Expert Systems with Applications xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An expansion of fuzzy information granules through successive refinements of their information content and their use to system modeling

Abdullah Balamash [c,*], Witold Pedrycz [a,c,b], Rami Al-Hmouz [c], Ali Morfeq [c]

[a] Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2V4, AB, Canada
[b] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
[c] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Granular Computing
Information granules
Fuzzy clustering
Successive refinements
Information content
Diversity

ABSTRACT

This study is concerned with a fundamental problem of expanding (refining) information granules being treated as functional entities playing a pivotal role in Granular Computing and ensuing constructs such as granular models, granular classifiers, and granular predictors. We formulate a problem of refinement of information granules as a certain optimization task in which a selected information granule is refined into a family of more detailed (precise, viz. more specific) information granules so that a general partition requirement becomes satisfied. As the ensuing information granules are directly linked with the more general information granule positioned at the higher level of hierarchy, the partition criterion is conditional by being implied (conditioned) by the description of the granule positioned one level up in the hierarchy. A criterion guiding a refinement of information granules is formulated and made fully reflective of the nature of the problem (being of regression-like or of classification character), which leads to a distinct way in which the diversity of information granules is articulated and quantified. With regard to the detailed algorithmic setting, we discuss the use of a so-called conditional Fuzzy C-Means and show how information granules (fuzzy sets) are formed in a successive manner. The method helps highlight the ensuing calculations of the resulting membership functions and reveal how the detailed structure of the data is captured. A number of numeric studies in the realm of system modeling are provided to demonstrate the performance of the approach and highlight the nature of the resulting information granules along with the performance of the fuzzy models in which these information granules are used.

1. Introductory notes

Information granules realize an abstract view at detailed, quite often numeric data. Information granules are essential building functional blocks using which a number of constructs of Granular Computing (Pedrycz, 2013b; Zadeh, 1997, 2011) are realized. We may refer here to a variety of ensuing granular models, predictors, classifiers, and data descriptors, etc. The level of abstraction captured by information granules can be effectively controlled in the sequel giving rise to the formation of a sound and legitimate trade-off between generality (relevance) and precision (which, as a matter of fact, is a succinctly expressed by Zadeh as a principle of incompatibility (Zadeh, 1975)). Information granules are building functional blocks being present in a broad spectrum of models, especially those rule-based oriented (Ayouni, Ben Yahia, & Laurent, 2011; Yen, Wang, & Gillespie, 1998) where information granularity, modularity of constructs, distributed functionality become apparent. We can refer here to rule-based systems where the granularity of information present there and giving rise to rule bases of different sizes helps strike a sound balance between generality (readability, transparency) of the underlying construct and the precision of the produced results. Models of different level of generality require information granules of a suitable level of specificity. Admitting a top–down strategy of model formation, we are often faced with a task of refining information granules and splitting some of them into more detailed entities.

It is instructive to position the approach presented here with some other studies reported in the literature, especially in the setting of regressing models. Information granules have been found useful in regression analysis. A recent application of information granules in the formation of regression models comes under the name of granular box regression (Peters, 2011) where a fuzzy graph is determined by embedding a given data set into a predefined number of boxes. These boxes realize information granules

* Corresponding author.
E-mail addresses: asbalamesh@kau.edu.sa (A. Balamash), wpedrycz@ualberta.ca (W. Pedrycz), ralhmouz@kau.edu.sa (R. Al-Hmouz), morfeq@kau.edu.sa (A. Morfeq).

http://dx.doi.org/10.1016/j.eswa.2014.11.027
0957-4174/

and are defined in a way to minimize their total volume. The work reported in Grzegorzewski (2013) shows a development of a formal model of granular regression as a generalization of the least square method. The main idea was to generate granular partition for the given data set based on multi-distances (i.e., the Fermat multi-distance (Martín & Mayor, 2011)). These two regression approaches are in some sense related to a class of regression algorithms called interval regression or fuzzy regression (Chen & Hsueh, 2009; Kahraman, Beşkese, & Bozbura, 2006; Yang & Lin, 2002), but the main difference is that the intervals (granules) in most of these works are defined based on the dependent variable. Our scheme is different than the schemes discussed in Grzegorzewski (2013) and Peters (2011) in a way that although these two schemes consider the independent variable in deciding about the information granules, they seem to combine both variables as if they are a single variable with an extra dimension and based on that the partitioning of the data is completed. In contrast, our work focuses on clustering realized for independent variables (more specifically, Fuzzy C-Means clustering) while being navigated (controlled) by the variance of the dependent variable reported for each cluster.

