

Table Representations of Granulations Revisited

Pre-topological Information Tables

I-Jen Chiang¹, Tsau Young Lin², and Yong Liu³

¹ Graduate Institute of Medical Informatics, Taipei Medical University,
205 Wu-Hsien Street, Taipei, Taiwan 110

ijchiang@tmu.edu.tw

² Department of Computer Science, San Jose State University,
San Jose, California 95192, USA

tylin@cs.sjsu.edu

³ College of Computer Science, Zhejiang University,
Hangzhou 310027, China

cckaffe@yahoo.com.cn

Abstract. This paper examines the knowledge representation theory of granulations. The key strengths of rough set theory are its capabilities in representing and processing knowledge in table format. For general granulation such capabilities are unknown. For single level granulation, two initial theories have been proposed previously by one of the authors. In this paper, the theories are re-visited, a new and deeper analysis is presented: Granular information table is an incomplete representation, so computing with words is the main method of knowledge processing. However for symmetrical granulation, the pre-topological information table is a complete representation, so the knowledge processing can be formal.

1 Introduction

Relational database theory is designed to model the real world of a long duration [10]. For each instance the interactions among entities can be very different from the other instance. So the relational model takes the common denominator and ignores the interactions among entities. So it assumes the universe of all entities and various attribute domains are Cantor sets, in which no interactions among elements are modeled. However, in data mining or data analysis the modeling is instance based. So the assumption that the underlying structure of the universe of entities is a classical set is an over simplified one.

1.1 Models of Real World Entities

A more realistic modeling is needed. What should be the proper model of real world entities? There are two views available. The first one is from the model theory of first order logic, where a model is a Cantor set together with the relational structure. The second one is from granulation.

2 Granulation and Relational Structure

Let us revise our arguments in [14]. According to Lotfi Zadeh [20]:

- “information granulation involves partitioning a class of objects (points) into granules, with a granule being a clump of objects (points), which are drawn together by indistinguishability, similarity or functionality.”
- The phrase “drawn together by indistinguishability, similarity or functionality,” in general, can be expressed mathematically by relations.

If the group of drawn together consists of n objects, then the relation is n -ary; we may refer to such structure as n -ary granulation. In general, for every n , there may be several, even infinitely many, n -ary relations. In granular computing, n can be any cardinal number. In relational structure of the model theory, they are all finite; first order logic does not use predicates of non-finite places. In real life, we go seldom beyond initial few n , taking topological space into consideration, $n = 2$ is adequate.

2.1 Basic Granulation and Relational Structure

As a first step, we have considered the simplest granulation or the simplest relational structure (of first order logic). In other words, the underlying structure is a binary granulation or equivalently binary relational structure [14], [10], [7], [6]. In this paper, we take a deeper and new view on its representation theory.

“Drawn together,” in the binary cases, implies certain level of symmetry. If p is drawn towards q , then q is also drawn towards p . Such symmetry, we believe, is imposed by impreciseness of natural language. To avoid such implications, we will rephrase it to “drawn towards an object p ,” so that it is clear the reverse may or may not be true.

In binary relation, “drawn together” can be viewed as a special case of “drawn towards p ,” since p may vary through every object of a granule. So for each p , we will use $B(p)$ to denote the group of objects that are drawn toward p . Now we have a localized version of Zadeh’s word:

Definition 1. *By a binary granulation we mean the association of an object $p \in V$ with a granule $B(p) \subseteq V$ (neighborhood), where p varies through all objects of the universe V . This association is a mapping $V \longrightarrow 2^V$, called a basic or binary granulation (BG).*

2.2 Geometric and Algebraic Views

It will be helpful to visualize the granulation, for this goal, we will use geometric terminology. We will refer to a granule as a neighborhood of p , and the collection,

$$\{B(p) \mid p \text{ varies through } V\}$$

is called the basic (binary) neighborhood system (BNS) of V . Note that it is possible that $B(p)$ is an empty set. In this case we will simply say p has no

neighborhood by abuse of language; to be very correct, we should say p has an empty neighborhood.

