Information Sciences 382–383 (2017) 150–169

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Grouping granular structures in human granulation intelligence

Yuhua Qian[a,*], Honghong Cheng[a], Jieting Wang[a], Jiye Liang[a], Witold Pedrycz[b], Chuangyin Dang[c]

[a] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Shanxi, China
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada
[c] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 16 June 2016
Revised 2 November 2016
Accepted 26 November 2016
Available online 30 November 2016

Keywords:
Granular computing
Granulation intelligence
Granular structure
Knowledge representation
Granular structure distance

ABSTRACT

Human granulation intelligence means that people can observe and analyze the same problem from various granulation points of view, which generally acknowledge an essential feature of human intelligence. Each granulation view can generate a granular structure through dividing a cognitive target into some meaningful information granules. This means that a large number of granular structures can be generated from the cognitive target. However, people can group these granular structures and select some representative ones for problem solving. This leads to an interesting research topic: how to efficiently and effectively group a family of granular structures. To address this issue, we first introduce a granular structure distance to measure the difference between two granular structures within a unified knowledge representation. Then, we propose a framework for grouping granular structures, called GGS algorithm, which is used to efficiently partition them. Moreover, two indices denoted as *DIS* and *APD* are also designed for evaluating the performance of a grouping result of granular structures. Finally, experiments carried out for nine data sets show that the GGS algorithm comes as a sound solution from perspectives of its convergence, effectiveness and scalability. In this way we have proposed and experimented with the general framework for discovering the structure inherent in granular structures, which can be afterwards used to simulate intelligent behavior of human's abilities of granular structure selection.

1. Introduction

Artificial Intelligence refers to ability of a machine to perform tasks through work like a person, and the purpose of its research is to let this machine work like person. It can be roughly divided into two aspects from the viewpoint of cognition level. One focuses on understanding and simulating the mechanism of perception including vision, touch, smell, hearing, and so on. The other concerns on understanding and simulating the advanced cognitive mechanism, such as human learning, reasoning, decision and solving ability for complex problems.

* Corresponding author: Fax: +86 0351 7010566/8176.
E-mail addresses: jinchengqyh@sxu.edu.cn, jinchengqyh@126.com (Y. Qian), chhsxdx@163.com (H. Cheng), jietingw@163.com (J. Wang), ljy@sxu.edu.cn (J. Liang), pedrycz@ee.ualberta.ca (W. Pedrycz), mecdang@cityu.edu.hk (C. Dang).

http://dx.doi.org/10.1016/j.ins.2016.11.024
0020-0255/

As one of human advanced cognitive mechanisms, granular computing (GrC) is becoming a timely and innovative topic in artificial intelligence, information processing, data mining and knowledge discovery. Granular computing, as a special term, is coined by Zadeh and Lin [19–21,61]. In 1979, Zadeh firstly argued that fuzzy information granulation plays a fundamental role in human reasoning. Three basic issues in granular computing are information granulation, organization and causation. As it was pointed out in [59–62], information granulation involves decomposition of whole into parts; the organization involves integration of parts into whole; and the causation involves association of causes with effects. In 1985, Hobbs [12] introduced a concept of granularity to characterize granulation degree used by an observer. In 1992, Zhang and Zhang [63] pointed out that people can observe and analyze the same problem from various granulation viewpoints, which is a generally acknowledged feature of human intelligence. In 1998, Yager and Filev [51] stated that human observation, measurement, conceptualization and reasoning are realized by taking information granularity into account. In particular, in 1997, Zadeh [60] proposed a so-called granular computing as a new independent research area in intelligent information processing. In order to come up with a uniform terminology, we call the above discussed cognition abilities *granulation intelligence*. To date, the granulation intelligence-based information processing has been applied to various fields that include rough set theory, clustering, machine learning, complex network, data mining and knowledge discovery, and others [3,4,7,28,40,43,44,54,56,57].

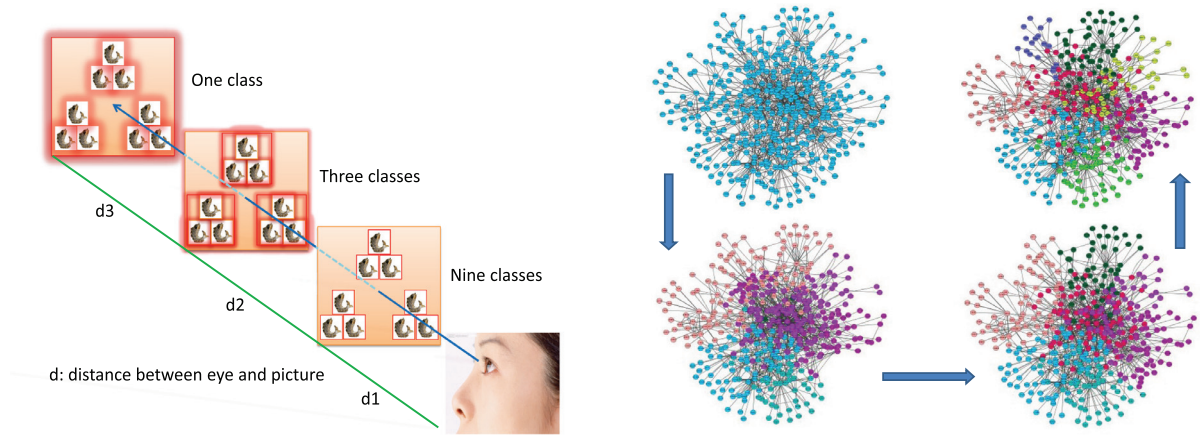
To date, there are several focal points in granular computing, which include measure of granularity [12,17,33,42,48,55,64], information processing includes databases [13,19,20], framework of granular computing [21,24,25,35,41,62], problem solving based on "granulate and conquer" principle and quotient theory [63], multigranulation view [31,34,38,39,45–47,50,52], and their applications. Just having this quick look, one can note that these developments evolve into a field of cross-disciplinary study.

In human granulation intelligence, a granule is regarded as a clump of objects drawn together by indistinguishability, similarity, connectivity and proximity of functionality [17,19,26,27,60]. Granulation of an object set leads to a collection of granules. In categorical data analysis, objects with the same attribute-values can not be differentiated each other, which can be putted into a granule according to the indistinguishability. In social network analysis, some potential societies can be discovered through analyzing connection relationships among members, which are looked forward as base granules in a social network. The granules construct base computing units in the granulation intelligence-based information processing.

A *granular structure* is a mathematical structure of the collection of granules, in which the inner structure of each granule is visible (a granule is a white box) and the interactions among granules are detected by the visible structures [19–21,33]. In granular computing, we often granulate a set of objects into some granules based on a feature subset through a feature selection algorithm [5,11]. The granulation of objects induced by an equivalent relation is a set of equivalence classes, in which each equivalence class can be regarded as an (Pawlak) information granule [25,29,32]; the granulation of objects induced by a tolerance relation generates a set of tolerance classes, in which each tolerance class also can be seen as a tolerance information granule [16,30,32]. By using a general binary relation, objects are granulated into a set of information granules, called a binary *granular structure* [14,19,20,35].

In granular computing, one often needs to measure the granulation degree of objects in a given data set, called *information granularity*. In general, the information granularity represents discernibility ability of information in a granular structure [37,42]. The smaller the information granularity, the stronger its discernibility ability[22]. As mentioned by Prof. Zadeh, in general, information granularity should characterize the granulation degree of objects from the viewpoint of hierarchy [23,60]. This provides a point of view that an information granularity should characterize hierarchical relationships among granular structures. How to calculate the information granularity of a granular structure has been an important issue. To date, several forms of information granularity have been proposed according to various views and targets [17,18,29,32,33,42,48,55]. Recently, Qian and Liang [33] presented an axiomatic definition of information granularity in a knowledge base, in which several existing forms of information granularity arise as special cases. From the axiomatic definition, the size of information granularity does not depend on the sizes of information granules, but some array of them. In a sense, the information granularity could be used to distinguish two granular structures. However, when the information granularity of one granular structure equals to the one characterizing another structure, one can say that these two granular structures possess the same granulation degree, however these two granular structures may be not equivalent. This means that an effective distance between two granular structures becomes imperative; this will be also quite apparent in this study.

In human granulation intelligence, a cognitive subject can observe various granular structures of an object from multiple perspectives or scales. This intelligence can be illustrated by two examples shown in Fig. 1. In Fig. 1(a), it can be seen that a cognitive subject can granulate a picture considering different partitions when the distance between the person and the picture varies from d_3 to d_1. In Fig. 1(b), we can see that a society could be divided into several much smaller societies while the connection constraint is strengthened [49]. However, these observed granular structures are not all necessary for problem solving or the requirement of a cognitive subject, which can be induced to two reasons. One is that the lifetime of some of granular structures observed is very short, while a cognitive subject is interested in those robust granular structures with much longer life spans. The other is that a cognitive subject does not need all granular structures for a practical solving task. In Fig. 1(a), it is easy to see that the granular structure with 1 class, that with 3 classes and that with 9 nine classes are observed when the distance between the eye and the picture are d_3, d_2 and d_1, respectively, which are three robust and stable granular structures for a cognitive subject. For a society discovery task, although various society partitions with the same number of societies can be obtained by different methods or parameters, we just need to know which partitions



(a) Picture granulation with different visual distance (b) Society granulation with different connection constraints

Fig. 1. Granular structures granulated by a cognitive subject in human granulation intelligence.

are highly representative. Fig. 1(b) gives their representatives when the number of societies are 3, 5 and 8, respectively. In fact, people can observe and analyze the same object from these robust and stable granular structures, which is one of core factors of human granulation intelligence. From the above analysis, it is very important how to search and select representative granular structures from all observed ones for simulating human granulation intelligence. This leads to an interesting issue of grouping granular structures.

The main motivation of this study is to investigate how to effectively group granular structures. As we know, there exists many forms for representing a granular structure based on different binary relations. To obtain a general method, in this paper, we first present a unified knowledge representation method for representing a granular structure. Based on the knowledge representation, we then introduce a method and its corresponding algorithm (just GGS) to rapidly group granular structures observed from a given data set, which is based on a so-called granular structure distance. The GGS algorithm not only ensure the convergence of its iteration process, and also can obtain a local optimal solution. The granular structure with the minimum knowledge distance to the center of a cluster can be selected as the representative of granular structures in this cluster. In order to evaluate the effectiveness of a program of grouping granular structures, we further give two evaluation indices: dispersion and approximation degree. The dispersion is used to characterize the inner dispersion of a set of granular structures. A grouping program with much smaller dispersion can be thought as a far better one. The approximation degree characterizes the average approximation ability of a grouping scheme for a given decision partition. The higher the value of the approximation degree of a grouping scheme, the better the group scheme is. Finally, numerical experiments completed for nine real data sets from UCI show the effectiveness and scalability of the proposed GGS algorithm.

The paper is organized as follows. In Section 2, through reviewing Pawlak granular structures, tolerance granular structures and neighborhood granular structures, we give a unified knowledge representation method of a granular structure. Section 3 proposes a so-called knowledge distance to measure difference between any two granular structures and gives some of its interesting properties. Section 4 presents an effective method for grouping granular structures (just GGS), and analyzes its time complexity and convergence. In Section 5, we present dispersion and approximation degree two evaluation indices for evaluating the performance of a grouping scheme. In Section 6, we test the convergence, effectiveness and scalability of the GGS algorithm through employing nine real data sets. Finally, Section 7 gives the concluding remarks.

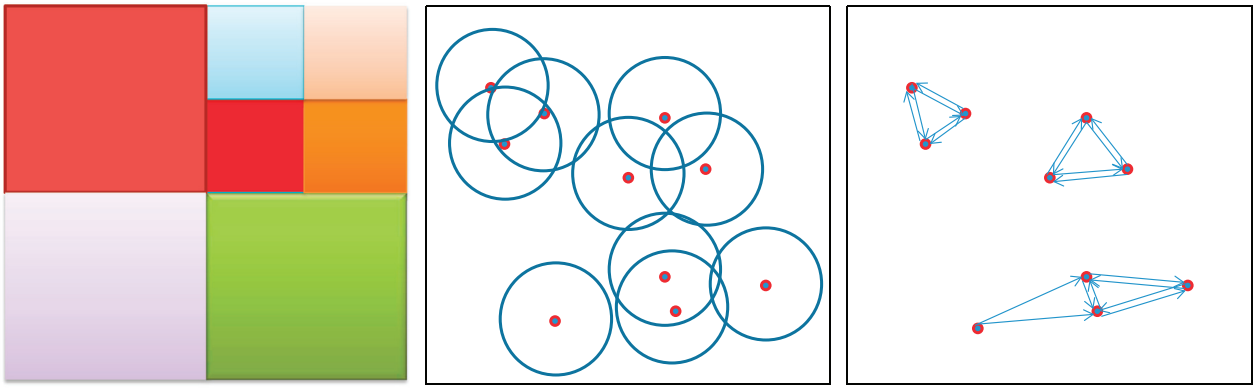
2. Knowledge representation of granular structures

In this section, through analyzing the process that human granulate a complex problem into some granules, we establish a unified knowledge representation method of granular structures. Throughout this paper, we suppose that the set of granular structures observed by a cognitive subject forms a finite nonempty set.

