Granular Computing Based on Graph Theory

Sujie Guan, Min Li and Shaobo Deng*

Scholar of Information Engineering, Nan Chang Institute of Technology, Nanchang, Jiangxi 330099, PR China
Email: houjiyuan2002@163.com

Abstract. The rough set theory begins with the concepts of the equivalence relations and equivalence classes. Then, we will use the equivalence relation to research granulation and granulation computing. In current paper, we present the concepts of elementary granulation and granulation, and describe how to convert a decision table into a granular network which has the characteristics of simple and intuitive for describe a decision table. We also define a granular network for a decision table. On this basis, we present a concrete method for converting a decision table into a granular network and give an example for illustrate purpose. Further, we describe a type of data reduction algorithm based on granular network. Lastly, we prove that the data reduction based on data analysis method is equivalent to the data reduction based on granular network. The data reduction based on granular network has the characteristics of simple and intuitive. Theoretical analysis and experimental results indicate that the proposed approach for the data reduction is efficient and feasible.

1. Introduction

What is a granulation; how shall we denote granulations? how shall we describe relation of granulations? and how do we construct some proper operations for granulations? All of these are the main research content in Granular computing [1-2]. Granulation and Granular computing have become a hot research topic. An underlying idea of granular computing is the use of groups, classes, or clusters of elements called granules [3]. we can analyze many properties of granulation or some potential rules about the granulations by using proper operations.

Lin researched granulation and granular computing using a partition, coverage and neighborhood system [1]. In Ref. [4], a study of the granular computing on binary relations is presented. The result is applied to the analysis of conflict of interests’ relation and the Chinese wall security policy. Yao has assumed Three-way decisions of granulation and granular computing [5-25]. Skowron and Stepaniuk discuss information granulation in knowledge discovery, where the information granules be used to deal with the problem of stable patterns extraction [6, 26]. For the granular computing a set-theoretic framework is proposed [7-8], where each element of a universe is associated with a nonempty family of neighborhoods. In Ref. [9], the notion of machine oriented data modeling is explored, where an attribute value in a relational model is a meaningful label of a set of granule. In a granular computing framework, the Ref [10] discusses approximation space which generalizes the approaches to concept approximation existing in rough theory. Pei had given two kinds of models for granulation [11], where the knowledge of fuzzy set theory was applied to the framework of granular computing. Ref. [12] proposed a data reduction method for consistent decision table, however, this method has some redundancy in the calculation and it lacks the soundness and completeness proof for the granulations. Ref. [18] proposed a granular network for an inconsistent decision table, but does not consider a granular network for
consistent decision table and lacks experiment analysis.

The rough set theory begins with the concept of the equivalence relation and equivalence classes. So we shall use equivalence relations to discuss granulations or granulation computing. This paper presents concepts for elementary granulations and granulations, then describes converting decision tables into granular networks which has the characteristics of simple and intuitive for description of decision table. So we describe a type of data reduction algorithm based on granular networks, this, in turn, makes the procedure of data reduction are more simple and intuitive.

This paper is organized as follows. Section 2 presents some concepts of granulation. In Section 3 this paper introduces concepts of granular networks and relevant propositions will be discussed. Section 4 introduces data reduction algorithm based on granular networks followed by an example for illustrative purposes. In section 5, “soundness” and “completeness” of granular network are proved. Section 6, concludes the paper and identifies future research directions.

2. Basic Knowledge
In this section, we present some basic concepts of rough set theory, which can be found in Refs. [13-17]. In rough set theory, given \( S=(U,A) \) and \( T=(U,A,C,D) \), where \( C,D \subseteq A \), \( U \) denotes a entity set and \( A \) denotes an attribute set.

Definition (1): Let \( S=(U,A) \) and \( a \in A \). Define a mapping \( f: U \rightarrow V_a \), where \( V_a \) is called \( \text{dom}(a) \). Let \( V=U \ast V_a \) and \( V \) is the value range of \( a \).

