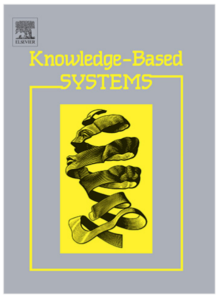
The final publication is available at Elsevier via https://dx.doi.org/10.1016/j.knosys.2018.10.015 © 2019. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/

Accepted Manuscript



Granularity measures and complexity measures of partition-based granular structures

Matthew X. Yao

PII: S0950-7051(18)30501-X
DOI: https://doi.org/10.1016/j.knosys.2018.10.015
Reference: KNOSYS 4534

To appear in: *Knowledge-Based Systems*

Received date : 1 June 2018
Revised date : 5 October 2018
Accepted date : 9 October 2018

Please cite this article as: M.X. Yao, Granularity measures and complexity measures of partition-based granular structures, *Knowledge-Based Systems* (2018), https://doi.org/10.1016/j.knosys.2018.10.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

*Revised Manuscript (Clean Version)
Click here to view linked References

Granularity measures and complexity measures of partition-based granular structures

Matthew X. Yao

Department of Mechanical and Mechatronics Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

Abstract

Granular computing is an emerging field of study in which the complexity of problem solving is reduced through granulation. Researchers have proposed various granularity measures of partitions to quantify the effects of granulation with respect to simplification. However, two important issues still remain and require careful investigation. The first issue is that a partition is only a simple two-level granular structure, which may not be sufficient for the full scope of granular computing. The second issue is a clarification of the differences between granularity and complexity. Although they are related to each other, they represent different things. To address the two issues, this paper makes three contributions. First, we extend the partition granulation scheme into multilevel granular structures based on progressive partitioning. Second, we propose a complexity measure of a partition that incorporates both the block-level interactions (interactions within a block) and the partition-level interactions (interactions between blocks of the partition). Third, we generalize the complexity measure to multilevel granular structures generated from a progressive partitioning process.

Keywords: Granularity Measure, Complexity Measure, Granular Structure, Granular Computing, Progressive Partitioning

1. Introduction

Granular computing is a problem solving technique in which a complex problem is subdivided into smaller components or granules to facilitate in-

Email address: mxyao@uwaterloo.ca (Matthew X. Yao)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

formation processing [1, 2, 8, 10, 11, 20, 23, 25, 28]. An overarching theme of granular computing is that decomposing a large amount of data into a smaller number of chunks would reduce the complexity of the analysis. Each chunk can then be further decomposed into smaller chunks. The repeated subdivision of the data would result in a hierarchical structure, providing a mechanism for complexity reduction. Although the hierarchical nature of granular computing and the resultant reduction of complexity have been accepted, there is still a lack of comprehensive theoretical or empirical studies on the notion of the complexity of granular structures.

This paper investigates measures of complexity of granular structures in an attempt to establish a sound basis for supporting granular computing. To contextualize our current study, we first invoke Simon's famous parable of two watchmakers. In 1962, Simon [26] published a seminal paper on hierarchy (i.e., a multilevel structure) as the architecture of complexity. He used two watchmakers, Hora and Tempus, to demonstrate his idea. Hora hierarchically organizes a watch as subassemblies of about ten elements or sub-subassemblies whereas Tempus does not use such an organization. With a hierarchical organization, Hora only needs to consider interactions of elements inside the same subassembly and does not need to consider interactions with elements in different subassemblies. In contrast, Tempus must consider the interactions between all elements. If both watchmakers must put down their assembly and start from scratch when interrupted, then Hora is able to assemble watches at a much faster rate than Tempus. From this parable, we can draw two important implications for granular computing. One is that a hierarchy may be a useful granular structure to support granular computing [31, 33]. The other is that the complexity of granular structure is determined, to some degree, by the interaction of elements and granules. We review existing studies and propose new complexity measures of a granular structure based on these two observations.

Influenced by the theory rough sets proposed by Pawlak [18], many studies on granular computing consider partitions as granular structures [30]. A very important concept is the granularity of partitions that reflects a coarsening-refinement relation on partitions. Researchers have proposed numerous granularity measures of partitions [4, 5, 12, 14, 15, 16, 21, 27, 29, 34, 35, 36]. As shown by the left branch of Figure 1, a measure of the granularity of a partition is defined based on the granularity of a set. The latter is defined, in turn, based on the cardinality of a set or the number of pairs in a set, as indicated by the dashed lines. Feng et al. [7] defined a measure of the granularity

ACCEPTED MANUSCRIPT

of a partition based on the cardinality of a set X, namely, $|X|$. Beaubouef et al. [3], Düntsch and Gediga [6], Miao and Wang [17], Wierman [27], and Yao [29] have used the Shannon entropy to define granularity measures of a partition. These measures can be expressed in terms of the Hartley [9] entropy $\log|X|$ of a set X. Miao and Fan [16], Liang and Shi [13], Qian and Liang [21], and Liang et al. [14] introduced granularity measures based on the number of pairs in a set, as given by $\binom{|X|}{2}$ and $\log\binom{|X|}{2}$, respectively, in Figure 1. By summarizing these measures, Yao and Zhao [34] derived a general class of granularity measures based on the expectation of granularities of all the blocks within a partition, as given by the first box in the left branch of the figure, namely, $G(\pi) = \sum_{X\in\pi} p(X)m(X)$, where π denotes a partition and $p(X)$ is the probability of the block X in the partition π.

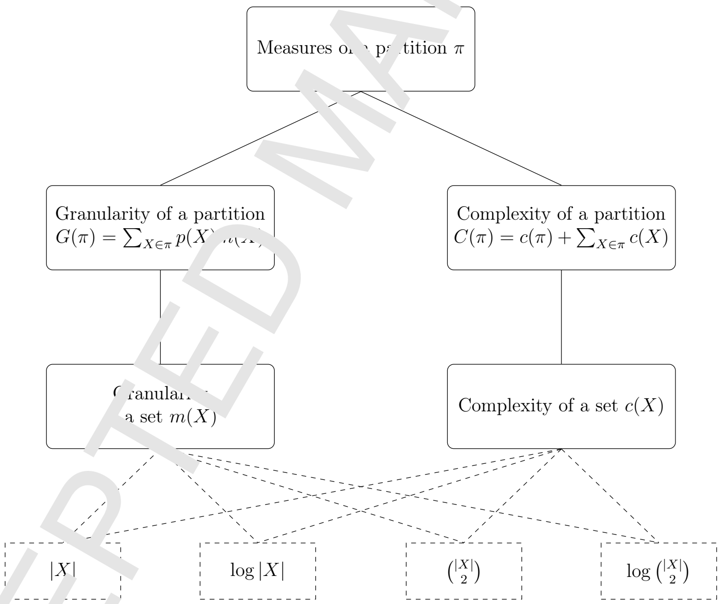


Figure 1: Granularity and complexity measures of a partition

A granularity measure of a set may also be used to define a complexity measure of the set, which depends on only the cardinality of the set. By partitioning a set, we do not need to consider the interaction of elements in different blocks. The complexity of a partition comes from two sources: the interaction of elements within a block as determined by size of the block and the interaction of blocks of a partition as determined by the number

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

of blocks in a partition. Since the granularity measures do not consider the number of blocks, they are significantly different from the complexity measures of a partition. For this reason, this paper introduces a class of complexity measures of a partition based on complexity measures of a set. This class is shown in the first box in the right branch of Figure 1, namely, $C(\pi) = c(\pi) + \sum_{X \in \pi} c(X)$.

The proposed complexity measures of a partition are defined by using a granularity/complexity measure of a set. As shown in Figure 1, existing granularity measures of a set consider only independence or pairwise interactions of elements in a set. They do not account for the fact that the complexity of a system could also be related to higher order interactions. To account for the various degrees of interaction, in this paper we introduce an i-th order complexity measure to capture interaction of i elements within in the set. By summing the complexity of $1, \ldots, k$ order interactions, we introduce a cumulative k-th order complexity measure.

A partition is a special type of two-level granular structure. Blocks in a partition can be further subdivided into finer levels of abstraction. A process of repeated subdivision or progressive partitioning would necessitate a hierarchical multilevel structure. The proposed complexity measure of a partition is generalized to be applicable to hierarchical granular structures. This is done through a recursive summation adhering to the progressive partitioning. Figure 2 provides an overview of the proposed class of complexity measures granular structures. The bottom level with dashed boxes represent examples of complexity measures of a set. The third level represents granularity measures of a set. In general, these can be defined by functions of the number of elements, pairs, triplets, etc. in the set. The second level constructs the i-th order complexity measure of a set based on the granularity measures of a set. The first level defines cumulative k-th order complexity measures by summing the various possible orders of interaction, including independence, pairwise interactions, triplet interactions, and so forth.