Let us recall that a granular structure is a mathematical structure of the collection of granules, in which the inner structure of each granule is visible (a granule is a white box) and the interactions among granules are detected by the visible structures [19–21,33]. The granulation of objects induced by a given binary relation has a special representation form itself, in which its granules have special sematic and meaning themselves. In what follows, we analyze several forms of granular structures in granular computing.

1. Granular structure induced by an equivalence relation

We assume the set of objects $U = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ defined by a set of features $A = \{a_1, a_2, \ldots, a_m\}$. Each feature a_j describes a domain of values, denoted by $V(a_j)$, associated with a defined semantic and a data type. A domain $V(a_j)$ is defined



(a) Granular structure with an equivalent relation (b) Granular structure with radius δ (c) Granular structure with k(=2) neighborhood

Fig. 2. Three kinds of granular structures.

as categorical if it is finite and unordered, i.e., $V(a_j) = \{a_j^{(1)}, a_j^{(2)}, \ldots, a_j^{(n_j)}\}$, where n_j is the number of categories of feature a_j, $1 \le j \le m$. If each feature in A is categorical, then U is called a categorical data set. Let $U = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ be a set of n objects, in which $\mathbf{x_i}$ is represented as a vector $(x_{i,1}, x_{i,2}, \ldots, x_{i,m})$, $x_{i,j} \in V_{a_j}$, $1 \le j \le m$. One says that $\mathbf{x_i} = \mathbf{x_k}$ if $x_{i,j} = x_{k,j}$ for each $j \le m$. The relation $\mathbf{x_i} = \mathbf{x_k}$ does not mean that $\mathbf{x_i}$ and $\mathbf{x_k}$ are the same object in the real-world database, but rather that the two objects have equal values in features $a_1, a_2, \ldots, a_m$. In Fig. 2(a), square frames with the same color can be arranged as the same granules, and all square frames can be granulated to form seven granules by using the equivalence relation R induced by seven colors, which give rise to an equivalence granular structure. Formally, U/R constitutes a partition of U, called a granular structure on U, and every equivalence class is called an equivalence information granule [6,8,10,24,29,36]. In general, we denote the equivalence granular structure as

$$U/R = \{E_R(x) : x \in U\}. \tag{1}$$

2. Granular structure induced by a neighborhood relation

If R_i $(i = 1, 2, \ldots, m)$ is a neighborhood relation, then $(U, \mathbf{R}) = (U, R_1, R_2, \ldots, R_m)$ can be called a neighborhood knowledge base [17]. A neighborhood relation R on the universe is a relation matrix $M(R) = (r_{ij})_{n\times n}$, where

$$r_{ij} = \begin{cases} 1, & d(x_i, x_j) \le \delta, \\ 0, & \text{otherwise}, \end{cases}$$

where d is a distance [14,65] between x and y, and δ is a nonnegative number. Let A, $B \subseteq AT$ be categorical and numerical attributes, respectively. The neighborhood granules of objects x induced by A, B, $A \cup B$ are defined as

1. $N_A(x) = \{x_i \in U \mid d_A(x, x_i) = 0\}$;
2. $N_B(x) = \{x_i \in U \mid d_B(x, x_i) \le \delta\}$;
3. $N_{A\cup B}(x) = \{x_i \in U \mid d_A(x, x_i) = 0 \wedge d_B(x, x_i) \le \delta\}$.

In Fig. 2(b), we can see that all objects within the radius δ of an object are regarded as its corresponding neighborhood. Each object has its neighborhood itself, and ten objects induce 10 neighborhood granules, which constitutes a covering of U.

Beside the above definition of a neighborhood, k-neighborhood method is also a very important definition of a neighborhood in the existing research, which is only effective for numeric data. The k-neighborhood granules of objects x induced by B are defined as

$N_B(x) = \{x_i,\ i \le k\}$, where x_i is the ith object of k nearest-neighbor objects.

Fig. 2(c) shows 10 neighborhood granules with $k = 2$, in which each object has two nearest neighborhoods and its 2-neighborhood granules has three objects including itself. This also constitutes a covering of U.

In general, we denote $N(U)$ to be a neighborhood granular structure on U, and a member $N_R(x)$ from $N(U)$ will be called a neighborhood granule. Formally, we denote the neighborhood granular structure as follows

$$N(U) = \{N_R(x) : x \in U\}. \tag{2}$$

3. Granular structure induced by a tolerance relation

It may happen that some of the attribute values for an object are missing. For example, in medical information systems there may exist a group of patients for which it is impossible to perform all the required tests. These missing values can be represented by the set of all possible values for the attribute. To indicate such a situation, a distinguished value (the

so-called null value) is usually assigned to those attributes. Generally, we denote the null value by $*$. We define a binary relation on U by

$$SIM(P) = \{(u, v) \in U \times U |\ \forall a \in P, a(u) = a(v) \ \text{ or } \ a(u) = * \ \text{ or } \ a(v) = *\}.$$

In fact, $SIM(P)$ is a tolerance relation on U. The concept of a tolerance relation has a wide variety of applications in classifications [16,30,32]. It can be easily shown that $SIM(P) = \bigcap_{a\in P} SIM(\{a\})$. Let $S_P(u)$ denote the set $\{v \in U | (u, v) \in SIM(P)\}$. Then, $S_P(u)$ is the maximal set of objects which are possibly indistinguishable by P with u. The set

$$SIM(P) = \{S_P(x) : x \in U\}. \tag{3}$$

is called the tolerance granular structure induced by R, and a member $S_R(u)$ from this set is called a tolerance information granule.

Moreover, similar to the above granular structures, some of other binary relations with a reflexive property also can induce the corresponding granular structures. For example, if R_i $(i = 1, 2, \ldots, m)$ is a dominance relation, then $(U, \Re) = (U, R_1, R_2, \ldots, R_m)$ is called a dominance knowledge base. The set $\{S_R(u) | u \in U\}$ is called the dominance granular structure induced by R, and a member $S_R(u)$ from this set is called a dominance information granule.

Above modes of granular structures in which the granules are Boolean (B-granular) play important roles in a wide variety of methods, approaches and techniques. Important though it is, these modes of granular structures have not a unified representation form, which brings a challenge for studying granular computing from the same perspective. Hence, a unified representation to granular structures is necessary.

Given a universe U and R a general binary relation on U, we denote its corresponding granular structure in terms of the following matrix

$$GS(R) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}, \tag{4}$$

where $p_{ij} \in \{0, 1\}$, where $p_{ij} = 1$ denotes that x_j is positioned within the granule determined by x_i, while $p_{ij} = 0$ means that x_j is not included in the granule determined by x_i. Furthermore, through using granules determined by every object, the binary granular structure of the universe can be represented as

$$GS(R) = (G_R(x_1), G_R(x_2), \ldots, G_R(x_n)), \tag{5}$$

where $G_R(x_i) = p_{i1}/x_1 + p_{i2}/x_2 + \ldots + p_{in}/x_n$. $G_R(x_i)$ is the granule induced by x_i. Here, "+" means the union of elements.

The cardinality of the information granule $G_R(x_i)$ can be calculated with

$$|G_R(x_i)| = \sum_{j=1}^{n} p_{ij}, \tag{6}$$

which appears to be a natural generalization of the cardinality of a set.

No matter any binary relation used, this knowledge representation method of granular structures dose not loose any information, which can be degenerated to its corresponding original version of granular structures.

In human granulation intelligence, a cognitive subject can observe various granular structures of an object from multiple perspectives or scales. In these granular structures, those robust and stable granular structures can be captured and recognized, which is one of core factors of human granulation intelligence. For further investigation in this paper, we focus the representation of multiple granular structures as follows.

Given a family of binary granular structures **GS** on U, denoted by

GS $= \{GS_1, GS_2, \ldots, GS_m\}$,where $GS_i = (G_{i1}, G_{i2}, G_{i3}, \ldots, G_{in})$ is a granular structure, and $G_{ij} = \{\frac{p_{ij}^1}{x_1} + \frac{p_{ij}^2}{x_2} + \frac{p_{ij}^3}{x_3} + \ldots + \frac{p_{ij}^n}{x_n}\}$ is an information granule.

Throughout this paper, we will use the above unified representation of multiple granular structures. Based on this representation, the paper explores how to effectively group granular structures.

3. Difference between granular structures

In the literature [53], Yao presented the concept of set closeness between two sets to measure the degree of the sameness (being the same) between sets. For A and B being two finite sets, this measure is defined by $H(A, B) = \frac{|A\cap B|}{|A\cup B|}$ $(A \cup B \neq \emptyset)$ [53]. Obviously, the formula $1 - H(A, B) = 1 - \frac{|A\cap B|}{|A\cup B|} = \frac{|A\cup B| - |A\cap B|}{|A\cup B|}$ $(A \cup B \neq \emptyset)$ can characterize the difference between two finite sets.

Given two granular structures $GS(P) = \{G_P(x) \mid x \in U\}$ and $GS(Q) = \{G_Q(x) \mid x \in U\}$, where the information granules $G_P(x) = \{\frac{p_1}{x_1} + \frac{p_2}{x_2} + \frac{p_3}{x_3} + \ldots + \frac{p_n}{x_n}\}$ and $G_Q(x) = \{\frac{q_1}{x_1} + \frac{q_2}{x_2} + \frac{q_3}{x_3} + \ldots + \frac{q_n}{x_n}\}$. Based on the view of set closeness outlined above, we can define the difference between the two information granules $G_P(x)$ and $G_Q(x)$ in the form

$$\frac{|G_P(x) \bigtriangleup G_Q(x)|}{|U|} = \frac{|G_P(x) \cup G_Q(x)| - |G_P(x) \cap G_Q(x)|}{|U|},$$

where $G_P(x) \cup G_Q(x) = \{\frac{p_1 \vee q_1}{x_1} + \frac{p_2 \vee q_2}{x_2} + \frac{p_3 \vee q_3}{x_3} + \ldots + \frac{p_n \vee q_n}{x_n}\}$, $G_P(x) \cap G_Q(x) = \{\frac{p_1 \wedge q_1}{x_1} + \frac{p_2 \wedge q_2}{x_2} + \frac{p_3 \wedge q_3}{x_3} + \ldots + \frac{p_n \wedge q_n}{x_n}\}$, and "$\vee$" is a maximum operator, "$\wedge$" is a minimum operator. Simply, we also use the notation $d(GS(P), GS(Q)) = |G_P(x) \bigtriangleup G_Q(x)|$.

Based on the above considerations, we introduce a so-called granular structure distance for measuring the difference between two granular structures from the same universe as follows.

Definition 1. Given a family of binary granular structures **GS** on U, $GS(P)$, $GS(Q) \in$ **GS**. Granular structure distance between $GS(P)$ and $GS(Q)$ is defined as

$$D(GS(P), GS(Q)) = \sum_{i=1}^{|U|} \frac{|G_P(x_i) \bigtriangleup G_Q(x_i)|}{|U|}, \tag{7}$$

where $|G_P(x_i) \bigtriangleup G_Q(x_i)| = |G_P(x_i) \cup G_Q(x_i)| - |G_P(x_i) \cap G_Q(x_i)|$.

The granular structure distance represents the measure of difference between two granular structures coming from the same universe. Obviously, $0 \le D(GS(P), GS(Q)) \le |U| - 1$.

In what follows, we need to verify whether the granular structure is a distance metric prior to using it in the process of grouping granular structures.

To address this issue, we first observe several cases existing in the context of sets.

Let A, B, C be three sets. We consider $A \cup B \cup C$ as a new set, and denote a certain array of all objects coming from this set by $Array = (t_1, t_2, \ldots, t_{|A\cup B\cup C|})$, $t_i \cap t_j = \emptyset$, t_i, $t_j \in A \cup B \cup C$. Using this array, we can represent the set A as $A' = \{a_1, a_2, \ldots, a_{|A\cup B\cup C|}\}$ by expressing

$$a_i = \begin{cases} 1 & \text{if } t_i \in A, \\ 0 & \text{else,} \end{cases}$$

for $t_i \in Array$.