Definition (2): Let \( S=(U,A) \) and \( a \in A \) and \( v \) be its value. The set \( (a)_v=\{x \in U \mid a(x)=v \land (\forall a \in A) \land (\forall v \in V_a) \} \) is called elementary granulation.

Definition (3): Let \( S=(U,A) \), and attribute \( a \in A \), the set \( (a)_{\forall} = \{ (a) \mid a \in A \text{ and } v \in V_a \} \) is called a granulation.

Definition (4): Let \( S=(U,A) \) and \( a \in A \). The set \( IG= \{ (a) \mid (\forall a \in A) \} \) is called information granulation.

Lemma (1): For an information table \( S=(U,A) \), any elementary granulation is correspondent to a certain equivalence class.

Lemma (2): For an information table \( S=(U,A) \), each granulation \( (a) \) is the set of equivalence classes.

3. Granular Network
Given any \( S=(U,R) \). Then \( S \) can be shown with granulations and elementary granulations. However, it is not intuitive. This section introduces the concepts for granular networks which can be intuitively shown.

Definition (5): Given \( S=(U,A) \). The granular network for \( S=(U,A) \) is defined by the following steps:

Step 1. Any elementary granulation for any granulation is presented by a point, which is called granular node of granular network.

Step 2. For any two elementary granulations \( (a_v) \) and \( (a_w) \), there is no edge connecting \( (a_v) \) and \( (a_w) \), where \( (a_v) \in (a) \text{ and } (a_w) \in (a), a \in A \).

Step 3. As for any two elementary granulations \( (a_v) \) and \( (b_w) \), if \( (a_v) \cap (b_w) \neq \emptyset \), then there is an edge connecting them, where the edge is also called a granulation edge, and then the weight value for the granulation edge is shown by \( (a_v) \cap (b_w) \), where \( (a_v) \in (a) \text{ and } (b_w) \in (b) \).

Step 4. The construction of the granular network for \( S \) is finished if all elementary granulations of the information granular IG have been handled by (1)-(3).

Definition (6): For the granular network is converted from a decision table. If each granulation edge is a directed edge.

By Definitions (5) and (6), a granular network for any decision table can be generated by Algorithm 1:

Algorithm 1: Obtain a granular network
Input: \( T=(U,A,C,D) \)
Output: a granular network
Step 1: Number each granulation (each attribute) from 0, and then number each elementary granulation of any granulation from 0.

Step 2: As for any elementary granulation \((a_i)\), depicting it with a point \((a_i)\) and its coordinates \((i, j)\), where \(i\) is the serial number of granulation \((a)\) and \(j\) is the serial number of elementary granulation \((a_i)\).

Step 3: Drawing an edge to connect any two adjacent elementary granulations from left to right according to the step 2 of definition (6).

Step 4: For each granulation edge, giving theirs weighted values according to the step 3 of the definition (6).

Step 5: Repeat Step1-Step3 until all elementary granulations have been handled. And then, we give the granular network.

Definition (7): In a granular network for \(S=(U,A)\), each path is made up of granulation edges with the equal weighted value from the beginning to the end of a elementary granulation. That can be named decision path, denoted by \(\theta_i\), where \(i\) is weighted value for a granulation.

4. Granular Computing

If granular networks are constructed for decision tables, then this section can further give some granular computing.

4.1. Data Reduction

Theorem (1): [13-14]: Let \(S=(U,A,C,D)\) and \(x \in U\). if for each \(y \neq x\), \(dx|C=dy|C\rightarrow dx|D=dy|D\) holds, then the rule \(dx\) is consistent; otherwise, that is inconsistent.

Theorem (2): Let \(T=(U,A,C,D)\) and \(a \in C\). if “a” does not have any core values, then “a” can be reduced, otherwise that can not be reduced.