When using granular structures, it is necessary to consider both the structural and semantic information. The structural information is application independent, whereas the semantic information is application dependent. Although semantic complexity is important, we cannot study it without the context of an application. Thus, we restrict the present study to the structural complexity of partition-based granular structures. Figure 1 and Figure 2 outline the contributions of this paper in the context of existing research, namely, to propose a class of complexity measures of a partition

ACCEPTED MANUSCRIPT

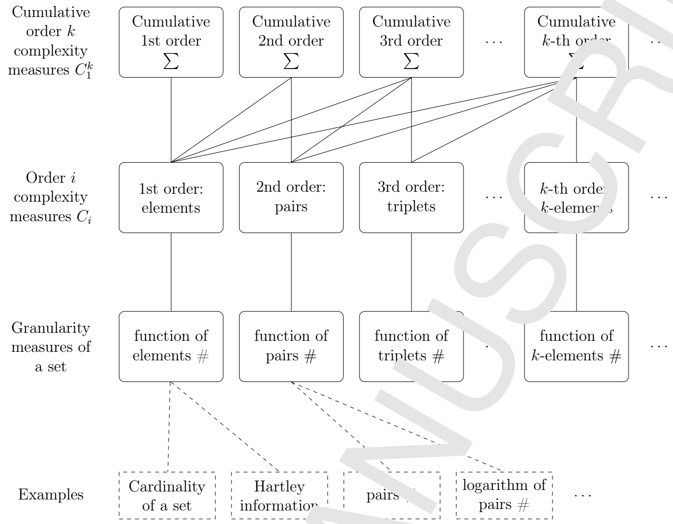


Figure 2: Cumulative k-th order complexity measures

and to generalize these measures to be applicable to hierarchical granular structures generated from progressive partitioning. To achieve these goals, the rest of the paper is organized as follows. Section 2 introduces the construction of hierarchical granular structures derived through progressive partitioning. Section 3 introduces the complexity measure of a set in the context of granularity measures of a set. Sections 4 and 5 discuss methods of quantifying partitions, including granularity measures of partitions and the newly proposed complexity measures of partitions. Finally, Section 6 is the extension of the complexity measure to a hierarchical granular structure induced by progressive partitioning, enabling the complexity measure to be used for quantifying multilevel structures.

2. Partitions and hierarchical granular structures by progressive partitioning

A partition of a universal set provides the simplest granular structure consisting of only two levels. Through a progressive partitioning process, it is possible to obtain a multilevel hierarchical granular structure.

ACCEPTED MANUSCRIPT

2.1. Partitions

The granulation of an information system involves subdividing or grouping certain elements together into smaller chunks of information referred to as granules. Because the granulation depends on the needs and goals of the user, the data can be grouped in any number of ways to help facilitate analysis. One special case of granulation which has been studied by many researchers is based on partitions or equivalence relations. In a partition based granular computing model [30], a universal set U is divided into smaller non-empty subsets called blocks. Each element from the original set is a member of only one block and all blocks are pairwise disjoint.

Definition 1. *A partition of a finite set U is a family of subsets of U, $\pi = \{X_1, \ldots X_m\}$, if and only if:*

$$
\begin{array}{rl}
\text{(i)} & X_i \neq \emptyset, \\
\text{(ii)} & \displaystyle\bigcup_{i=1}^{m} X_i = U, \\
\text{(iii)} & X_i \cap X_j = \emptyset, \quad \text{where} \quad i \neq j.
\end{array}
$$

Each subset X_i is called a block of the partition.

There is a one-to-one correspondence between the set of all partitions of U and the set of all equivalence relations on U. If $E \subseteq U \times U$ is an equivalence relation on U, namely, E is reflexive, symmetric, and transitive, then the family of equivalence classes of E is a partition $U/E = [x]_E \mid y \in U$, where $[x]_E = \{y \in U \mid xEy\}$ is the equivalence class containing $x \in U$. Conversely, given a partition π, an equivalence relation can be defined by $xE_\pi y$ if and only if x and y are in the same block of π.

In the context of granular computing, each block of a partition may be interpreted as a granule. Since each element or piece of information is contained in only one subset, a partition is often considered to be one of the simplest granulation schemes. Consequently, partition based granular computing has been studied by many researchers in the context of rough set theory [18, 27, 30].

2.2. Progressive partitioning

To further decompose or granulate a problem, a partition can be refined through further subdivisions of its blocks. Each further subdivision of a block

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

also adheres to the properties of a partition. When a block is subdivided, the resulting sub-blocks are a partition of the original block. Thus, every refinement of a partition is also a partition of the original partition. The process of refining a partition successively is referred to as progressive or recursive partitioning. The result is a multilevel granular structure that is more suitable for representing the various levels of detail required in granular computing.

It is helpful to think of the progressive refinement as a tree structure. The root node is the original information, and every level consists of a refinement of the previous level. An example of such a structure is shown in Figure 3. The multilevel structure preserves the relationship between successive refinements and allows for the creation of functions to search for appropriate levels of granularity. In this case we could define a look-up function to progress up the tree into coarser granulations (less detail) and a look-down function to progress down the tree for finer granulations (more detail).

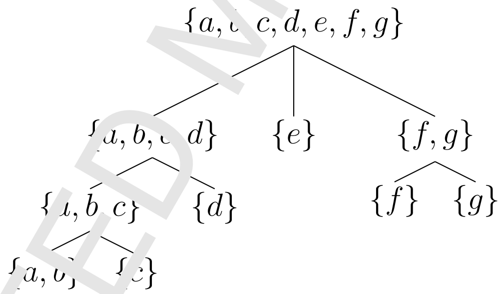


Figure 3: Tree representation of a granular structure from progressive partitioning

In order to clearly and explicitly represent different types of granules produced in a progressive partitioning process, we introduce the notions of granules and granular structures. There are two types of granules for forming a granular structure. A simple granule is a nonempty subset of a universal set and a composite granule consists of other granules (either simple or composite). A composite granule contains structural information regarding its constituent granules. Formally, we define atomic (i.e., simple) and composite (i.e., non-atomic) granules recursively.

Definition 2. *Let U be a finite and non-empty universal set. The family of granules can be constructed recursively with the following rules:*

(i) An atomic granule g is any non-empty subset of U, that is, $\emptyset \neq g \subseteq U$;

ACCEPTED MANUSCRIPT

(ii) *A non-atomic granule is a finite non-empty set G of granules, that is, each element of G is either an atomic granule or a non-atomic granule.*

By definition, a composite granule contains granules that, in turn, may contain smaller granules. A composite granule in fact represents a hierarchical structure with multiple levels of granules. In this paper, we will refer to a granule as a granular structure and use these terms interchangeably. In many situations, we are interested in the family of all elements of U that appear in all atomic granules used to form a granule.

Definition 3. *Given a granule G, the set of elements that appear in G is recursively defined as follows: for an atomic granule g and a composite granule G,*

$$\begin{aligned} &\text{(i)} \qquad e(g) = g, g \subseteq U, \\ &\text{(ii)} \qquad e(G) = \bigcup_{F \in G} e(F). \end{aligned} \tag{1}$$

Definition 4. *A granule G is called a nested granule if (i) G is an atomic granule, or (ii) G is composed of a family of nested granules and the sets of elements of granules of G are pairwise disjoint, that is, for $F, F' \in G$, if $F \neq F'$, then $e(F) \cap e(F') = \emptyset$.*

The atomic granules are the smallest components of U and cannot be further decomposed. The progressive partitioning process creates a nested structure as the atomic granules and nested granules are contained within larger nested granules. The nested structure highlights the hierarchical nature of the partition-based structures. By moving up and down the various levels of this progressive partitioning, the data can be viewed at different resolutions. In order to show the granular structure in a similar form, we use a dot-representation of a composite granule as a unit which is then connected to its constituent granules. Figure 4 shows an example of the nested granular structure resulting from progressive partitioning, which corresponds to Figure 3. An internal dot node represents a nested granule, where the contents of the granule at that level are not of interest and consequently hidden. If further detail is required, one can move down a level to unpack and observe the contents within the nested granule. The leaf nodes represent atomic granules that are subsets of U.

ACCEPTED MANUSCRIPT

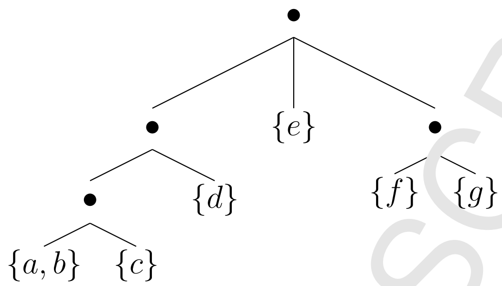


Figure 4: A dot-representation of the nested granular structure of Figure 3: $G = \{\{\{\{a, b\}, \{c\}\}, \{d\}\}, \{e\}, \{\{f\}, \{g\}\}\}$

A partition is a special case of a composite granule, as shown in Figure 5. It can be seen that a partition is a two-level granular structure, namely, the level of the family of blocks represented by the dot and the level of individual block represented by subsets of U.

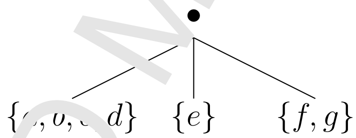


Figure 5: Granule structure induced by a partition: $\pi = \{\{a, b, c, d\}, \{e\}, \{f, g\}\}$

In developing granularity and complexity measures of granular structures, it is of interest to discuss the meaning and implications of independence and interaction between elements of a set. An element or granule is said to be independent of another element or granule if they can be considered in isolation. A pair of elements or granules are dependent if they must be considered in tandem because of their interactions. In general, a set of elements or granules are dependent if any one is related to the rest. The complexity of any natural or artificial system is determined by how individual components of the system interact with each other. Systems with independent (non-interacting) elements are simpler, and systems with highly interacting components are complex. In reality, while some components are highly dependent, some other components are only loosely dependent. By exploiting the weaker dependencies, we may decompose a complex system into subsystems, so that we do not need to consider the minor interactions between elements in different subsystems. In this way, we may turn complexity into simplicity.

ACCEPTED MANUSCRIPT

The earlier mentioned parable of two watchmakers is a good example to show how to reduce complexity of any natural or artificial system. If we have a watch which is subdivided into various subassemblies, we only need to consider the interaction of elements in the same subassembly and the interaction of subassemblies, and we do not need to consider interactions between elements in different subassemblies. Since the number elements in each subassembly and the number of subassemblies are small, we reduce the complexity of watchmaking. We may use software system development to further illustrate this point. Consider a large software system which is divided into numerous subroutines. Naturally, if subroutines are fully independent of one another, the code is less complex than if each subroutine was required to interact with numerous other subroutines. It is not surprising that large software systems are organized into hierarchical structures in order to reduce its complexity.

