: It is obvious that $|X_n|$ monotonically increases in this iterative operations above, where $|X_n|$ denotes the number of objects of X_n. Because the universe U is a finite set, there must exist $X_{n'} = X_{n'+1} \subseteq U$. Then, it follows that $X_{n'} = X_{n'+1} = X_n \cup f(Gr_{n'})$. And by the operations above, one further has that $Gr_{n'} = g(X_{n'})$ and $Gr_{n'} = Gr_{n'} \cup g(X_{n'}) = g(X_{n'})$. Since $X_{n'} = X_{n'+1}$, then $Gr_{n'} = g(X_{n'+1})$. And it follows that $X_{n'+1} = f(Gr_{n'+1})$. Hence, we can obtain that $(Gr_{n'}, f(Gr_{n'}))$ is a child granule node.

Definition 13: Given an information system IS = (U, A), let (U', G', I) be a granular space corresponding to a combination granule Gr on IS. Suppose that $R = \{(Gr, f(Gr))|Gr \subseteq G', Gr = g(f(Gr))\}$ is a set which is constructed by all child granule nodes of the granular space, then there must exist only one partial relation set $(R, \prec)$ corresponding to R. And there exists only one minimal parent granule node (also called infimum) and one maximal child granule node (also called supremum) for each child granule node in $(R, \prec)$. A knowledge granular structure generated by this partial relation set is called a projection of the Gr on IS.

According to Theorem 2 and Definition 13, in the granular space corresponding to a combination granule, there always exists only one minimal infimum and one maximal supremum for each child granule node. Then we can obtain the projection corresponding to the combination granules in information systems. It shows that this projection can describe the detail hierarchical structure of the characteristic granules in information systems. Here, in accordance with the principles that the child node is beneath the parent one and two nodes are connected with a line, then we can obtain a hierarchical structure chart of the characteristic granules. In what follows, we give an example to illustrate the performances of the definitions and theorems above.

Example 2: Consider an information system IS = (U, A), described in Table 2, where the user requirements are $V_a>0$, $V_b<5$, $V_c = 1$ and $V_e>0$.

From Table 2, the user requirements can be described as a characteristic granule $Gr = (\omega, h(\omega))$, where $\omega = \{(a, V_a>0), (b, V_b<5), (c, V_c = 1), (e, V_e>0)\}$. Then, by computing, it follows from Definition 11 that we can obtain the granular space (U', G', I) corresponding to Gr, where $U' = \{x_1, x_2, x_3, x_5, x_6\}$, $G' = \{Gr_1, Gr_2, Gr_3, Gr_4\}$. One has that $Gr_1 = ((a, V_a>0), h(a, V_a>0))$, $Gr_2 = ((b, V_b<5), h(b, V_b<5))$, $Gr_3 = ((c, V_c = 1), h(c, V_c = 1))$ and $Gr_4 = ((e, V_e>0), h(e, V_e>0))$. Thus, we have all the child granule nodes as follows: $(\{Gr_1, Gr_2, Gr_3, Gr_4\}, \{x_5\})$ $(\{Gr_1, Gr_2, Gr_4\}, \{x_1, x_5, x_6\})$ $(\{Gr_2, Gr_3, Gr_4\}, \{x_2, x_5\})$ $(\{Gr_1, Gr_2\}, \{x_1, x_3, x_5, x_6\})$ $(\{Gr_2, Gr_4\}, \{x_1, x_2, x_5, x_6\})$ $(\{Gr_2\}, \{x_1, x_2, x_3, x_5, x_6\})$. So that we can draw a hierarchical structure chart of granular space corresponding to the characteristic granule Gr, shown in Fig. 1.

Table 2: Information system

U	a	b	c	d	e
x_1	3	1	0	0	1
x_2	-2	3	1	0	3
x_3	5	0	-1	-1	0
x_4	0	7	0	0	0
x_5	2	1	1	1	1
x_6	1	2	2	2	1

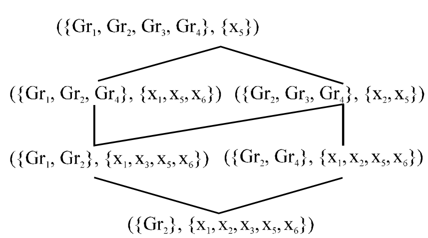


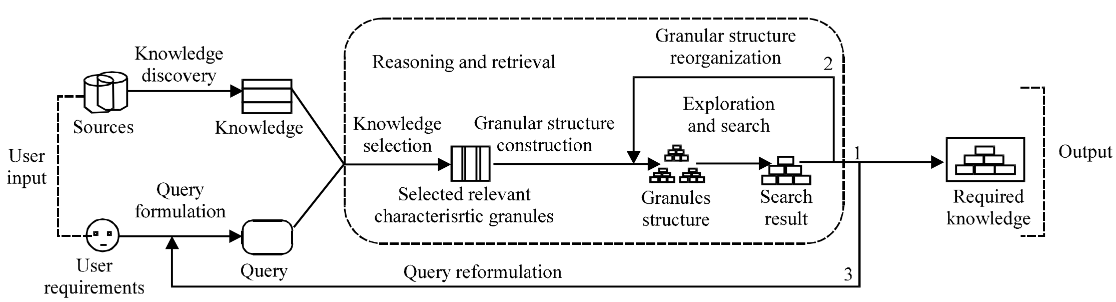
Fig. 1: Hierarchical structure corresponding to characteristic granules

