Granularity-Based User-Centric Multi-Strategies and Application in Knowledge Retrieval

Lin Sun[1,2], Ning Zhong[1], Jiucheng Xu[2]
[1]International WIC Institute, Beijing University of Technology, Beijing 100124, P. R. China
[2]College of Computer & Information Engineering, Henan Normal University, Xinxiang 453007, P. R. China
Email: linsunok@gmail.com

***Abstract*—Granular computing is a general computing paradigm of problem solving for effectively using granules in problem solving. From the viewpoint of granularity, this paper presents a new granular computing data cycle model in which information granule construction, granular space, information granule operations, and knowledge acquisition approach are proposed respectively. To do a more user oriented way, a multi-strategies model, namely, Base-Level learning strategy, user interest retention strategy, and starting point strategy, introduces to unify search and reasoning for effective problem solving. Then, we discuss granularity-based unification using three strategies mentioned above and show three-levels of granularity from human problem solving to a wide variety of user needs satisfied. Furthermore, on the basis of user-centric multi-strategies and granular information processing, we develop a conceptual framework of granularity-based knowledge retrieval model, which enlarges the application areas of granular computing.**

***Index Terms*—granular computing, information granule, granularity, multi-strategies, knowledge retrieval**

I. Introduction

With rapid development of the Internet and the Internet of Things, a new world, called hyper world, is emerging by coupling and empowering humans in the social world, information/computers in the cyber world, and things in the physical world [1]. The notion of Wisdom Web of Things (W2T) is a novel vision for computing and intelligence in the post-WWW era with data, connection and service explosions, recently put forward by a group of leading researchers from the fields of Web Intelligence, Ubiquitous Intelligence, Brian Informatics, and Cyber Individual, and then a W2T data cycle system is proposed to drive the cycle, namely, from things to data, information, knowledge, wisdom, services, humans, and back to things [2]. Furthermore, the emerging Web of Things provides a transmission channel of data cycle. Thus, the core problem of data cycle construction mentioned above is to develop a highly efficient data cycle system. Fast-evolving Web intelligence research and development initiatives are now moving toward understanding the multifaceted nature of intelligence in depth and incorporating it at Web scale and in a ubiquitous environment [2]. As we know, many existing Web intelligence technologies mainly focus on the conversion and utilization of data to provide users with more intelligent services on the Internet/Web or SEA-nets. However, most of these studies are limited in some specific technologies about data conversion mechanism. Then, using only these existing technologies cannot fully solve the problems of enormous data and realize holistic intelligence for the harmonious symbiosis of humans, computers and things in the emerging hyper world.

In recent years, one of the major problems is that acquiring all the relevant data is hard when the data goes to Web scale. By using the approach unifying reasoning and search [3], the search will help to gradually select a small right set of data that is further provided for reasoning. Thus, how to search for a good or more relevant subset of data and do reasoning on it becomes a very important problem for us. Meanwhile, since user's backgrounds and expectations may be differential, the same strategy cannot meet the diversity of their needs. Therefore, it is desirable to propose an approach using various strategies to unify search and reasoning for user needs.

Granulation is a natural problem-solving methodology deeply rooted in human thinking [4]. It is well known that granular computing is a general computing paradigm of problem solving for efficiently using granulation in problem solving. The basic principles and ideas of granular computing consist of granulation that granulates problems and computations into granules. Then, 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 views, ignoring irrelevant information in order to achieve effective problem solving [5]. A granule is a set of elements that are drawn together by their equality, similarities, indistinguishability from some aspects [6-8]. Granules can be grouped into multiple levels to form a hierarchical granular structure, and the hierarchy can also be built from multiple perspectives [5]. Then, some granular computing models [9,10] are studied. Although these models which copy the human instinct of information processing can increase classification performance, they use the concept of sub-attributes to describe information granules which are collections of objects arranged together based on their similarity, functional adjacency and indistinguishability [10-13].

doi:10.4304/jcp.8.8.2101-2109

Following the perspective of granularity, we present a new granular computing data cycle model which can discover knowledge from information granules. The model involves information granule construction, rough set-based granular space, information granule operations, and knowledge acquisition approach.

Basically speaking, reflecting the limited ability of human brains, perceptions are inaccurate, and in more concrete terms, perceptions are granular [10]. It means that the boundaries of perceived classes are not sharp, and the values of attributes are granulated [14]. This is especially true when a problem involves vague, uncertain, or incomplete information. It may be sometimes difficult to differentiate distinct elements, and so one is forced to consider information granules [10,13-16]. Zadeh [17] summarized the reasons/situations why we need to process perception-based information. Many researchers paid lots of attentions on uncertainty/vagueness in human decision making, such as fuzzy sets, rough sets, granular computing, etc. [14]. However, these researches are not intended to replace traditional measurement information-based methods which operate numerical data, and their purposes are to let the developed computational theories refer to reality [10]. Thus, an advanced/integrated mechanism should be provided to imitate human ability of processing information, such as extracting knowledge from information granules and making a decision based on them. Furthermore, various hierarchical processing can be performed on the organized structures, and the scattered large-scale data can be grouped together as granules in different levels or under different views to meet various user needs [18]. Following the above-mentioned inspirations, we propose a user-centric approach to refine and process user's query based on cognitive science and granular computing. From the viewpoints of granularity, we extent Base-Level Learning [19,20], user interest retention [18], and starting point [18] as multi-strategies model for unifying user driven search and reasoning under time granule constraints. Then, inspired by user-centric and granular information processing for human problem-solving, we develop a conceptual framework of granularity-based knowledge retrieval model.