In forming granular structures, dependent elements must be placed in the same granule and elements in different granules are independent. Thus, we choose the notion of interactions as a basis for our study of complexity of a granular structure. We propose a family of complexity measures which covers the spectrum of independence and varying levels of dependence. One may select a specific complexity measure from the class according to the actual levels of dependence for a particular application.

2.3. Deriving multilevel granular structures in information tables

A commonly used representation of data in rough set theory and granular computing is the information table [19, 32]. An example of an information table from Quinlan [22] is presented in Table 1. The information table contains information about various objects, in this case, people, with regards to certain attributes, in this case, height, hair colour, and eye colour. The objects in the information table can be partitioned based on any one of the attributes, for example, partitioning the people based on their height producing the sets of short and tall people, respectively.

Besides being useful for creating a single partition, the attributes of an information table can be used to create a multilevel granular structure through a progressive partitioning. Each block of a partition can be further refined using the other attributes. An example of a hierarchical multilevel structure induced by such a progressive partitioning is shown in Figure 6. In this example, we form a multilevel structure by partitioning the information table in the following order of attributes: height, eye colour, and hair colour. We

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1: Information Table Example

Object	Height	Eyes	Hair	Class
a	short	blue	blond	+
b	short	brown	blond	-
c	tall	blue	red	+
d	tall	blue	dark	-
e	tall	blue	dark	-
f	tall	blue	blond	+
g	tall	brown	dark	-
h	short	brown	blond	-

stop the progressive partitioning if a set of objects have the same class label. Note that the figure represents the multilevel structure for this progressive partitioning scheme only. It is also possible to progressively partition with a different order of attributes.

The decision tree given in Figure 6 can be used to derive a set of classification rules. For example, "**if** height is short and eye colour is blue, **then** the class is +." Naturally, the complexity of the tree determines the complexity of the set of classification rules. As an application, the proposed measures of complexity may be applied in machine learning and data analysis for constructing a suitable decision tree.

3. Granularity and complexity of a set

According to Definition 2, an atomic granule is a subset of a universal set and a composite granule is a set of granules. By the fact that a granular structure is recursively defined by sets of granules, a study of granularity measures and complexity measures of a granular structure necessarily stems from a study of the corresponding measures of a set. There can be two categories of granularity and complexity measures, depending on whether elements in a set are assumed to be independent or interdependent. This leads to the two classes of independence-based and interaction-based measures. If the elements are assumed to be independent, the granularity and complexity measures become the same and will depend only on the number of elements within the set. If elements are assumed to be interdependent, a complexity measure must also take into consideration the interactions of elements.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

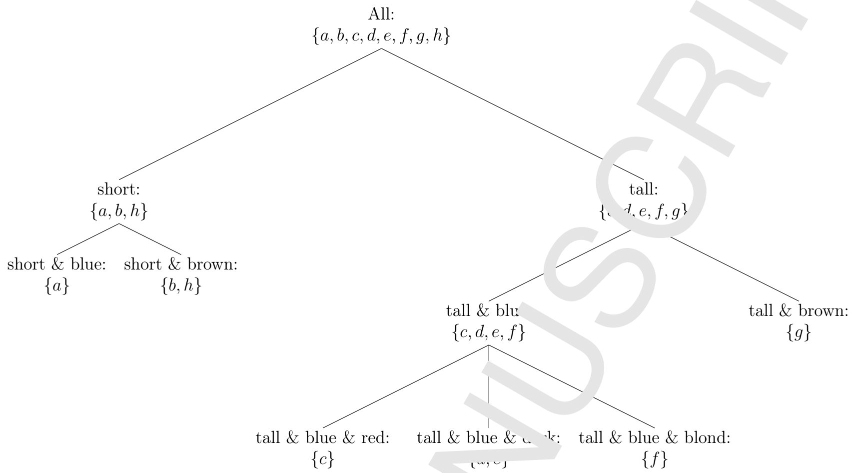


Figure 6: Progressive partitioning of the information table in Table 1.

3.1. *Granularity and complexity measures assuming the independence of elements*

A granule is a cluster of individuals or a chunk of information that can be studied as a single unit. Intuitively speaking, the concept of the granularity of a granule is used to describe our perception of the sizes of granules. That is, a measure of granularity should reflect the sizes of different granules. We can order granules according to their granularity.

In a set-theoretic setting, we use a finite set of objects to represent an atomic granule and use a set of granules to represent a composite granule. A larger set has more elements and, thus, has a larger granularity. This relationship immediately suggests the use of the cardinality (i.e., the number of elements in a set) for defining a simple granularity measure $m_1(X)$ of a set $X \subseteq U$:

$$m_1(X) = |X|, \tag{2}$$

where $|\cdot|$ denotes the cardinality of the set. The measure m_1 simply counts the number of elements of in a set. Based on the cardinality of a set, more granularity measures have been proposed by researchers to capture various semantics interpretations of granularity.

One type of granularity measures of a set is derived from an information-theoretic consideration. A subset X of a finite universal set is viewed as possible alternatives or choices. For an element randomly picked from X,

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

we have limited information about whether it is a specific element. There is more uncertainty when the element is picked from a larger subset compared to when the element is picked from a smaller subset. The information uncertainty of X is defined by:

$$m_h(X) = \log(|X|), \tag{3}$$

which is the Hartley entropy [9] of the set X. It is used as a measure of the nonspecificity inherent in the set X. When $X = \{x\}$ is a singleton subset, we have $\log(|\{x\}|) = 0$, that is, we are sure that the element is x if it is picked from $\{x\}$. When $X = U$, we have the maximum value, indicating that we are most unsure about which one is the element randomly picked from U. This offers a very interesting interpretation of granularity. A set of a larger granularity leads to a higher nonspecificity when we choose an element from the set.

By definition, we have $1 \leq m_1(X) \leq |U|$ and $0 \leq m_h(X) \leq \log|U|$ for a non-empty subset X of U. In some situation, it may be more convenient to use a normalized measure. By normalization, we have the following normalized measures:

$$\begin{aligned} \overline{m}_1(X) &= \frac{|X|}{|U|}, \\ \overline{m}_h(X) &= \frac{\log|X|}{\log|U|}. \end{aligned} \tag{4}$$

The values of the two normalized measures are in the unit interval $[0, 1]$.

It is interesting to note that measures m_h, $\overline{m}_1$, and $\overline{m}_h$ are all monotonic increasing transformations of the cardinality of a set. This class of granularity measures of a set is systematically studied by Yao and Zhao [34].

Consider now the connection of the granularity of a set and the complexity of a set. It is reasonable to assume the complexity of a set also depends on the number of elements in the set. In other words, a granularity measure may serve as a measure of the complexity of a set. At the same time, we must point out an assumption implicitly made. When measuring the granularity of a set by a monotonic increasing transformation of the cardinality of a set, we simply count the number of elements in a set. In the context of measuring the complexity of a set, this is based on an assumption that the elements of the set are independent. If elements are interdependent, simply counting the number of elements may not be sufficient.

ACCEPTED MANUSCRIPT

3.2. Complexity measures assuming the interdependence of elements

If we assume that elements in a set are independent, the complexity of the set only depends on the number of elements in the set. In other words, the granularity measure m_1 of a set can be used as a measure of the complexity of the set:

$$c_1(X) = m_1(X) = |X|. \tag{5}$$

If the elements in the set are dependent, we must consider the interactions of elements. The simplest type of interaction is pairwise interaction. Miao and Fan [16] and Qian and Liang [21] used the number of pairs in a set to define the granularity of a set. This gives the following measure of pairwise complexity of a set:

$$c_2(X) = \binom{|X|}{2} = \frac{|X|(|X|-1)}{2}. \tag{6}$$

The measure c_2 is a monotonic increasing transformation of the cardinality of the size. Although one may interpret c_2 as a measure of granularity [21, 34], it may be more meaningful to interpret c_2 as a measure of complexity, rather than a measure of granularity. In other words, we use the number of elements in a set to quantify the granularity, which is a measure of the first-order, element level complexity. The number of pairs in a set is a measure of the second-order, pairwise complexity.

If X contains three or more elements, we may consider interactions of three elements. The number of triplets in X contributes to a measure of the complexity of the third order:

$$c_3(X) = \binom{|X|}{3} = \frac{|X|(|X|-1)(|X-2|)}{3 \times 2}. \tag{7}$$

By following the same ideas, we can have a measure of i-th order complexity, where $1 \leq i \leq |X|$.

Definition 5. *For a finite set X, a measure of i-th order complexity is given by:*

$$c_i(X) = \binom{|X|}{i} = \frac{|X|!}{i!(|X|-i)!}. \tag{8}$$

That is, $c_i(X)$ is the number of i elements in X. It is assumed that $\binom{|X|}{i} = 0$ if $|X| < i$. The normalized measure of i^{th} order complexity is given by:

$$\overline{c}_k(X) = \frac{\binom{|X|}{i}}{\binom{|U|}{i}} \tag{9}$$

ACCEPTED MANUSCRIPT

and is bounded by the interval $[0, 1]$.

In defining a measure of i-th order complexity, we only consider the interaction of exactly i elements. For a given set, we may argue that its complexity is determined jointly by the number of elements (i.e., first-order complexity), the pairwise interactions (i.e., the second-order complexity), triplet interactions (i.e., third-order complexity), and so on. Therefore, a complexity measure must account interactions for all orders up to a number k, in order to fully quantify the complexity arising from all possible interacting elements. By summing up complexity measures of orders from 1 to k, we define a cumulative complexity measure $c_1^k(X)$ for all orders of interaction.