At this point, the data reduction can be performed by the following two algorithms:

Algorithm 2: data reduction algorithm based on granular networks of consistent decision tables
Input: any consistent decision table \(T=(U, A, C, D)\)
Output: core values.
Step 1: By Definition (5) and Algorithm 1, construct a granular network \(G\) for \(T=(U, A, C, D)\).
Step 2: For \(G\), check each path and recording each of core elementary granulations and weighted values.
Step 3: Record each of core elementary granulation and theirs weighted value.
Step 4: For each attribute if it has no core values, then the attribute can be omitted; otherwise, it can not.
End.

Algorithm 3: the data reduction based on the granular network for an inconsistent decision table
Input: an inconsistent decision table \(T=(U, A, C, D)\)
Output: core values.
Step 1: By the above definitions and algorithm 1, construct a granular network \(G\) for \(T\).
Step 2: For \(G\), check each path and recording core elementary granulation and weighted value.
Step 3: Write each core elementary granulation and weighted value.
Step 4: For some paths have core values on each condition elementary granulation, omit them.
Step 5: Omitting any attributes which have no core values.
End.

4.2. Equivalence Theorem

In the section, we discuss and prove that he data reduction based on granular network is equivalent to the data reduction based data analysis method.

Theorem (3) [Sufficient Theorem]: Given any decision table \(T\), the data reduction based on granular network for \(T\) corresponds with the data reduction based data analysis method.

Theorem (4) [Consistency theorem]: Given any decision table \(T\), the data reduction based on data analysis method corresponds with the data reduction based on granular network.
Proof: From theorems (2, 4) to theorems (1, 3), we can easily prove the theorem.

Theorem (5) [Equivalence theorem]: Given any decision table T, the data reduction based on data analysis method is equivalent to the data reduction based on granular network.

Compared with other related methods, the reduction method proposed in this paper has the following characteristics:

(1) Data reduction method in the paper is based on granular network. Elementary granulations, weighted values (individuals) and granulations are applied to describe decision tables and its decision rules. The data reduction about granular networks has a simply and intuitively expressive. We can easily obtain all of core elementary granulations and core values. There does not exist plenty of comparison operations.

(2) The algorithms 2 and 3 describe the process of attribute-reduction and attribute-value-reduction. Step 2 is a main reduction procedure. The algorithms 2-3 do not need various comparison operations, but need to scan the whole granular network only once to get the core value. In the whole procedure of data reduction in this paper, this paper avoids many comparison operators.

(3) Step 2 of Algorithms 2 and 3 is the main procedure of data reduction. This step is equivalent to the procedure of attribute reduction, and to the attribute value reduction. Time complexity is within \( O(|m| |D_1| \ldots |D_n| |A||U|) \), where \(|U|\) is the number of the individuals of a decision table. The algorithms 2-3 need more \(|A||U|\) storage spaces to store weighted value of granulation edges, which is also the need of the storage space to accommodate a decision table.

(4) Compared with [12]. This paper has much more improvement and is entirely different from Ref. [12]. The later only has handled consistent decision tables and does not intuitively describe and discuss decision rules. However, this paper gives some meaningful properties. Firstly, this paper intuitively describes the decision table. Secondly, this paper intuitively describes each decision rule. Finally, this paper not only gives the proof of the soundness about granular network, but also gives the proof of completeness. The method based on Granular Network, in this paper, is soundness and completeness.

(5) Compared with Ref. [18]. Ref. [18] proposed a granular network for an inconsistent decision table, but does not consider a granular network for consistent decision table and lacks experiment analysis. However, this paper gives some meaningful properties based on Ref. [18]. Firstly, this paper intuitively describes decision tables including consistent or inconsistent decision table. Secondly, this paper intuitively describes each decision rule. Finally, this paper proves that the data reduction based on data analysis method is equivalent to that based on granular network and gives some experiment analysis.

5. Experiment
To demonstrate the effectiveness of our proposed algorithms, we conduct experiments on four UCI (http://archive.ics.uci.edu/ml/datasets.html) data sets. The information of these data sets is shown in table 1.

Firstly, we introduce the experiment situation-system: window sp3, cpu frequency: 1.7GHZ and memory: 768M. Because these data sets have no null values, we discretize them by the OrthogonalFileScaler method of the Rosetta software.

