

Information Granulation and Granular Relationships

JingTao Yao, *Senior Member, IEEE*

Abstract—As an emerging research method to deal with information and knowledge processing, various topics of granular computing have recently received more attention by researchers. The foundations of granular computing involves general principles of many disciplines that are explored for many years such as divide and conquer, interval computing, fuzzy sets, rough sets and so on. Granules, granulations and relationships are some of the key issues in the study of granular computing. This paper aims to understand mainly granular computing theory from the perspective of information granulation and granular relationships.

I. INTRODUCTION

Granular computing (GrC) is an umbrella term to cover any theories, methodologies, techniques, and tools that make use of granules in problem solving [12], [17]. The first appearance of the concept was in 1979 under the name of *information granularity* in L.A. Zadeh's pioneer paper [15]. The term "granular computing" came to life with a suggestion from T. Y. Lin in the discussion of BISC Special Interest Group on Granular Computing [18]. Basic ingredients of granular computing are granules such as subsets, classes, and clusters of a universe. GrC is a new term for problem solving with computer science and may be viewed more on the philosophical rather than technical level.

Although the term is new, the basic notions and principles of granular computing occur under various guises in a wide variety of fields [14], [15]. Some example fields are belief functions, artificial intelligence, cluster analysis, chunking, data compression, databases, decision trees, divide and conquer, interval computing, machine learning from examples, structure programming, quantization, quotient space theory, and rough set theory.

Much research has been conducted recently in various aspects of granular computing. For example, granular computing is presented in binary relation models [1], [2]. A mathematic binary neighborhood system was formed from an informal granulation definition. Many studies on granular computing are focused on concrete models with

soft computing systems such as fuzzy systems [8], neural networks [20], rough sets [7], [13], quotient space [19], machine learning [10] and rough logic reasoning [3].

Researchers have also paid more attention to granular computing issues and solutions on problem solving [6], [12], [14]. A framework of granular computing can be established by applying its own principles from two perspectives, granular computing as structured thinking and structured problem solving [14]. From the philosophical perspective or the conceptual level, granular computing focuses on structured thinking based on multiple levels of granularity. The implementation of such a philosophy in the application level deals with structured problem solving [14]. There are many issues of granular computing that need to be studied in order to understand the theory of granular computing. These issues include basic components of granular computing (granule, granulated views and levels, hierarchies, and granular structures) and their usage (granulation, computing with granules). This paper aims to understand granular computing theory from information granulation and granular relationships. We mainly focus on granular relationships.

II. BASIC CONCEPTS OF GRANULAR COMPUTING

Key issues to understand any system are understanding the basic elements and the operations to them. The basic ingredients of granular computing are granules. The operation on granules is called granulation. We will give brief explanations to these two concepts in this section.

A. Granule

A granule is defined as "*a small particle; especially, one of numerous particles forming a larger unit*" in Merriam-Webster's Dictionary [4]. The meaning of granule in granular computing is very similar to the one in the above definition, i.e., any subset, class, object, or cluster of a universe is called a granule. These granules are composed of finer granules that are drawn together by distinguishability, similarity, and functionality [16]. A group of concepts or objects can be considered as a granule by their spatial neighborhood, closeness, cohesion, etc. Although granular computing is intended to deal with imprecision, uncertainty and partial truth [18],

J. T. Yao is with the Department of Computer Science, University of Regina, 3737 Wascana Parkway, Regina, SK, Canada S4S 0A2 (email: jtyao@cs.uregina.ca)

the granules may be of crisp or fuzzy format. A granule may have a different format and meaning when used in a particle model. For example, in a set-theoretic setting, such as rough sets and cluster analysis, a granule may be interpreted as a subset of a universal set while in programming, a granule can be a program module [14]. Granules at the lowest level are composed of elements or basic particle of the particular model that is used. For instance, the finest granules are words in an article universe. They are formed with basic particles, i.e., letters. They may be considered as singleton granules in some special cases.

B. Granulation

The original meaning of granulation from dictionaries is the act or process of forming something into granules [4]. It is a process to make a larger object to smaller ones. Zadeh adopts this idea to decompose a universe to granules in one of his early articles on granular computing in 1996 [16]. To quote Zadeh's definition, "*granulation involves a decomposition of whole into parts. Conversely, organization involves an integration of parts into whole.*" Based on this definition, there will be two operations in granular computing, namely, granulation and organization.

