International Journal of Approximate Reasoning 53 (2012) 1061–1079

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

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

Granular fuzzy models: A study in knowledge management in fuzzy modeling

Witold Pedrycz [a,b,c,*], Mingli Song [a]

[a] Department of Electrical & Computer Engineering, University of Alberta, Edmonton, Canada T6R 2V4 AB
[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

ARTICLE INFO

Article history:
Available online 29 May 2012

Keywords:
Information granules
Knowledge management
Granular fuzzy model
Principle of justifiable granularity
Information specificity
Data coverage

ABSTRACT

In system modeling, knowledge management comes vividly into the picture when dealing with a collection of individual models. These models being considered as sources of knowledge, are engaged in some collective pursuits of a collaborative development to establish modeling outcomes of global character. The result comes in the form of a so-called *granular* fuzzy model, which directly reflects upon and quantifies the diversity of the available sources of knowledge (local models) involved in knowledge management. In this study, several detailed algorithmic schemes are presented along with related computational aspects associated with Granular Computing. It is also shown how the construction of information granules completed through the use of the principle of justifiable granularity becomes advantageous in the realization of granular fuzzy models and a quantification of the quality (specificity) of the results of modeling. We focus on the design of granular fuzzy models considering that the locally available models are those fuzzy rule-based. It is shown that the model quantified in terms of two conflicting criteria, that is (a) a coverage criterion expressing to which extent the resulting information granules "cover" include data and (b) specificity criterion articulating how detailed (specific) the obtained information granules are. The overall quality of the granular model is also assessed by determining an area under curve (AUC) where the curve is formed in the coverage-specificity coordinates. Numeric results are discussed with intent of displaying the most essential features of the proposed methodology and algorithmic developments.

1. Introduction

In system modeling, we augment a plethora of existing modeling methodologies and algorithms by moving from a concept of single, individual, and local models to a family of models in which individual models collaborate with others with an ultimate intent of forming more abstract, holistic model of the underlying phenomenon, process or a system delivering a global, albeit less detailed view at the reality. Along with the associated tangible benefits and better rapport with reality (distributed systems, various modeling perspectives), this shift brings a number of new challenges irrespectively from the development technologies one has started with. There is no surprise that in fuzzy modeling and computing with words [21] with its plethora of design techniques, see [1–3,6,10–12,18] involving criteria of accuracy and interpretability [4,9] and invoking promising methods of global optimization [17], this concept has to translate into sound concepts, methodology, design strategies, and finally detailed algorithms.

* Corresponding author at: Department of Electrical & Computer Engineering, University of Alberta, Edmonton, Canada T6R 2V4 AB.
E-mail address: wpedrycz@ualberta.ca (W. Pedrycz).

http://dx.doi.org/10.1016/j.ijar.2012.05.002

Before proceeding with the conceptual and algorithmic considerations, it is beneficial to make some observations of a general character, which helps motivate the study and focus on the essence of the required design features.

Let us consider a system for which formed is a series of models. The system can be perceived from different, quite diversified points of view. It can be observed over some time periods and analyzed at different levels of detail. Subsequently, the resulting models are built with different objectives in mind. They offer some particular albeit useful views at the system. We are interested in forming a holistic model of the system by taking advantage of the individual sources of knowledge – models, which have been constructed so far. When doing this, we are obviously aware that the sources of knowledge exhibit diversity and hence this diversity has to be taken into consideration and carefully quantified. No matter what the local models look like, it is legitimate to anticipate that the global model (at the higher level of hierarchy) is more general, abstract. From the perspective of the concept of the global model, it is fair to assume that the model has to engage a sort of formalism of information granules. Information granules [14,19,20] are a genuine generalization of numeric information. As such, information granules offer a unique way of quantifying a diversity of sources of knowledge under consideration and expressing this aspect in the form of the level of granularity (specificity).

The objective of this study is to develop the concepts of *granular* fuzzy models – a new modeling paradigm in fuzzy modeling in which we form a hierarchy of modeling constructs starting with fuzzy models at the lower level of the hierarchy and ending up with a granular fuzzy model emerging at the higher level.