Definition 6. *A measure of cumulative complexity of orders 1 to k is defined by:*

$$c_1^k(X) = \sum_{i=1}^{k} c_i(X) = \sum_{i=1}^{k} \binom{|X|}{i}. \tag{10}$$

The measure $c_1^k(X)$ *represents the total complexity of the set with respect to interactions from 1 to k elements.*

The cumulative complexity measure $C_k(X)$ for $1 \leq k \leq 3$ are:

$$\begin{aligned} c_1^1(X) &= c_1(X) = |X|, & (11)\\ c_1^2(X) &= c_1(X) + c_2(X) = \frac{|X|(|X|+1)}{2}, & (12)\\ c_1^3(X) &= c_1(X) + c_2(X) + c_3(X) = \frac{|X|(|X|^2+5)}{6}. & (13) \end{aligned}$$

It is interesting to note that all are monotonic increasing transformations of the cardinality of X. While c_1^1 is the granularity of X, c_1^2 and c_1^3 reflect the complexity of X. In general, we do not have a simple formula to compute the cumulative complexity of a set. In real world applications, it may be sufficient to use the cumulative measures of a small k.

In order to understand the physical meaning of the cumulative complexity measure, we may consider X to be the set of components or parts of a system. A system with more interacting parts is considered complex for two main reasons. First, a system with more parts is more complex than another system with less parts. Secondly, the interaction between various parts further increases the complexity. Whereas current studies consider

ACCEPTED MANUSCRIPT

only independence or pairwise interactions, the newly proposed complexity measure accounts for all orders of complexity. We consider the interaction-based measures of a set to be the complexity measures of a set.

4. Granularity measures of a partition

A partition is set of subsets that represents a structured subdivision of a universal set. Each block is an atomic granule and the partition is a composite granule. For a partition, measures of granularity and complexity are closely related. Measures of granularity of a partition have been studied by many researchers [4, 16, 21, 27, 29, 34]. Current granularity measures fall under two categories: information-theoretic measures and pairwise interaction-based measures. They can be unified within a family of expected granularity measures [34].

4.1. Information-theoretic measures

One common type of granularity measures are the information-theoretic granularity measures, which use the Shannon and Hartley entropy as a basis. Consider a partition $\pi = \{X_1, \ldots, X_n\}$. The partition has a probability distribution $P(\pi) = (p(X_1), \ldots, p(X_n))$ where $p(X_i) = |X_i|/|U|$. The Shannon entropy of the partition is defined as:

$$\begin{aligned} H(\pi) &= -\sum_{i=1}^{n} p(X_i) \log(p(X_i)) \\ &= -\sum_{i=1}^{n} \frac{|X_i|}{|U|} \log\left(\frac{|X_i|}{|U|}\right). \end{aligned} \tag{14}$$

It represents the amount of information generated by a given probability distribution [24]. For any block of the partition π, the Hartley entropy [9] is defined as:

$$H_0(X_i) = \log|X_i| \tag{15}$$

and measures the non-specificity of a set, which is the granularity measure m_h of a set.

ACCEPTED MANUSCRIPT

The Shannon entropy of a partition can be rearranged in terms of the Hartley entropy as follows:

$$\begin{aligned} H(\pi) &= -\sum_{i=1}^{n} \frac{|X_i|}{|U|} \log\left(\frac{|X_i|}{|U|}\right) \\ &= \log|U| - \sum_{i=1}^{n} \frac{|X_i|}{|U|} \log(|X_i|). \end{aligned} \tag{16}$$

It has been used as a measure of granularity of a partition by many authors [4, 3, 6, 17, 27]. Miao and Wang [17] and Düntsch and Gediga [6] first used the Shannon entropy as a measure of roughness or granularity and Wierman [27] was the first to call it a granularity measure.

Although the full Shannon entropy of a partition has been considered as a granularity measure, from Equation (16) it can be seen that $\log|U|$ is the constant Hartley entropy $H_0(U)$, independent of the partition. As a result, some authors [29] have used only the second term in Equation (16) as a granularity measure $M_h(\pi)$:

$$M_h(\pi) = \sum_{i=1}^{n} \frac{|X_i|}{|U|} \log|X_i| = \sum_{i=1}^{n} \frac{|X_i|}{|U|} m_h(X_i), \tag{17}$$

which is the mathematical expectation of m_h. The value of M_h is bound within the limits of 0 and $\log|U|$, which represent the finest (singleton blocks) and coarsest (one block) partitions, respectively. The information-theoretic class of granularity measures do not consider interactions between the elements of a block or between the blocks themselves, which makes them unsuitable as a measure of complexity.

4.2. Pairwise interaction-based measures

The pairwise interaction-based measures of granularity are built upon the idea of counting the number of interacting pairs of elements of U under a partition [34].

Let $\pi = \{X_1, \ldots, X_n\}$ be a partition of the universal set U and E_π be the corresponding equivalence relation. Miao and Fan [16] proposed the following interaction based granularity measure:

$$\overline{M}_1(\pi) = \frac{|E_\pi|}{|U \times U|} = \frac{\sum_{i=1}^{n}|X_i \times X_i|}{|U \times U|} = \sum_{i=1}^{n} \frac{|X_i|}{|U|}\frac{|X_i|}{|U|} = \sum_{i=1}^{n} \frac{|X_i|}{|U|}\overline{m}_1(X_i), \tag{18}$$

ACCEPTED MANUSCRIPT

where $U \times U$ is the Cartesian product of U and U. The numerator is the number of pairs in the equivalence relation and the denominator is the number of all pairs in the coarsest equivalence relation $U \times U$. The maximum value of the measure is therefore 1. Equivalently, the measure M_1 can be interpreted in terms of the partition. The term $|X_i \times X_i|$ corresponds to the number of possible pairwise interactions between elements in the block X_i, where the order of the two elements in a pair is considered. The term $|U \times U|$ is the number of pairwise interactions in the coarsest partition $\{U\}$. A coarser relation allows for a larger number of interactions, so the granularity measure increases with the coarsening of partitions.

We have two interesting observations of the measure M_1. First, we can re-express M_1 as follows:

$$\begin{aligned}
\overline{M}_1(\pi) &= \frac{\sum_{i=1}^{n}|X_i \times X_i|}{|U \times U|} \\
&= \frac{1}{|U \times U|}\sum_{i=1}^{n} 2\left(\frac{|X_i|(|X_i|-1)+|X_i|}{2}\right) \\
&= \frac{1}{|U \times U|}\sum_{i=1}^{n}\left(2\binom{|X_i|}{2}+|X_i|\right) \\
&= \frac{1}{|U \times U|}\sum_{i=1}^{n}\left(2c_2(X_i)+c_1(X_i)\right) \\
&= \frac{1}{|U \times U|}\sum_{i=1}^{n}\left(c_2(X_i)+c_1^2(X_i)\right). \qquad (19)
\end{aligned}$$

That is, $\overline{M}_1$ can be written as a combination of both c_1 and c_2, which is related to cumulative complexity of orders 1 to 2. Second, the measure M_1 is similar to the information-theoretic measure M_h in the sense of mathematical expectation. The term $\overline{c}_1(X_i) = |X_i|/|U|$ depends only on the number of elements within block X_i. Therefore, the measure M_1 is not a fully interaction-based measure.

Qian and Liang [21] proposed a true interaction-based granularity measure M_q, called the combination entropy. They used combinations to represent the possible number of pairwise interactions:

$$\overline{M}_2(\pi) = \sum_{i=1}^{n}\frac{|X_i|}{|U|}\frac{\binom{|X_i|}{2}}{\binom{|U|}{2}} = \sum_{i=1}^{n}\frac{|X_i|}{|U|}\overline{c}_2(X_i), \qquad (20)$$

ACCEPTED MANUSCRIPT

which is a mathematical expectation of $\overline{c}_2(X_i)$. The combination entropy is bounded by 0 and 1, that is, $0 \leq M_q(\pi) \leq 1$.

4.3. A class of expected granularity measure

Let $\pi = \{X_1, \ldots, X_n\}$ denote a partition of a finite and nonempty universe U. In order to see the connection between the three measures, $M_h, \overline{M}_1$ and $\overline{M}_2$, we first consider the notions of measures of the granularity and complexity of a block of π or a subset of of U. Given two subsets $A, B \subseteq U$ with $A \subseteq B$, the granularity of the smaller set A is less than or equal to the granularity of the larger set B. The size of a set as defined by its cardinality provides a good measure of the granularity of the set. In fact, any positive monotonic increasing transformation of the cardinality of the block X_i may serve as a measure of granularity [34].

The partition π defines a probability distribution $P(\pi) = (p(X_1), \ldots, p(X_n))$, where $p(X_i) = |X_i|/|U|$. The three measures of granularity of a partition, discussion in the last subsection, can be expressed as mathematical expectations as follows:

$$\begin{aligned} M_h(\pi) &= \sum_{i=1}^{n} p(X_i) m_h(X_i), \\ \overline{M}_1(\pi) &= \sum_{i=1}^{n} p(X_i) \overline{m}_1(X_i), \\ \overline{M}_2(\pi) &= \sum_{i=1}^{n} p(X_i) \overline{c}_2(X_i), \end{aligned} \tag{21}$$

where the corresponding three measures the granularity and complexity of a set are given by [34].

$$\begin{aligned} m_h(X_i) &= \log(|X_i|), \\ \overline{m}_1(X_i) &= \frac{|X_i|}{|U|}, \\ \overline{c}_2(X_i) &= \frac{\binom{|X_i|}{2}}{\binom{|U|}{2}}. \end{aligned} \tag{22}$$

That is, the granularity of a partition is given as the expectation of the granularity or complexity of all blocks of the partition.