The rest of this paper is organized as follows. Section II reviews granular computing, presents a new granular computing data cycle model, and proposes information granule construction, granular space, information granule operations, and knowledge acquisition approach. Section III extends Base-Level Learning strategy, user interest retention strategy, and starting point strategy. Then, granularity-based unification of multi-strategies and three-levels of granularity are discussed. In Section IV, a conceptual framework of granularity-based knowledge retrieval model is described. Finally, the conclusions are described in Section V.

II. GRAUNLAR COMPUTING

Granular computing deals with the processing of complex problems from two different aspects [21]. The first aspect is a problem granulation that granulates a complex problem into granules, including the construction, representation, and interpretation of granules, while the second aspect is to compute and reason these granular entities in terms of intra-relationships within these entities and interrelationships or interactions between them. Castellano and Fanelli [13] indicated that the main issues of granular computing are how to construct the information granules and how to describe information granules. Another particular question that arises is how to determine the level of granularity. To represent information granules and determine the level of granularity, Bargiela and Pedrycz [11] proposed the "hyper box" and ''inclusion and compatibility" to measure information granules. Su et al. [22] presented ''sub-features" to describe information granules and ''*H*-index and *U*-ratio" to determine the level of granularity. However, most of granular computing related researches focused more on such fundamental issues. Therefore, if we want to acquire knowledge from information granules, we must try to solve these questions mentioned above. As a concrete approach to problem-solving based on large-scale data, this process contains two major steps, namely, the search of relevant data and problem-solving based on searched data from the perspective of granularity-based multi-strategies. Then, this study presents a new granular computing data cycle model which can be helpful to implement the W2T data cycle system [2]. Our proposed methodology follows the procedure shown in Fig. 1 which involves ten steps. The details of these steps will be provided in the following subsections.

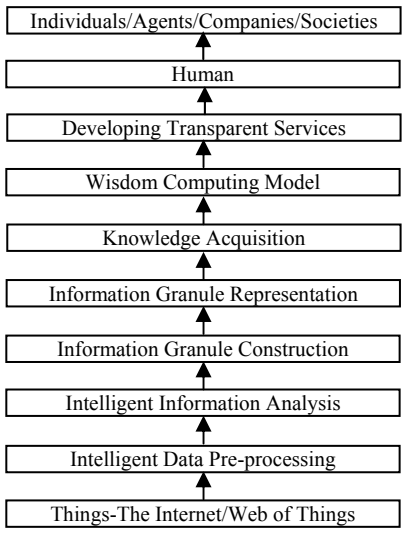

Figure 1. A new granular computing data cycle model

A. Information Granule Construction

The set-theoretic interpretation of granules can be formally developed based on the notion of concepts, a basic unit of human thought [23]. 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 [24]. The intension describes the

intrinsic properties or features shared by all objects, and the extension of a concept is a set of objects which are concrete examples of a concept.

Let U be a nonempty domain consisting of a set of individuals, denoted by $U = \{x, y, \ldots\}$. 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 [23]. If ϕ is a formula, the set $m(\phi)$ defined as: $m(\phi) = \{x \in U \mid x \models \phi\}$, is called the meaning of the formula ϕ [15]. The meaning of a formula ϕ is therefore 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 $m(\phi)$. Thus, a connection between formals and subsets of U is established.

An information system (*IS*) (or information table) represents all available information and knowledge. That is, objects are only perceived, observed, or measured by using a finite number of properties. Hence, a concept definable in an information system can be expressed by a pair $(\phi, m(\phi))$. More specifically, ϕ is a description of $m(\phi)$ in the information system, the intension of concept $(\phi, m(\phi))$, and $m(\phi)$ is the set of objects satisfying ϕ, the extension of concept $(\phi, m(\phi))$.

It is noted that a primitive notion of granular computing is that of granules. Granulation of a universe involves dividing the universe into subsets or grouping individual objects into clusters. A granule is a subset of the universe. A family of granules that contains every object in the universe is called a granulation of the universe. A granule contains both an intension (description of the granule properties) and an extension (elements that conform to that description) [25]. Then, we investigate the formulation representation of granules in information systems as follows.

Definition 1. An information granule (*IG*) is defined as the tuple $IG = (\varphi, m(\varphi))$, where φ refers to the intension of information granule *IG*, and $m(\varphi)$ represents the extension of information granule *IG*.

B. Granular Space

As we know, rough sets analysis studies relationships between objects and their feature values in an information system. An information system provides a convenient way to describe a finite set of objects by a finite set of features [19]. 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 features; V is the union of feature domains such that $V = \bigcup_{a \in A} V_a$ and V_a denotes the value domain of feature $a \in A$; f: $U \times A \rightarrow V$ is an information function which associates a unique value of each feature with every object belonging to U, so that for any $a \in A$ and $u \in U$, $f(u, a) \in V_a$. Also, $IS = (U, A, V, f)$ can be written more simply as (U, A).

Let $IS = (U, A)$ be an information system with $B = \{b_1, b_2, \ldots, b_k\} \subseteq A$. Suppose that $V_{b_i} = \{V_{b_{i,1}}, V_{b_{i,2}}, \ldots, V_{b_{i,k}}\}$ is the domain of feature b_i, and each $V_{b_{i,j}}$ may be viewed as a concept. Then, there must exist $\varphi = \{I_1, I_2, \ldots, I_k\}$ such that $I_i \in V_{b_i}$ is a set of feature values corresponding to B. Then, the intension of an information granule can be denoted by $\varphi = \{I_1, I_2, \ldots, I_k\}$, and the extension can be denoted by $m(\varphi) = \{u \in U | f(u, b_1) = I_1 \wedge f(u, b_2) = I_2 \wedge \ldots \wedge f(u, b_k) = I_k, b_i \in B, i \in \{1, 2, \ldots, k\}\}$. Here, $m(\varphi)$ describes the internal structure of the information granule.