It is worth noting that there have been some developments along the line of collaboration, consensus-building, and knowledge exchange and its effective usage in decision-making [5,7,8,13,15] and fuzzy models [16], however the role and unavoidable emergence of information granules as a vehicle to quantify a diversity of sources of knowledge has not been fully exploited.

The organization of this paper is reflective of the main objectives of the overall study.

We briefly review the main developments exiting in the literature and contrast the proposed approach with the main pursuits that have been reported so far (Section 2). This comparative view is of particular interest as it helps highlight the main differences between the proposal put forward in this study. Next, the principle of justifiable granularity (Section 3) is discussed in detail with intent to having an effective and parametrically flexible vehicle to form information granules. In Section 4, we elaborate on the concept of granular fuzzy models while in Section 5 we move onto more focused discussion in which local fuzzy models are those formed by rules. By doing this, it becomes possible to reveal all essential design steps as well as present a way of quantifying the performance of the resulting construct and highlight the multiobjective character of the design in which tradeoffs exist between the specificity of the results and their relevance (expressed by a coverage index). A suite of experimental studies is covered in Section 6. Some main conclusions are summarized in Section 7.

2. An aggregation of fuzzy models

An idea of using a collection of models studied *en block* has been around in the literature for some time. It is commonly known under the term ensemble of learning machines in machine learning, mixture of experts, and alike. There has been a compelling evidence demonstrating (both in theory and through experimental studies) that the performance of such ensembles of learning machines (say, weak classifiers) is substantially better than a single model or a classifier; refer to [25–28]. To offer a glimpse at the diversity and a surprising similarity among of the existing approaches, a list of selected models along

Table 1
Selected models and their aggregation.

Reference	Local models	Aggregation mechanism	Aggregation result
[22]	"n" numeric values in [0,1] of "n" criteria $\{A_1, A_2, \ldots, A_n\}$,	OWA (ordered weighted averaging):	Numeric value
[23]	"n" numeric values in [0,1] of "n" criteria $\{A_1, A_2, \ldots, A_n\}$,	$G(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{q} \alpha_i F w_i([B_i]) B_i$ is a subset of relevant criteria, M_i provides a description of what portion of the criteria in need be satisfied in order for the module to be satisfied and α_i indicates the value of satisfying this module	Numeric value
[24]	"n" interval-valued fuzzy sets	$Y = \frac{\sum_{i=1}^{m} W_i X_i}{\sum_{i=1}^{m} W_i}$,	Interval – valued fuzzy set
[25]	n CART trees in an ensemble	EP (Expectation propagation) pruning algorithm	A subset of ensemble members (bagging, Adaboost, and Random Forest used to generate ensemble)
[26]	n nearest-neighbor classifiers	majority voting	Numeric class label
[27]	"n" MLP classifiers	Linear function combiner with optimal values of the classifier weights	Numeric class label
[28]	"n" decision trees	Each classifier in the ensemble is replaced by a mini ensemble of a pair of subclassifiers with a random linear oracle to choose between the two.	An ensemble; numeric output
[29]	"n" neural network (NN) classifiers	No aggregation; each NN is an element of the ensemble; the smallest error on a validation set.	An ensemble of "n" neural networks; numeric output

with their essential features is presented in Table 1. Further on, we refer to the same table and the main findings resulting from it to highlight the key differences with the approach presented in this study.

In spite of the visible diversity present in the table, there is also a profoundly visible similarity: the result of aggregation is positioned at the same level of granularity as the results produced by the individual models involved in the aggregation.

3. The principle of justifiable granularity and design of fuzzy sets

The principle of justifiable granularity [14] is concerned with a formation of a meaningful representation of a collection of numeric values (real numbers), say $\{x_1, x_2, \ldots, x_N\}$. Such a representation comes in the form of a certain information granule rather than a single numeric entity, no matter how such a single numeric individual has been selected. What is being done in statistics is an example of this principle realized in the language of probabilistic information granules. A sample of numeric data is represented not only by its mean or median (which is a very rough description) but also by the standard deviation. Both the mean and the standard deviation imply a realization of a certain probabilistic information granule, such as e.g., a Gaussian one. The probabilistic information granules are just one of the possibilities to build an information granule to represent a collection of numeric data. Some other formal approaches that could be engaged here involve sets, fuzzy sets, rough sets and others. Formally, we can view the process of granulation of information as a transformation $\mathcal{g}$ operating on the set of data $\{x_1, x_2, \ldots, x_N\}$ positioned in **R** and resulting in a certain information granule $\boldsymbol{G}$