Using this expression, one obtain the array expressions of A, B and C as follows

$A' = \{a_1, a_2, \ldots, a_{|A\cup B\cup C|}\}$,

$B' = \{b_1, b_2, \ldots, b_{|A\cup B\cup C|}\}$ and

$C' = \{c_1, c_2, \ldots, c_{|A\cup B\cup C|}\}$.

This representation is illustrated by the following example.

Example 1. Given three sets $A = \{1, 2, 3\}$, $B = \{2, 3, 4\}$ and $C = \{3, 4, 5\}$. We compute the new expressions of A, B and C through using the above method.

Obviously, we obtain $A \cup B \cup C = \{1, 2, 3, 4, 5\}$. Assume that the array is $Array = (1, 2, 3, 4, 5)$.

For A, one has $A' = (1, 1, 1, 0, 0)$. Similarly, $B' = (0, 1, 1, 1, 0)$ and $C' = (0, 0, 1, 1, 1)$.

Based on these denotations, we then express the distance between two sets in the form

$$d(A, B) = \sum_{i=1}^{|A\cup B\cup C|} (a_i \oplus b_i), a_i \in A', b_i \in B', \tag{8}$$

where $\oplus$ is an exclusive OR operator.

Analogously, one has that $d(B, C) = \sum_{i=1}^{|A\cup B\cup C|} (b_i \oplus c_i)$ and $d(A, C) = \sum_{i=1}^{|A\cup B\cup C|} (a_i \oplus c_i)$.

From the above, we formulate the following lemma.

Lemma 1. *Let A, B, C be three sets, then $d(A, B) + d(B, C) \ge d(A, C)$, $d(A, B) + d(A, C) \ge d(B, C)$ and $d(A, C) + d(B, C) \ge d(A, B)$.*

Proof. Suppose that $A' = \{a_1, a_2, \ldots, a_{|A\cup B\cup C|}\}$, $B' = \{b_1, b_2, \ldots, b_{|A\cup B\cup C|}\}$ and $C' = \{c_1, c_2, \ldots, c_{|A\cup B\cup C|}\}$. From $(a_i \oplus b_i) + (b_i \oplus c_i) \ge a_i \oplus c_i$, it follows that

$$\begin{aligned} d(A, B) + d(B, C) &= \sum_{i=1}^{|A\cup B\cup C|} (a_i \oplus b_i) + \sum_{i=1}^{|A\cup B\cup C|} (b_i \oplus c_i) \\ &= \sum_{i=1}^{|A\cup B\cup C|} ((a_i \oplus b_i) + (b_i \oplus c_i)) \\ &\ge \sum_{i=1}^{|A\cup B\cup C|} (a_i \oplus c_i) = d(B, C). \end{aligned}$$

Similarly, $d(A, B) + d(A, C) \ge d(B, C)$ and $d(A, C) + d(B, C) \ge d(A, B)$. □

If we use $G_P(x_i) \cup G_Q(x_i) \cup G_R(x_i)$ to re-express every information granule through the above expressions and denotations, one easily knows that

$$D(GS(P), GS(Q)) = \sum_{i=1}^{|U|} \frac{|G_P(x_i) \bigtriangleup G_Q(x_i)|}{|U|} = \sum_{i=1}^{|U|} \frac{d(G_P(x_i), G_Q(x_i))}{|U|}.$$

Table 1
Granular structures induced by binary relations P, Q and R respectively.

	x_1	x_2	x_3	x_4	x_5
$GS(P)$	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$	$\{x_4, x_5\}$
$GS(Q)$	$\{x_1, x_2\}$	$\{x_1, x_2\}$	$\{x_3\}$	$\{x_4, x_5\}$	$\{x_4, x_5\}$
$GS(R)$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_3\}$	$\{x_4\}$	$\{x_5\}$

Based this expression and Lemma 1, the following theorem is obtained.

Theorem 1. *Let* $\mathbf{GS}(U)$ *be the set of all granular structures induced by* U*, then* $(\mathbf{GS}(U), D)$ *is a distance space.*

Proof.

1. One can obtain easily that $D(K(P), K(Q)) \geq 0$ from Definition 1.
2. It is obvious that $D(GS(P), GS(Q)) = D(GS(Q), GS(P))$.
3. For the proof of the triangle inequality, one only needs to prove that for any $GS(P)$, $GS(Q)$, $GS(R) \in \mathbf{GS}(U)$, $D(GS(P), GS(Q)) + D(GS(P), GS(R)) \geq D(GS(Q), GS(R))$, $D(GS(R), GS(Q)) + D(GS(P), GS(R)) \geq D(GS(Q), GS(P))$ and $D(GS(R), GS(Q)) + D(GS(P), GS(Q)) \geq D(GS(P), GS(R))$ hold.

From Lemma 1, it is obvious that for any $x_i \in U$, $d(G_P(x_i), G_Q(x_i)) + d(G_P(x_i), G_R(x_i)) \geq d(G_Q(x_i), G_R(x_i))$, $d(G_P(x_i), G_Q(x_i)) + d(G_Q(x_i), G_R(x_i)) \geq d(G_P(x_i), G_R(x_i))$ and $d(G_P(x_i), G_R(x_i)) + d(G_Q(x_i), G_R(x_i)) \geq d(G_P(x_i), G_Q(x_i))$. Hence, one has

$$
\begin{aligned}
&D(GS(P), GS(Q)) + D(GS(P), GS(R)) \\
&= \sum_{i=1}^{|U|} \frac{|G_P(x_i) \bigtriangleup G_Q(x_i)|}{|U|} + \sum_{i=1}^{|U|} \frac{|G_P(x_i) \bigtriangleup G_R(x_i)|}{|U|} \\
&= \sum_{i=1}^{|U|} \frac{d(G_P(x_i), G_Q(x_i))}{|U|} + \sum_{i=1}^{|U|} \frac{d(G_P(x_i), G_R(x_i))}{|U|} \\
&= \sum_{i=1}^{|U|} \frac{1}{|U|} (d(G_P(x_i), G_Q(x_i)) + d(G_P(x_i), G_R(x_i))) \\
&\geq \sum_{i=1}^{|U|} \frac{d(G_Q(x_i), G_R(x_i))}{|U|} \\
&= D(GS(Q), GS(R)).
\end{aligned}
$$

Similarly, one can obtain that $D(GS(R), GS(Q)) + D(GS(P), GS(R)) \geq D(GS(Q), GS(P))$, $D(GS(R), GS(Q)) + D(GS(P), GS(Q)) \geq D(GS(R), GS(P))$.

Therefore, $(\mathbf{GS}(U), D)$ is a distance space. □

The above theorem is illustrated with the use of the following example.

Example 2. Given a set $U = \{x_1, x_2, x_3, x_4, x_5\}$ and $GS(P)$, $GS(Q)$, $GS(R)$ be three granular structures induced by binary relations P, Q, R on the set U, which are shown in Table 1.

Through computing their granular structure distances, we obtain
$D(GS(P), GS(Q)) = \frac{1}{5}(1 + 1 + 2 + 0 + 0) = \frac{4}{5}$,
$D(GS(P), GS(R)) = \frac{1}{5}(1 + 1 + 2 + 1 + 1) = \frac{6}{5}$ and
$D(GS(Q), GS(R)) = \frac{1}{5}(0 + 2 + 0 + 1 + 1) = \frac{4}{5}$.
Hence, one has $\frac{4}{5} + \frac{6}{5} = \frac{10}{5} > \frac{4}{5}$, $\frac{4}{5} + \frac{4}{5} = \frac{8}{5} > \frac{6}{5}$. It is clear that $D(GS(P), GS(Q)) + D(GS(P), GS(R)) \geq D(GS(Q), GS(R))$, $D(GS(P), GS(Q)) + D(GS(Q), GS(R)) \geq D(GS(P), GS(R))$ and $D(GS(Q), GS(R)) + D(GS(P), GS(R)) \geq D(GS(P), GS(Q))$.
From the above discussions, it can be seen that the proposed granular structure distance D is a distance metric on $\mathbf{GS}(U)$, which can be afterwards used to group granular structures.

4. Algorithm for grouping granular structures

In human granulation intelligence, a cognitive subject can observe various granular structures of an object from multiple perspectives or scales. However, these observed granular structures are not all necessary useful in problem solving, which need to be organized for some groups. One of objectives of grouping granular structures is to find several "natural" subsets so that granular structures within the same clusters are close to each other and those coming from different clusters are dissimilar from each other according to the predefined similarity metric [1,2]. In this study, the predefined dissimilarity measure is assigned as the proposed granular structure distance D.

Based on the granular structure distance D, we can construct a corresponding objective function for a grouping algorithm for grouping granular structures. An algorithm for grouping granular structures aims to search for a partition of a given granular structure set into k clusters that minimizes the objective function. Therefore, the objective of clustering a set of n

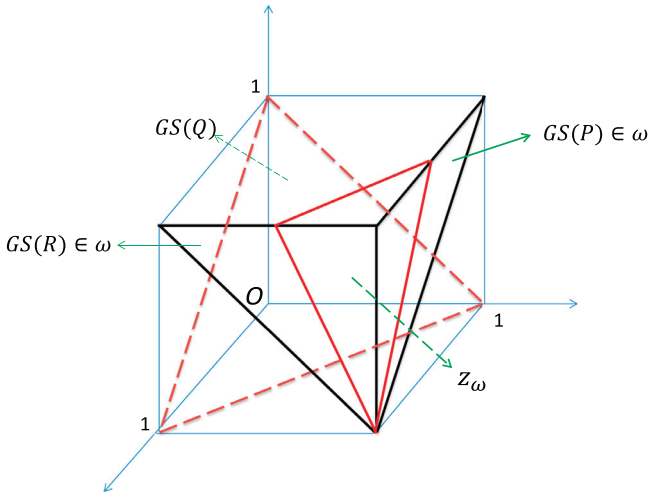


Fig. 3. Geometric semantics of granular structures in a cluster and its center.

granular structures into k clusters is to find Z that minimizes

$$F(Z) = \sum_{j=1}^{k} \sum_{GS \in \omega_j} D(GS, \mathbf{z_j}), \tag{9}$$

where $k(\leq n)$ is a certain number of clusters, and $Z = \{\mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z_k}\}$, $\mathbf{z_j}$ is the jth cluster center.

In what follows, we propose a granular structure distance-based algorithm for grouping a set of granular structures, which uses the same diagram of classic K-means clustering algorithm [1].

Algorithm 1. The granular distance-based algorithm for grouping granular structures (GGS)
Input: A set of granular structures $\mathbf{GS} = \{GS_1, GS_2, \ldots, GS_m\}$ and the number of groups k;
Output: k clusters.

1. In the granular structures, randomly choose k granular structures as initial cluster centers: $z_1^0, z_2^0, \ldots, z_k^0$. Let $l = 0$.
2. If $D(GS_i, z_c^{(l)}) = \min_j\{D(GS_i, z_s^{(l)})\}$, $i = 1, 2, \ldots, m$, then allocate GS_i to the cluster $\omega_c^{(l+1)}$, which generates new clusters $\omega_s^{(l+1)}$ $(s = 1, 2, \ldots, k)$.
3. Compute the center of each new cluster as
$z_s^{(l+1)} = \{G_{s1}^{(l+1)}, G_{s2}^{(l+1)}, \ldots, G_{sn}^{(l+1)}\}$, $s = 1, 2, \ldots, k$,
where
$$G_{sj}^{(l+1)} = \left\{ \frac{\frac{1}{n_s^{(l+1)}} \sum_{GS_i \in \omega_s^{(l+1)}} p_{i1}}{x_1} + \frac{\frac{1}{n_s^{(l+1)}} \sum_{GS_i \in \omega_s^{(l+1)}} p_{i2}}{x_2} + \ldots + \frac{\frac{1}{n_s^{(l+1)}} \sum_{GS_i \in \omega_s^{(l+1)}} p_{in}}{x_n} \right\}, \ s = 1, 2, \ldots, k,$$
where $n_s^{(l+1)}$ is the number of granular structures in the cluster $\omega_s^{(l+1)}$.
4. If $z_s^{(l+1)} = z_s^{(l)}$ $(s = 1, 2, \ldots, k)$, then the algorithm stops; otherwise, $l = l + 1$, goto (2).

In the above algorithm, the time complexity of determining cluster labels of n granular structures at each iteration is $O(kn)$. Let the number of iterations be l, we then conclude that the time complexity of the GGS algorithm is $O(knl)$, which is a linear time complexity like the classic K-means clustering algorithm. This implies that the GGS algorithm can efficiently group a given set of granular structures, which will be also seen in the scalability analysis completed in Section 6.3.