ACCEPTED MANUSCRIPT

By generalizing these measures, Yao and Zhao [34] introduced a family of expected granularity measures:

$$E_m(\pi) = \sum_{X \in \pi} p(X)\, m(X) = \sum_{i=1}^{n} p(X_i)\, m(X_i), \tag{23}$$

where $m(X_i)$ is a measure of granularity or complexity of a subset of U. It is required that the measure of granularity for a subset of U is a positive monotonic increasing transformation of its cardinality. We can obtain many different measures as special cases of the family of expected granularity measures. For example, if we use the cardinality of a block and the number of all pairs produced by a block, respectively, as measures of the granularity of a block, that is, $m_1(X_i) = |X_i|$ and $c_2(X_i) = \binom{|X_i|}{2}$, we have two new measures of the granularity of a partition:

$$\begin{aligned} M_1(\pi) &= \sum_{i=1}^{n} p(X_i) m_1(X_i) = \sum_{i=1}^{n} p(X_i)|X_i|, \\ M_2(\pi) &= \sum_{i=1}^{n} p(X_i) c_2(X_i) = \sum_{i=1}^{n} p(X_i) \binom{|X_i|}{2}. \end{aligned} \tag{24}$$

They are, respectively, the average block size and the average number of pairs in a block induced by the partition π. Semantically, they capture different aspects of a partition; the former does not consider interaction of elements within a block and the latter considers pairwise interactions of elements within a block.

4.4. *Axiomatic characterization of a measure of granularity and the implications for defining a complexity measure*

Several authors have investigated an axiomatic foundation for measuring the granularity of a partition by suggesting properties that must be satisfied [21, 27, 34]. They discussed the rationale behind a measure of granularity based on two key notions, namely, a refinement-coarsening relation and size-isomorphism on the family of all partition Π on U.

Definition 7. *A refinement-coarsening relation $\preceq$ on the family of all partitions Π of a set U is defined as follows:*

$$\pi \preceq \pi' \Longleftrightarrow E_\pi \subseteq E_{\pi'}. \tag{25}$$

ACCEPTED MANUSCRIPT

Equivalently, π is a refinement of π' or π' is a coarsening of π if every block of π is a subset of a block of π'.

Definition 8. *A partition π is size-isomorphic to another partition π', if there exists a bijection $f : \pi \longrightarrow \pi'$ such that $\forall X \in \pi$, $|f(X)| = |X|$.*

The refinement-coarsening relation is a partial ordering of partitions, that is, $\preceq$ is reflexive, anti-symmetric, and transitive. A measure of granularity must reflect the refinement-coarsening relation such that a refined partition has a lower granularity and coarsened partition has a higher granularity. When defining the refinement-coarsening relation, the composition of blocks in π and π' must be examined in terms of set inclusion. Consider two partitions on a universal set $U = \{a, b, c, d, e, f\}$, $\pi = \{\{a, b\}, \{c, d, e, f\}\}$ and $\pi' = \{\{a, c\}, \{b\}, \{d\}, \{e\}, \{f\}\}$. It is reasonable to say that π' has a smaller granularity than π. However, since π' is not a refinement of π, we cannot use the relation $\preceq$ to reflect comparison of granularity. Therefore, we must have an element-independent relation between partitions. For this purpose, Wierman [27] proposed the notion of size-isomorphisms in Definition 8 by considering only the structure of partitions, instead of individual elements. The size-isomorphism relation $\cong$ is an equivalence relation on partitions which means a measure of granularity must have the same value for partitions of the same structure. In the previous example, π' is size-isomorphic to $\pi'' = \{\{a, b\}, \{c\}, \{d\}, \{e\}, \{f\}\}$ and $\pi'' \preceq \pi$. If π' and π'' have the same granularity and the granularity of π' is less than that of π, then the granularity of π'' is less than that of π.

In summary, according to the refinement-coarsening relation and the size-isomorphism, a granularity measure must satisfy at least the following two properties:

$$\begin{aligned} \pi \preceq \pi' &\implies M(E_\pi) \le M(E_{\pi'}), \\ \pi \cong \pi' &\implies M(E_\pi) = M(E_{\pi'}). \end{aligned} \tag{26}$$

The first property states that a refined partition has a lower granularity. The second property states that two partitions having the same structure must have the same granularity. All of the previously discussed granularity measures of a partition satisfy these two requirements.

As mentioned in the introduction, the present study considers only the structural complexity of granular structures. Two size-isomorphic partitions

ACCEPTED MANUSCRIPT

have the same structural complexity. However, they may have different semantic content. As pointed out by a reviewer of this paper, "in data analysis or decision making, two granularity structures with size-isomorphisms may yield completely different results." In practice, it is necessary to combine the structural complexity and the semantic content into a common framework. While structural complexity can help us identify simple structures, the semantic content is important for deriving useful results.

5. Complexity measures of a partition

This section examines measures of complexity of a partition. The results provide a basis for introducing a family of complexity measures for a special type of hierarchical granular structures induced by progressive partitioning.

5.1. From granularity measures to complexity measures

Studies on measures of the granularity serve a starting point for defining complexity of granular structure in general and partitions in specific. A question that naturally arises is whether the refinement-coarsening and size isomorphic relations are good criteria for defining a measure of complexity. When it comes to complexity induced by a partition, the notion of size-isomorphism is still applicable. Two partitions with the same structure must have the same complexity. On the other hand, the refinement-coarsening relation is no longer appropriate. The complexity of a partition may not be the same as its granularity.

Consider the two extreme cases of partitions, namely, the coarsest partition $\{U\}$ and the finest partition $\{\{x\} \mid x \in U\}$. For $\{U\}$, we have only one block, which is easy to process. However, we have $|U|$ number of elements to process in the single block, which is complex to process. In other words, $\{U\}$ is simple in terms of the number of blocks, but is complex in terms of number of elements within the block. For the finest partition, all blocks are singleton subsets, each of which is easy to process. However, the number of blocks are the maximum among all possible partitions, which is complex to process. That is, the finest partition is complex in terms of the number of blocks and is simple in terms of number of elements within each block. The numbers of blocks and the number of elements in the blocks both contribute to the complexity of a partition. In general, the complexity of a partition is a balance of these two sources of complexity. The complexity of a partition decreases with subdivisions of U, but increases as the blocks of the partition

ACCEPTED MANUSCRIPT

approach singleton subsets. Since measures of granularity only consider the elements within the blocks, the granularity monotonically decreases as the partition is refined towards singleton blocks. They therefore do not reflect the nature of complexity. In this vein, it is challenging to define a relation for characterizing the complexity partitions, as there is still a lack of accepted definition and interpretation of the complexity of partitions. In the rest of this paper, we attempt to provide a partial solution to this important problem.

When proposing a measure of complexity of a partition, measures of granularity of a partition do provide useful hints. First, while the information-theoretic measure does not consider interaction between different elements in a block, the interaction-based measure considers only pairwise interaction. The complexity may be caused by higher order interactions, for example, triplet interactions, which these measures do not account for. A measure of complexity measure may need to consider higher level interaction. Second, measures of granularity do not consider the number of blocks or the interaction between blocks. A measure of complexity may need to consider interactions between blocks of a partition. Third, a partition is only a two-level granular structure. In general, we need to consider multilevel granular structures induced by progressive partitioning. These observations motivate the introduction of measures of granular structures based on results from measures of granularity of a partition.

A partition is a basic element for interpreting a hierarchical granular structure produced by progressive partitioning. As a prerequisite for studying the complexity of multilevel granular structures, we first examine the complexity of a partition as a two-level granular structure shown in Figure 5. As a basis for determining a complexity measure of a partition, we must first consider a meaningful understanding of complexity that a complexity measure will reflect.

The two-level interpretation of a partition suggests that the complexity of a partition has contributions from both levels. At the partition-level, a partition is a set of blocks where each block is considered as one element. The partition-level describes the structure of the partition. A larger number of blocks can have more interactions and is consequently more complex. Since the partition is a set of blocks, we can use the complexity measure of a set $c_i(\pi)$ to measure the i-th order interactions between the blocks in the partition. At the block-level, the complexity of any block is also determined by the number of possible interactions within the block. If we consider the

ACCEPTED MANUSCRIPT

block as a set, the complexity of any block $X \in \pi$ can also be measured by the i-th order complexity measure $c_i(X)$ of a set.

Definition 9. *Suppose c is a complexity measure of a set. A complexity measure C of a partition $\pi = \{X_1, \ldots, X_n\}$, induced by c, is defined by:*

$$C(\pi) = c(\pi) + \sum_{X \in \pi} c(X). \tag{27}$$

In $c(\pi)$, the complexity measure of c is applied to the partition π as a set of blocks.

The granularity of a partition depends on the granularity measure of a set. Similarly, the complexity measure of a partition depends on the complexity measure of the set. However, the measures are applied differently. Whereas the granularity of a partition is the expectation of the granularity of its comprising blocks, the the complexity of a partition considers a combined contribution from the two levels of the partition. By definition, the meaning of the proposed measure of complexity depends on the meaning of the complexity measure c of a set. Although different complexity measures may be uniformly represented as positive monotonic increasing transformations of the cardinality of a set, they have different interpretations. In turn, they lead to different interpretations of complexity measures defined by Definition 9.

