Applied Soft Computing 9 (2009) 264–273

Contents lists available at ScienceDirect

Applied Soft Computing

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

# The design of fuzzy information granules: Tradeoffs between specificity and experimental evidence

Fusheng Yu [a], Witold Pedrycz [b,c,*]

[a] School of Mathematical Sciences, Beijing Normal University, Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing 100875, China
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada T6G 2G7
[c] System Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland

A R T I C L E I N F O

*Article history:*
Received 10 October 2005
Received in revised form 23 July 2007
Accepted 8 October 2007
Available online 22 April 2008

*Keywords:*
Information granules
Information granulation
Granular time series
Specificity-experimental evidence tradeoffs
Membership functions

A B S T R A C T

Granulation of information has become an underlying concept permeating a vast array of pursuits. In this study, we address a fundamental issue of the design of "meaningful" information granules. The underlying principle guiding their development realizes a tradeoff between a specificity of information granule (which we want to keep as high as possible) and the associated experimental evidence (which needs to be maintained high as well). Recognizing the fact that these are directly in conflict, we formulate the problem as a certain optimization task. In the sequel, we discuss the solutions and analyze their properties. In particular, we will deal with an array of parametric optimization dealing with various commonly encountered types of membership functions.



## 1. Introductory comments and problem statement

Information granulation and information granules [19] are intuitively appealing concepts that play important role in human cognition, processing and communication [17–20]. Information granulation plays a central role in granular computing [10], computing with words, and the computational theory of perceptions. Information granules are the generic conceptual and computing objects of granular computing. There are several frameworks in which information granules are built: set theory and interval analysis, fuzzy sets, rough sets, shadowed sets, probabilistic sets and probability, higher-level granular constructs [1,2,9,11,12,17,18]. As information granules are generic building blocks using which we realize all processing, their construction requires careful attention.

In this study, when building information granules we follow a general two-phase procedure. During the first phase we form a collection of segments of numeric data that establish a certain level of specificity we choose to look at the data. During the second phase we form a granular representation of the data falling within the individual segments, refer to Fig. 1.

Let us discuss these two phases in more detail. The segmentation imposes a certain level of granularity we consider to be of relevance in the given problem at hand. The higher the level of details we are interested in, the smaller the segments should be. Technically, segments are represented by sets and could overlap or may be non-overlapping. When dealing with time series, we form segments, called time windows, which embrace a series of consecutive elements of the series. There could be some overlap between the consecutive windows or they may not share any elements with the neighboring window. Likewise, we can think of segments of digital images (say windows of $8 \times 8$ pixels), etc. Refer to Fig. 2 for some illustrative examples. It is worth stressing that the segmentation mechanisms are well documented in the literature and there are a number of algorithmic means to support their formation.

Each segment embraces some numeric data. We are interested in finding a representative of the data falling within the scope of the segment. There are two general alternatives. One, which is commonly used, deals with a determination of a numeric representative of the elements belonging to **W**. This could be a modal value of the elements, mean or median. The second alternative, which is far less known is concerned with the construction of the granular representation of the elements of **W**. Our ultimate objective is to construct an information granule

* Corresponding author at: Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada T6G 2G7. Tel.: +1 780 492 3333; fax: +1 780 492 1811.
*E-mail addresses:* yufusheng@263.net (F. Yu), pedrycz@ece.ualberta.ca (W. Pedrycz).

1568-4946/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2007.10.026

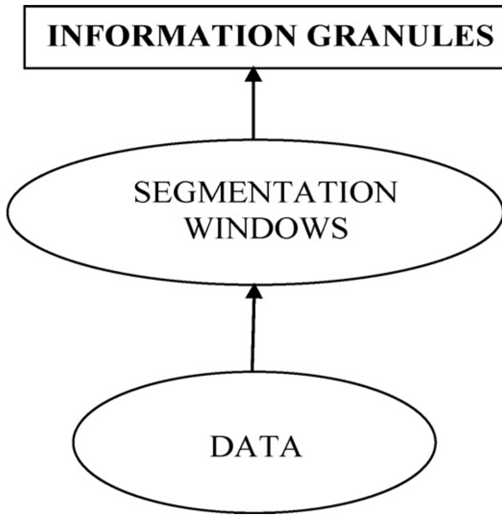


**Fig. 1.** Granulation of numeric information—two main development phases.

that captures the essence of numeric data of **W**. The problem requires a careful formulation as there are a number of viable alternatives that are worth pursuing. In this study, we will form it through an optimization of a tradeoff of specificity of information granule and its experimental evidence.