Definition 2. Given an information system $IS = (U, A)$, its granular space (*GS*) can be defined as a quadruple on concepts:

$$GS = (U, S = \{(\varphi_n, m(\varphi_n)) | n \in \Gamma^+\}, \{R_i | i \in \Gamma^+\}, T),$$

where U is the universe, $S = \{(\varphi_n, m(\varphi_n)) | n \in \Gamma^+\}$ denotes the set of involved information granules, R_i denotes the relations between information granules and Γ^+ the set of positive integers, and $T = (\varphi_n, m(\varphi_n)) \times R_i \times (\varphi_{n'}, m(\varphi_{n'}))$ is called a function carrying on the junction between two concepts and binary relations.

Form Definition 2, it follows that different levels of relations among the above concepts induce a hierarchical structure called a granular structure. In particular, multi-levels granular structure needs to be considered, namely, internal structure of information granules, collective structure of a family of granules, hierarchical structure of from data to information and knowledge and so on. Meanwhile, there are additional requirements to make the granular space more practical. It is known that 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 the universe, and it also requires that the union of all the elementary granules covers the universe [23].

C. Information Granule Operations

Definition 3. Given an information system $IS = (U, A)$, let $IG = (\varphi, m(\varphi))$ be an information granule, and its size can be defined as the cardinality of the extension of the *IG*, i.e., $|m(\varphi)|$. Namely, the size may be interpreted as the degree of abstraction or concreteness.

Definition 4. Given an information system $IS = (U, A)$, let $IG = (\varphi, m(\varphi))$ be an information granule. If $\varphi = \{V_{b_{i,j}}\}$, where $V_{b_{i,j}} \in V_{b_i} = \{V_{b_{i,1}}, V_{b_{i,2}}, \ldots, V_{b_{i,k}}\}$ is a categorical value, and V_{b_i} is the domain of feature $b_i \in A$, then *IG* is called an elementary information granule of feature b_i, or an elementary granule for short. Namely, $m(\varphi) = \{u \in U | f(u, b_i) = V_{b_{i,j}}, b_i \in A\}$.

Definition 5. Given an information system $IS = (U, A)$, let $I = \{I_1, I_2, \ldots, I_k\}$ be a k-itemset, where $I_i \in V_{b_i}$ $(i = 1, 2, \ldots, k)$ is a feature value of feature b_i, and $B = \{b_1, b_2, \ldots, b_k\} \subseteq A$. Then, information granule $IG = (I, m(I))$

is called a k-itemset granule, where $m(I) = \{u \in U | f(u, b_1) = I_1 \wedge f(u, b_2) = I_2 \wedge \ldots \wedge f(u, b_k) = I_k, b_i \in B, i \in \{1, 2, \ldots, k\}\}$.

It should be ensured that a 1-itemset granule is an elementary granule satisfying the given conditions. From Definition 5, we can obtain the following properties.

Property 1. If $I = \{I_1, I_2, \ldots, I_k\}$ is a k-itemset, then $m(I) = m(\{I_1\}) \cap m(\{I_2\}) \cap \ldots \cap m(\{I_k\})$.

Property 2. If $I \subseteq V'$, $J \subseteq V'$, and $I \subseteq J$, then $m(J) \subseteq m(I)$.

Property 2 states that the size of information granule (I, $m(I)$) is greater than or equal to that of information granule (J, $m(J)$), or(I, $m(I)$) is more abstract than (J, $m(J)$) as far as their concepts are concerned. So that there exists inclusion relation between (I, $m(I)$) and (J, $m(J)$), denoted by $\preceq$, i.e., $(J, m(J)) \preceq (I, m(I))$.

Definition 6. Given a k_1-itemset granule (I, $m(I)$) and a k_2-itemset granule (J, $m(J)$), if $I \subseteq J \subseteq V'$, then $m(J) \subseteq m(I)$, i.e., the intension of (J, $m(J)$) is more concrete than that of (I, $m(I)$), denoted by $m(I) \prec m(J)$. All these granules lead to a hierarchical structure by using $\prec$ order, called a multi-dimensional granular hierarchy.

Definition 7. Given an information system $IS = (U, A)$, let a function $g(a, M)$ denote a object set whose feature value is equal to M in the IS, where $M \subseteq V_a$, and $a \in A$, i.e., $g(a, M) = \{u | f(u, a) \in M, x \in U\}$. Then a information granule of the IS can be also defined as follows: $IG = ((a, M), g(a, M))$, where (a, M) refers to the intension of information granule IG, and $g(a, M)$ represents the extension of information granule IG. Here, if $M = V_a$, then IG is called an elementary information granule of feature a on IS.

In an information system $IS = (U, A)$, let $B = \{b_1, b_2, \ldots, b_k\} \subseteq A$ be a subset of features and $\mu = \{(b_1, M_1), (b_2, M_2), \ldots, (b_k, M_k)\}$ such that $M_i \in V_{b_i}$ is a set of feature values corresponding to B. Then, the intension of a information granule can be defined as $\mu = \{(b_1, M_1), (b_2, M_2), \ldots, (b_k, M_k)\}$, and the extension can be defined as $g(\mu) = \{u | f(u, b_1) = M_1 \wedge f(u, b_2) = M_2 \wedge \ldots \wedge f(u, b_k) = M_k, u \in U, b_i \in B, i \in \{1, 2, \ldots, k\}\}$. So that $IG = (\mu, g(\mu))$ is called a combination information granule. Here, $g(\mu)$ describes the internal structure of the information granule. In the following, we introduce three types of operations to study some properties of information granules in an information system as follows:

(1) The decomposition operation of IG is denoted by $D(IG) = (Gr_1, Gr_2, \ldots, Gr_n)$, where $Gr_i = ((a_i, M_i), g(a_i, M_i))$, $1 \le i \le n$.