The complexity measures m_h and m_1 consider only the number of elements in a set. In some sense, they implicitly assume that elements in a set are independent. If the elements in the set are assumed to be independent, then these measures can also serve as the first-order complexity measure of a set. On the other hand, measures $\overline{c}_2$ and c_2 consider the number of pairs produced by a set and implicitly assume that the elements of a set have pairwise interaction. As such, these measures can be used to quantify the second-order complexity of the set. These measures reflect different sources of complexity, resulting from either the size of the set or from the number of interacting pairs within a set. Beyond second-order measures, higher-order measurements can also be used to measure the complexity of the partition. For cumulative complexity measure c_1^k of a set given by Definition 6, a cumulative complexity measure of a partition would be:

$$C_1^k(\pi) = c_1^k(\pi) + \sum_{X \in \pi} c_1^k(X) = \sum_{i=1}^{k} \binom{|\pi|}{i} + \sum_{X \in \pi} \sum_{i=1}^{k} \binom{|X|}{i}, \tag{28}$$

ACCEPTED MANUSCRIPT

which is simply an example of the class of complexity measures defined by Definition 9.

In the rest of this section, we use several examples to study the meaning of different complexity measures.

Example 1. *Consider the first-order complexity measure of a set X, $c_h(X) = m_h(X) = \log(|X|)$, used as an information-theoretic measure of granularity. For the universe, we have $c_h(U) = \log |U|$. For a partition $\pi = \{X_1, \ldots, X_n\}$, according to Definition 9, the complexity of the partition can be computed as:*

$$C_h(\pi) = \log(|\pi|) + \sum_{X \in \pi} \log(|X|) = \log(|\pi|) + \sum_{i=1}^{n} \log(|X_i|), \tag{29}$$

where $|\pi|$ is the number of blocks in π. For the coarsest partition $\{U\}$, we have $C_m(\{U\}) = \log(|\{U\}|) + \log(|U|) = \log(|U|)$, which is in fact the granularity of the universal set U. That is, the coarsest partition has the same complexity as the set U itself. This is reasonable, if we treat the coarsest partition as U. For the finest partition, we have $C_h(\{\{x\} \mid x \in U\}) = \log(|\{\{x\} \mid x \in U\}|) + \sum_{x \in U} \log(|\{x\}|) = \log(|U|)$. Again, it has the same complexity as the universe U. This result is consistent with the common practice of treating the finest partition as U.

Given a universe U with $|U| > 2$, suppose we partition U into two blocks $\pi = \{X_1, X_2\}$. The complexity of π is computed as:

$$\begin{aligned} C_h(\pi) &= \log(|\pi|) + \log(|X_1|) + \log(|X_2|) \\ &= \log(2) + \log(|X_1|) + \log(|X_2|) \\ &> \log(|U|). \end{aligned}$$

That is, according to the complexity measure, the partitioning of U into two blocks in fact increases the complexity. This is consistent with the assumption that elements in U are independent. In other words, if elements are independent, granulation does not provide any advantage with regards to complexity reduction.

Example 2. *Consider a first-order complexity measure of a set X, $c_1(X) = m_1(X) = |X|$. The complexity of U is given by $C_1(U) = |U|$. For a partition $\pi = \{X_1, \ldots, X_n\}$, the complexity of the partition can be computed as:*

$$C_1(\pi) = |\pi| + \sum_{X \in \pi} |X| = |\pi| + \sum_{i=1}^{n} |X_i| = |\pi| + |U|, \tag{30}$$

ACCEPTED MANUSCRIPT

which is greater than the complexity of U*. For the coarsest partition, the complexity is* $C_1(\{U\}) = |\{U\}| + |U| = 1 + |U|$*, which is greater than the complexity of* U*. This reflects the fact that* U *is only a single-level granular structure, while* $\{U\}$ *is a two-level granular structure. The latter is more complex than the former. For the finest partition, the complexity is* $C_{m_1}(\{\{x\} \mid x \in U\}) = |\{\{x\} \mid x \in U\}| + \sum_{x \in U} |\{x\}| = 2|U|$*. This complexity measure is a sum of the cardinality at two levels of the partition under the assumption that all elements are independent. The measure clearly shows that, when elements are independent, partitioning the universe in fact increases the complexity. That is, one can simply consider elements in* U *one-by-one, instead of clustering them into groups.*

Example 3. *Consider a second-order complexity measure of a set* X*,* $c_2(X) = \binom{|X|}{2}$*, where* $\binom{|X|}{2} = 0$ *if* $|X| < 2$*. It is a measure that accounts for pairwise interactions within a set. For the universe, we have* $c_2(U) = \binom{|U|}{2}$*. For a partition* $\pi = \{X_1, \ldots, X_n\}$*, the complexity of the partition can be computed as:*

$$C_2(\pi) = \binom{|\pi|}{2} + \sum_{X \in \pi} \binom{|X|}{2} = \binom{|\pi|}{2} + \sum_{i=1}^{n} \binom{|X_i|}{2}. \qquad (31)$$

For the coarsest partition, we have $C_2(\{U\}) = \binom{|\{U\}|}{2} + \binom{|U|}{2} = \binom{|U|}{2}$*, which is the same as the complexity of* U*. For the finest partition, we have* $C_2(\{\{x\} \mid x \in U\}) = \binom{|\{\{x\}|x \in U\}|}{2} + \sum_{x \in U} \binom{|\{x\}|}{2} = \binom{|U|}{2}$*, which is again the same complexity of* U*. Given a universe* U *with* $|U| > 2$*, suppose we partition* U *into two blocks* $\pi = \{X_1, X_2\}$*. The complexity of* π *is computed as:*

$$\begin{aligned} C_2(\pi) &= \binom{|\pi|}{2} + \binom{|X_1|}{2} + \binom{|X_2|}{2} \\ &= \binom{2}{2} + \binom{|X_1|}{2} + \binom{|X_2|}{2} \\ &< \binom{|U|}{2}. \end{aligned}$$

According to this complexity measure, the partitioning of U *into two blocks decreases the complexity. If the elements of a set are assumed to have pairwise interaction, then subdivision of* U *provides advantages with regards to complexity reduction.*

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Examples 1 to 3 demonstrate that the complexity of a partition is a multifaceted notion, depending on the processing elements under different assumptions. If elements and blocks of a partition are assumed to be independent, using a partition indeed leads to an increase of complexity. On the other hand, if the elements and blocks of a partition are assumed to be pairwise dependent, the introduction of partition results in complexity reduction. The complexity of a partition is a balance of the partition-level and block-level complexity. A finer partition will have a higher partition-level complexity and a lower block-level complexity. In the case of independence, a partition of U reduces the block-level complexity, but increases the partition-level complexity through the creation of non-interacting blocks. In the case of pairwise interaction, a partition of U reduces the number of possible interactions by isolating related elements into their respective blocks. Although the partition-level complexity is increased, it can be outweighed by the reduction of the block-level complexity. An interaction-based complexity measure enables us to search for the right level of granularity that produces the least complexity in processing.

6. Complexity measures of granular structures induced by progressive partitioning

A partition is a special type of granular structures. To quantify the effect of the numerous components in a partition, an interaction-based complexity measure was proposed in the last section to account for the different orders of interactions between blocks and between elements within a block. By extending the same argument, we can study complexity measures of a granular structure in general.

6.1. Complexity measures for nested granular structures

A partition is a two-level granular structure. Given a complexity measure of a set, in Definition 9 we define the induced complexity of a partition by summing the two components in Equation (27). The two terms in the equation correspond to the two levels. For a general granular structure, we need to consider multiple levels. According to the recursive definition of a multi-level granular structure in Definition 2, we can apply the complexity measure of a set recursively to compute the complexity of a granular structure.

ACCEPTED MANUSCRIPT

Definition 10. *Suppose c is a complexity measure of a set. A complexity measure of a granular structure, as given by Definition 2, can be recursively defined based on c as follows: for an atomic granule g and a composite granule G,*

$$\begin{aligned} C(G) &= c(g), \quad G = g \subseteq U \text{ is an atomic granule}, \\ C(G) &= c(G) + \sum_{F \in G} C(F), \quad G \text{ is a composite granule}, \end{aligned} \tag{32}$$

where $c(G)$ is applied to G by treating G as a set of granules.

According to the definition, the complexity of a composite granule is the sum of the two sources of complexity given by a granule. The first term, $c(G)$, is the complexity of the granule G as a family of granules. The second term, $\sum_{F \in G} C(F)$, is the total complexity of all granules in G. That is, the complexity of a granule combines both types of complexity. By using this general definition, we can obtain various complexity measures of the complexity of a granular structure by using different complexity measures of the complexity of a set.

We have shown that the number of interactions of different orders contribute to the complexity of a set. By considering all orders of the possible interactions in a granular structure with respect to a cumulative complexity measure of a set, we can have a complexity measure that offers a holistic view of the granular structure. With respect to the i-th order interaction complexity and cumulative order k complexity, we have two families of complexity measures of a granular structure.

Definition 11. *Let $c_i(X) = \binom{|X|}{i}$ denote the complexity measure defined by the number of interactions of exact i elements in a set. A complexity measure of a granular structure, induced by c_i, is defined as follows: for an atomic granule g and a composite granule G,*

$$\begin{aligned} C_i(G) &= c_i(g) = \binom{|g|}{i}, \\ C_i(G) &= c_i(G) + \sum_{F \in G} C_i(F) = \binom{|G|}{i} + \sum_{F \in G} C_i(F), \end{aligned} \tag{33}$$

where $c_i(G)$ is the complexity of G as a set of granules.