Secondly, we need to determine what kinds of data structures can be used for a graph network. In our experiments, a sparse matrix is selected. We do not need to store the information of all nodes, but only the information of adjacent nodes.

| Names   | Number of attributes | Number of instances | miss values? |
|---------|---------------------|---------------------|--------------|
| Iris    | 5                   | 150                 | No           |
| Abalone | 9                   | 4177                | No           |

Table 1. Description of data sets.
Thirdly, our algorithms need construct the sparse matrix for $T=(U,A,C,D)$ with $|U|$ rows and $|A|$ columns.

Then, we would give the time of construction sparse matrix.

Now, the granular networks for the four decision tables are as show in figures 1a-1d.

**Figure 1.** (a) The granular network for iris data set. (b) The granular network for Abalone data. (c) The granular network for mushroom data set. (d) The granular network for P- R-H-Digits data set.

In figures 1a-1d, the horizontal axis represents granulations, the vertical axis represents elementary granulations, and $a_1$, $\ldots$, $a_i$ denote the attribute names of the data sets. If a data set has small scale data, the figure is succinct, such as, figure 1a. Figure 1a describes 150 instances but only has 28 edges. Figure 1b has 279 edges. Figure 1c has 285 edges, and figure 1d has only 484 edges.

Fourthly, we need to use a sparse matrix to store a granular network and record the execution time of data reduction based on granular network.

Table 2 lists out the execution time of construction sparse matrix, the execution time of data reduction, the number of the core condition attributes, and the execution time of drawing granular networks. Table 2 gives also the number of the nodes (the elementary granulations) of the last condition attribute and the number of nodes (the elementary granulations) of decision attribute, which be discussed in the following.
From table 2, we can see once the sparse matrix for a granular network have been constructed, data reduction procedure become simple and do not spend plenty of time. But it would cost much time to draw the granular network for a decision table, such as the 10.484 seconds for Mushroom and 14.016 seconds for P-R-H-Digits. However, in practice, we do not need the graph but its sparse matrix in most cases. In here, the graph only gives us the intuitively perspective.

Table 2. Information of granular computing on data sets list in table 1.

| Names          | Execution time (s) | Number |
|----------------|-------------------|--------|
|                | Construction sparse matrix | Data reduction | Drawing granular networks | Condition attributes after reduction | Nodes of the last condition attribute | Nodes of the decision attribute |
| Iris           | 0.031             | 0.0016  | 0.25 | 4 | 3 | 3 |
| Abalone        | 0.672             | 1.391   | 2.078 | 8 | 8 | 29 |
| Mushroom       | 4.438             | 0.844   | 10.484 | 19 | 7 | 2 |
| P-R-H-Digits   | 5.609             | 0.657   | 14.016 | 15 | 5 | 2 |

6. Conclusion
We presented the definitions of elementary granulation, granulation, granular network, respectively, and described an algorithm for how to convert a decision table into a granular network which has the characteristics of simple and intuitive expression. Some properties of the granular network were analyzed. On the basis of this, we presented a type of data reduction method based on granular network, which makes the procedure of data reduction are more simple and straightforward. In addition, we give the soundness and completeness proofs for a granular network, which shows the method of granular network is sound and complete. The experimental result shows that the proposed approach for data reduction is efficient and feasible. Future research work will include how to find some relevant properties and give some applications of granular network, as well as how to improve the efficiency of the construction of sparse matrix.

Acknowledgment
This work is partially supported by the grants from the grant from Jiangxi Education Department (No. GJJ161109), the National Science Foundation of China (No. 61363047).