Now, we clarify the semantics of cluster centers, which could be helpful for better understanding working mechanism of the GGS algorithm. To address this issue, we give a geometric representation of a granular structure. For example, there are three granular structures within the same cluster ω as follows:

$$GS(P) = \left(\left\{\frac{1}{x_1} + \frac{1}{x_2} + \frac{0}{x_3}\right\}, \left\{\frac{1}{x_1} + \frac{0}{x_2} + \frac{1}{x_3}\right\}, \left\{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}\right\}\right),$$
$$GS(Q) = \left(\left\{\frac{1}{x_1} + \frac{0}{x_2} + \frac{0}{x_3}\right\}, \left\{\frac{0}{x_1} + \frac{1}{x_2} + \frac{0}{x_3}\right\}, \left\{\frac{0}{x_1} + \frac{0}{x_2} + \frac{1}{x_3}\right\}\right) \text{ and}$$
$$GS(R) = \left(\left\{\frac{1}{x_1} + \frac{1}{x_2} + \frac{0}{x_3}\right\}, \left\{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}\right\}, \left\{\frac{0}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}\right\}\right).$$

If these three objects be seen as the three corresponding coordinate x_1, x_2 and x_3, respectively, then the three granular structures can be represented as three triangles, which is shown in Fig. 3. In terms of this representation, the cluster center

of them will be also a triangle (see the red triangle Z_ω in Fig. 3). Although the center $Z_\omega = (\{\frac{1}{x_1} + \frac{1}{x_2} + \frac{0}{x_3}\}, \{\frac{1}{x_1} + \frac{0.5}{x_2} + \frac{1}{x_3}\}, \{\frac{0.5}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}\})$ is not a real neighborhood granular structure, it can be regarded as their cluster center. The cluster centers can be used to analyze the entire similarity within a cluster.

In what follows, we will investigate the convergence of the GGS algorithm. As a prerequisite, we first formulate two lemmas.

Lemma 2. *Given two information granules $G_P(x)$ and $G_Q(x)$, $x \in U$. Then, $d(G_P(x), G_Q(x))$ is a Minkowsky metric.*

Proof. Given a n dimensional vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, one has that

$l_p: \ \|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, \ i = 1, 2, \ldots, n,$

$l_1: \ \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|, \ i = 1, 2, \ldots, n.$

From Eq. (7), we have that

$|G_P(x) \triangle G_Q(x)|$

$= |G_P(x) \bigcup G_Q(x)| - |G_P(x) \bigcap G_Q(x)|$

$= |max\{G_P(x), G_Q(x)\}| - |min\{G_P(x), G_Q(x)\}|$

$= ||G_P(x) - G_Q(x)||$

$= \|G_P(x) - G_Q(x)\|_1$ where $||G_P(x) - G_Q(x)||$ indicates the cardinality of the absolute value of $G_P(\cdot) - G_Q(\cdot)$.

Hence, $d(G_P(x), G_Q(x))$ is a Minkowsky metric with l_p, where $p = 1$, i.e. l_1 metric between two samples. □

Example 3. (Continuation of Example 2). From Example 2, we get $G_P(x_1) = (1, 1, 1, 0, 0)$ and $G_Q(x_1) = (1, 1, 0, 0, 0)$, then

$|G_P(x_1) \triangle G_Q(x_1)|$

$= |G_P(x_1) \bigcup G_Q(x_1)| - |G_P(x_1) \bigcap G_Q(x_1)|$

$= |max\{G_P(x_1), G_Q(x_1)\}| - |min\{G_P(x_1), G_Q(x_1)\}|$

$= ||(1, 1, 1, 0, 0) - (1, 1, 0, 0, 0)||$

$= |(0, 0, 1, 0, 0)|$

$= 1$

$= \|G_P(x_1) - G_Q(x_1)\|_1.$

Lemma 3. *Given a n dimensional vector sequence $\{\mathbf{x}^m\}$, $m = 1, \ldots, \infty$. If its l_2 norm converges to a constant C, then its l_1 norm converges to $\sqrt{n}C$.*

Proof. The l_1 norm and the l_2 norm of the n dimensional vector sequence $\{\mathbf{x}^m\}$, $m = 1, \ldots, \infty$, can be described as

$l_1: \|\mathbf{x}^m\|_1 = \sum_{i=1}^n |x_i^m|, \ i = 1, 2, \ldots, n,$ and

$l_2: \|\mathbf{x}^m\|_2 = \left(\sum_{i=1}^n |x_i^m|^2\right)^{\frac{1}{2}}, \ i = 1, 2, \ldots, n.$

Then, one has

$\|\mathbf{x}^m\|_1^2 = (|\mathbf{x}_1^m| + |\mathbf{x}_2^m| + \ldots |\mathbf{x}_n^m|)^2 \le n(|\mathbf{x}_1^m|^2 + |\mathbf{x}_2^m|^2 + \ldots |\mathbf{x}_n^m|^2) = n\|\mathbf{x}^m\|_2^2.$

If

$\lim_{m\to\infty} \|\mathbf{x}^m\|_2 = C \Leftrightarrow \|\mathbf{x}^m\|_2 - C \to 0, \ m \to \infty,$ then

$\|\mathbf{x}^m\|_1 \le \sqrt{n}\|\mathbf{x}^m\|_2 \Leftrightarrow \|\mathbf{x}^m\|_1 - \sqrt{n}C \le \sqrt{n}(\|\mathbf{x}^m\|_2 - C) \to 0, \ m \to \infty,$ namely,

$\lim_{m\to\infty} \|\mathbf{x}^m\|_1 = \sqrt{n}C.$

Therefore, the l_1 norm of the sequence $\{\mathbf{x}^m\}$, $m = 1, \ldots, \infty$, will converge to $\sqrt{n}C$. This completes the proof. □

Example 4. Consider a n dimensional vector sequence $\{\mathbf{x}^m\}$, where $\mathbf{x}^m = (1 + \frac{1}{m}, 1 + \frac{1}{m}, \ldots, 1 + \frac{1}{m})$, $m = 1, \ldots, \infty$. Its l_2 norm converges to $\sqrt{n}$ and its l_1 norm converges to n, which are as follows

$\|\mathbf{x}^m\|_2 = \sqrt{n|1 + \frac{1}{m}|^2} \to \sqrt{n}$

and

$\|\mathbf{x}^m\|_1 = n|1 + \frac{1}{m}| \to n.$

Through using the above two lemmas, we prove the convergence of the GGS algorithm based on the granular structure distance.

Theorem 2. *The GGS algorithm based on the granular structure distance is convergent.*

Proof. First, we give a brief analysis that the GGS algorithm is convergent in the sense of l_2 metric.

From Lemma 2, we know that granular structure distance can be induced to a Minkowsky metric. Hence, we can assume that cluster w_j has n_j granular structures, and the entire distance between the granular structures and the center z_j of the cluster w_j is

$f_j = \sum_{GS \in w_j} \|GS - z_j\|^2, \ j = 1, 2, \ldots, k.$

When assigning the granular structure $GS_i, i \in \{1, 2, \ldots, n_j\}$ from cluster w_j to its nearest cluster w_k, then cluster w_j is changed to $\tilde{w}_j$ and the cluster center z_j is changed to $\tilde{z}_j$. Similarly, cluster w_k and its center is also changed in the same manner. Obviously,

$$\tilde{z_j} = \frac{n_j z_j - GS_i}{n_j - 1} = z_j - \frac{1}{n_j - 1}(GS_i - z_j),$$
$$\tilde{z_k} = \frac{n_k z_k + GS_i}{n_k + 1} = z_k + \frac{1}{n_k + 1}(GS_i - z_k).$$

Then, one has $\tilde{f}_j = \sum_{GS\in\tilde{w}_j} \|GS - \tilde{z_j}\|^2$

$$\begin{aligned}
&= \textstyle\sum_{GS\in\tilde{w}_j} \|GS - z_j + \frac{1}{n_j - 1}(GS_i - z_j)\|^2 \\
&= \textstyle\sum_{GS_i\in\tilde{w}_j} \|GS - z_j\|^2 + 2\sum_{GS_i\in\tilde{w}_j}(GS - z_j)\cdot\frac{1}{n_j - 1}(GS_i - z_j) + \sum_{GS\in\tilde{w}_j}\frac{1}{(n_j-1)^2}\|GS_i - z_j\|^2 \\
&= f_j - \|GS_i - z_j\|^2 + \frac{2}{n_j - 1}(GS_i - z_j)\left[\textstyle\sum_{GS_i\in\tilde{w}_j} GS - (n_j - 1)z_j\right] + \frac{1}{n_j - 1}\|GS_i - z_j\|^2 \\
&= f_j + \frac{2 - n_j}{n_j - 1}\|GS_i - z_j\|^2 + \frac{2}{n_j - 1}(GS_i - z_j)[(n_j - 1)(\tilde{z_j} - z_j)] \\
&= f_j + \frac{2 - n_j}{n_j - 1}\|GS_i - z_j\|^2 - \frac{2}{n_j - 1}(GS_i - z_j)^2 \\
&= f_j - \frac{n_j}{n_j - 1}\|GS_i - z_j\|^2
\end{aligned}$$

In the similar way, we obtain the following

$\tilde{f}_k = f_k + \frac{n_k}{n_k + 1}\|GS_i - z_k\|^2$.

According to the grouping mechanism, one knows that $\|GS_i - z_j\|^2 \geq \|GS_i - z_k\|^2$. Thus,

$\frac{n_k}{n_k + 1}\|GS_i - z_k\|^2 < \frac{n_j}{n_j - 1}\|GS_i - z_j\|^2$.

Summarizing the above, we have that $\tilde{f}_j + \tilde{f}_k < f_j + f_k$.

It shows that in the sense of l_2 norm, using the principle that assigning a granular structure to its nearest cluster center with the granular structure distance D can make the value of objective function $F(Z)$ converge to the minimum F^*. Combing this principle and Lemma 3 together, we have

$$\begin{aligned}
F(Z) &= \textstyle\sum_{j=1}^{k}\sum_{GS\in w_j} D(GS, z_j) \\
&= \textstyle\sum_{j=1}^{k}\sum_{GS\in w_j} \|GS - z_j\|_1 \\
&= \textstyle\sum_{j=1}^{k}\sum_{GS\in w_j}\sum_{i=1}^{n} \|GS(x_i) - z_j(x_i)\|_1 \\
&\leq \textstyle\sum_{j=1}^{k}\sum_{GS\in w_j}\sum_{i=1}^{n} \sqrt{n}\|GS(x_i) - z_j(x_i)\|_2 \\
&= \textstyle\sqrt{n}\sum_{j=1}^{k}\sum_{GS\in w_j}\sum_{i=1}^{n} \|GS(x_i) - z_j(x_i)\|_2 \to \sqrt{n}F^*
\end{aligned}$$

that is $F(Z) \to \sqrt{n}F^*$, where n is the number of granular structures to be grouped. This completes the proof. □

Thanks to the convergence of the GGS algorithm, the minimization of F in (9) can be finished when adopting the granular structure distance D. Though The GGS algorithm often terminates at a local optimum, it is very efficient when processing a large number of granular structures.

5. Evaluation on performance of grouping granular structures

In existing clustering analysis, there are several known evaluation indices for evaluating the performance of a clustering result, such as clustering accuracy [2], category utility function [9], adjusted rank index [15], which are based on class labels of objects. If class label of each of granular structures is beforehand given, then these evaluation indices can be directly used to evaluate the performance of grouping granular structures. However, because grouping granular structures arise as a new research issue, there are no open granular structure set to be employed for evaluating performance of an algorithm for grouping granular structures. Hence, there are no known evaluation method to be used for this objective. In this section, we explore how to evaluate the effectiveness of the proposed method.

5.1. Dispersion

As mentioned in Section 4, the objective of grouping granular structures is to find several natural subsets so that granular structures within the same clusters are close to each other and those from different clusters are dissimilar from each other according to the predefined similarity metric. This implies that those granular structures within each cluster should have high cohesion and low dispersion. To quantitatively characterize the inner dispersion of a grouping result of granular structures, we introduce a so-called dispersion measure.

We analyze the case of a cluster including some granular structures. Fig. 4 visualizes a cluster ω_u and the granular structures within it. To measure the dispersion of this cluster, one first regards the center of this cluster as a reference point and then computes the sum of granular structure distances between every granular structure GS_i and this center Z_u. The smaller the value of the sum, the higher the dispersion of this cluster, and the better its clustering performance.