(2) The intersection operation of IGs is denoted by $Gr_i \wedge Gr_j = (\mu_k, g(\mu_k))$, where $Gr_i = (\mu_i, g(\mu_i))$, $Gr_j = (\mu_j, g(\mu_j))$, $i \neq j$, $\mu_i = \{(b_1, M_1), (b_2, M_2), \ldots, (b_n, M_n)\}$, $\mu_j = \{(b_1, N_1), (b_2, N_2), \ldots, (b_n, N_n)\}$, $\mu_k = \{(b_1, M_1 \wedge N_1), (b_2, M_2 \wedge N_2), \ldots, (b_n, M_n \wedge N_n)\}$.

(3) The union operation of IGs is denoted by $Gr_i \vee Gr_j = (\mu_k, g(\mu_k))$, where $Gr_i = (\mu_i, g(\mu_i))$, $Gr_j = (\mu_j, g(\mu_j))$, $i \neq j$, $\mu_i = \{(b_1, M_1), (b_2, M_2), \ldots, (b_n, M_n)\}$, $\mu_j = \{(b_1, N_1), (b_2, N_2), \ldots, (b_n, N_n)\}$, $\mu_k = \{(b_1, M_1 \vee N_1), (b_2, M_2 \vee N_2), \ldots, (b_n, M_n \vee N_n)\}$.

D. Knowledge Acquisition Approach

It is known that one of the key advantages in computing with granules is that our understanding of a particular problem can differ depending on the level of granulation [23]. However, what level of granularity is suitable? This work employs objective indexes, H-index and U-ratio, developed by Su et al. [22] to solve this question. Meanwhile, a large feature set often contains redundant and irrelevant information, and can actually degrade the performance of the classifier [26]. Therefore, one needs techniques to reduce the dimension of examples and should use either feature extraction, feature selection or a combination of the both [27]. Then, on the basis of Fig. 1, we further design a new methodology through information granulation, feature extraction, and feature selection to acquire knowledge. Briefly, our proposed method involves two major parts as follows: First, we introduce information granulation to reduce the data size. Second, the feature extraction and feature selection techniques are employed to reduce dimensions of features and then we acquire knowledge from these constructed information granules. The main advantage of our proposed method is to reduce both the size of features and the size of the data. Fig. 2 shows the procedure of the proposed methodology.

III. User-Centric Multi-Strategies Model

A. Base-Level Learning Strategy

Availability of human memories for specific items is sensitive to frequency and recency [28]. The Adaptive Control of Thought-Rational (ACT-R) theory is a cognitive architecture in which a production system coordinates the activity of modules associated with perception, memory, and action, and it has been particularly successful and influential [20]. The declarative memory in ACT-R is organized in chunks, and each chunk has a base-level activation (BLA) reflecting the frequency and recency of its use. Each new use of the chunk adds another term to the sum, which then decays independently with rate d, and the total count so far is denoted by n, and t_i is the time since the *i-th* use of the item [28]. Then, the Base-Level Learning equation is denoted by

$$B = \ln(\sum_{i=1}^{n} t_i^{-d}).$$

However, one serious practical drawback is that it is very expensive to compute, and ACT-R models have ground to a halt because of its complexity [28]. The Base-Level Learning equation is analytically intractable and its implementation requires a complete record of the exact time stamps of each use of each chunk. Therefore, a memory-efficient approximation is needed. Then, we introduce Z-test technique [29] to construct a new Base-Level Learning equation as follows.

Definition 8. Let U be a nonempty domain consisting of a set of individuals, $B = \{b_1, b_2, \ldots, b_n\}$, any $b_j, b_k \in B$. A new Base-Level Learning equation is defined as

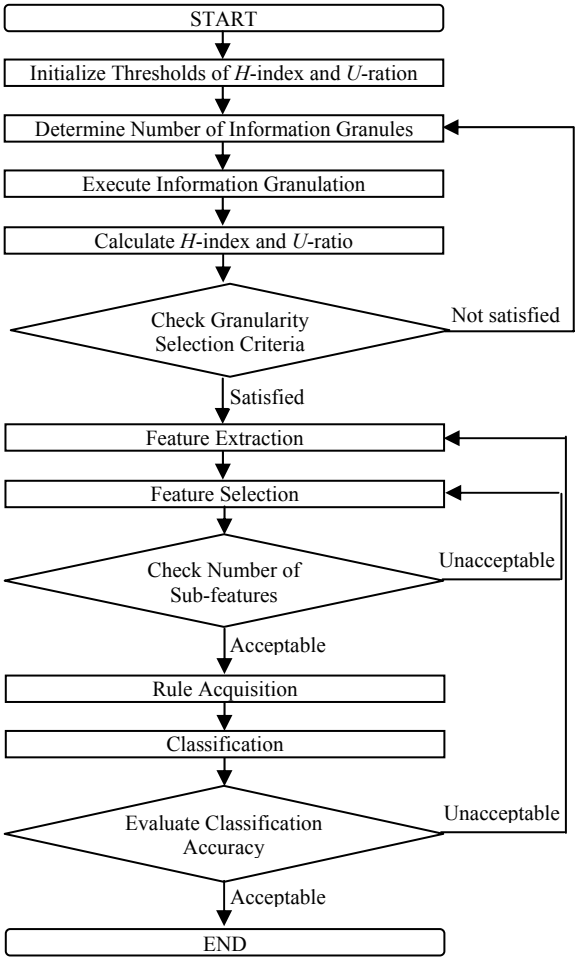

Figure 2. The procedure of the proposed methodology of knowledge acquisition

$$BLL(i,j,k) = \ln(\sum_{i=1}^{n} t_i^{-d}) + \sum_j W_j Z_{jk}$$
$$= \ln(\sum_{i=1}^{n} t_i^{-d}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}},$$