References
[1] Lin T Y 2006 Granular computing on partitions, coverings and neighborhood systems The Proceedings International Forum on Theory of GRC from Rough Set Perspective pp 1-7.
[2] Zadeh L A 1997 Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic Fuzzy Sets and Systems 90 (2) 111-127.
[3] Yao Y Y 2000 Granular computing: Basic issues and possible solutions The Proceedings of the 5th Joint Conference on Information Sciences pp 186-189.
[4] Lin T Y 1998 Granular computing on binary relations I: data mining and neighborhood systems, II: rough set representations and belief functions Rough Sets in Knowledge Discovery1 ed Polkowski L and Skowron A (Heidelberg: Physica-Verlag) pp 107-140.
[5] Yao Y Y 2006 Three Perspectives of Granular Computing, The Proceedings International Forum on Theory of GRC from Rough Set Perspective (IFTGrcCRSP2006) pp16-21.
[6] Skowron A and Stepaniuk J 1999 Towards discovery of information granules Principles of Data Mining and Knowledge Discovery (Berlin Heidelberg: Springer) pp 542-547.
[7] Yao Y Y 1999 Rough Sets, neighborhood systems, and granular computing Proc. 1999 IEEE Canadian Conf. Electrical and Computer Engineering (Edmonton,Canada) pp 1553-1558.
[8] Yao Y Y 1999 Granular computing using neighborhood systems Advances in Soft Computing Engineering Design and Manufacturing ed Roy R, Furuhashi T and Chawdhry P K (London Springer-Verlag) pp 539-553.

[9] Young T 2000 Data mining and machine oriented modeling: A granular computing approach Applied Intelligence 13 (2) 113-124.

[10] Skowron A, Świniarski R and Synak P 2005 Approximation spaces and information granulation Transactions on Rough Sets III (Berlin Heidelberg: Springer) pp 175-189.

[11] Pei D 2006 Two classes of models of granular computing The Proceedings International Forum on Theory of GRC from Rough Set Perspective (IFTGrCRSP2006) pp 52-55.

[12] Deng S B, Li M and Guan S J 2007 The structure of granular network based on granular computing and its application in data reduction The IEEE International Conference on Granular Computing (2007 GRC) pp 89-93.

[13] Liu Q 2001 Rough Set and Rough Inference. (Beijing Science Press).

[14] Zhang W, Wu W and Liang J 2001 Rough Set Theory and Method (Beijing Science Press).

[15] Wang G 2001 Rough Set Theory and Its Knowledge Acquisition (Xi’an: Xi’an Jiaotong University Press).

[16] Li M, Shang C, Feng S, et al. 2014 Quick attribute reduction in inconsistent decision tables Information Sciences 254 155-180.

[17] Li M, Deng S, Feng S, et al. 2014 Fast assignment reduction in inconsistent incomplete decision systems. Journal of Systems Engineering and Electronics 25 (1) 83-94.

[18] Guan S 2013 Granular computing for inconsistent decision table 2013 IEEE International Conference on Granular Computing (GrC) pp 126-131.

[19] Liu Y, Xiong R and Chu J 2009 Quick attribute reduction algorithm with hash Chinese Journal of Computers 32 (8)1493-1499.

[20] Xu Z Y, Liu Z P and Yang B R 2006A quick attribute reduction algorithm with complexity of max(\(|C||U|\), \(|C||U/C|||\)) Chinese Journal of Computers 29 (3) 391-399.

[21] Liu P, Ali A and Rehman N 2019 Multi-granulation fuzzy rough sets based on fuzzy preference relations and their applications IEEE Access 7 147825-147848.

[22] Hu C X, Zhang L, Wang B J, Zhang Z and Li F Z 2019 Incremental updating knowledge in neighborhood multigranulation rough sets under dynamic granular structures Knowledge-Based Systems 163 811-829.

[23] Li M and Wang G Y Approximate concept construction with three-way decisions and attribute reduction in incomplete contexts. Knowledge-Based Systems, 2016, 91:165-178.

[24] Wang X and Li J 2018 Three-way decisions, concept lattice and granular computing: Editorial International Journal of Machine Learning and Cybernetics 9 (11) 1765-1766.

[25] Su R, Panoutsos G and Yue X 2020 Data-driven granular computing systems and applications Granular Computing 1-2.

[26] Wang X and Li J 2020 New advances in three-way decision, granular computing and concept lattice International Journal of Machine Learning and Cybernetics 11 (5) 945-946.