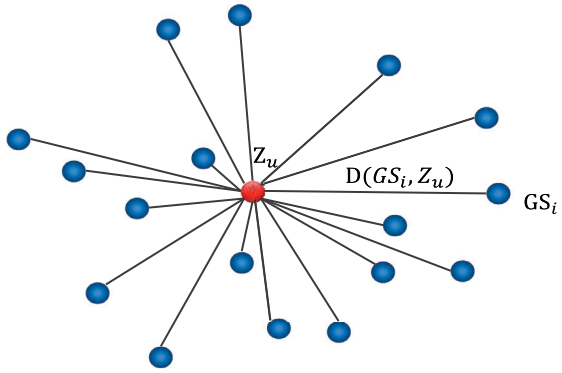


Fig. 4. Inner dispersion of granular structures in the same cluster.

Based on this consideration, we can present the following definition of the dispersion of a grouping result of granular structures.

Definition 2. Given a grouping result $\omega_1, \omega_2, \ldots, \omega_k$ of a set of granular structures. The dispersion of the grouping result is defined as

$$Dis(Z) = \frac{1}{k}\sum_{s=1}^{k}\frac{|\omega_s|}{n}\sqrt{\frac{1}{|\omega_s|}\sum_{t=1,\ GS_t\in\omega_s}^{|\omega_s|} D^2(GS_t, Z_{\omega_s})}, \tag{10}$$

where $D(GS_t, Z_{\omega_s}) = \frac{1}{|U|}\sum_{i=1}^{|U|}|GS_t(x_i) \bigoplus Z_{\omega_s}(x_i)|$ and $|\omega_s|$ is the number of granular structures within the cluster ω_s.

From the above definition, we state that for a given k, if the value of the dispersion $Dis(Z)$ of a grouping result is much smaller, one can say this result have much better grouping performance.

5.2. Approximation degree

Rough set theory, originated by Pawlak [24,25], has become a well-established mechanism for uncertainty management in a wide variety of applications related to artificial intelligence [6,8,10,24,29]. One of the strengths of rough set theory is that all its parameters are obtained from the given data. The numerical value of imprecision is not pre-assumed, but is calculated on the basis of approximations which are the fundamental concepts used to express imprecision of knowledge. In other words, the rough set data analysis utilizes solely the granular structure of the given data, expressed as classes of suitable binary relations. Based on the idea of approximation in rough set theory, in this subsection, we propose a so-called approximation degree (APD) for characterizing the average approximation ability of a group of granular structures for a given decision.

Let U be a finite and non-empty set called the universe and $R \subseteq U \times U$ a binary relation on U. Then $K = \langle U, R\rangle$ is called an approximation space [6,8,10,24,29]. The binary relation R divides the set U into some subsets, which constructs a granular structure of U, denoted by $GS(R) = (G_R(x_1), G_R(x_2), \ldots, G_R(x_n))$.

Given an approximation space $K = \langle U, R\rangle$ and an arbitrary subset $X \subseteq U$, one can construct a rough set of the set on the universe by elemental information granules in the following definition:

$$\begin{cases} \underline{R}X = \{x \mid G_R(x) \subseteq X,\ x \in U\}, \\ \overline{R}X = \{x \mid G_R(x) \cap X \neq \emptyset,\ x \in U\}. \end{cases}$$

where $\underline{R}\ X$ and $\overline{R}X$ are called R-lower approximation and R-upper approximation with respect to R, respectively. The order pair $\langle \underline{R}X,\ \overline{R}X\rangle$ is called a rough set of X with respect to the binary relation R.

There are two kinds of attributes for a classification problem, which can be characterized by a decision table $S = (U, C \cup D)$ with $C \cap D = \emptyset$, where an element of C is called a condition attribute, C is called a condition attribute set, an element of D is called a decision attribute, and D is called a decision attribute set. The attribute set D determines an equivalence relation in the following way:

$R_D = \{(x, y) \in U \times U \mid d(x) = d(y),\ \forall d \in D\}$,

where $d(x)$ and $d(y)$ denote the values of objects x and y under a decision attribute d, respectively. This equivalence relation R_D partitions U into some equivalence classes (also called a decision) given by

$U/R_D = \{[x]_D \mid x \in U\}$, for simplicity, U/R_D will be replaced by U/D,

where $[x]_D$ denotes the equivalence class determined by x with respect to D, i.e., $[x]_D = \{y \in U \mid (x, y) \in R_D\}$.

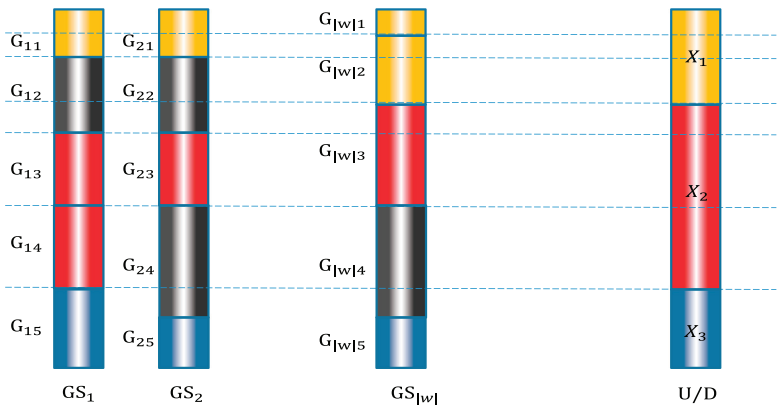


Fig. 5. Approximation ability of granular structures in the same cluster.

Assume the objects are partitioned into r mutually exclusive crisp subsets $\{X_1, X_2, \ldots, X_r\}$ by the decision attributes D. Given any binary relation R and its granular structures, then one can define the lower and upper approximations of the decision attributes D as

$$\begin{cases} \underline{R}D = \{\underline{R}X_1, \underline{R}X_2, \ldots, \underline{R}X_r\}, \\ \overline{R}D = \{\overline{R}X_1, \overline{R}X_2, \ldots, \overline{R}X_r\}. \end{cases}$$

Denoted by

$POS_R(D) = \bigcup_{i=1}^{r} \underline{R}X_i,\ 1 \le i \le r,$

it is called the positive region of D with respect to the binary relation R.

In rough set theory, the dependency function (or level of consistency [24]) is used to characterize the dependency degree of an attribute subset with respect to a given decision [13,29]. Given a decision table $S = (U, C \cup D)$, the dependency function of condition attributes C with respect to the decision attribute D is formally defined as $\gamma_C(D) = |POS_C^U(D)|/|U|$. Using this idea, it can be easily extended to measure the approximation ability of a granular structure for approximating a given decision with the following form.

$apr(GS, D) = \frac{|POS_{GS}(D)|}{|U|}$,

where $|\cdot|$ denotes the cardinality of a set and $0 \le apr(GS, D) \le 1$.

In this study, we call the above measure an approximation degree. The approximation degree reflects a granular structure GS's power to approximate a given decision D. When $apr(GS, D) = 1$, one says D completely depends on the granular structure GS. This means that the decision can be precisely described by the information granules from the granular structure GS. This approximation degree can be used to design a measure for evaluating the performance of a grouping result of granular structures.

Definition 3. Given a grouping result $\omega_1, \omega_2, \ldots, \omega_k$ of a set of granular structures, and a decision D. The approximation degree of the grouping result is defined as

$$APD(Z) = \frac{1}{k}\sum_{s=1}^{k}\frac{|\omega_s|}{n}\sqrt{\frac{1}{|\omega_s|}\sum_{t=1,\ GS_t\in\omega_s}^{|\omega_s|}(|POS_{GS_t}(D)| - \frac{1}{|\omega_s|}\sum_{t=1,\ GS_t\in\omega_s}^{|\omega_s|}|POS_{GS_t}(D)|)^2}, \tag{11}$$

where $|POS_{GS_t}(D)|$ is the number of objects in the positive region of D with respect to the granular structure GS_t.

Fig. 5 gives $|\omega|$ granular structures in a cluster ω and a decision U/D. From the principle of the GGS algorithm, these granular structures in the same cluster should be close to each other. This indicates that these granular structures could have similar ability for approximating a given decision. This implies the variance of their positive regions may be small. For a result of grouping granular structures, we have the same observation. Hence, the approximation degree $APD(Z)$ also can be seen an index of evaluating the performance of a grouping result of granular structures.

From the above discussion and analysis, we state that for a given k, if the value of the approximation degree $APD(Z)$ of a grouping result is much smaller, one can say this result have much better grouping performance.

In the experimental analysis in next section, the dispersion $DIS(Z)$ and the approximation degree $APD(Z)$ will be employed for displaying the performance of grouping results of granular structures.

It deserves to point out that because grouping granular structures is a new research issue, the dispersion DIS and the approximation degree APD are only two tentative indexes, which may not be perfect measures for evaluating clustering performance of a grouping result of granular structures. It should be said that this in itself is a very interesting and important research problem, which can be further studied in the future.

Table 2
Description of nine publicly available data sets from UCI.

	Data sets	Number of samples	Number of features	Number of granular structures
1	Soybean-small	47	35	157
2	Breast cancer	699	9	106
3	Glass	214	9	404
4	Haberman	306	3	822
5	Iris	150	4	731
6	Hayes–Roth	132	4	45
7	Seeds	210	7	568
8	Wine	178	13	248
9	Zoo	101	16	69

6. Experimental studies

The primary objective of the grouping algorithms is to discover the grouping structures inherent in granular structures. As the assumption is that a certain structure may exist in a given data set, a clustering algorithm is used to verify the assumption and recover the inherent structure. If the algorithm can discover the structure, then it may form a good solution. In this section, we aim to test the convergence, effectiveness and scalability of the GGS algorithm for grouping granular structures from a data set coming with various granulations.

The nine real data sets used in the experiments are outlined in Table 2, which were all downloaded from UCI repository of machine learning databases [66].

In order to perform experimental analysis, one needs to generate a collection of granular structures from each data set. Without loss of generality, we adopt a neighborhood relation to a collection. As we recall that the neighborhood relation R on the universe is a relation matrix $M(R) = (r_{ij})_{n\times n}$, where

$$r_{ij} = \begin{cases} 1, & d(x_i, x_j) \le \delta, \\ 0, & \text{otherwise}, \end{cases}$$

where d is a certain distance [14,65] between x and y, δ is a nonnegative number. Let $B \subseteq AT$ be numerical attributes, the neighborhood granules of objects x induced by B can be determined by

$N_B(x) = \{x_i \in U \mid d_B(x, x_i) \le \delta\}$,where δ varies from 0 to 0.5 with the step 0.0005. Through granulating the data set with this step, we can obtain 1000 granular structures of a given data set. Through removing the duplicate granular structures, we produce a set of different granular structures from each data set, which is outlined as Table 2. These granular structures from every data set are used in experimental analysis in the experiments. All experimental work was performed by a high-performance computer (Inter(R) Xeon(R) CPU, 2.00 GHz (Double) and 16 GB memory) in this section.

6.1. Convergence analysis

Let us investigate and test the convergence of the proposed GGS algorithm with two dissimilarity measures d_1 and d_2.

In the experiment, we carried out 100 runs of the algorithm on the set of granular structures coming from each of the nine data sets, respectively. In each run, the initial cluster prototypes of the GGS algorithm were randomly generated. The convergence behavior of the algorithm is displayed in Fig. 6. Here we show the 100 curves, where each curve refers to the values of the objective function obtained in the successive iterations of the algorithm.