With the development of granular constructs comes an important design question about a successive formation of refined, more detailed information granules. In a concise manner one can capture the essence of this pursuit as follows: by starting with a collection of information granules defined in some multivariable space of real numbers $\boldsymbol{X}$, $\boldsymbol{X} = \boldsymbol{R}^n$ and denoted here by $A_1, A_2, \ldots, A_r$, we refine them successively by choosing the most "suitable" candidate, say A_i and on its basis develop a family of c refined information granules $A_{i1}, A_{i2}, \ldots, A_{ic}$ so that A_{ij}s become more specific and realize a "conditional" partition induced by A_i in the sense of the satisfaction of the following equality

$$A_i = \bigcup_{j=1}^{c} A_{ij} \tag{1}$$

Namely

$$\sum_{j=1}^{c} A_{ij} = A_i \tag{2}$$

Thus the general partition requirement (Abonyi, Feil, Nemeth, & Arva, 2005; Bezdek, 1981; Pedrycz, 2005, 2013b; Pedrycz & de Oliveira, 2008) imposed on $A_1, A_2, \ldots, A_r$ can be rewritten when proceeding with the refinement of information granule A_i as follows

$$A_1 + A_2 + \cdots + A_{i-1} + \sum_{j=1}^{c} A_{ij} + A_{i+1} + \cdots + A_r = 1 \tag{3}$$

In the above expression (expansion) we encounter a standard normalization condition (all membership values sum up to 1). In other words, we have $A_i = \sum_{j=1}^{c} A_{ij}$ Further refinements of the kth information granule A_k result in a series of more detailed information granules $A_{k1}, A_{k2}, \ldots, A_{kc}$ and as a result of this granular expansion the following general expression arises

$$A_1 + A_2 + \cdots + A_{i-1} + \sum_{j=1}^{c} A_{ij} + A_{i+1} + \cdots + A_{k-1} + \sum_{j=1}^{c} A_{kj} + A_{k+1} + A_r = 1 \tag{4}$$

Of course, in the next expansion step one might refine one of A_{ij} (say, A_{il}) and in such a way we encounter further expansion realized in the following form

$$A_1 + A_2 + \cdots + A_{i-1} + \sum_{\substack{j=1 \\ j \neq l}}^{c} A_{ij} + \sum_{j=1}^{c} A_{ilj} + A_{i+1} + \cdots + A_{k-1} + \sum_{j=1}^{c} A_{kj} + A_{k+1} + A_r = 1 \tag{5}$$

with the partition-like requirement expressed as $\sum_{j=1}^{c} A_{ilj} = A_{il}$

The refinements (specification) of information granules are carried out with regard to some criterion that characterizes the diversity (information content) of the granules. The granule of the highest diversity is a candidate for further refinements. The content of information granules formed in a given multivariable space of independent variables $\boldsymbol{R}^n$ could exhibit a different nature. In particular, two general and practically viable scenarios are worth distinguishing:

(a) if the dependent variable is continuous (regression problems), the content of information granule is directly related with its diversity (variance or any other measure of dispersion) reported for the associated output variable,
(b) if the dependent variable is a class label (as encountered in classification problems), the content of information granule is associated with the error rate (misclassification) occurring for the patterns localized within the scope of a given information granule.

The objective of the study is to establish concepts of expansion (refinements) of information granules and propose a detailed algorithmic setting in which such refinements are carried out in a systematic manner, which are characterized by the highest information content (diversity). The task being fundamental to the realization of a main agenda of Granular Computing exhibits a significant level of novelty as well as supports a sound methodology of designing of constructs involving information granules. To augment further investigations by all required algorithmic settings, we use Fuzzy C-Means (FCM) (Bezdek, 1981) as a detailed algorithmic framework. We demonstrate that its variant, a so-called context-based FCM (or conditional clustering) (Pedrycz, 1998) helps link information granules formed in the process of successive refinements. This has to be stressed, though, that the findings of this study and the developed methodology go beyond this particular mechanism of fuzzy clustering.

In the context of this study, it is essential to look at the research being conducted so far as the ideas of both specialization (refinement) and generalization as these two concepts have been investigated in fuzzy sets and rough sets. One may refer here to the studies reported in Pedrycz and Sosnowski (2000) however in contrast to the approach proposed here, the concept of *conditional* partitioning in the refinement process has not been fully investigated. In Roh, Pedrycz, and Ahn (2014), the concept of information granules is used in classification problems. The FCM was used to form information granules and then the Particle Swarm Optimization (PSO) algorithm was applied to improve the clustering results by adjusting the prototypes. In Zeng and Dong (2014), the authors proposed a scheme to attenuate noise from heart signal (HS) where the short-Fourier transform was used to decompose each HS cycle into fragments (granules) where the noise can be easily removed from each fragment and then the fragments are remerged to from the clean signal. In Gacek (2013), the authors suggested alternative granular representation of time series and using the principle of justifiable granularity (Pedrycz & Homenda, 2013) to form adjustable time windows. In Lu, Pedrycz, Liu, Yang, and Li (2014), a human perceivable time series modeling technique was proposed. The method is based on dividing the time series into intervals of equal size and representing each interval by a fuzzy (linguistic) granule using the concept of justifiable granularity. The goal was to trade off precision with computational overhead, which is acceptable for decision-making applications where accurate numeric values of prediction are not necessary. In Wang, Liu, and Pedrycz (2013) and Wang, Liu, Pedrycz, and Shao (2014), the authors suggested a partitioning algorithm that can improve forecasting by partitioning a given data set into unequal intervals