Also we should note that many different points, $p, q, \dots$ may have the same neighborhood (granule) $B(p) = B(q)$. The set of all q such that $B(q)$ is equal to $B(p)$, is called the center set $C(p)$ of the granule $B(p)$; each element in $C(p)$ is called a center. The collection of the center sets forms a partition on V .

To help us manipulating the granulation, we also reformulate it algebraically:

$$R = \{(p, v) \mid v \in B(p) \text{ and } p \in V\}$$

is a binary relation (BR) defined by BG.

Proposition 1. *A basic (binary) neighborhood system (BNS), a basic (binary) granulation (BG), and a binary relation (BR) are equivalent.*

3 Knowledge Representations

First we setup a convention.

Convention: A symbol is a string of "bits and bytes." Regardless of whether that symbol may or may not have the intended real world meaning, no real world meaning participates in the formal processing. A symbol is termed a word, if the intended real world meaning *participates* in the formal processing.

Please note that "symbol" here is equivalent to the "word" in group theory.

The main idea here is to extend representation theory of rough sets to granular computing, in which granules have overlapping semantics. Real world granulation often cannot be expressed by equivalence relations. For example, the notions of "near", "similar", and "conflict" are not equivalence relations. The granulation of human body by body, leg and head, and etc is not a partition. So there are intrinsic needs to generalize the knowledge representation theory of partition (rough set theory) to more general settings (granular computing).

We will re-interpret and refine some earlier works; see the latest overview in [7]. Here are the three main topics:

1. Relational Table: This is the classical rough set representation of partitions. The basic idea is to assign a meaningful name to each equivalence class of an equivalence relation (partition). These names are independent from each other, since equivalence classes are mutually disjoint. The representation, in rough set theory, has been called an information system, a knowledge representation system, an information table, or a data table. In the relational database theory, it is called a *bag relation*.
2. Granular Table: This is the first representation theory of granulation observed [12], [15], [14]. The basic idea is to assign a meaningful name to each granule of a granulation (binary relation.) These names are not independent from each other, since granules may overlap; in other words, these names have non-trivial interactions. We capture these interactions partially. We represent it by the binary relation of the intersections of granules.

3. Topological Table: The representation is similar, but deeper than granular table. In this representation, we also consider the induced partition. Each equivalence class, called the center set, is uniquely associated with a granule. Hence, we assign a meaningful name to each granule, as well as equivalence class; so we have a granular table and a partition table. They are algebraically isomorphic [11]. In the partition table, we capture the interactions of granules by a pre-topology [7], [8], [9]. If the granulation (binary relation) is symmetric, the representation is complete in the sense we can recapture the binary relations from the pre-topological table.

4 Relational Tables - Representations of Partitions

A *partition* is a collection of pairwise disjoint subsets whose union is V . The corresponding algebraic concept is an equivalence relation. So each subset is called an equivalence class in mathematics; to synchronize with granulation, we may call it granule. It is the simplest kind of granulation.

Pawlak (1982) [18] and Tony Lee (1983)[5] observed that a relational table is a knowledge representation of a universe of entities. Each column induces an equivalence relation (partition) on the universe; n columns induce n partitions. More generally, they observed:

Proposition 2. *A subset B of attributes of a relational table K , in particular a single attribute, induces an equivalence relation Q^B on V .*

To do the knowledge representation, we will explore the converse. We shall recall some of the analysis in [12], [13], [15], [14].

Definition 2. *The pair (V, Q) is a granular data model (GDM), if V is a classical set of entities, and Q is a finite family of equivalence relations on V .*

Pawlak called it a knowledge base. As the latter one often has different meaning, we will use GDM. By reversing Pawlak and Tony Lee's observation, we assign a word (meaningful to human) to each equivalence class. Such an assignment can be expressed in a table format. We will illustrate the idea by example.