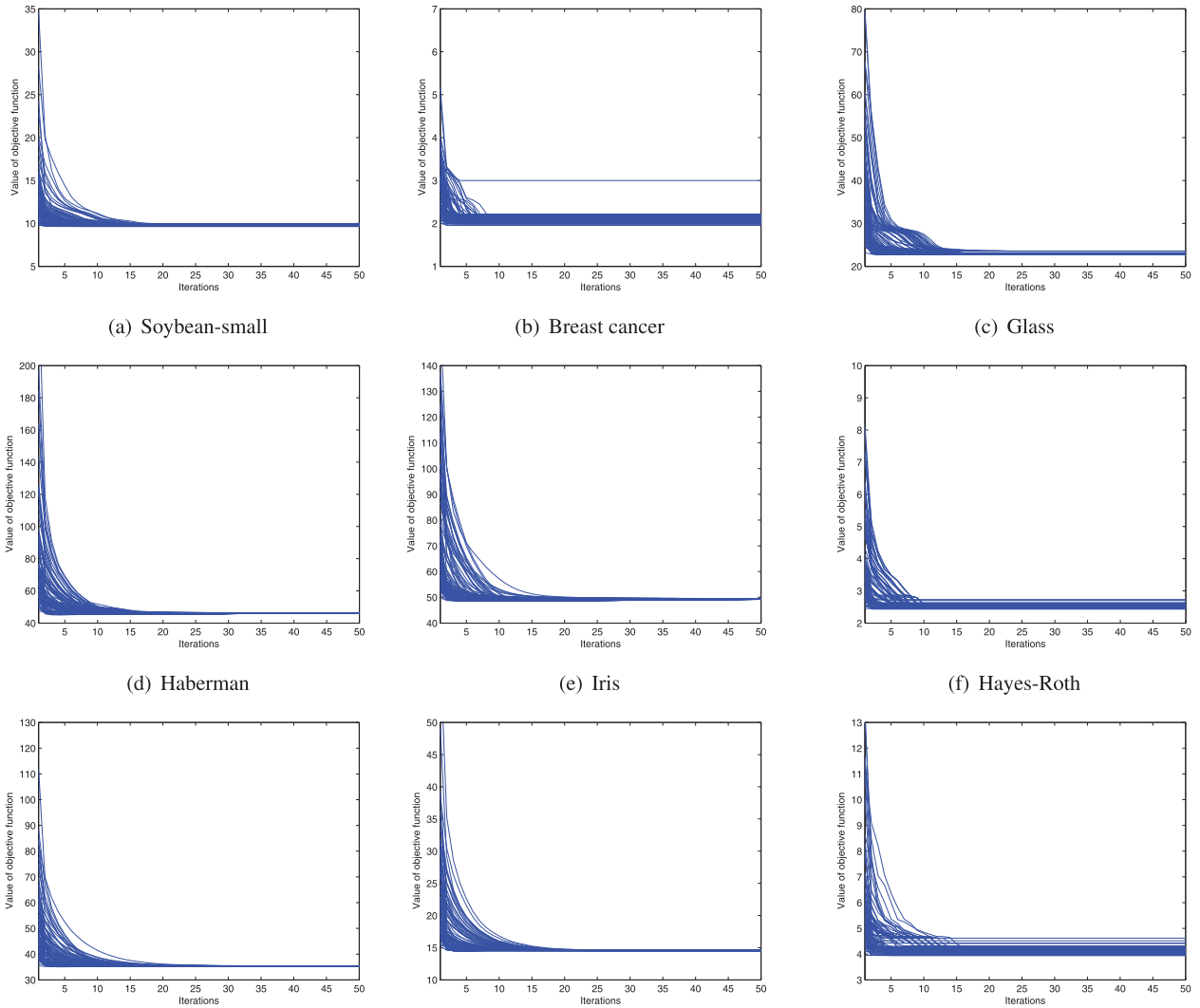
It can be seen from Fig. 6, that values of the objective function decrease over the progression of the algorithm. The main progress is noted at the initial iterations and afterwards the method converges and does not result in further reduction of the values of the objective function. This shows that the GGS algorithm is convergent, which can be used safely for grouping granular structures in human granulation intelligence.

6.2. Effectiveness analysis

To evaluate the performance of a grouping result of the granular structures, two validity indices were proposed in Section 5, namely dispersion and approximation degree. In this section, we test the grouping performance of the GGS algorithm.

Firstly, we test the performance of the GGS algorithm using the index *DIS* on the nine data sets in Table 2. The dispersion index is used to measure the dispersion degree of the grouping result of granular structures. The lower the value of this index on a grouping result, the better the clustering performance of this result. In this experimental analysis, we let the numbers of clusters vary from 2 to $\sqrt{n}$ [1]. The obtained experimental results are shown in Fig. 7.

[1] In the literature [58], Yu argued that possible number of clusters often lies within the interval $[2, \sqrt{n}]$, where n is the number of objects in the data set to be clustered.



(a) Soybean-small (b) Breast cancer (c) Glass

(d) Haberman (e) Iris (f) Hayes-Roth

(g) Seeds (h) Wine (i) Zoo

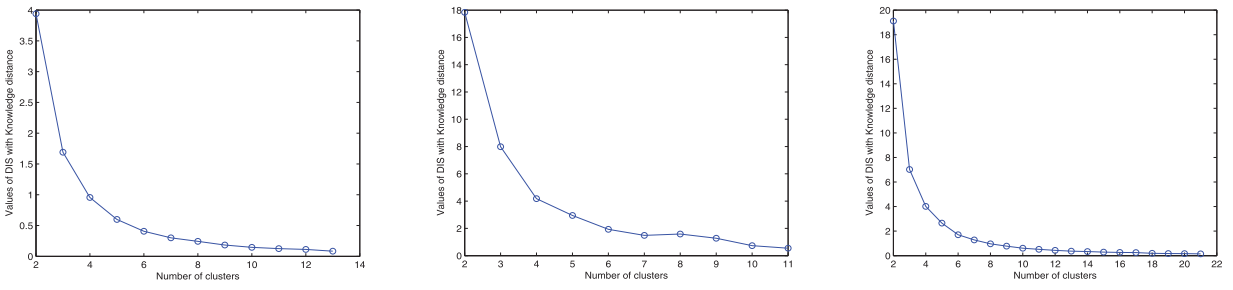
Fig. 6. The values of objective function in against iterations with different initial cluster centers.

It is easy to see from Fig. 7 that the *DIS* index consistently decreases as the number of clusters increases for the same set of granular structures, which is able to quickly achieve a smaller value. This implies that the GGS algorithm can obtain a better grouping result for a given number of clusters.

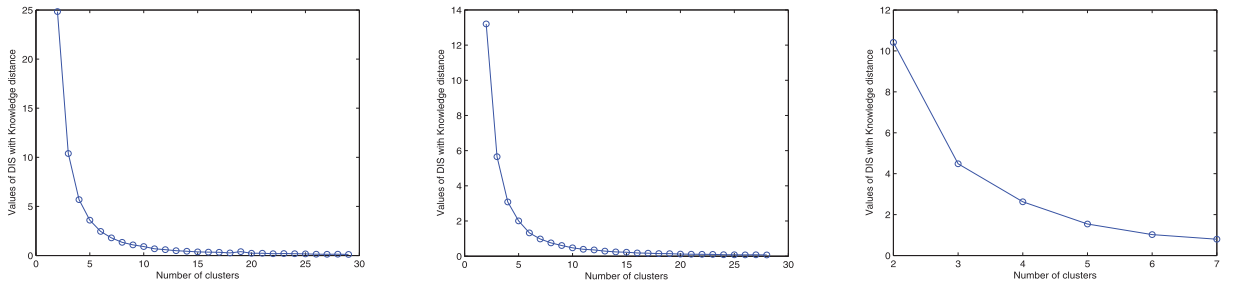
In what follows, we consider another index *APD* for evaluating the performance of the GGS algorithm from a different point of view. The approximation degree is employed to characterize the mean approximation degree of every clusters from a grouping result. Like the *DIS*, if a grouping result is better, then the value of the *APD* index becomes smaller. The values of the index *APD* are presented in Fig. 8.

From Fig. 8, the values of the *APD* index tends to decrease as the number of clusters increases for the same set of granular structures, which is also able to quickly achieve a smaller value. Like the *DIS* index, this means that the GGS algorithm can obtain a better grouping result for a given number of clusters. However, this trend is not strictly monotonic. This is because grouping the same granular structures into different number of clusters generates different grouping results, which may have different positive regions for a given decision.

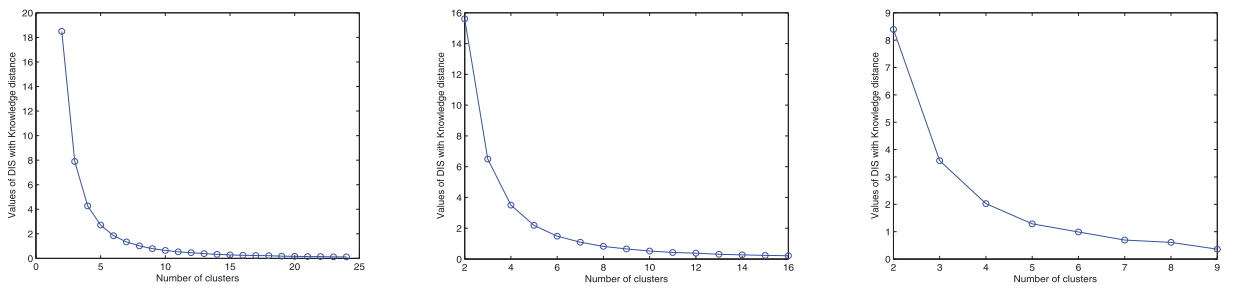
Given the above two indices, the performance of the GGS algorithms induced by different dissimilarities can be compared when dealing with the same number of clusters. If there exists a new dissimilarity measure between granular structures such that the value of *DIS* index is much smaller than that being generated by other measures, then this dissimilarity measure may lead to a better selection.



(a) Soybean-small (b) Breast cancer (c) Glass



(d) Haberman (e) Iris (f) Hayes-Roth



(g) Seeds (h) Wine (i) Zoo

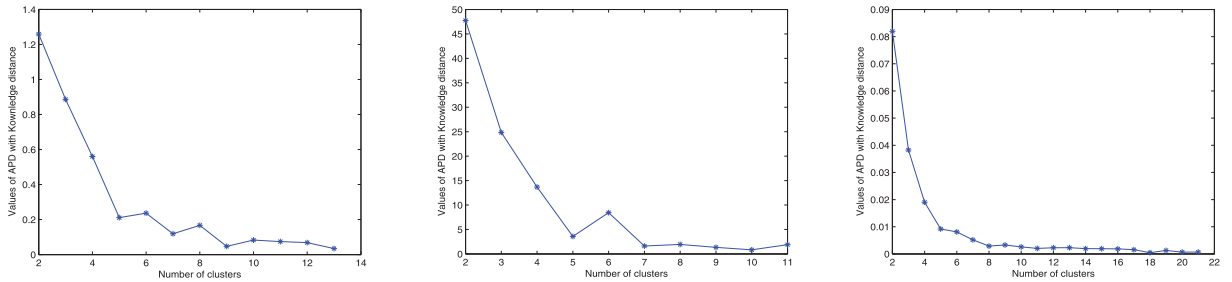
Fig. 7. Values of *DIS* index in against number of clusters.

6.3. Scalability analysis

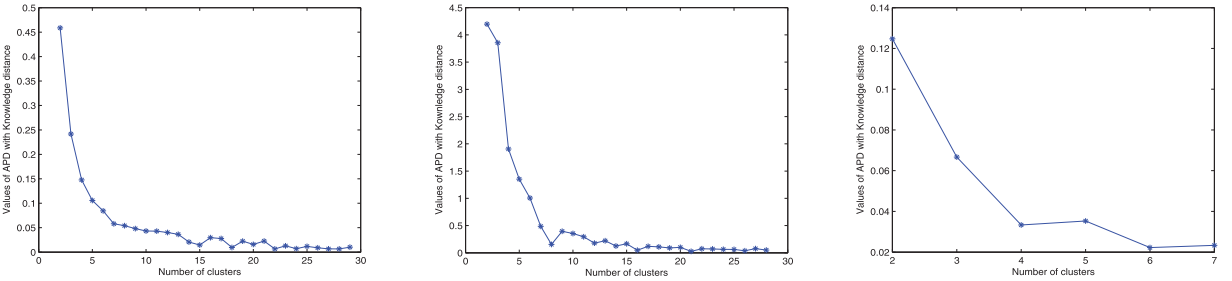
The purpose of this experiment is to test the scalability of the GGS algorithm.

We tested two types of scalability of the GGS algorithm on a family of granular structures. The first one is the scalability against the number of granular structures for a given number of clusters, and the second is the scalability against the number of clusters for a given number of granular structures. Fig. 9 shows the results of using the GGS algorithm to group different numbers of granular structures into 5 clusters. Fig. 10 displays the results of using the GGS algorithm to the obtained granular structures from each of the nine data sets into different numbers of clusters. In Fig. 10, the numbers of clusters also vary from 2 to $\sqrt{n}$, where n is the number of different granular structures induced by a given data set with the step of 0.0005. The plots show the computational time of the GGS algorithm for every grouping task.