We adopt a more general view of granulation in granular computing. That is, granulation involves the process of two directions: construction and decomposition. The construction involves the process of forming a larger and higher level granule with smaller and lower level sub-granules. This is a bottom-up process. The decomposition involves the process of dividing a larger granule into smaller and lower level granules, which is similar to the dictionary and Zadeh's definitions of granulation. This is a top-down process. The reason that we consider a more general and broad view of granulation is that construction and decomposition are tightly related. When one chooses a particular granulation in an application, the benefits and efficiency of one direction is correlated to its opposite direction. If we consider a decomposition operation without the consideration of construction we may end up with a very efficient decomposition operation and a very inefficient construction.

III. RELATIONSHIPS AMONGST GRANULES

Relationships amongst granules may be classified into two types, interrelationship and intrarerelationship. Granulation, regardless of direction, is dealing with relationships between granules. Decomposition concerns breaking down a larger granule into smaller granules from which a larger granule can still be formed with

construction. Construction concerns grouping smaller granules that share similarity, indistinguishability, and functionality to a larger granule. The relationship involved in former granulation is considered as interrelationship, the later intrarerelationship.

A. Intrarerelationship

The very first definition of granule as "*being a clump of points drawn together by similarity, indistinguishability, and functionality*" [16] emphasizes why a collection of sub-granules or objects form a granule. It was argued that granulation and computation are two important and related issues of granular computing research [12]. Granulation deals with the construction, interpretation, and representation of granules. Computation deals with the computing and reasoning with granules and granular structures. In other words, interrelationship is the basis of grouping small objects together.

One of the key issues of interrelationship is constructing a granule based on similarity, indistinguishability, and functionality. Generally, different measures may be used for this purpose. For instance, intention and extension of concepts can be used in granular tree construction [11]. Distance measures are used in cluster analysis to determine whether or not objects belong to a cluster. We will not discuss on construction of granules in details in this article. The rest part of this article will mainly focus on interrelationship and leave the detailed study of interrelationship in other venues.

B. Refinement and Coarsening

Granular computing involves structured human thinking. A high level granule represents a more abstract concept and a low level a more specific concept. The level of abstraction may be represented in terms of coarse and fine relationships.

A granule o_1 is a refinement of another granule o_2 , or equivalently, o_2 is a coarsening of o_1 , denoted by $o_1 \preceq o_2$, or $o_2 \succeq o_1$ if every sub-granule or object of o_1 is contained in some sub-granules of o_2 . We call \preceq a fine relationship, and \succeq a coarse relationship. These relationships are also known as form and enclosure, i.e., o_1 forms o_2 and o_2 encloses o_1 .

For instance, Section II is a granule of this article. Subsection B is a granule of Section II. Granule Section II is coarser than granule Subsection B and granule Subsection B is finer than granule Section II. The relationship may be presented in formulas as,

$$\begin{aligned} \text{Section II} &\succeq \text{Subsection B} \\ \text{Subsection B} &\preceq \text{Section II}. \end{aligned}$$

C. Granulations as Partitions and Coverings

Partitions and coverings are two simple and commonly used granulations of a universe. They can be defined in the following way if a set-theoretical approach is adopted [10]. A partition of a finite universe U is a collection of non-empty and pairwise disjoint subsets of U whose union is U . Each subset in a partition is also called a block or an equivalence class.

When U is a finite set, a partition $\pi = \{X_i \mid 1 \leq i \leq m\}$ of U consists of a finite number m of blocks. In this case, the conditions for a partition can be simply stated by: (i). for all i , $X_i \neq \emptyset$, (ii). for all $i \neq j$, $X_i \cap X_j = \emptyset$, and (iii). $\bigcup\{X_i \mid 1 \leq i \leq m\} = U$.

There is a one-to-one correspondence between the partitions of U and the equivalence relations (i.e., reflexive, symmetric, and transitive relations) on U . Each equivalence class of the equivalence relation is a block of the corresponding partition.

A covering of a finite universe U is a collection of non-empty subsets of U whose union is U . The subsets in a covering are called covering granules. When U is a finite set, a covering $\tau = \{X_i \mid 1 \leq i \leq m\}$ of U consists of a finite number m of covering granules. In this case, the conditions for a covering can be simply stated by: (i). for all i , $X_i \neq \emptyset$ and (ii). $\bigcup\{X_i \mid 1 \leq i \leq m\} = U$.