ACCEPTED MANUSCRIPT

Definition 12. *Let $c_1^k(X) = \sum_{i=1}^{k} c_i(X)$ be the cumulative k-th order ($k \geq 1$) complexity measure of a set X. The complexity of a granule, induced by c_1^k is recursively defined as: for an atomic granule g and a composite granule G,*

$$
\begin{aligned}
C_1^k(g) &= c_1^k(g) = \sum_{i=1}^{k} c_i(g) = \sum_{i=1}^{k} \binom{|g|}{i}, \\
C_1^k(G) &= \sum_{i=1}^{k} C_i(G) = \sum_{i=1}^{k} \left(c_i(G) + \sum_{F \in G} C_i(F) \right) \\
&= \sum_{i=1}^{k} \binom{|G|}{i} + \sum_{F \in G} C_1^k(F). \qquad (34)
\end{aligned}
$$

Example 4. *When interactions are considered, a decomposition of U can be shown to reduce its complexity. Consider the example shown in Figure 4, that is, a composite granule $G = \{\{\{\{a,b\},\{c\}\},\{d\}\},\{e\},\{\{f\},\{g\}\}\}$. We can calculate the values of different complexity measures as follows:*

$$
\begin{aligned}
C_1^1(U) &= C_1(U) = 7, \\
C_1^1(G) &= C_1(G) = 16, \\
C_1^2(U) &= C_1^1(U) + C_2(U) = 7 + 21 = 28, \\
C_1^2(G) &= C_1^1(G) + C_2(G) = 16 + 7 = 23, \\
C_1^3(U) &= C_1^2(U) + C_3(U) = 28 + 35 = 63, \\
C_1^3(G) &= C_1^2(G) + C_3(G) = 23 + 1 = 24.
\end{aligned}
$$

If the elements of the granule are assumed to be independent, measured by C_1^1, it can be seen that subdividing the granule into more blocks actually increases the complexity. If the elements are assumed to have pairwise interactions, measured by C_1^2, subdividing the granule reduces the complexity. If the elements have triplet interactions, measured by C_1^3, the complexity is further reduced.

Definition 10 gives rise to a family of complexity measures generated by complexity measures of a set. We can generate many specific measures of complexity. Consider the information-theoretic measure of granularity as a

ACCEPTED MANUSCRIPT

measure of complexity $c_h(X) = m_h(X) = \log(|X|)$. For an atomic granule g and a composite granule G, we have:

$$\begin{aligned} C_h(g) &= c_h(g) = \log(|g|), \\ C_h(G) &= c_h(G) + \sum_{F \in G} C_h(F) \\ &= \log(|G|) + \sum_{F \in G} C_h(F) \end{aligned} \quad (35)$$

When G is a partition, we obtain Equation (29) as a special case.

We now examine properties of specific types of complexity measures. In C_1 we do not consider any interaction between different elements or granules in a granule, namely, all components of a granule are independent. It can be easy verified that the granule U has the minimum value of C_1. This suggests that further decomposition of U into smaller granules will increase the complexity, which is consistent with the assumption of independence. That is, under the independence assumption, U is the only structure that we need when considering individual elements separately and we do not need to consider other granules.

Consider a universe U and a partition $\pi = \{X_1, \ldots, X_n\}$. The cumulative k-th order complexity of U is:

$$C_1^k(U) = \sum_{i=1}^{k} \binom{|U|}{i}. \quad (36)$$

The cumulative k-th order complexity of the partition is

$$C_1^k(\pi) = \sum_{i=1}^{k} \binom{|\pi|}{i} + \sum_{i=1}^{k} \sum_{j=1}^{|\pi|} \binom{|X_j|}{i}. \quad (37)$$

The size of the granules and the number of granules is highly coupled in a partition. Some observations can be made regarding the nature of the complexity of a partition as measured by C_1^k, $k \geq 2$. The first term represents the contribution to the complexity from the interaction between blocks of the partition and the second term represents contribution from the interactions within each block for all blocks. As the number of blocks increases, the average size of the blocks decreases. When $|\pi| \geq 1$, we have $|X_i| < |U|$. It follows that in each block we do not have some higher order interactions as

ACCEPTED MANUSCRIPT

we do in U. In addition, we do not need to consider interactions of objects in different blocks. In other words, any interaction considered by the second term in $C_1^k(\pi)$ is considered by $C_1^k(U)$, but the reverse is not true. Therefore, for $k \geq 2$, we have the following inequality:

$$\sum_{i=1}^{k} \binom{|U|}{i} > \sum_{i=1}^{k} \sum_{j=1}^{|\pi|} \binom{|X_j|}{i}. \tag{38}$$

The quantity,

$$\sum_{i=1}^{k} \binom{|U|}{i} - \sum_{i=1}^{k} \sum_{j=1}^{|\pi|} \binom{|X_j|}{i}, \tag{39}$$

is the reduction of complexity at the level of individual blocks. On the other hand, a partition also introduces complexity at the partition level, as given by the first term in $C_1^k(\pi)$, namely $\sum_{i=1}^{k} \binom{|\pi|}{i}$. Therefore, whether a partition increases the overall simplicity is determined by a trade-off of the reduction and the increment of complexity.

The discussion suggests that we can re-express the comparison of the complexities of U and π,

$$\sum_{i=1}^{k} \binom{|U|}{i} > \sum_{i=1}^{k} \sum_{j=1}^{|\pi|} \binom{|X_j|}{i} + \sum_{i=1}^{k} \binom{|\pi|}{i}, \tag{40}$$

equivalently as follows:

$$\sum_{i=1}^{k} \binom{|U|}{i} - \sum_{i=1}^{k} \sum_{j=1}^{|\pi|} \binom{|X_j|}{i} > \sum_{i=1}^{k} \binom{|\pi|}{i}. \tag{41}$$

The left hand side is the reduction of complexity and the right hand side is the increment of complexity. By using a larger number of smaller blocks, we have a higher reduction and a higher increment. By using a lower number of larger blocks, we have a lower reduction and a lower increment. The combined results is undetermined, that is, may be either favourable or unfavourable. The coarsest and finest partitions represent two extreme cases. According to measure C_1^k, their complexities are given by

$$\begin{aligned} C_1^k(\{U\}) &= 1 + C_1^k(U), \\ C_1^k(\{\{x\} \mid x \in U\}) &= C_1^k(U) + |U|. \end{aligned} \tag{42}$$

ACCEPTED MANUSCRIPT

They are, in fact, greater than the complexity of U. The finest partition has a higher complexity than that of the coarsest partition which is the reverse order given by the granularity of partitions. To have the complexity lower than the complexity of U, we need to use a partition that is somewhere in the middle.

A nested granular structure is obtained by progressive partitioning. The analysis of the complexity of a partition can be extended to a granular structure. As the complexity of each granule composing the nested granular structure decreases with decomposition or subdivision the overall complexity of the nested granular structure will also decrease. According to the measure $C_1^k(G)$, when higher degrees of interaction or dependency are considered, it is highly possible to construct a multilevel granular structure with a complexity less than the complexity of U. That is, we use granular structures when we decompose larger granules into smaller granules in order to eliminate higher order dependency. The analysis is consistent with the argument given by Simon [26] regarding the use of hierarchical systems to reduce complexity.

6.2. Axiomatic characterization of a measure of complexity

The multilevel granular structures induced by the progressive partitioning necessitate additional considerations. The definition of size-isomorphism of partitions can be generalized to nested granular structures given by Definition 4. However, a mapping between two blocks cannot easily generalized. We recast the definition of size isomorphism in terms of a bijection over the set of objects, which allows us to generalize to the case of multilevel structures.

Definition 13. *A partition π is isomorphic to another partition π', written $\pi \cong \pi'$, if we can obtain π from π' by a bijection $f : U \longrightarrow U$, and vice versa. That is, two objects $f(a)$ and $f(b)$ are in the same block of π if and only if a and b are in the same block of π'.*

The atomic granules which make up a nested granule form a partition of U. In order for two nested granular structures to be isomorphic, the corresponding two partitions must be isomorphic according to Definition 13. In addition, the two trees expressed as a dot-representation such as that shown in Figure 4 must be isomorphic, that is, the two trees have the same structure. The definition of two isomorphic nested granular structures immediately follows.

ACCEPTED MANUSCRIPT

Definition 14. *A granular structure G is isomorphic to another granular structure G', written $G \cong G'$, if we can obtain G from G' by a bijection $f : U \longrightarrow U$, and vice versa. That is, the partition formed by atomic granules in G is isomorphic to the partition formed by the atomic granules in G' and the two trees of G and G' are isomorphic.*

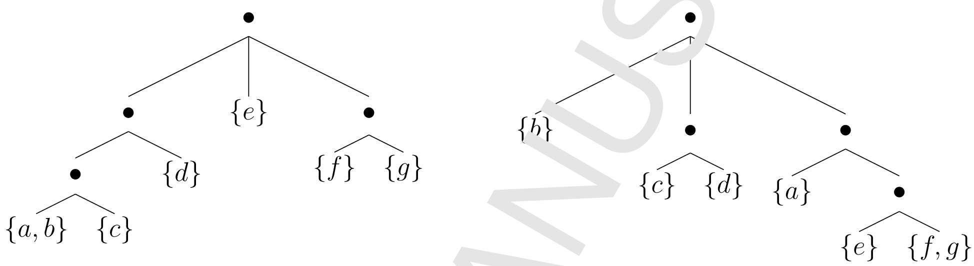


Figure 7: Two isomoprhic nested granular structures G (left) and G' (right)

Example 5. *Consider a universal set of objects $U = \{a, b, c, d, e, f, g\}$. Figure 7 shows an example of two isomorphic granular structures. It can be seen that we can obtain $G' = \{\{\{\{a, b\}, \{c\}\}, \{d\}\}, \{e\}, \{\{f\}, \{g\}\}\}$ from $G = \{\{b\}, \{\{c\}, \{d\}\}, \{\{a\}, \{\{e\}, \{f, g\}\}\}\}$ by using the bijection $F(a) = f, F(b) = g, F(c) = e, F(d) = a, F(e) = b, F(f) = d, F(g) = c$, and vice versa. The leaf nodes of G form a partition $\{\{a, b\}, \{c\}, \{d\}, \{e\}, \{f\}, \{g\}\}$ that is, according to Definition 13, isomorphic to the partition $\{\{b\}, \{c\}, \{d\}, \{a\}, \{e\}, \{f, g\}\}$ formed by the leaf nodes of G' under the bijection F. Furthermore, the structure of the two granules are the same. In defining the isomorphic relation between granules, the order of granules are not important, that is, the order of subtrees is not important.*

If two granular structures are isomorphic to each other, that is, they have the same structure, then they must also have the same complexity. A complexity measure must satisfy the following requirement:

$$G \cong G' \Longrightarrow C(G) = C(G'). \tag{43}$$

The complexity measures introduced in this paper are based on the structure of a granule, and do not depend on the the names of the elements within the granules. For an atomic granule, the complexity is measured solely by the

ACCEPTED MANUSCRIPT

number of elements contained inside. For a composite granule, the complexity is measured by the number of granules and their respective structures. A bijection f which defines isomorphic granules can be viewed as the renaming of elements and granules while preserving the structure. Therefore, all complexity measures satisfy the requirement given by Equation (43).