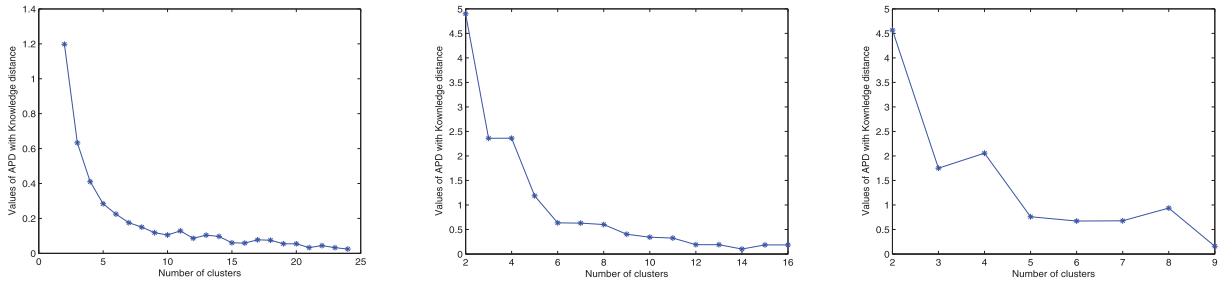
Referring to Fig. 9, one important observation is that the computing time of the GGS algorithm tends to increase linearly as the number of granular structures increases. This observation is consistent with the linear time complexity of the GGS algorithm. Of course, in some cases (Fig. 9(d) and (g)) grouping less granular structures into a given number of clusters took more time than grouping more granular structures into the same number of clusters. This reason is that the former clustering may take more iterations to converge than the latter one. In addition, as one of the important advantages of the GGS algorithm, we also can see that the GGS algorithm is very efficient. For example, the computational time shown in Fig. 9(f) is less than 0.4 s on the data set Hayes–Roth, while the computational time as shown in Fig. 9(i) is less than 0.22 s on the data set Zoo.



(a) Soybean-small (b) Breast cancer (c) Glass



(d) Haberman (e) Iris (f) Hayes-Roth



(g) Seeds (h) Wine (i) Zoo

Fig. 8. Values of *APD* index in against number of clusters.

From Fig. 10, we observe that the computing time of the GGS algorithm tends to increase with the number of clusters becoming larger. But this trend is not especially monotonic, as the grouping the same granular structures into different number of clusters may take different number iterations and it could take different computing time to converge. In particular, we see that the GGS algorithm is very efficient for various numbers of cluster.

Finally, we also tested scalability of the GGS algorithm on a very large set of granular structures. An artificial data set with 100 objects is employed for this experiment. We first randomly generate 50,000 granular structures on this artificial data set. Then, we observe the scalability against the number of granular structures for three given numbers of clusters, which are set as $k = 3, 5, 7,$ respectively. Fig. 11 shows the results of using the GGS algorithm to group different numbers of granular structures into 3, 5 and 7 clusters, respectively. In this figure, the $x-$coordinate axis means the number of granular structures, and the $y-$coordinate axis is the computational time of the GGS algorithm for every grouping task.

It can be seen from Fig. 11 that the computational time of the GGS algorithm also tends to increase linearly with the number of granular structures varying from 10,000 to 50,000. Another observation is that grouping the same number of granular structures into a larger number of clusters took more computing time than grouping them into the lower number of clusters. These two observations are consistent with the linear time complexity of the GGS algorithm. In particular, we also can see that the GGS algorithm is very efficient for grouping a large-scale granular structure set. For example, the GGS algorithm only spent almost 250 s when $k = 7$. Hence, we can say that the GGS algorithm is an effective and efficient approach to discover grouping property inherent in various granular structures.

Summarizing the above experimental analysis, the obtained results tell us that the proposed GGS algorithm does not only guarantee to be convergent but also becomes efficient in the determination of better clustering results. The *DIS* and *APD* are

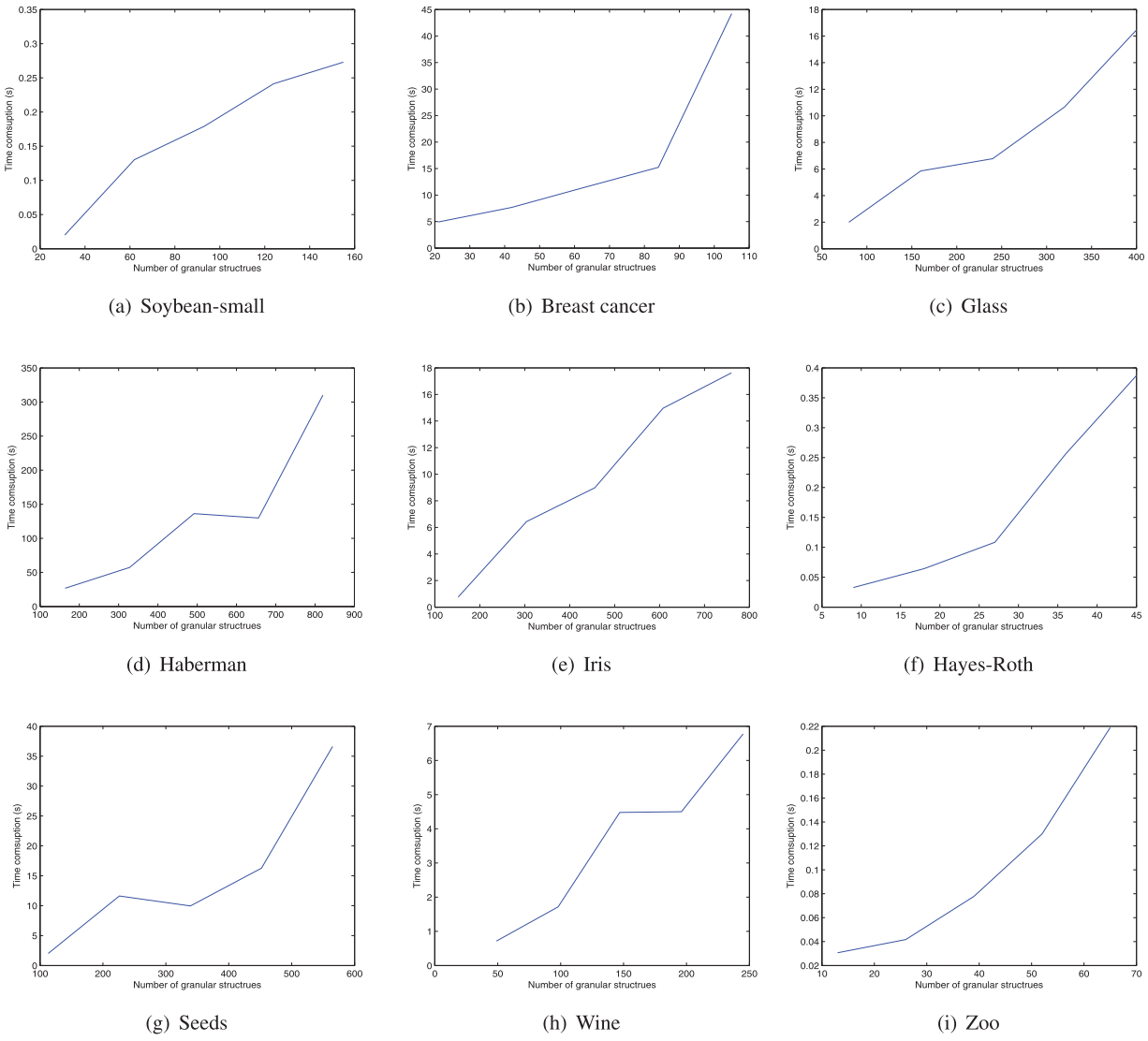
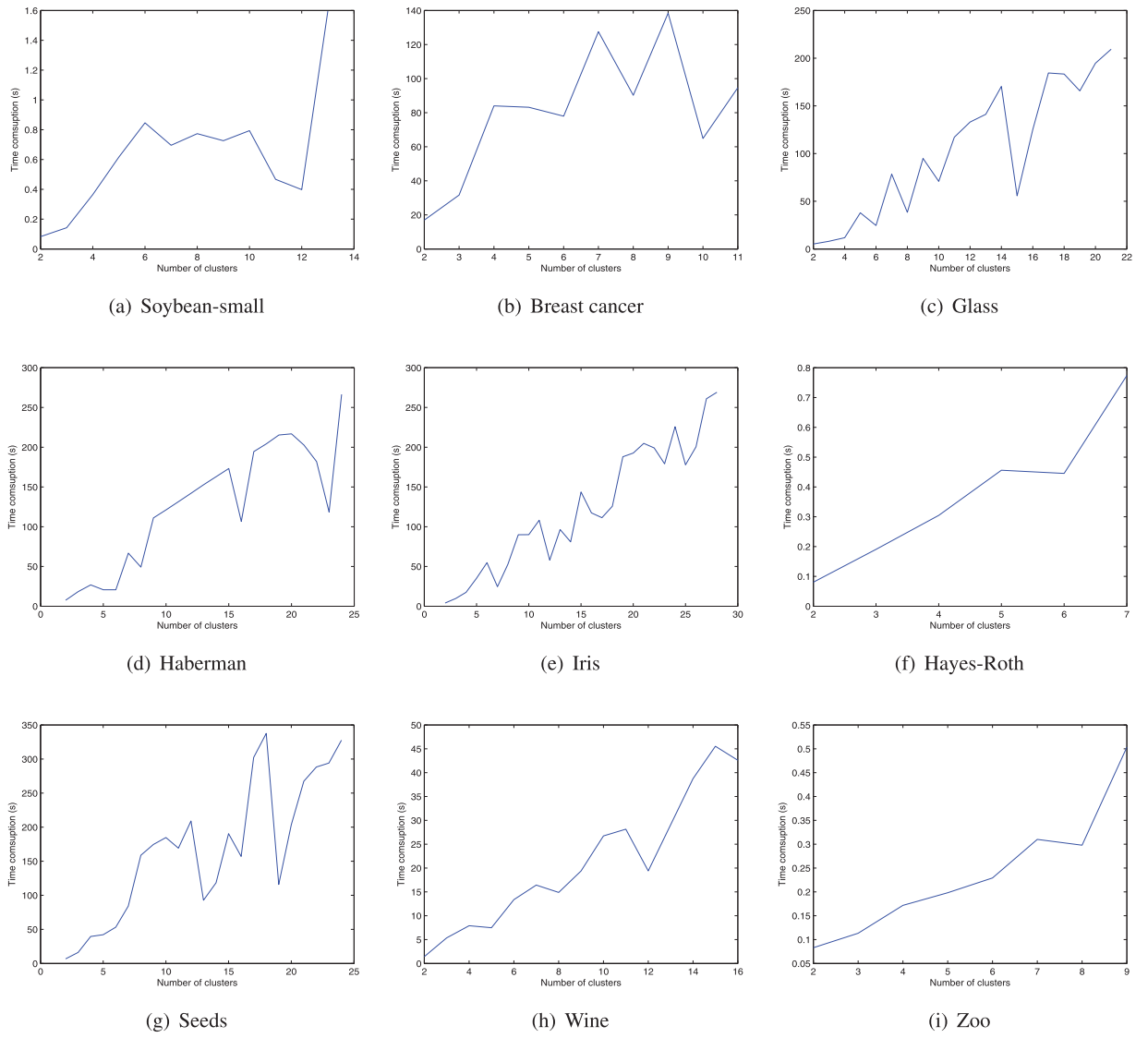


Fig. 9. Computing time of GGS algorithm in against numbers of granular structures.

effective indices to evaluate the performance of grouping granular structures. It is worth noting that, the novel point of the GGS algorithm is to propose a framework for discovering the grouping structure inherent in granular structures, which can be used to simulate intelligent behavior of human's quickly selecting information granularity.

7. Conclusions

From different granulation viewpoints, people can observe and analyze the same problem, which is called human granulation intelligence. For a given data set, through adopting different scales, a large number of granular structures can be generated. However, people can group these granular structures and select some representative ones for problem solving. This leads to an important problem of how to efficiently and effectively group a family of granular structures. As the main motivation of this paper, we have done a tentative research from five steps. The first step has given a unified knowledge representation for unifying various granular structures induced by different binary relations, which is helpful for establishing a general framework for grouping granular structures. Based on this unified knowledge representation, in second step, a so-called granular structure distance has been introduced to calculate the difference between two granular structures. Some properties of this distance have been analyzed to show its rationality. In third step, a general framework (GGS) has been developed to quickly group a set of granular structures. To evaluate the performance of a grouping result of granular structures, in fourth step, we have designed two indices: the dispersion *DIS* and the approximation degree *APD*. Finally, the fifth step has employed nine real data sets and one artificial set of granular structures for testing the performance of the GGS



(a) Soybean-small (b) Breast cancer (c) Glass

(d) Haberman (e) Iris (f) Hayes-Roth

(g) Seeds (h) Wine (i) Zoo

Fig. 10. Computing time of GGS algorithm in against numbers of clusters.

algorithm. The experimental results from convergence, effectiveness and scalability, have shown that the GGS algorithm is a good solution, which can be safely used to discover the grouping structure inherent in granular structures. This framework would provide a strategy for simulating intelligent behavior of human's grouping granular structures.