$$\mathcal{g} : \{x_1, x_2, \ldots, x_N\} \rightarrow \boldsymbol{G} \tag{1}$$

The realization of the transformation $\mathcal{g}$ becomes a crux of the design of information granules. The result – information granule $\boldsymbol{G}$ depends which formalism of information granularity has been adopted in advance. For instance, $\boldsymbol{G}$ could be a characteristic function in case of sets defined over $\mathbf{R}, \boldsymbol{G} \in \boldsymbol{P}(\mathbf{R})$, a fuzzy set, $\boldsymbol{G} \in \boldsymbol{F}(\mathbf{R})$, etc. Recall that $\boldsymbol{P}(\mathbf{R})$ and $\boldsymbol{G} \in \boldsymbol{F}(\mathbf{R})$ stand for a family of sets (intervals) and fuzzy sets, respectively, defined over all real numbers.

In case of other formalisms of information granulation, the development of the corresponding granules is guided by a certain optimization criterion, which becomes a crux of the principle of justifiable granularity. In general, in such criteria, we manage two conflicting requirements. The one is about forming an information granule of sufficiently high level of experimental evidence that is accumulated behind it and in this way supports its existence (and usage). The second one is about maintaining high specificity of the resulting information granule. In what follows, we show a construction of interval-based information granules in presence of some numeric evidence, see Fig. 1.

The construction of Ω is realized in two phases. We start with a determination of the numeric representative of the data. Denote it by "m". A sound representative is its median, med, whose advantage is that it is a robust estimator and typically comes as one of the elements of the data. Another option is to consider a mean value as a numeric representative.

The construction of the bounds of the interval (a and b) is realized independently. We concentrate on the optimization of the upper bound (b). The calculations of the lower bound (a) are carried out in an analogous fashion.

We span the numeric interval Ω (= [a, b]) in such a way that (i) the numeric evidence accumulated within the bounds of Ω is as high as possible. We quantify this requirement by counting the number of data falling within the bounds of Ω, that is card $\{x_k | x_k \in \Omega\}$. This count has to be maximized. At the same time, we require that (ii) the support of Ω is as low as possible, which makes Ω specific (detailed) enough. These two requirements are in conflict. To come up with a single performance index V taken as a product of two functionals f_1 and f_2, $V = f_1 * f_2$ where f_1 is an increasing function of the cardinality of the elements falling within the bounds of the information granule Ω and f_2 is a decreasing function of the support (length) of $|b - m|$. To address the two requirements formulated above, our intent is to maximize Q with respect to the boundary of the interval.

One of the interesting alternatives is to consider the following two functionals:

$$f_1(u) = u \tag{2}$$

and

$$f_2(z) = \exp(-\alpha z) \tag{3}$$

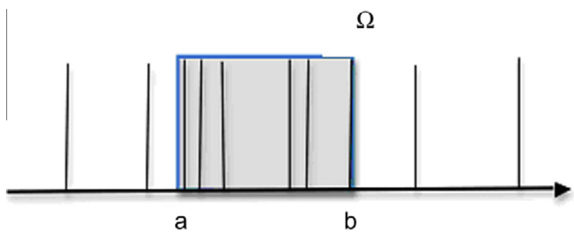

Fig. 1. Realization of the principle of justifiable granularity for numeric data and an interval type of information granules. Note that the information granule is distributed around a certain numeric representative (m); $a < m < b$.

where $z = |b - m|$ and α is a positive parameter controlling the changes of the function and subsequently offering some flexibility in the produced an information granule. The maximization of the expression

$$V(b) = \text{card}\{x|x > m \text{ and } x < b\} * \exp(-\alpha|m - b|) \quad (4)$$

is carried out with respect to unknown upper bound "b". Here "m" stands for the numeric representative of the numeric data. Likewise, we proceed with the maximization of the expression (objective function) $V(b) = \text{card}\{x|x \in \Omega\} * \exp(-\alpha|m - b|)$ with "b" being the upper bound of the information granule. It is worth noting that (4) comes with some parameter (α), which is essential to the further usage of granular constructs.