Let $U = \{id_1, id_2, \dots, id_9\}$ be a set of nine balls with two partitions:

1. $\{\{id_1, id_2, id_3\}, \{id_4, id_5\}, \{id_6, id_7, id_8, id_9\}\}$
2. $\{\{id_1, id_2\}, \{id_3\}, \{id_4, id_5\}, \{id_6, id_7, id_8, id_9\}\}$

We label the first partition COLOR and the second WEIGHT. They are the best summarizations of the given partitions from the view of human. Next, we will name each equivalence class by its real world characteristic: We name the first equivalence class Red, because each ball of this group has red color (appears to human). Note that this name reflects human's view and human only. For example, physical characteristics, such as wave length are *not* implemented and stored in the system. In AI, such terms, COLOR and Red, are called semantic primitive [1]. They are primitives (undefined terms) from the view of computer systems, but they do have the intent to represent human perceived semantics. We will summarize the previous analysis as follows:

1. $id_1 \longrightarrow (\{id_1, id_2, id_3\}) \longrightarrow \text{Red}$
 The first $\longrightarrow$ says that id_1 belongs to the equivalence class $[id_1]$ and the second $\longrightarrow$ says that the equivalence class has been named Red.
2. $id_2 \longrightarrow (\{id_1, id_2, id_3\}) \longrightarrow \text{Red}$
- ...
4. $id_4 \longrightarrow (\{id_4, id_5\}) \longrightarrow \text{Orange}$
- ...
9. $id_9 \longrightarrow (\{id_6, id_7, id_8, id_9\}) \longrightarrow \text{Yellow}$

Similarly, we have names for all WEIGHT-classes. We have constructed Table 1:

Table 1. Constructing a relational table by naming each granule

U	COLOR	WEIGHT
id_1	Red	W1
id_2	Red	W1
id_3	Red	W2
id_4	Orange	W3
id_5	Orange	W3
id_6	Yellow	W4
id_7	Yellow	W4
id_8	Yellow	W4
id_9	Yellow	W4

Note that each word represents an equivalence class of a partition. So the words within a column have no overlapping semantics; Each word is independent from each other. So these words can be treated as symbols. In the table processing of rough set theory, they have been regarded as symbols. Their intended semantics can only be carried out in the presence of human operators.

4.1 Granular Tables

The representation of a partition is rested on two properties:

1. Each object p belongs to an equivalence class (the union of equivalence class covers the whole universe)
2. No objects belong to more than one equivalence class (equivalence classes are pairwise disjoint)

We need a similar property in binary granulation: Let B be a binary granulation

1. Each object, $p \in V$, is assigned to one and only one B -granule
2. No objects are assigned to more than one granule. Note that we are *not* using the memberships; we are considering the binary granulation, that is, the association between object and its granule.

Next we assign each B -granule a unique meaningful name. Such an association allows us to represent *a finite set of binary granulations* by a "relational table", called a *granular table*.

Let us recall the illustration in [15]. In binary granulation, each p is associated with a unique binary neighborhood B_p , which consists of balls that have certain color component:

1. $B_{id_1} = B_{id_2} = B_{id_3} = \{id_1, id_2, id_3, id_4, id_5\}$ is the set of all balls that have red color component in their color coating.
2. $B_{id_4} = B_{id_5} = \{id_1, id_2, id_3, id_4, id_5, id_6, id_7, id_8, id_9\}$ is the set of all balls that have red or yellow color components in their color coating.
3. $B_{id_6} = B_{id_7} = B_{id_8} = B_{id_9} = \{id_4, id_5, id_6, id_7, id_8, id_9\}$ is the set of all balls that have yellow color components in their color coating.

Then to each binary granule (neighborhood), we assign a word (not a symbol).