where t_i is the time granule since the *i-th* use of the item, W_j denotes the weight of feature b_j, Z_{jk} (Z-test) denotes the correlation intensity between feature b_j and feature b_k, $I(j,k) = \log\frac{p(j,k)}{p(j)p(k)} = \log\frac{p(k|j)}{p(k)}$ represents the mutual information between feature b_j and feature b_k, $E_j = \frac{1}{n}\sum I(j,k)$ represents the average value of the mutual information, and $\mu_j = \frac{1}{n}\sum(I(j,k) - E_j)^2$ presents the standard deviation.

The new Base-Level Learning equation uses time granule, mutual information and Z-test technologies, and takes two-step method to execute feature selection. The first step is to predicate selection, and the second step is to choose relevant dataset corresponding to each predicate. Thus, the new Base-Level Learning equation particularly estimates the decay effect during quiescent periods following bursts of activity through introducing correlation intensity. Hence, the improved equation proposed takes this critical information into account.

When a new use of the chunk occurs, it assumes index $i = 1$, all previous uses shift one index up, and time stamp t_{m+1} passes into oblivion [28]. Thus, based on the ideas of the new improved approximation in [28], the formula of new Base-Level Learning equation mentioned above can be transformed as follows:

$$BLL(i,j,k)$$
$$= \ln(\sum_{i=1}^{n} t_i^{-d}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}}$$
$$= \ln(\sum_{i=1}^{m} t_i^{-d} + \sum_{i=1}^{n-m} t_{m+i}^{-d}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}}$$
$$\approx \ln(\sum_{i=1}^{m} t_i^{-d} + \sum_{i=1}^{n-m} (t_m + \frac{t_n - t_m}{n-m} i)^{-d}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}}$$
$$\approx \ln(\sum_{i=1}^{m} t_i^{-d} + \int_0^{n-m} (t_m + \frac{t_n - t_m}{n-m} x)^{-d} dx) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}}$$
$$\approx \ln(\sum_{i=1}^{m} t_i^{-d} + \frac{(n-m)(t_n^{1-d} - t_m^{1-d})}{(1-d)(t_n - t_m)}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}}.$$

The reformulated equation formalizes these ideas to keep the exact timing of a few most recent events and ignore the details of the distant past. Then, the improved formula is nearly perfect even for $m = 1$. Furthermore, when the special case for $d = 0.5$ follows from the reformulated equation, then we can obtain that

$$\begin{aligned} & BLL(i,j,k) \\ & \approx \ln(\sum_{i=1}^{m} t_i^{-0.5} + \frac{(n-m)(t_n^{1-0.5} - t_m^{1-0.5})}{(1-0.5)(t_n - t_m)}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}} \\ & = \ln(\sum_{i=1}^{m} \frac{1}{\sqrt{t_i}} + \frac{2(n-m)(\sqrt{t_n} - \sqrt{t_m})}{(\sqrt{t_n} - \sqrt{t_m})(\sqrt{t_n} + \sqrt{t_m})}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}} \\ & = \ln(\sum_{i=1}^{m} \frac{1}{\sqrt{t_i}} + \frac{2(n-m)}{\sqrt{t_n} + \sqrt{t_m}}) + \sum_j W_j \frac{I(j,k) - E_j}{\sqrt{\mu_j}}. \end{aligned}$$

The reformulated equation above optimizes the new Base-Level Learning equation for the important special case d =0.5, which is the default decay parameter in ACT-R theory. Therefore, it follows that the new improve formula can track the theoretical primal Base-Level Learning equation closely, and reduce effectively the computational complexity. Thus, it can be helpful for allowing the ACT-R architecture theory to scale up to more realistic memory sizes and more prolonged learning periods [28].

B. User Interest Retention Strategy

User interests can be mined from either user logs or users public information, and described as a set of concepts that users are familiar with [18]. Since interest retentions may be related to user recent interests and possible queries from users, we introduce new user interest retention model based on the interest retention models [18] to track users' retained interests and examine them. Then, an extended model of measuring interest retentions based on the cumulative interest value is introduced as follows:

$$TI(i,n) = \sum_{j=1}^{n} AN(i,j),$$

where n is the number of time granule intervals (e.g. day, week, month, year) considered, $j \in \{1, 2, ..., n\}$ is a variable that is used to index each time granule interval, $AN(i, j)$ is the appearance number of the interest i in the time granule interval j, and $TI(i, n)$ reflects the value of total interest on i, namely, how many time granules have an interest appeared in the considered time granule intervals. However, the computation formula mentioned above may not correctly reflect user's current interests.

