Applied Soft Computing 13 (2013) 4209–4218

Contents lists available at ScienceDirect

Applied Soft Computing

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

Building the fundamentals of granular computing: A principle of justifiable granularity

Witold Pedrycz[a,b,c,*], Wladyslaw Homenda[d]

[a] Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB, Canada
[b] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[c] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
[d] Faculty of Mathematics and Information Science, Warsaw University of Technology, pl. Politechniki 1, 00-660 Warsaw, Poland

ARTICLE INFO

Article history:
Received 25 November 2012
Received in revised form 1 May 2013
Accepted 19 June 2013
Available online 3 July 2013

Keywords:
Granular computing
Information granules
Design of granules
Justifiable granularity
Fuzzy sets
Intervals

ABSTRACT

The study introduces and discusses a principle of justifiable granularity, which supports a coherent way of designing information granules in presence of experimental evidence (either of numerical or granular character). The term "justifiable" pertains to the construction of the information granule, which is formed in such a way that it is (a) highly legitimate (justified) in light of the experimental evidence, and (b) specific enough meaning it comes with a well-articulated semantics (meaning). The design process associates with a well-defined optimization problem with the two requirements of experimental justification and specificity. A series of experiments is provided as well as a number of constructs carried for various formalisms of information granules (intervals, fuzzy sets, rough sets, and shadowed sets) are discussed as well.

1. Introduction

Granular computing [1,2,24,25] embraces a spectrum of concepts, methodologies, algorithms and applications, which dwell upon information granules and their processing. Information granularity is a fundamental concept that permeates this entire area. Information granules are building blocks using which a problem is represented, its model is constructed and ensuing decisions are constructed. Granular computing concentrates on processing information granules and constructively develops a holistic view at the discipline and incorporates the exisiting technologies and formalisms of sets (interval analysis), fuzzy sets [7,14,15], rough sets [11–13], probabilistic granules, shadowed sets [16–18] and alike. The fundamental quest arising in granular computing concerns a systematic way of forming information granules. With this regard, it becomes ultimate to develop a general way of designing information granules irrespectively from the formalism within which information granules are expressed. In brief, the crux of the problem can be highlighted as follows: given a collection of pieces of data or information granules (pieces of experimental evidence), form a representative information granule, which reflects the nature of the available experimental evidence. The ultimate objective of this study is to introduce a concept of justifiable granularity supporting a way of realizing an information granule. In addition to the concept itself, we form the underlying optimization problem in which the requirements of the proposed construct are expressed as well-defined optimization objectives so that a solution can be formed in a formal fashion. The formulation of the problem is independent from the way in which information granules are presented (say, as fuzzy sets or intervals).

In retrospect, while granular computing has enjoyed a significant growth, a comprehensive discussion on the realization of information granules is evidently lacking. Needless to say that there have been a lot of studies in which information granules were constructed in conjunction with some specific applications and in the context of the use of a certain formal vehicle of information granules, say fuzzy sets or rough sets. With this regard, different techniques of clustering are often studied. Clustering is viewed as a means of building a collection of information granules – clusters [18,20]. Depending on the formal mechanisms of clustering (where we encounter clustering, fuzzy clustering, rough clustering), formed is a collection of information granules such as sets, fuzzy sets or rough sets. The realization of information granules is associated with the underlying performance index (objective function) used in the clustering method. In contrast with the concept and the ensuing methodology introduced here, clustering lacks this unified approach (so finally information granules are reflective of quite diversified rationale behind a certain method being used). Furthermore in clustering we form a collection of information

* Corresponding author at: Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2G7 AB, Canada. Tel.: +1 7804398731.
E-mail address: wpedrycz@ualberta.ca (W. Pedrycz).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.06.017

granules; in this study we are concerned with a formation of a single information granule. Likewise the existing collection of user-driven methods aimed at the design of information granules exhibits a visible diversity and the method there do not offer a unified treatment of the concept of information granularity and support a general way of their construction.

In a nutshell, it is intuitively appealing to assume that an essence of the collection of experimental data (evidence) can be captured in a form a certain representative whose nature should be more general (abstract) in comparison with the experimental data to be dealt with. In case **D** is a collection of numeric data, the representative has to be granular (set, fuzzy set, etc.). In situation when **D** comprises information granules $\mathbf{D} = \{G_1, G_2, \ldots, G_N\}$, one envisions that the representative of **D** is elevated at the higher level of abstraction than the original entities. Symbolically, we can use here the notation G^2 (GG) to describe this construct of interest. For instance, we anticipate that in case of **D** coming as a collection of fuzzy sets, the resulting granular representative could be formed as a fuzzy set of higher type, say interval-valued fuzzy set, type-2 fuzzy set, probabilistic set, etc.

In a formal manner, we can pose the problem as follows:

Given some experimental evidence (typically, of numerical nature), construct a single information granule, which is (a) experimentally justifiable, and (b) exhibits a significant level of specificity.

This formulation of the problem is general in many different ways and when moving with an algorithmic realization of the principle, one has to express the underlying optimization objectives (what the requirements (a)–(b) are) and decide upon the formalism using which information granules are expressed.