1. Having-RED = Name(B_{id_1}) = ... = Name(B_{id_3}); The name indicates that all the balls in this granule has red color component in its coating.
2. Having-RED+YELLOW = Name(B_{id_4}) = Name(B_{id_5}); The name indicates that all the balls in this granule has red and yellow color component (orange color is a mixture of red and yellow)
3. Having-YELLOW = Name(B_{id_6}) = ... = Name(B_{id_9})

These words have human-perceived semantics attached and will participate in formal processing. The non-empty intersections among granules imply that there are non-trivial logical interactions among these words; and such interactions will be respected during data processing. By considering the following map:

- Entities $\rightarrow$ Granules $\rightarrow$ Words, we have
 1. $id_1 \rightarrow B_{id_1} \rightarrow$ Having-RED
 - ...
 4. $id_4 \rightarrow B_{id_4} \rightarrow$ Having-RED+YELLOW
 5. $id_5 \rightarrow B_{id_5} \rightarrow$ Having-RED+YELLOW
 - ...
 9. $id_9 \rightarrow B_{id_9} \rightarrow$ Having-YELLOW
 - ... similar statements for WEIGHT

They are summarized in granular table Table 2. To process such a table, we need computing with words (respecting the semantics). In Table 3, we express the binary relation, called granular binary relation, among these words. The binary relation only partially captures the interactions among words (in a column). Note that this binary relation reflects the non-empty intersection of granules: For example (Having-Red, Having-Red+Yellow) $\in B_{COLOR}$ if and only if the two granules have non-empty intersection.

Perhaps, we should stress again that attribute values have overlapping semantics; so the interactions among these words have to be properly handled.

Definition 3. *The name of a granule is binary related to the name of another granule, if they have non-empty intersection.*

Such a binary relation is described in Table 3; but we need to stress that the binary relation does not adequately describe the relationships among granules. We need computing with words to deal with the information on semantic level.

Table 2. Granular Table: There are interactions among the words

BALLs	Granulation 1	Granulation 2
id_1	Having-RED	W1
id_2	Having-RED	W1
id_3	Having-RED	W2
id_4	Having-RED+YELLOW	W3
id_5	Having-RED+YELLOW	W3
id_6	Having-YELLOW	W4
id_7	Having-YELLOW	W4
id_8	Having-YELLOW	W4
id_9	Having-YELLOW	W4

Table 3. A Symmetric Binary Relation for Color Attributes

Having-RED	Having-RED
Having-RED	Having-RED
Having-RED	Having-RED+YELLOW
Having-RED+YELLOW	Having-RED
Having-RED+YELLOW	Having-RED+YELLOW
Having-RED+YELLOW	Having-YELLOW
Having-YELLOW	Having-RED+YELLOW
Having-YELLOW	Having-YELLOW

5 Topological Tables

Note that the binary granulation $B : V \rightarrow 2^U; p \mapsto B(p)$ is a map whose inverse images $C(p) = B^{-1}(B(p))$ induce an equivalence relation E_B on V . The equivalence class is called the center set of $B(p)$. Let the center set be:

$$C_w = B^{-1}(B_p), \tag{1}$$

where $w=\text{Name}(B_p)$. Verbally, C_w consists of all objects that have the same B -granule B_p . We use the granule's names to index the family of the center sets

$$\begin{aligned} C_{\text{Having-RED}} &\equiv \text{Center of } B_{id_1} = \text{Center of } B_{id_2} \\ &= \text{Center of } B_{id_3} = \{id_1, id_2, id_3\} \\ C_{\text{Having-RED+YELLOW}} &\equiv \text{Center of } B_{id_4} = \text{Center of } B_{id_5} \\ &= \{id_4, id_5\} \\ C_{\text{Having-YELLOW}} &\equiv \text{Center of } B_{id_6} = \dots = \text{Center of } B_{id_9} \\ &= \{id_6, id_7, id_8, id_9\} \end{aligned}$$

For $B(p) = \emptyset$, $C(p) = \{x \mid B(x) = \emptyset\}$. We call the collection of $\{C(p) \mid p \in V\}$ topological partition with the understanding that there is a neighborhood $B(p)$ for each equivalence class $C(p)$. The neighborhoods capture the interaction among equivalence classes. Such a family $\{C(p)\}$ is a partition in BNS-spaces. Now, we will define the topological binary relation.