With regard to the above maximization (4) one can note an interesting and intuitively appealing relationship between "b" and the values of the parameter α. Consider that the data larger that the numeric representative are $x_1, x_2, \ldots, x_N$ Recall that "m" is a certain numeric representative. In spite of the values of α, the location of optimal "b" is positioned in-between min x_i and max x_i Given the form of (4) if b falls below the bound min x_i, then the performance index is equal to zero. If b moves beyond max x_i, the performance index decreases monotonically. Indeed, take b_1 and b_2 where $b_1 < b_2$. In this case, the performance index $V(b_1)$ is higher that $V(b_2)$ which indicates that moving towards higher values of "b" after exceeding max x_i leads only to the reduced values of the performance index.

In a similar way we can develop an information granule in the form of a certain fuzzy set. Here the only difference is that the cardinality of the set of data contained within the interval has to be replaced by the sum of membership degrees of the data belonging to the fuzzy set.

The algorithm realizing the principle of justifiable granularity produces an information granule (either an interval or a fuzzy set) based on a collection of numeric data. The nature of the numeric data themselves can be quite different. Two situations of significant practical relevance are worth highlighting here:

(a) The numeric data could result from measuring some variables present in the system. In this case, information granules are treated as non-numeric data, which can be then used in the design of the model and highlight the structure of a large number of numeric data.
(b) The numeric data are just membership values of some fuzzy sets reported for a certain point of the universe of discourse. The granular representation resulting from the discussed construct gives rise to the information granule of higher type, fuzzy set of type-2, to be more specific. Depending on the nature of the information granule formed here, we construct an interval-valued type-2 fuzzy sets or type -2 fuzzy sets. It is worth stressing that in this manner, we have arrived at a constructive way of designing of type-2 fuzzy sets – the area that has been very much left neglected in the existing studies.

The algorithms supporting the design of information granules presented so far have been concerned with one-dimensional data. In situation of multivariable data, the developed method can be applied to individual variables separately and then combined by taking a Cartesian product of the one-dimensional information granules.

4. The emergence and a concept of granular fuzzy models

As highlighted in the introductory section, one can acknowledge that the component of granularity brings a desired level of flexibility required to reflect and quantify an inherent diversity of individual fuzzy models we have to deal with. The underlying concept is visualized in Fig. 2. The unavoidable granularity effect manifests at the resulting model formed at the higher level of abstraction, hence the construct obtained there comes as a generalization of fuzzy models present at the lower level of the hierarchy, namely *granular* fuzzy models.

A certain system or phenomenon is perceived from different points of view (perspectives). The fuzzy models emerging there, denoted as FM-1, FM-2, ..., FM-p use locally available data D-1, D-2, ..., D-p. In general, we can envision fuzzy models to be of different nature (say, rule-based, neurofuzzy model, fuzzy cognitive map), however for the sake of the clarity of overall presentation we assume that all of them are fuzzy rule-based models. The models FM-1, FM-2, ..., FM-p are now brought together by forming a global model at the higher level of hierarchy. Not engaging into more specific discussion, some general and intuitively convincing observations could be made. The model at the higher level is definitely more abstract (general) than those at the lower level as it tries to embrace a variety of sources of knowledge (fuzzy models). As such we envision that the output of such fuzzy model is less specific than the numeric output generated by any of the fuzzy models at the lower level. This is not surprising at all. Imagine that the input is equal to $\mathbf{x}$ with $\mathbf{x}$ being a numeric vector of inputs coming from the system. Each fuzzy model produces the numeric outputs FM-1($\mathbf{x}$), FM-2($\mathbf{x}$), ..., FM-p ($\mathbf{x}$). In virtue of the different perception of the system itself, it is very likely that these inputs are different; we do not rule out that they could be close to each other. If the fuzzy model at the higher level is to capture this diversity of the sources of knowledge (viz. fuzzy models), intuitively for the same $\mathbf{x}$ it should return an output that is more abstract that a single numeric entity, that is a certain information granule. Hence such fuzzy models producing granular outputs will be referred to as *granular* fuzzy models.