KNOWLEDGE RETRIEVAL BASED ON GRANULE STRUCTURE

Knowledge represented in a structured way is consistent with human thoughts and is easily understandable. Sometimes, users do not know exactly what they want or are lack of contextual awareness (White *et al*., 2006). If knowledge can be provided visually in a structured way, it will be very useful for users to explore and refine the query (LaBrie, 2004). One of the key features of knowledge retrieval is that knowledge are visualized in a structured way so that users could get contextual awareness of related knowledge and make further retrieval (Yao *et al*., 2007). Therefore, our knowledge retrieval system based on granular structure concentrates on how to provide and use knowledge in more convenient ways. Then, this method of problem solving is to match query requirements of users with selected relevant characteristic granules of information systems using corresponding knowledge selection algorithms and then return the search results which are required knowledge to the users.

To this end, all objects are formalized as an information system and the query requirements of users can be formalized as characteristic granule $Gr = (\omega, h(\omega))$. Then, it follows from Definitions 11-13 that we can obtain the projection corresponding to the combination granules in the information system. We need to point out that the full process contains three parts of feedbacks. First is to obtain a set of objects of characteristic granule Gr which completely satisfies the query requirements of user, i.e., $h(\omega)$. That is, the part is the certain query results. Second is to delete a set of objects which does not completely satisfy the query requirements of user. Third is to obtain a set of objects of characteristic granule Gr which incompletely satisfies the query requirements of user. That is, the last part is an approximate query results which are obtained by rank. Thus, a unit of knowledge may be decomposed into a family of smaller units. Their decomposition and relationship represent the internal structures. The collective structures of a family of knowledge units describe the relations of knowledge units in the same level. Different levels of knowledge units form a partial ordering. The hierarchical structures describe the integrated whole of a web of knowledge units from a very high level of abstraction to the very nest details. So our granular structure discovered and constructed by a knowledge retrieval should be in multilevels and multiviews. Therefore, this paper constructs a new granular structure and improves the adaptability of the granule and then a knowledge retrieval system based on the novel granular structure model is proposed, a process of which is shown in Fig. 2.

Figure 2 shows a conceptual framework of a knowledge retrieval system based on the granular structure model. The main process can be described as follows: (1) Knowledge Discovery: Discovering knowledge from sources by data mining, machine learning, knowledge acquisition and other methods. (2) Query Formulation: Formulating queries from user requirements by user inputs. (3) Knowledge Selection: Selecting the range of possible related knowledge based on user query and knowledge discovered from data/information sources and formalizing the query and knowledge as granules to select relevant characteristic granules. (4) Granular Structure Construction: Reasoning and retrieval according to different views of knowledge, domain knowledge, user background, etc. in order to form granular structures. Domain knowledge can be provided by expert systems. User background and preference can be provided by user logs. And all these relevant characteristic granules form granular structures of different granularity levels. (5) Exploration and Search: Exploring the granular structure to get general awareness and refine the search. Through understanding the relevant granular structures, users can search into details on what they are interested in to get the required knowledge. (6) Granular Structure Reorganization:



1: User requirements are satisfied 2: Views of granuler structure need to be changed 3: Query needs to be reformulated

Fig. 2: Knowledge retrieval based on granular structure

Reorganizing granular structures if users need to explore other views of selected granules. (7) Query Reformulation: Reformulating the query if the constructed structures cannot satisfy user requirements.

CONCLUSIONS

Granular computing, a field of study that aims at extracting the commonality of human and machine intelligence from the viewpoint of granularity, emphasizes that human can always focus on appropriate levels of granularity and perspectives, ignoring irrelevant information in order to achieve effective problem solving. Granules can be grouped into multiple levels to form a hierarchical granular structure and the hierarchy can also be built from multiple perspectives. And various hierarchical processing can be performed on the organized structures. Thus, this study improves the formulation representation of granules, so that the granular computing model is more practical. Based on that, we construct granular spaces which are projected to information systems and then obtain the granular structure model. From the viewpoints of user interests and granular information processing, a conceptual framework of knowledge retrieval system and its main process are described detailedly to further enlarge the application areas of granular computing.