Ebbinghaus [30] proposed a forgetting curve for describing the forgetting mechanism of memory. Loftus suggests that the forgetting function satisfies an exponential formula $P = Ae^{-bT}$, where P represents the performance measure of memory retention, A and b are two parameters for the model, and T is the delay time [19,30,31]. The total activation of a chunk determines the time taken to retrieve the chunk from declarative memory, i.e. $t = Fe^{-A}$, where t is the retrieval time, A is the chunk's activation, and F is a parameter that scales retrieval time [32]. Then, within a time granule interval, an interest may appear several times, and each appearance of the interesting topic will have activation for the overall retained interests. Therefore, the retained interest value in user interest retention model is a sum of retention values from each previous appearance. Thus, based on the ideas of the upper formulas mentioned above, the user interest retentions can be defined as:

$$\begin{aligned} UIR(i,n) &= TI(i,n)P(i,j)t(i,j) \\ &= \sum_{j=1}^{n} AN(i,j) \times Ae^{-(b+1)T_{i,j}} \times Fe^{-T_{i,j}} \\ &= \sum_{j=1}^{n} AN(i,j) \times AFe^{-(b+1)T_{i,j}}, \end{aligned}$$

where $T_{i,j}$ denotes the delay time of the interest i starting from the time granule interval j. For each time granule interval j, i might appear $AN(i, j)$ times, and each time will contribute a value $AFe^{-(b+1)Ti,j}$ to the total retentions of an interest, where A, F and b are constants.

Thus, $UIR(i, n)$ concentrates on both frequency and recency. Although the accuracy of the proposed model above is not ideal, it is easy to prove that user interest retentions are one of the major factors to construct the new granular computing data cycle model mentioned above, which is related to users' current interests. Therefore, it can be utilized to efficiently accelerate and refine the search process of data cycle system.

C. Starting Point Strategy

Psychological experiments have proofed that human prefers to solve problems using terms from the basic level (the ones that are used more frequently than others [33]), and in this way the problem-solving process can be accelerated [19,34]. In addition, concepts in a basic level are used more frequently than others [33]. Zeng et al. [18] considered a starting point was composed of a user ID and several nodes, which reflected the user interests. The starting point was used for refining the unification of retrieval and reasoning process in the form that the user might prefer. Based on the hierarchical granular structure and the new granular computing data cycle model, inspired by the basic level for human problem-solving in cognitive science, we extend the starting point in [18], which consists of a user identity and a set of nodes served as the background for the user. Furthermore, the retained user interests that are obtained through user interest retention model proposed above can be considered as some nodes in the starting point model.

Following the above ideas, a starting point strategy that unifies retrieval and reasoning together is introduced. The starting point strategy is outlined as the following major steps:

Step 1: Select the biggest values of interest retentions based on the proposed user interest retention model with $UIR(i, n)$, and construct a starting point based on selected interest retentions.

Step 2: Judge whether the query input by the user contains all the information through using the starting point strategy. If yes, go to Step 4.

Step 3: Reset the vague or incomplete query using the starting point strategy.

Step 4: Query the dataset using the query with Familiarity-Driven or Novelty-Driven principles in [18].

Thus, to obtain more satisfied reasoning and retrieval results, the strategies of Base-Level Learning, user interest retention, and starting point can be integrated together, which will provide reasoning results based on starting point in every level of specificity to produce a more user-preferred form of results.

D. Granularity-based Unification Using Multi-strategies

Although three strategies introduced above focus on providing some possible problem-solving solutions, and are designed to meet one type of user needs, we would like to emphasize that the extended strategies can be integrated together and unified in a more user oriented way from the viewpoint of granularity in order to provide users with refined results or satisfy more complex user needs. Then, under the term of granular reasoning, our goal is how to unify the retrieval and reasoning process from the viewpoint of granularity, namely, how to search for a good subset of the original dataset, and do reasoning on the selected dataset based on the idea of granularity. We also need to point out that although the strategies introduced in this paper are inspired by some basic strategies in granular computing, the granular structures more specifically granular knowledge structures that are mentioned in this paper are different from previous studies [34,35]. In granular computing, granules are organized hierarchically from larger grain sizes to smaller ones (or the other way around), and the granules in coarser levels contain the ones in finer levels. Thus, Fig. 3 shows three-levels of granularity based on multi-strategies model as follows.