Definition 4. *The name of a granule is topologically binary related to the name of another granule, if the first granule has non-empty intersection with the center set of second granule. We regard the name of the second granule as a member of the neighborhood of the name of first granule.*

Thus, for example to define the topological binary relation B_{COLOR} we have

$$(\text{Having} - \text{RED}, \text{Having} - \text{RED} + \text{YELLOW}) \in B_{\text{COLOR}}$$

if $B_{id_1} \cap C_{\text{Having-RED+YELLOW}} \neq \emptyset$ and $id_i \in C_{\text{Having-RED}}$ etc. Note that the B-granule is definable by the induced partition, if B is symmetric [8], [9].

Proposition 3. *If $B \subseteq V \times V$ is a symmetric binary relation, and E_B its induced equivalence relation, then each B-binary neighborhood is a union of E_B -equivalence classes.*

So B is definable on attribute domain (a quotient set of V) that consists of all center sets. So Table 4 and Table 5 completely defined by B and vice versa.

Note that such a binary structure cannot be deduced from the table structure. We are ready to introduce the notion of semantic property.

Definition 5. *A property is said to be semantic if it is not implied by the table structure. A property is said to be syntactic if it is implied by the table structure.*

The binary relation (Table 3) is not derived from the table structure (of Table 2) so it is a semantic property. This type of tables has been studied in [17,16] for approximate retrievals; and is called topological relations or tables. Formally:

Definition 6. *A relational table (e.g. Table 4) whose attributes are equipped with topological binary relations (e.g. Table 5 for COLOR attribute) is called a (pre-) topological relation.*

By replacing the names of binary granules with the center sets, Table 2 is transformed to Table 4; they are isomorphic. However, the topologies are different: Table 5 provides the topology of Table 4. Table 3 provides that of Table 2.

Table 4. Topological Table

BALLs	Granulation 1	Granulation 2
id_1	$C_{\text{Having-RED}}$	W1
...	...	...
id_3	$C_{\text{Having-RED}}$	W2
id_4	$C_{\text{Having-RED+YELLOW}}$	W3
...	...	...
id_9	$C_{\text{Having-YELLOW}}$	W4

Table 5. A Topological Binary Relation on the Center sets of COLOR

$C_{\text{Having-RED}}$	$C_{\text{Having-RED}}$
$C_{\text{Having-RED}}$	$C_{\text{Having-RED+YELLOW}}$
$C_{\text{Having-RED+YELLOW}}$	$C_{\text{Having-RED}}$
$C_{\text{Having-RED+YELLOW}}$	$C_{\text{Having-RED+YELLOW}}$
$C_{\text{Having-RED+YELLOW}}$	$C_{\text{Having-YELLOW}}$
$C_{\text{Having-YELLOW}}$	$C_{\text{Having-RED+YELLOW}}$
$C_{\text{Having-YELLOW}}$	$C_{\text{Having-YELLOW}}$

Table 6. Topological Table

BALLs	Granulation 1	Granulation 2
id_1	NAME($C_{\text{Having-RED}}$)	NAME(W1)
...	...	...
id_4	NAME($C_{\text{Having-RED+YELLOW}}$)	NAME(W3)
...	...	...
id_9	NAME($C_{\text{Having-YELLOW}}$)	NAME(W4)

Theorem 1. *Given a finite set of binary relations B , a finite set of equivalence relations E_B can be induced. The knowledge representation of B is a topological representation of E_B .*

Proof. (A Sketchy) As we have illustrated before, the knowledge representations of B and E_B are accomplished by giving meaningful names to the granules and its center sets respectively. Note that Table 4 is a table of equivalence classes of E_B and its knowledge representation in Table 6 is a table with symbolic names replacing the equivalence classes. By replacing the names of binary granules with those of the center sets we will have Table 2 transformed to Table 6. Therefore, syntactically, the knowledge representation of B and E_B is the same (isomorphic). We can directly impose an isomorphic binary relation on Table 6. Note that Table 5 and the imposed relation provide the same pre-topology. In other words, the isomorphism becomes a topological isomorphism; $\square$