Within the broad area of grouping granular structures, this study delivers preliminary albeit interesting and promising results. In future works, there are two interesting and important issues to be looked at. One is how to effectively measure the dissimilarity in-between granular structures. The other is how to design validity indices for evaluating the performance of a grouping result of granular structures. The studies along this line would significantly promote the research of granulation intelligence and information processing.

Acknowledgment

We would like to thank Research Assistants Rui Han and Yebin Li for performing a large number of numeric experiments for several months. The authors also thank Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education for the usage of one high-performance computer for realizing the empirical study.

This work was supported by the National Natural Science Fund of China (Nos. 61672332, 61322211, 61432011, U1435212), Program for New Century Excellent Talents in University (No. NCET-12-1031), National Key Basic Research and Development Program of China(973) (Nos. 2013CB329404, 2013CB329502), Program for the Outstanding Innovative Teams of Higher Learning Institutions of Shanxi, and Program for the Young San Jin Scholars of Shanxi.

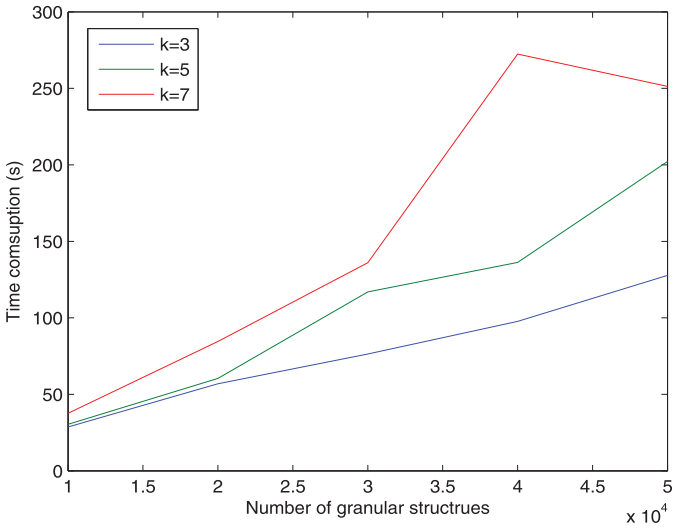


Fig. 11. Scalability of GGS algorithm against the number of granular structures with $k = 3, 5$ and 7.

References

[1] M.R. Anderberg, Cluster Analysis for Applications, Academic Press, 1973.
[2] L. Bai, J.Y. Liang, C.Y. Dang, F.Y. Cao, The impact of cluster representatives on the convergence of the k-modes type clustering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1509–1522.
[3] S. Calegari, D. Ciucci, Granular computing applied to ontologies, Int. J. Approx. Reasoning 51 (4) (2010) 391–409.
[4] D. Ciucci, D. Dubois, A map of dependencies among three-valued logics, Inf. Sci. (Ny) 250 (2013) 162–177.
[5] M. Dash, H. Liu, Consistency-based search in feature selection, Artif. Intell. 151 (2003) 155–176.
[6] I. Düntsch, G. Gediga, Uncertainty measures of rough set prediction, Artif. Intell. 106 (1998) 109–137.
[7] D. Dübois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (1990) 191–209.
[8] G. Gediga, I. Düntsch, Rough approximation quality revisited, Artif. Intell. 132 (2001) 219–234.
[9] M.A. Gluck, J.E. Corter, Information, uncertainty, and the utility of categories, in: Proceedings of the 7th Annula Conference of the Cognitive Science Society, Lawrence Erlbaum Associates, Irvine, CA, 1985, pp. 283–287.
[10] J.W. Guan, D.A. Bell, Rough computational methods for information systems, Artif. Intell. 105 (1998) 77–103.
[11] I. Guyon, A. Elisseeff, An introduction to variable feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[12] J.R. Hobbs, Granularity, proceedings of IJCAI, Los Angeles (1985) 432–435.
[13] X.H. Hu, N. Cercone, Learning in relational databases: a rough set approach, Comput. Intell. 11 (2) (1995) 323–338.
[14] Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. (Ny) 178 (18) (2008) 3577–3594.
[15] L. Hubert, P. Arabie, Comparing partitions, J. Classif. 2 (1) (1985) 193–218.
[16] M. Kryszkiewicz, Rough set approach to incomplete information systems, Inf. Sci. (Ny) 112 (1998) 39–49.
[17] J.Y. Liang, Y.H. Qian, Information granules and entropy theory, Sci. China Ser. F 51 (10) (2008) 1427–1444.
[18] J.Y. Liang, F. Wang, C.Y. Dang, Y.H. Qian, Incremental approach to feature selection based on rough set theory, IEEE Trans. Knowl. Data Eng. 26 (2) (2014) 294–308.
[19] T.Y. Lin, Granular computing on binary relations i: data mining and neighborhood systems, in: A. Skowron, L. Polkowski (Eds.), Rough Sets In Knowledge Discovery, Physica-Verlag, 1998a, pp. 107–121.
[20] T.Y. Lin, Granular computing on binary relations II: Rough set representations and belief functions, in: A. Skowron, L. Polkowski (Eds.), Rough Sets In Knowledge Discovery, Physica-Verlag, 1998b, pp. 121–140.
[21] T.Y. Lin, Granular computing: fuzzy logic and rough sets, in: L.A. Zadeh, J. Kacprzyk (Eds.), Computing with words in information/intelligent systems, Physica-Verlag, 1999, pp. 183–200.
[22] J.M. Ma, W.X. Zhang, Y. Leung, X.X. Song, Granular computing and dual galois connection, Inf. Sci. (Ny) 177 (2007) 5365–5377.
[23] M. Niepert, M. Gyssens, B. Sayrafi, D.V. Gucht, On the conditional independence implication problem: a lattice-theoretic approach, Arficial Intelligence 202 (2013) 1–132.
[24] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.
[25] Z. Pawlak, A. Skowron, Rudiments of rough sets, Inf. Sci. (Ny) 177 (2007) 3–27.
[26] W. Pedrycz, Relational and directional aspects in the construction of information granules, IEEE Trans. Syst. Man Cybern. 32 (5) (2002) 605–614.
[27] W. Pedrycz, A. Bargiela, Granular clustering: a granular signature of data, IEEE transactions on system, Man Cybern. 32 (2) (2002) 212–224.
[28] W. Pedrycz, Granular Computing: Analysis and Design of Intelligent Sytems, CRC Press/Francis Taylor, Boca Raton, 2013.
[29] Y.H. Qian, J.Y. Liang, W. pedrycz, C.Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (2010a) 597–618.
[30] Y.H. Qian, J.Y. Liang, C.Y. Dang, Incomplete multigranulation rough set, IEEE transactions on systems, Man Cybern. 40 (2) (2010b) 420–431.
[31] Y.H. Qian, J.Y. Liang, Y.Y. Yao, C.Y. Dang, MGRS: a mulit-granulation rough set, Inf. Sci. (Ny) 180 (2010c) 949–970.
[32] Y.H. Qian, J.Y. Liang, W. pedrycz, C.Y. Dang, An efficient accelerator for attribute reduction from incomplete data in rough set framework, Pattern Recognit. 44 (2011a) 1658–1670.
[33] Y.H. Qian, J.Y. Liang, W.Z. Wu, C.Y. Dang, Information granularity in fuzzy binary grc model, IEEE Trans. Fuzzy Syst. 19 (2) (2011b) 253–264.
[34] Y.H. Qian, S.Y. Li, J.Y. Liang, Z.Z. Shi, F. Wang, Pessimistic rough set based decisions: a multigranulation fusion strategy, Inf. Sci. (Ny) 264 (2014a) 196–210.
[35] Y.H. Qian, H. Zhang, F.J. Li, Q.H. Hu, J.Y. Liang, Set-based granular computing: a lattice model, Int. J. Approx. Reason. 55 (3) (2014b) 834–852.
[36] R. Quinlan, Induction of decision rules, Mach. Learn. 1 (1) (1986) 81–106.
[37] C.E. Shannon, The mathematical theory of communication, Bell Syst. Tech. J. 27 (3–4) (1948) 373–423. 623–656
[38] Y.H. She, X.L. He, On the structure of the multigranulation rough set model, Knowl. Based Syst. 36 (2012) 81–92.

[39] B.K. Tripathy, M. Nagaraju, Topological properties of incomplete multigranulation based on rough fuzzy sets, Commun. Comput. Inf. Sci. 270 (2012) 92–102.
[40] E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J.W.T. Lee, Attributes reduction using fuzzy rough sets, IEEE Trans. Fuzzy Syst. 16 (5) (2008) 1130–1141.
[41] W. Pedryez, A.V. Vasilakos, Linguistic models and linguistic modeling, IEEE Trans. Syst. Man Cybern. Part B 29 (6) (1999) 745–757.
[42] M.J. Wierman, Measuring uncertainty in rough set theory, Int. J. Gen. Syst. 28 (4) (1999) 283–297.
[43] W.Z. Wu, Y. Leung, J.S. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Trans. Knowl. Data Eng. 21 (10) (2009a) 1461–1474.
[44] W.Z. Wu, Y. Leung, J.S. Mi, On generalized fuzzy belief functions in infinite spaces, IEEE Trans. Fuzzy Syst. 17 (2) (2009b) 385–397.
[45] W.Z. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi-scale decision tables, Inf. Sci. (Ny) 181 (2011) 3878–3897.
[46] W.Z. Wu, Y. Leung, Optimal scale selection for multi-scale decision tables, Int. J. Approx. Reason. 54 (2013) 1107–1129.
[47] W.Z. Wu, Y.H. Qian, T.J. Li, S.M. Gu, On rule acquisition in incomplete multi-scale decision tables, Information Sciences doi:10.1016/j.ins.2–16.03.041.
[48] B.W. Xu, Y.M. Zhou, H.M. Lu, An improved accuracy measure for rough sets, J. Comput. Syst. Sci. 71 (2005) 163–173.
[49] D.J. Watts, S.H. Strogatz, Collecive dynamics of 'samll-world' networks, Nature 393 (6684) (1998) 440–442.
[50] W.H. Xu, Q.R. Wang, S.Q. Luo, Multigranulation fuzzy rough sets, J. Intell. Fuzzy Syst. 26 (3) (2014) 1323–1340.
[51] R.R. Yager, D. Filev, Operations for granular computing: mixing words with numbers, in: Proceedings of 1998 IEEE International Conference on Fuzzy Systems, 1998, pp. 123–128.
[52] X.B. Yang, Y.S. Qi, X.N. Song, J.Y. Yang, Test cost sensitive multigranulation rough set: model and minimal cost selection, Inf. Sci. (Ny) 250 (2013) 184–199.
[53] Y.Y. Yao, Information granulation and rough set approximation, Int. J. Intell. Syst. 16 (1) (2001) 87–104.
[54] Y.Y. Yao, Interpreting concept learning in cognitive informatics and granular computing, IEEE Trans. Syst. Man Cybern. Part B 39 (4) (2009) 855–866.
[55] Y.Y. Yao, L.Q. Zhao, A measurement theory view on the granularity of partitions, Inf. Sci. (Ny) 213 (2012) 1–13.
[56] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989.
[57] D.S. Yeung, D.G. Chen, E.C.C. Tsang, J.W.T. Lee, X.Z. Wang, On the generalization of fuzzy rough sets, IEEE Trans. Fuzzy Syst. 13 (3) (2005) 343–361.
[58] J. Yu, General c-means clustering model, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1197–1211.
[59] L. Zadeh, Fuzzy logic equals computing with words, IEEE Trans. Fuzzy Syst. 4 (2) (1996) 103–111.
[60] L. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets Syst. 90 (2) (1997) 111–127.
[61] L. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems, Soft Comput. 2 (1998) 23–25.
[62] L. Zadeh, A new direction in AI-toward a computational theory of perceptions, AI Mag. 22 (1) (2001) 73–84.
[63] B. Zhang, L. Zhang, Theorey and Applications of Problem Solving, North-Holland: Elsevier Science Publishers, 1992.
[64] H.Y. Zhang, W.X. Zhang, C.L. Mei, Entropy of interval-valued fuzzy sets based on distance and its relationship with similarity measure, Knowl. Based Syst. 22 (6) (2009) 449–454.
[65] P. Zhu, Q. Hu, Adaptive neighborhood granularity selection and combination based on margin distribution optimization, Inf. Sci. (Ny) 249 (2013) 1–12.
[66] UCI Machine Learning Repository, http://www.ics.uci.edu/mlearn/MLRepository.html, 2010.