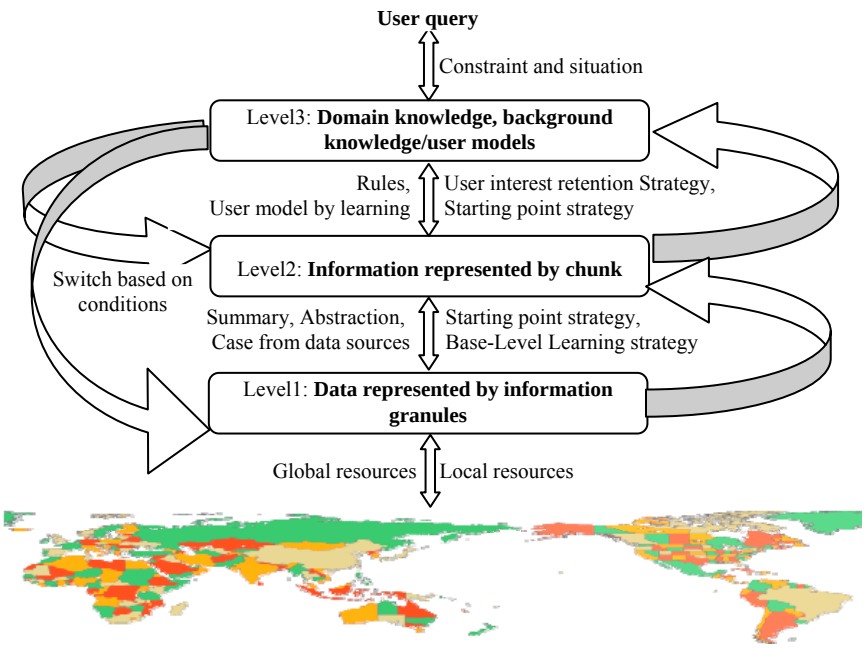

Figure 3. Three-levels of granularity based on multi-strategies

IV. Granularity-Based Knowledge Retrieval

In information sciences, the diagram of data-information-knowledge-wisdom hierarchy is used to describe different levels of abstraction in human centered information processing. Knowledge retrieval is a typical example of human-centered computing, and its evaluation is more related to personal judgments, which makes a balance between computer automation and user intervention [36]. In the data level, data retrieval systems are used to find relevant data, and in the information level, information retrieval systems are used to acquire knowledge. Then, knowledge retrieval mainly focuses on the knowledge level. Thus, knowledge retrieval systems aim at constructing the granular construct of the semantics and knowledge organization. Therefore, if the information retrieval systems can be enhanced from solving problems on the information level to the knowledge level, better results can be expected to be retrieved for user requirements. Some works [37,38] proposed concept-based methods to refine and expand queries. The developments, however, are concentrated on the context of a submitted query but not a user's background knowledge, in order to capture an information need. Furthermore, one of the key features of knowledge retrieval is that knowledge is visualized in a structured way so that users could get contextual awareness of related knowledge and make further retrieval [36]. Therefore, our granularity-based knowledge retrieval system concentrates on how to provide and use knowledge in more convenient ways. This method of problem-solving is to match query requirements of users with selected relevant

characteristic granules of information systems using corresponding knowledge acquisition algorithms, and then return the search results which are required knowledge to the users. Then this paper constructs a new granular computing model and improves the adaptability of the granule. Thus, a knowledge retrieval system based on the novel granular computing model is proposed, a process of which is shown in Fig. 4.

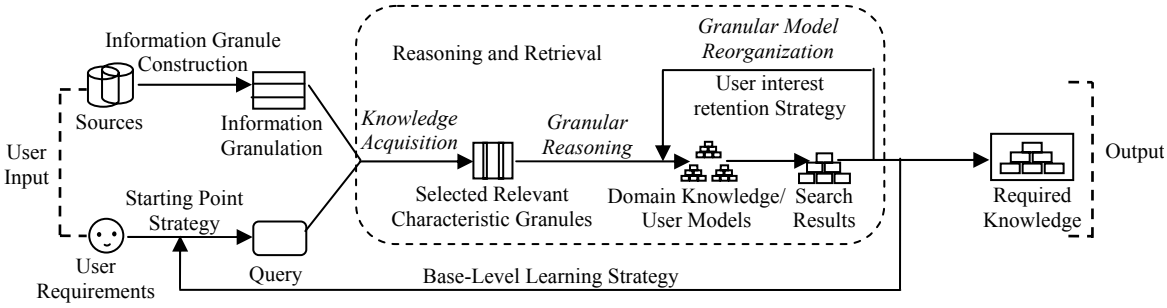

Figure 4. Granularity-based knowledge retrieval

V. Conclusions

Since the unification of Web scale search and reasoning from the viewpoint of granularity brings many human problem-solving strategies to reasoning, it can be considered as an effort towards Web intelligence. To do a more user oriented way, from the granularity point of view, a new granular computing data cycle model is presented. Based on the hierarchical granular structure and inspired by the basic level for human problem-solving in cognitive science, a multi-strategies model is introduced to unify search and reasoning for effective problem-solving. On the basis of user-centric multi-strategies and granular information processing, a conceptual framework of granularity-based knowledge retrieval model is developed. Furthermore, the knowledge retrieval model attempts to enhance the existing information retrieval systems from solving problems to the knowledge level. The proposed model contributes to the development to the next generation of retrieval systems.

Acknowledgment

We are highly grateful to the anonymous reviewers, referees and Editor-in-Chief for their valuable comments and hard work. This work was supported by the National Natural Science Foundation of China (Nos. 60873104, 61040037, 60905027), 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, 13A520529), the Beijing Natural Science Foundation (No. 4102007), the CAS/ SAFEA International Partnership Program for Creative Research Teams, the Education Fund for Youth Key Teachers of Henan Normal University, and the Postgraduate Science and Technology Program of Beijing University of Technology (No. ykj-2012-6765).

References

[1] N. Zhong, J. M. Liu, and Y. Y. Yao, *Advances in Web Intelligence*. Higher Education Press, 2011.

[2] N. Zhong, J. H. Ma, R. H. Huang, J. M. Liu, et al., "Research challenges and perspectives on wisdom web of things", *Journal of Supercomputing*, DOI 10.1007/s11227-010-0518-8, 2010.

[3] Y. Zeng, Y. Wang, Z. S. Huang, and N. Zhong, "Unifying web-scale search and reasoning from the viewpoint of granularity", in AMT 2009, LNCS 5820, 2009, pp.418-429.

[4] T. Y. Lin, "Granular computing and modeling the human thoughts in web documents", in IFSA 2007, LNAI 4529, 2007, pp.263-270.

[5] Y. Y. Yao, "A unified framework of granular", in Handbook of Granular Computing, Wiley, 2008, pp.401-410.

[6] Y. Y. Yao, "Perspectives of Granular Computing", in Proceedings of IEEE International Conference on granular computing, 2005, pp.85-90.

[7] L. Yan and S. Yan, "Granular reasoning and decision system's decomposition", *Journal of Software*, vol.7, pp.683-690, 2012.

[8] L. Sun, J. C. Xu, Y. W. Hu, and L. N. Du, "Granular space-based feature selection and its applications", *Journal of Software*, in press.

[9] C. T. Su, L. S. Chen, and T. L. Chiang, "A neural network based information granulation approach to shorten the cellular phone test process", *Computers in Industry*, vol.57, pp.412-423, 2006.