6 Conclusions

In the series of our papers, we have literally taken Zadeh's intuitive description of clumps as a formal mathematical notion of granulation. It is essentially a mild generalization of binary relations and neighborhood systems in (pre-)topological spaces [19], [17], [16], [14], [15], [7], [6]. By giving a (meaningful) name to each granule, we have a representation theory. The processing of this kind of representations has to be relied on computing with words; there are unformalized interactions among the attributes values (names of overlapping granules); the interactions need further investigation. However, the topological view, in the case of symmetric binary relational structure, does capture the representation quite "completely" in the sense that the interactions among granules can be specified formally by binary relations (of center sets which are equivalence classes).

References

1. Barr, A., Feigenbaum, E.A.: The handbook of Artificial Intelligence, Willam Kaufmann (1981)
2. Birkhoff, G., MacLane, S.: A Survey of Modern Algebra, Macmillan (1977)
3. Brualdi, R.A.: Introductory Combinatorics, Prentice Hall (1992)
4. Cai, Y.D., Cercone, N., Han, J.: Attribute-oriented induction in relational databases. In: Knowledge Discovery in Databases. AAAI/MIT Press, Cambridge, MA (1991) 213–228
5. Lee, T.T.: Algebraic Theory of Relational Databases. The Bell System Technical Journal, 62(10), December (1983) 3159–3204
6. Lin, T.Y., Liao, C.J.: Granular Computing and rough sets: An Incremental Development. In: Data Mining and Knowledge Discovery Handbook, Springer, to appear (2005)
7. Lin, T.Y.: Granular Computing from Rough Set Prospect. IEEE Computational intelligence Society Newsletter, November (2004)
8. Lin, T.Y.: Chinese Wall Security Policy Models: Information Flows and Confining Trojan Horses. In: Data and Applications Security XVII: Status and Prospects, Kluwer Academic Publishers (2003) 275–297
9. Lin, T.Y.: Granular Computing on Binary Relations-Analysis of Conflict and Chinese Wall Security Policy. In : Proceedings of RSCTC'2002. LNAI 2475, Springer (2002) 296–299
10. Lin, T.Y., Louie, E.: Modeling the Real World for Data Mining: Granular Computing Approach. In: Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference (2001) 3044–3049
11. Lin, T.Y.: Generating Concept Hierarchies/Networks: Mining Additional Semantics in Relational Data. In: Advances in Knowledge Discovery and Data Mining. LNAI 2035, Springer (2001) 174–185
12. Lin, T.Y.: Data Mining and Machine Oriented Modeling: A Granular Computing Approach. Journal of Applied Intelligence 13(2), Kluwer, (2000) 113–124
13. Lin, T.Y.: Granular Computing: Fuzzy Logic and Rough Sets. In: Computing with words in information/intelligent systems, Springer-Verlag (1999) 183–200
14. Lin, T.Y.: Granular Computing on Binary Relations I: Data Mining and Neighborhood Systems. In: Rough Sets In Knowledge Discovery, Springer-Verlag (1998) 107–121
15. Lin, T.Y.: Granular Computing on Binary Relations II: Rough Set Representations and Belief Functions. In: Rough Sets In Knowledge Discovery, Springer-Verlag (1998) 121–140
16. Lin, T.Y.: Neighborhood systems and approximation in database and knowledge base systems. In: Proceedings of the Fourth International Symposium on Methodologies of Intelligent Systems (1989) 75–86
17. Lin, T.Y.: Neighborhood Systems and Relational Database. In: Proceedings of 1988 ACM Sixteen Annual Computer Science Conference, February 23–25 (1988) 725
18. Pawlak, Z.: Rough sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers (1991)
19. Sierpinski, W., Krieger, C.: General Topology. University of Toronto Press (1956)
20. Zadeh, L.: The Key Roles of Information Granulation and Fuzzy Logic in Human Reasoning. In: 1996 IEEE International Conference on Fuzzy Systems, September 8–11, 1 (1996)