According to the definition, a partition consists of disjoint subsets of the universe, and a covering consists of possibly overlapping subsets. Partitions are a special case of coverings. Granular trees can be generated with with partition or covering operations [11].

D. Partial Ordering

When comparing granules with fine or coarse relationships, every sub-granule within a finer granule is contained in the coarse granule. However, it may not be the case for all granules. In some cases, not *every* but only *some* sub-granules of a fine granule are contained in the coarse granule. Therefore the fine and coarse relationship are not fully true but partially true. We call these relationships partial fine (p-fine) and partial coarse (p-coarse) relationships. We can simply change the word *every* in Subsection B to *some* to define p-fine and p-coarse relationships.

A granule o_1 is p-refinement of another partition o_2 , or equivalently, o_2 is a p-coarsening of o_1 , denoted by $o_1 \sqsubseteq o_2$, or $o_2 \sqsupseteq o_1$ if some sub-granules of o_1 are contained in some sub-granules of o_2 . We call \sqsubseteq a p-fine relationship, and \sqsupseteq a p-coarse relationship.

A fine relationship can be viewed as a special case of a p-fine relationship. With these relationships, a partial

ordering of granules can be formed instead of a granular tree. In many cases, a granule is not totally enclosed by another granule. Therefore, we cannot construct a perfect granular tree.

E. ISA Relationships

ISA relationships are used in extended entity relationship diagrams [9]. An ISA is basically a class hierarchy used in entity relationship diagrams. Two subclasses share majority properties of their superclass except a couple of properties. Overlap or covering constraints may apply to ISA hierarchies. Overlap constraints determine whether two lower level subclasses are allowed to contain the same entity. Covering constraints determine whether all elements in subclasses collectively include all elements in their superclass. The concept ISA is also adopted in the ontology community for ontology formulation when dealing with a specification of a conceptualization.

In granular computing, an ISA is defined as a special case of refinement. If $X \preceq Y$ and X inherit all properties of Y , we say X ISA Y . Suppose we have two granules: ‘hospital’ and ‘children’s hospital’. Both granules are representations of hospitals while children’s hospital is a special type of hospital. Therefore relationships,

$$\text{children's hospital } ISA \text{ hospital}$$

and

$$\text{children's hospital } \preceq \text{ hospital}$$

hold. Although ‘emergency department’ is a finer granule of ‘hospital’ granule, however, there is no ISA relationship between them.

F. Similarity Relationship

Similarity is a key to forming an intrarelation of a granule. It can also be used to measure closeness amongst granules. Various distance measures can be used to calculate how similar two granules are. Like in cluster analysis, similarity between two granules can be defined as an average distance between sub-granules. Similarities are usually normalized between 0 to 1. A similarity with a value 0 means that two granules are totally different with no overlap sub-granules. A similarity with value 1 means they are the same granules with identical sub-granules and structure, ordering, as well as the semantics of them are exactly the same. We call these values the degree of similarity.

A similarity of granules o_1 and o_2 can be defined as average similarity amongst similarities between sub-granules of o_1 and sub-granules of o_2 ,

$$Sim(o_1, o_2) = \frac{1}{m \times n} \sum_{i=1, j=1}^{m, n} Sim(o_{1,i}, o_{2,j})$$

where m, n are the numbers of granules in o_1 and o_2 , $o_{1,i}$ is the i th sub-granule of o_1 , and $o_{2,j}$ is the j th sub-granule of o_2 .

The similarity of semantics is perhaps the most difficult aspect of similarity relationships. For instance, it is relatively easy to measure the similarity of two research articles by analyzing their structure, ordering and words used. A Euclidean distance may be used for distance measures. CiteSeer measures document similarity at the sentence level [5]. However, it is hard to compare their contents based on meaning automatically. It is in fact a challenge for computing with words.

G. Fuzzy Relationships

As a concrete example of granular computing, a framework based on fuzzy set theory was studied recently [17]. Granules are constructed with generalized constraints. A granule o is defined as

$$o = \{X | X \text{ is } R\},$$

where X is a value taken from a universe and R is a constraining relation. Example constraints are equality, possibility, probability, fuzzy and verity.

Fuzzy relationships are represented in terms of fuzzy graphs or if-then rules. For example,

$$\text{if } X \text{ is } A \text{ then } Y \text{ is } B$$

represents a fuzzy relationship between $o_1 = \{X | X \text{ is } A\}$ and $o_2 = \{Y | Y \text{ is } B\}$. The method involved is important in computing with words.