Alluding to Table 1, we can contrast the proposed construct with the approaches existing in the literature, especially that on a surface there seem to be some similarities. There are several evident differences, though:

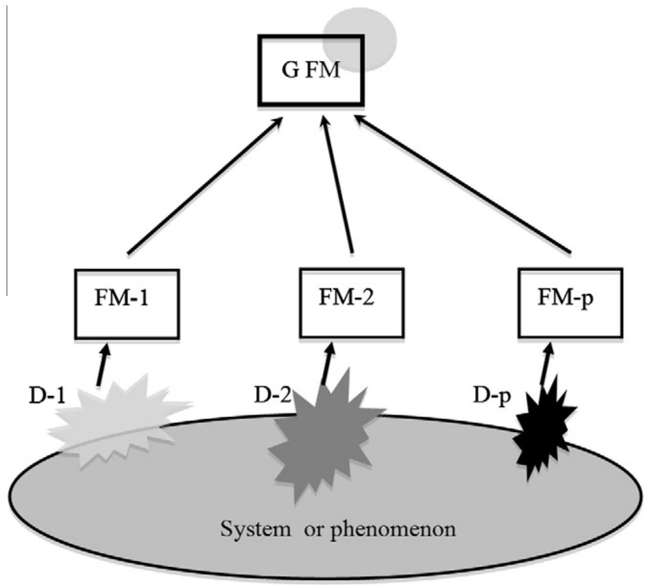

Fig. 2. The emergence of information granularity in the model at the upper level of the hierarchy as a result of dealing with the diversity of fuzzy models FM-1, FM-2, ..., FM-p. The emergence of information granularity is denoted by large grey dot placed next to the granular fuzzy model being formed there.

- The result of aggregation is positioned at the higher level of abstraction in comparison with the results generated by the individual models. For instance, the models involved in the aggregation produce numeric outputs and numeric outputs are also generated as a result of aggregation. Here, the outputs are granular (coming in the form of intervals, fuzzy sets, rough sets, etc.) even though the individual models generate numeric outputs. This can be viewed as a substantial advantage of the overall topology as it helps quantify the diversity of the individual sources of knowledge (viz. models). The quantification of this diversity is realized by running the principle of justifiable granularity.
- The data are not shared meaning that each model is constructed on a basis of locally available data; furthermore the variables at each data set need not be the same. This stands in sharp contrast with the scenario of bagging or boosting of a family of models.
- The individual models could be very different (say, neural networks, rule-based models, linear regression, etc.)

5. The detailed considerations: from fuzzy rule-based models to granular fuzzy models

In this section, we present in more detail one of the interesting scenarios of capturing sources of knowledge (fuzzy models) in case the individual sources of knowledge-fuzzy models are treated as a collection of fuzzy rule-based models. The transfer of knowledge realized here is focused on the use of information granules forming a backbone of the individual fuzzy models at the lower level. As the antecedents of the rules are formed on a basis of information granules, a realization of a certain model at the higher level of hierarchy calls for a formation of a collection of information granules to start with. Here we envision two main directions in the construction of information granules as being portrayed in Fig. 3:

(a) A selection of a suitable subset of information granules forming the individual models, Fig. 3(a). The prototypes of the information granules are selected in such a way so that they represent all prototypes of the models to the highest extent. This is a combinatorial optimization problem, which may call for techniques of Evolutionary Optimization (e.g., Genetic Algorithms, GAs) or population-based optimization (e.g., Particle Swarm Optimization, PSO) to arrive at solutions to the problem. The optimization criterion quantifies a reconstruction error of the prototypes at the lower level when being expressed in terms of the subset of the prototypes formed at the upper level.

The second approach, illustrated in Fig. 3(b), is concerned with clustering (granulation) of prototypes available at the lower level. The standard FCM can be used here. It operates on a family of the prototypes present in all fuzzy models positioned at the lower level of hierarchy and produces "*c*" prototypes at the upper level of the hierarchy. In light of the construction of the information granules, which have been built at the higher level (which are also sought as more abstract view at the information granules present at the lower level), we may refer to them as information granules of higher type, say type-2 information granules (and type-2 fuzzy sets, in particular).