The material is arranged in the following manner. Section 2 briefly introduces the main model of information granulation. Section 3 concentrates on the properties of the core of the information granules. Next studied are several parametric properties of the information granules and related optimization pursuits. In the sequel, we elaborate on information granulation realized in a certain probabilistic setting and augment the representation of information granules to address the issues of higher-order dynamics. The study includes a number of experimental studies. We consider here time series yet the same approach is fully functional in highly dimensional cases (such as e.g., images).

In what follows, we adhere to the standard notation used throughout models of signal processing and augment it by the terminology and notions specific to fuzzy sets. Numeric values of the time series are arranged as a sequence $x_1, x_2, \ldots, x_n$ with the indexes denoting consecutive discrete time moments.

## 2. The granulation of information as an optimization problem

Adhering to the specificity-experimental relevance principle, we start with a concise formulation of the problem of granulation of information by casting it in the framework of a certain well-defined optimization task. We consider a certain granulation (segmentation) window of size $k$ consisting of $k$ successive elements of the given time series $x_1, x_2, \ldots, x_n$. Such non-overlapping windows are denoted by **W**.

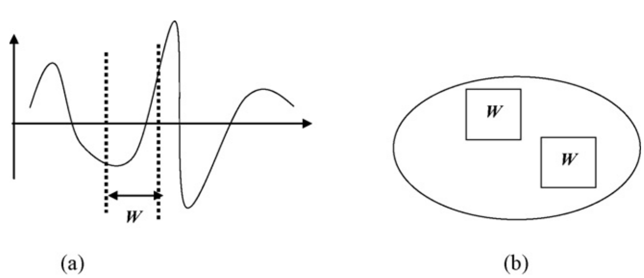


**Fig. 2.** Examples of overlapping and non-overlapping segments **W** occurring in the process of information granulation: (a) time series and (b) images.

### 2.1. The performance index

The development process of information granules can be posed as follows, cf. [1,14]:

> Given a collection of numeric data confined to the granulation window **W**, construct a fuzzy set (information granule) $A$ belonging to a certain family of fuzzy sets (say, triangular, parabolic, trapezoidal, etc.) such that it experimentally highly legitimate and retains high specificity.

The above formulation, which sounds quite convincing, involves two fairly conflicting requirements. First $A$ should embrace enough experimental data (to become experimentally valid). Second, it should be specific enough and this is accomplished by keeping its support as compact (small) as possible. These two requirements give rise to the maximization of the σ-count of the fuzzy set and the minimization of its support. The maximization realized in the form:

$$\text{maximize} \sum_{i=1}^{k} A(x_i)$$

assures us about enough experimental support behind $A$. Commonly, the sum $\sum_{i=1}^{k} A(x_i)$ can be treated as an overall level of "activation" of the information granule (which tells us about the cumulative experimental evidence gathered by $A$). In the sequel, the minimization of support (width) of $A$, that is

$$\text{supp}(A) = b - a$$

helps us to build a fuzzy set of high specificity. Here $a$ and $b$ are the left and the right bound of the support of the information granule (fuzzy set $A$).

Combining these two requirements we arrive at the following performance index that is a crux of our optimization task:

$$Q = \frac{\sum_{i=1}^{k} A(x_i)}{b - a}$$

We maximize $Q$ with respect to the parameters of the fuzzy set under consideration. The assumed performance index comes with a straightforward interpretation as shown in Fig. 3.

### 2.2. Granularity of information granules and a level of surprise

Alluding to the fundamental specificity-relevance tradeoff we addressed in the development of information granules, it is worth demonstrating an interesting effect of handling unexpectedness. What we have in mind concerns the following phenomenon. Given some numeric representative of **X**, call it $m$, we encounter a new

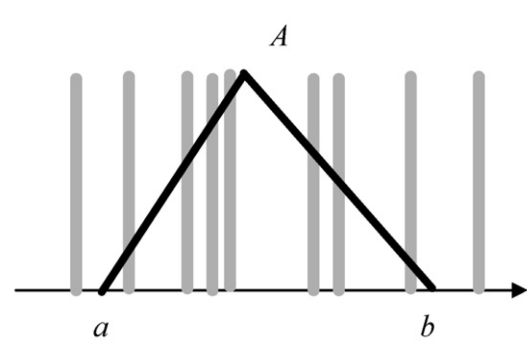


**Fig. 3.** The design of information granules (fuzzy sets) as an optimization process; experimental numeric data points falling under some granulation window are displayed as singletons. The sum of membership grades is maximized; the support of the fuzzy set is minimized (so its specificity is maximized).