ACKNOWLEDGMENTS

The authors wish to thank Editors and the anonymous reviewers for their hard work on this study. This work was supported by the National Natural Science Foundation of China (No. 60873104, No. 61040037), the Key Project of Science and Technology Department of Henan Province (No. 112102210194), the Science and Technology Research Key Project of Educational Department of Henan Province (No. 12A520027) and the Education Fund for Youth Key Teachers of Henan Normal University.

REFERENCES

Chen, T.S., J. Chen and Y.H. Kao, 2010. A novel hybrid protection technique of privacy-preserving data mining and anti-data mining. Inform. Technol. J., 9: 500-505.

Hobbs, J.R., 1985. Granularity. Proceedings of the 9th International Joint Conference on Artificial Intelligence, August, 1985, Los Angeles, pp: 432-435.

LaBrie, R.C., 2004. The impact of alternative search mechanisms on the effectiveness of knowledge retrieval. Ph.D. Thesis, Arizona State University.

Lin, T.Y., 2006. Granular computings II: infrastructures for AI-engineering. Proceedings of the IEEE International Conference on Granular Computing, May 10-12, 2006, USA., pp: 2-7.

Liu, Q. and Q. Liu, 2004. Granules and applications of granular computing in logical reasoning. Chin. J. Comput. Res. Dev., 41: 546-551.

Liu, Q., H. Sun and Y. Wang, 2007. Granulations based on semantics of rough logical formulas and its reasoning. Proceedings of the International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, May 14-16, 2007, Toronto, Canada, pp: 419-426.

Liu, Q., S.H. Liu and F. Zheng, 2001. Rough logic and its applications in data reduction. Chin. J. Software, 3: 415-419.

Ma, J.M., W.X. Zhang, Y. Leung and X.X. Song, 2007. Granular computing and dual Galois connection. Inform. Sci., 177: 5365-5377.

Miao, D.Q., G.Y. Wang and Q. Liu, T.Y. Lin and Y.Y. Yao, 2007. Granular Computing: Past, Present and the Future Perspectives. China Academic Press, Beijing..

Pal, S.K., B.U. Shankar and P. Mitra, 2005. Granular computing, rough entropy and object extraction. Pattern Recogn. Lett., 26: 2509-2517.

Pawlak, Z., 1991. Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht, Boston, ISBN 0792314727, Pages: 229.

Pei, D.W., 2007. Some models of granular computing. Proceedings of the International Conference on Granular Computing, November 2-4, 2007, San Jose, California, USA., pp: 11-16.

Qi, J.J. and L. Wei, 2008. Attribute reduction in consistent decision formal context. Inform. Technol. J., 7: 170-174.

Sun, L., J.C. Xu, Z.A. Xue and L.J. Zhang, 2011. entropy-based feature selection and its application. J. Inform. Comput. Sci., 8: 1525-1532.

White, R.W., B. Kules, S.M. Drucker and M.C. Schraefel, 2006. Supporting exploratory search. Commun. ACM, 49: 37-39.

Wu, W.Z., Y. Leung and J.S. Mi, 2009. Granular computing and knowledge reduction in formal contexts. IEEE Trans. Knowl. Data Eng., 21: 1461-1474.

Yan, L. and Q. Liu, 2007. A logical method of formalization for granular computing. Proceedings of the IEEE International Conference on Granular Computing, November 2-4, 2007, USA., pp: 22-27.

Yao, Y.Y. and J.T. Yao, 2002. Granular computing as a basis for consistent classification problems. Proceedings of PAKDD 2002 Workshop on Toward the Foundation of Data Mining, May 6-8, 2002, Taipei, Taiwan, pp: 101-106.

Yao, Y.Y., 2004a. A partition model of granular computing. Trans. Rough Sets, 1: 232-253.

Yao, Y.Y., 2004b. Concept formation and learning: A cognitive informatics perspective. Proceedings of the 3rd IEEE International Conference on Cognitive Informatics, August 16- 17, 2004, Victoria, Canada, pp: 42-51.

Yao, Y.Y., Y. Zeng, N. Zhong and X.J. Huang, 2007. Knowledge retrieval. Proceedings of the Interernational Conference on Web Intelligence, November 2-5, 2007, Silicon Valley, USA., pp: 729-735.

Yao, Y.Y., D.Q. Miao and F.F. Xu, 2009. Granular structures and approximations in rough sets and knowledge spaces. Rough Set Theory, 174: 71-84.

Zadeh L.A., 1997. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst., 90: 111-127.

Zeng, Y., N. Zhong, Y. Wang, Y. Qin and Z. Huanget *et al.*, 2011. User-centric query refinement and processing based on granularity related strategies. Knowl. Inform. Syst., 27: 419-450.

Zhang, L. and B. Zhang, 2003. The quotient space theory of problem solving. Framework 59: 11-15.