Given the collection of information granules (which can be represented as a family of the prototypes), we are at position to develop a model at the higher level. We note there is an inherent granularity of the associated model, which comes from the fact that for any prototype formed at the higher level $\mathbf{v}_i$, each fuzzy model at the lower level returns some numeric value, say, FM-1($\mathbf{v}1$), FM-2($\mathbf{v}_2$), ..., FM-p($\mathbf{v}_i$). It is very unlikely that all these values are the same. This set of data is subject to the

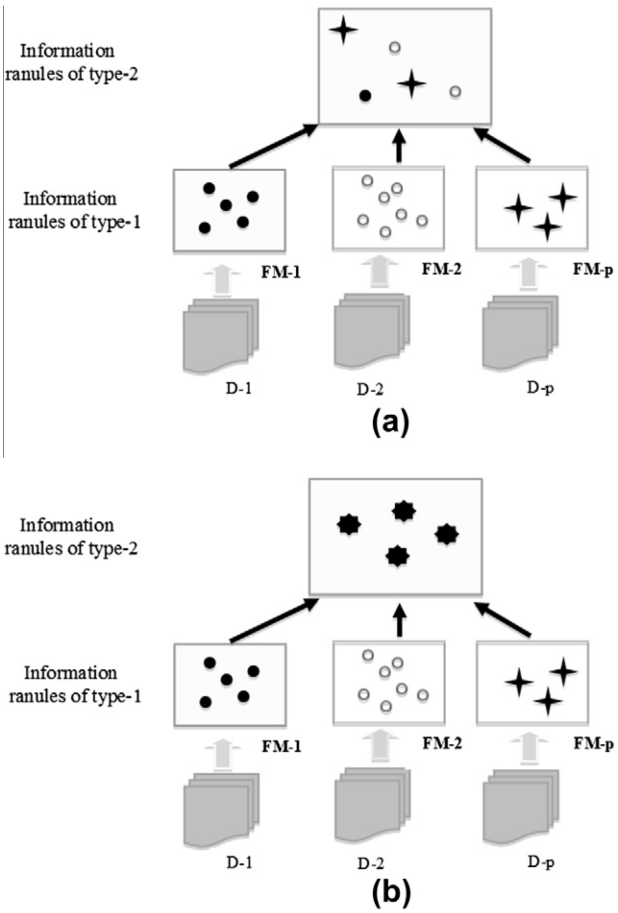

Fig. 3. Formation of the information granules at the higher level: (a) selection and (b) clustering of prototypes.

granulation process (with $\mathcal{g}$ denoting a granulation mechanism described in Section 2). The size (level of granularity) of the resulting information granule depends upon the predetermined value of the parameter α, which was used to construct information granules. We will be taking advantage of this flexibility in the realization of the model guided by two conflicting objectives (as this will be discussed later in more detail).

Overall, for "c" information granules built at the higher level, the available experimental evidence (originating at the lower level of hierarchy) arises in the form

$$\begin{aligned}&\{(\mathbf{v}_1, \mathcal{g}\{FM-1(\mathbf{v}_1), FM-2(\mathbf{v}_1), .., FM-p(\mathbf{v}_1)\}),\\&(\mathbf{v}_2, \mathcal{g}\{FM-1(\mathbf{v}_2), FM-2(\mathbf{v}_2), \dots, FM-p(\mathbf{v}_2)\}), \dots\\&(\mathbf{v}_i, \mathcal{g}\{FM-1(\mathbf{v}_i), FM-2(\mathbf{v}_i), \dots, FM-p(\mathbf{vi}_1)\}), \dots\\&(\mathbf{v}_c, \mathcal{g}\{FM-1(\mathbf{v}_c), FM-2(\mathbf{v}_c), \dots, FM-p(\mathbf{v}_c)\})\}\end{aligned} \tag{5}$$

and invokes an evident component of granularity as highlighted in Fig. 4. Taking these data into consideration, we construct a granular model. Preferably the model of this nature has to be structure free as much as possible. In this case, a technique of case-based reasoning (CBR), or being more specific, granular case based reasoning can be anticipated, see Fig. 5. The other option worth considering is a concept of fuzzy regression. Moving to the computational details, we compute a degree of activation of the cases by a certain input $\mathbf{x}$.

$$u_i(\mathbf{x}) = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|\mathbf{x}-\mathbf{v}_i\|}{\|\mathbf{x}-\mathbf{v}_j\|}\right)^{2/(m-1)}}, \quad m > 1 \tag{6}$$