The study is arranged in a top–down fashion. We start with the essence of the concept (Section 2). A series of illustrative experiments is reported in Section 3; here we show how information granules are designed in presence of discrete numeric data as well as data governed by some probability density functions. Next, we discuss a realization of information granules in the form of rough sets, look at the parametric aspects of the performance index and look at the weighted data. In Section 7, we show how with the aid of the representation theorem, the parameterized family of intervals can be arranged together to form a fuzzy set and a type-2 fuzzy set. A construction of a multidimensional information granule is presented in Section 8 while in Section 9 we elaborate on the realization of the principle of justifiable granularity in the presence of information granules rather than numeric evidence.

2. The concept of justifiable granularity and its algorithmic realization

As noted above, we are concerned with a development of a single information granule Ω based on some experimental evidence (data) coming in a form of a collection of a one-dimensional (scalar) numeric data, $\mathbf{D} = \{x_1, x_2, \ldots, x_N\}$ where $x_k \in \mathbf{R}$. In what follows, to focus the discussion, all information granules will be defined in **R**. This information granule is expressed in a certain formal framework of granular computing by being formed as an interval (set), fuzzy set, rough set, shadowed set and alike [22]. The essence of the principle of justifiable granularity is to form a meaningful information granule Ω based on available experimental evidence (data) **D** where we require that such a construct has to adhere to the two intuitively compelling requirements:

(i) *Experimental evidence*: The numeric evidence accumulated within the bounds of Ω has to be as *high* as possible. By requesting this, we anticipate that the existence of the information granule is well motivated (justified) as being reflective of the existing experimental data. For instance, if Ω is a set (interval) then the more data are included within the bounds of Ω, the better – in this way the set becomes more legitimate. Likewise in case of a fuzzy set, the higher the sum of membership degrees of the data in Ω, the higher the justifiability of this fuzzy set is.

(ii) *Semantic meaning:* At the same time, the information granule should be as *specific* as possible. This request implies that the resulting information granule comes with a well-defined semantics (meaning). In other words, we would like to have Ω highly detailed, which the information granule semantically meaningful (sound). This implies that the smaller (more compact) the information granule (lower information granule) is, the better. This point of view is in agreement with our general perception of knowledge being articulated through constraints (information granules) specified in terms of statements such as "x is A", "y is B" where A and B are constraints quantifying knowledge about the corresponding variables. Evidently, the piece of knowledge coming in the form "x is in [1,3]" is more specific (semantically sound, more supportive of any further action, etc.) than another piece of knowledge where we know only that "x is in [0,10]".

While these two requirements are appealing from an intuitive perspective, they have to be translated into some operational framework in which the formation of the information granule can be realized. This framework depends upon the accepted formalism of information granulation, viz. a way in which information granules are described as sets, fuzzy sets, shadowed sets, rough sets, probabilistic granules and others.

For the clarity of presentation and to focus on the nature of the construct, let us start with an interval (set) representation of information granule Ω. The requirement of experimental evidence is quantified by counting the number of data falling within the bounds of Ω. When a finite set of experimental data **D** is provided, we determine cardinality of elements falling within the bounds of Ω, namely $\text{card}\{x_k \mid x_k \in \Omega\}$. More generally, we may consider an increasing function of this cardinality, say $f_1(\text{card}\{x_k \mid x_k \in \Omega\})$ where f_1 is an increasing function of its argument. The simplest example is a function of the form $f_1(u) = u$.

The specificity of the information granule Ω associated with its well-defined semantics (meaning) can be articulated in terms of the length of the interval. In case of $\Omega = [a, b]$, any continuous non-increasing function f_2 of the length of this interval, say $f_2(m(\Omega))$ where $m(\Omega) = |b - a|$ can serve as a sound indicator of the specificity of the information granule. The shorter the interval (the higher the value of $f_2(m(\Omega))$), the better the satisfaction of the specificity requirement. It is evident that two requirements identified above are in conflict: the increase in the values of the criterion of experimental evidence (justifiable) comes at an expense of a deterioration of the specificity of the information granule (specific). As usual, we are interested in forming a sound compromise between these requirements.

Having these two criteria in mind, let us proceed with the detailed formation of the interval information granule. We start with a numeric representative of the set of data **D** around which the information granule Ω is created. A sound numeric representative of the data is its median, med(**D**). Recall that the median is a robust estimator of the sample and typically comes as one of the elements of **D**. Once the median has been determined, Ω (the interval $[a, b]$) is formed by specifying its lower and upper bounds, denoted here by "a" and "b", respectively; refer also to Fig. 1.

The determination of these bounds is realized independently. Let us concentrate on the optimization of the upper bound (b). The optimization of the lower bound (a) are carried out in an analogous fashion. For this part of the interval, the length of Ω or its nonincreasing function, as noted above. In the calculations of the cardinality of the information granule, we take into consideration the elements of **D** positioned to the right from the median, that is $\text{card}\{x_k \in \mathbf{D} \mid \text{med}(\mathbf{D}) < x_k \le b\}$. Again, in general, we can compute $f_1(\text{card}\{x_k \in \mathbf{D} \mid \text{med}(\mathbf{D}) < x_k \le b\})$, where f_1 is an increasing function.