[10] M. C. Chen, L. S. Chen, C. C. Hsu, and W. R. Zeng, "An information granulation based data mining approach for classifying imbalanced data", *Information Sciences*, vol.178, pp.3214-3227, 2008.

[11] A. Bargiela and W. Pedrycz, "Recursive information granulation: aggregation and interpretation issues", *IEEE Transactions on Systems, Man, and Cybernetics*, vol.33, pp.96-112, 2003.

[12] L. Sun, J. C. Xu, J. Y. Ren, T. H. Xu, and Q. Q. Zhang, "Granularity partition-based feature selection and its application in decision systems", *Journal of Information and Computational Science*, vol.9, pp.3487-3500, 2012.

[13] G. Castellano and A. M. Fanelli, "Information granulation via neural network-based learning", in IFSA World Congress and 20th NAFIPS International Conference, 2001, pp.3059-3064.
[14] L. A. Zadeh, "A new direction in AI: toward a computational theory of perceptions", *AI Magazine*, vol.22, pp.73-84, 2001.
[15] Y. Y. Yao and J. T. Yao, "Granular computing as a basis for consistent classification problems", in Proceedings of PAKDD'02 Workshop on Toward the Foundation of Data Mining, 2002, pp.101-106.
[16] L. A. Zadeh, "Fuzzy sets and information granularity", *Advances in Fuzzy Set Theory and Applications*, 1979, pp.3-18.
[17] L. A. Zadeh, "Toward a generalized theory of uncertainty (GTU) an outline", *Information Sciences*, vol.172, pp.1-40, 2005.
[18] Y. Zeng, N. Zhong, Y. Wang, et al., "User-centric Query Refinement and Processing Using Granularity Based Strategies", *Knowledge and Information Systems*, vol.27, pp.419-450, 2011.
[19] A. A. Petrov and J. R. Anderson, "The dynamics of scaling: a memory-based anchor model of category rating and absolute identification", *Psychological Review*, vol.112, pp.383-416, 2005.
[20] J. R. Anderson, D. Bothell, M. D.Byrne, et al., "An integrated theory of the mind", *Psychological Review*, vol.111, pp.1036-1060, 2004.
[21] J. H. Han and T. Y. Lin: "Granular computing: models and applications", *International Journal of Intelligent Systems*, vol.25, pp.111-117, 2010.
[22] C. T. Su, L. S. Chen, and Y. Yih, "Knowledge acquisition through information granulation for imbalanced data", *Expert System with Applications*, vol.31, pp.531-541, 2006.
[23] Y. Y. Yao, D. Q. Miao, and F. F. Xu, "Granular structures and approximations in rough sets and knowledge spaces", in Rough Set Theory: A True Landmark in Data Analysis, Springer, Berlin, 2009, pp.71-84.\
[24] R. L. Solso, M. K. Maclin, and O. H. Maclin, *Cognitive Psychology*, Boston: Allyn and Bacon, 2005.
[25] L. Sun, J. C. Xu, C. Wang, T. H. Xu, and J. Y. Ren, "Granular computing-based granular structure model and its application in knowledge retrieval", *Information Technology Journal*, vol.11, 1714-1721, 2012.
[26] O. Oyeleye and E. A. Lehtihet, "A classification algorithm and optimal feature selection methodology for automated solder joint defect inspection", *Journal of Manufacturing Systems*, vol.17, pp.251-262, 1998.
[27] A. K. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: a review", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, pp.4-37, 2000.
[28] A. A. Petrov, "Computationally efficient approximation of the base-level learning equation in ACT-R", in Proceedings of the Seventh International Conference on Cognitive Modeling, Trieste, Italy, 2006, pp.391-392.
[29] G. J. Song, S. W. Tang, D. Q. Yang, and T. J. Wang, "A spatial feature selection method based on maximum entropy theory", *Chinese Journal of Software*, vol.14, pp.1544-1150, 2003.
[30] H. Ebbinghaus, *Memory: A Contribution to Experimental Psychology Hermann Ebbinghaus*, Teachers College, Columbia University, New York, 1913.
[31] J. R. Anderson and L. J. Schooler, "Reflections of the environment in memory", *Psychological Science*, vol.2, pp.396-408, 1991.
[32] D. W. Schneider and J. R. Anderson, "Modeling fan effects on the time course of associative recognition", *Cognitive Psychology*, vol.64, pp.127-160, 2012.
[33] E. J. Wisniewski and G. L. Murphy, "Superordinate and basic category names in discourse: a textual analysis", *Discourse Processing*, vol.12, pp.245-261, 1989.
[34] T. Rogers and K. Patterson, "Object categorization: reversals and explanations of the basic-level advantage", *Journal of Experimental Psychology: General*, vol.136, pp.451-469, 2007.
[35] Y. Zeng and N. Zhong, "On granular knowledge structures", in Proceedings of the first International Conference on Advanced Intelligence, 2008, pp. 28-33.
[36] Y. Y. Yao, Y. Zeng, N. Zhong, and X. J. Huang, "Knowledge retrieval", in International Conference on Web Intelligence, 2007, pp.729-735.
[37] X. H. Tao, Y. F. Li, and R. Nayak, "A knowledge retrieval model using ontology mining and user profiling", *Integrated Computer-Aided Engineering*, vol.15, pp.313-329, 2008.
[38] F. A. Grootjen and T. P. Van Der Weide, "Conceptual query expansion", *Data & Knowledge Engineering*, vo.56, pp.174-193, 2006.

Lin Sun received his B.S. and M.S. degree in computer science and technology from Henan Normal University, Henan, China, in 2003, and 2007, respectively. He is a Ph.D. candidate at Beijing University of Technology. His main research interests include granular computing, rough set, Web intelligence, intelligent information processing.