There are many other relationships between granules. Many relationships are binary. For instance, o_1 is finer than o_2 , o_2 ISA o_3 , and o_2 is similar with degree of 0.8 to o_4 . Some relationships are n -ary. For instance, o_1, o_2, \dots, o_n forms o_m and o_1, o_2 and o_3 covers o_4 . Hyperlinks may be considered as a new type of granular relationship. A hyperlink of granule o_1 is an element but not a sub-granule of o_1 . The link it leads to is a granule, say o_2 . o_2 may also contain a hyperlink to o_1 . A granular network can be formed with these hyperlinks together with a group of new relationships within the network.

IV. CONCLUDING REMARKS

Granules, granulations and relationships are some of the key issues in the study of granular computing. We mainly study one of the important concepts in granular computing namely, relationships amongst granules. It is not the intention to study these issues completely and comprehensively in this article. It is hoped that more research on granular relationships will lead us to success in problem solving approaches.

REFERENCES

- [1] T.Y. Lin, Granular computing on binary relations I: Data mining and neighborhood systems, *Rough Sets In Knowledge Discovery*, A. Skowron and L. Polkowski (eds), Physica-Verlag, pp107-121, 1998.
- [2] T.Y. Lin, Granular computing on binary relations II: Rough set representations and belief Functions, same as [1], pp121-140, 1998.
- [3] Q. Liu and S. L. Jiang, Reasoning about information granules based on rough logic, *International Conference on Rough Sets and Current Trends in Computing*, pp139-143, 2002.
- [4] Merriam-Webster Online: <http://www.m-w.com/>
- [5] CiteSeer: Scientific Literature Digital Library, <http://citeseer.ist.psu.edu/>
- [6] A. Skowron, and J. Stepaniuk, Information granules: Towards foundations of granular computing, *International Journal of Intelligent Systems*, **16**(1), 57-85, 2001.
- [7] S. Tsumoto, S. Hirano, Classification of temporal sequences using multiscale matching and rough clustering, *Proceedings of SPIE Vol. 5812, Data Mining, Intrusion Detection, Information Assurance, And Data Networks Security*, 2005, Orlando, USA, B. V. Dasarathy (eds), pp23-30.
- [8] W. Pedrycz and G. Vukovich, Granular computing in the development of fuzzy controllers, *International Journal of Intelligent Systems*, **14**(4), 419-447, 1999
- [9] R. Ramakrishnan and J. Gehrke, *Database Management Systems*, McGraw-Hill, 2002.
- [10] J.T. Yao and Y.Y. Yao, Induction of classification rules by granular computing, *Proceedings of the Third International Conference on Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence*, pp331-338, 2002.
- [11] J.T. Yao and Y.Y. Yao, A granular computing approach to machine learning, *Proceedings of the 1st International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'02)*, Singapore, pp732-736, 2002.
- [12] Y.Y. Yao, Granular computing: basic issues and possible solutions, *Proceedings of the 5th Joint Conference on Information Sciences*, pp186-189, 2000.
- [13] Y.Y. Yao, Information granulation and rough set approximation, *International Journal of Intelligent Systems*, **16**(1), 87-104, 2001.
- [14] Y.Y. Yao, Granular Computing, *Computer Science*, **31**(10.A), 1-5, 2004.
- [15] L.A. Zadeh, Fuzzy sets and information granularity, *Advances in Fuzzy Set Theory and Applications*, M. Gupta, R.K. Ragade, R.R. Yager (eds), North-Holland Publishing Company, pp3-18, 1979.
- [16] L.A. Zadeh, Key roles of information granulation and fuzzy logic in human reasoning, Concept formulation and computing with words, *Proceedings of IEEE 5th International Fuzzy Systems*, p1, 1996
- [17] L.A. Zadeh, Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, **90**(2), 111-127, 1997.
- [18] L.A. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems, *Soft Computing*, **2**(1), 23-25, 1998.
- [19] L. Zhang and B. Zhang, The quotient space theory of problem solving, *Fundamenta Informaticae*, **59**(2-3), 287-298, 2004.
- [20] Y.Q. Zhang, M.D. Fraser, R.A. Gagliano and A. Kandel, Granular neural networks for numerical-linguistic data fusion and knowledge discovery, *IEEE Transactions on Neural Networks*, **11**(3), 658-667, 2000.