When studying a granularity measure of a partition, we also use a refinement-coarsening relation on partitions. As discussed earlier, the refinement-coarsening relation does not reflect our perception of the complexity of partitions. Thus, we need to introduce a "equally or less complex than," or equivalently, "not more complex than" relation $\preceq_c$. The relation $\preceq_c$ would allow for a partial ordering of various granular structures. Accordingly, a granule that is less complex than another granule would have a lower complexity value. In other words, we would have the following property of the complexity measure:

$$G \preceq_c G' \Longrightarrow C(G) \leq C(G'). \tag{44}$$

In conjunction with the isomorphic property, these two properties would form an axiomatic basis for defining a complexity measure of a granular structure.

Like the refinement-coarsening relation $\preceq$, the relation $\preceq_c$ is reflexive and transitive. However, unlike $\preceq$, it is difficult to define $\preceq_c$ based on the structures of granules. As shown by different cumulative orders of complexity measures, the complexity of granules is a multi-faceted notion. It may be impossible to define $\preceq_c$ in general. However, it may be possible to define $\preceq_c$ under a specific interpretation of complexity. This will be an interesting future research topic.

7. Conclusion

Granular computing is an interdisciplinary field that is growing in popularity. A common granulation scheme that researchers have studied is partition based granular computing. The present work addresses two main deficiencies with current studies. Firstly, researchers have focused on the partition as a two-level structure. The concept of partition-based granular structures is introduced to account for multilevel structures induced through progressive partitioning or refinement. The nested granular structure arises from the hierarchy of composite granules, which are composed of either atomic granules or composite granules. Secondly, a complexity measure is introduced to account for the complexity arising from the structure

ACCEPTED MANUSCRIPT

of a partition. Current studies consider only what is contained in individual blocks. The newly proposed measure considers the complexity arising from two sources, from the contents of each block and from the structure itself. Thus, we consider interactions between elements of a block and also between blocks of a partition. Whereas previous studies consider granularity measures of a partition as the expected granularity of its blocks, the proposed complexity measure considers the cumulative complexity arising from the blocks and the structure through summation. By increasing the number of blocks, the complexity of each block is increased, but the complexity resulting from the structure is increased. An avenue for future research is to study the axiomatization of specific classes of complexity measures.

The focus of this paper was to develop the theoretical foundation for the development of a newly proposed complexity measure. A main objective is to draw attention to the differences between the two notions of granularity and complexity of granular structures. While the former has been extensively studied, this paper is the first study of the latter. As such, the present work is focused on a theoretical investigation only. The proposed measures are applicable to any methods of granular computing where the complexity is of a main concern. While the general ideas are application independent, the choice of a particular k in the cumulative k-th complexity is application dependent. In order to avoid unnecessary distractions from focusing on any particular applications, we only consider simple examples to illustrate the computation of the values of the proposed measures. As a natural extension of this work, future work will include an experimental evaluation of the proposed complexity measures in various applications.

Acknowledgements

The author would like to thank the reviewers for their constructive and critical comments, Prof. Yiyu Yao for his fruitful discussions and insightful comments throughout the writing of the manuscript, and Prof. Jean-Pierre Hickey for his helpful comments and suggestions on improvements of the manuscript.

[1] A. Bargiela and W. Pedrycz. *Granular Computing: An Introduction.* Kluwer Academic Publishers, 2002.

[2] A. Bargiela and W. Pedrycz. *Human-centric information processing through granular modelling*, volume 182. Springer, 2008.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[3] T. Beaubouef, F. E. Petry, and G. Arora. Information theoretic measures of uncertainty for rough sets and rough relational databases. *Information Sciences*, 109(1-4):185–195, 1998.

[4] J. Dai and H. Tian. Entropy measures and granularity measures for set-valued information systems. *Information Sciences*, 240:72–82, 2013.

[5] J. Dai, W. Wang, Q. Xu, and H. Tian. Uncertainty measurement for interval-valued decision systems based on extended conditional entropy. *Knowledge-Based Systems*, 27:443–450, 2012.

[6] I. Düntsch and G. Gediga. Uncertainty measures of rough set prediction. *Artificial intelligence*, 106(1):109–137, 1998.

[7] Q. Feng, D. Miao, J. Zhou, and Y. Cheng. A novel measure of knowledge granularity in rough sets. *International Journal of Granular Computing, Rough Sets and Intelligent Systems*, 1(3):233–251, 2009.

[8] H. Fujita, T. Li, and Y. Yao. Advances in three-way decisions and granular computing. *Knowledge-Based Systems*, 91(2):1–3, 2016.

[9] R. V. Hartley. Transmission of information. *Bell Labs Technical Journal*, 7(3):535–563, 1928.

[10] Y. Jing, T. Li, C. Luo, S.-J. Horng, G. Wang, and Z. Yu. An incremental approach for attribute reduction based on knowledge granularity. *Knowledge-Based Systems*, 104:24–38, 2016.

[11] X. Kang and D. Miao. A study on information granularity in formal concept analysis based on concept-bases. *Knowledge-Based Systems*, 105:147–159, 2016.

[12] J. Liang, R. Li, and Y. Qian. Distance: A more comprehensible perspective for measures in rough set theory. *Knowledge-Based Systems*, 27:126–136, 2012.

[13] J. Liang and Z. Shi. The information entropy, rough entropy and knowledge granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(01):37–46, 2004.

ACCEPTED MANUSCRIPT

[14] J. Liang, J. Wang, and Y. Qian. A new measure of uncertainty based on knowledge granulation for rough sets. *Information Sciences*, 179(4):458–470, 2009.

[15] G. Lin, J. Liang, and Y. Qian. Uncertainty measures for multigranulation approximation space. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 23(03):443–457, 2015.

[16] D. Q. Miao and S. D. Fan. The calculation of knowledge granulation and its application. *Systems Engineering-theory & Practice*, 1:7–14, 2002.

[17] D. Q. Miao and J. Wang. On the relationships between information entropy and roughness of knowledge in rough set theory. *Pattern Recognition and Artificial Intelligence*, 1, 1998.

[18] Z. Pawlak. Rough sets. *International journal of computer & information sciences*, 11(5):341–356, 1982.

[19] Z. Pawlak. *Rough sets: Theoretical aspects of reasoning about data*, volume 9. Springer Science & Business Media, 2012.

[20] W. Pedrycz, A. Skowron, and V. Kreinovich. *Handbook of granular computing*. John Wiley & Sons, 2008.

[21] Y. Qian and J. Liang. Combination entropy and combination granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(02):179–193, 2008.

[22] J. R. Quinlan. Learning efficient classification procedures and their application to chess end games. In *Machine learning*, pages 463–482. Springer, 1983.

[23] S. Salehi, A. Selamat, and H. Fujita. Systematic mapping study on granular computing. *Knowledge-Based Systems*, 80:78–97, 2015.

[24] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[25] W. Shu and H. Shen. Multi-criteria feature selection on cost-sensitive data with missing values. *Pattern Recognition*, 51:268–280, 2016.

ACCEPTED MANUSCRIPT

[26] H. A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106:467–482, 1962.

[27] M. J. Wierman. Measuring uncertainty in rough set theory. *International Journal of General System*, 28(4-5):283–297, 1999.

[28] J. Yao. *Novel Developments in Granular Computing: Applications for Advanced Human Reasoning and Soft Computation*. IGI Global, 2010.

[29] Y. Yao. Probabilistic approaches to rough sets. *Expert systems*, 20(5):287–297, 2003.

[30] Y. Yao. A partition model of granular computing. In *Transactions on Rough Sets I*, pages 232–253. Springer, 2004.

[31] Y. Yao. Integrative levels of granularity. In *Human-Centric Information Processing Through Granular Modelling*, pages 31–47. Springer, 2009.

[32] Y. Yao. Interpreting concept learning in cognitive informatics and granular computing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(4):855–866, 2009.

[33] Y. Yao. A triarchic theory of granular computing. *Granular Computing*, 1(2):145–157, 2016.

[34] Y. Yao and L. Zhao. A measurement theory view on the granularity of partitions. *Information Sciences*, 213:1–13, 2012.

[35] P. Zhu. An improved axiomatic definition of information granulation. *Fundamenta Informaticae*, 120(1):93–109, 2012.

[36] P. Zhu and Q. Wen. Information-theoretic measures associated with rough set approximations. *Information Sciences*, 212:33–43, 2012.

ACCEPTED MANUSCRIPT