As the outputs are evidently granular, we determine the lower and upper bound of the granular model based on the bounds of y_i^- and y_i^+ where $[y_i^-, y_i^+] = \mathcal{g}\{FM-1(\mathbf{v}_i), FM-2(\mathbf{v}_i), \dots, FM-p(\mathbf{v}_i)\}$. We obtain the following expressions for the bounds:

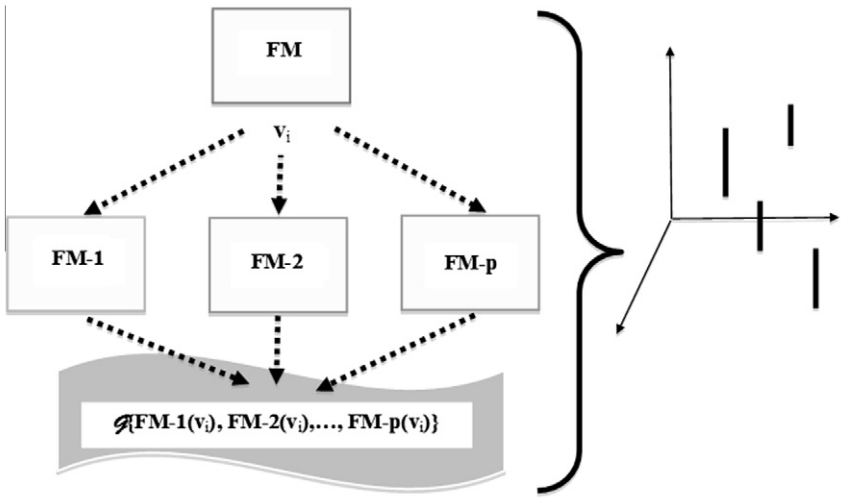

Fig. 4. Experimental evidence behind the formation of the granular model to be constructed at the higher level. Considered here is an interval format of resulting information granules.

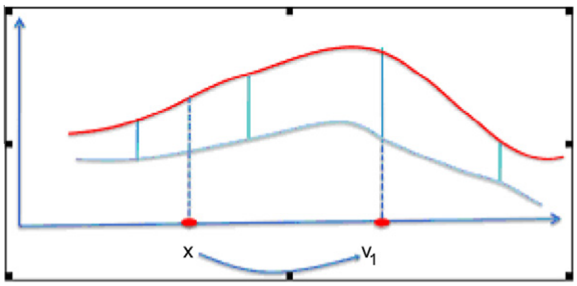

Fig. 5. The mechanism of granular case-based reasoning and a construction of lower and upper bounds.

lower bound

$$y^{-}(\mathbf{x}) = \sum_{i=1}^{c} u_i(\mathbf{x}) y_i^{-} \tag{7}$$

upper bound

$$y^{+}(\mathbf{x}) = \sum_{i=1}^{c} u_i(\mathbf{x}) y_i^{+} \tag{8}$$

An example of the granular "envelope" produced in this way is shown in Fig. 6. Note that the parameter "m" (fuzzification coefficient) affects the shape of the granular mapping (and this could be used as an additional parametric flexibility inbuilt into the granular model).

The granularity of the data themselves depends upon the value of α being used when running the process of justifiable granularity. We are faced with a two-objective optimization problem where one of the objectives is to make the envelope as narrow as possible (to achieve high specificity level, which is desirable) and at the same "cover" as much experimental evidence as possible (so a large number of data are included within the lower and upper bound making the granular model highly legitimized by the experimental evidence captured by the series of fuzzy models at the lower level of the hierarchy). Intuitively, the two requirements, which are evidently in conflict, are illustrated in Fig. 7.

The satisfaction of the two requirements and a way of achieving a certain compromise can be controlled by choosing a certain value of α. To proceed with a numeric quantification of these two criteria, we introduce two indicators that are closely reflective of the two criteria we identified above:

(a) The cumulative length of information granules (intervals) of the output of the granular fuzzy model described as

$$L = \sum_{i} L_i \tag{9}$$

which is reflective of the specificity (the level of detail) achieved in the granular model, and L_i is the length of an interval generated by the ith instance.

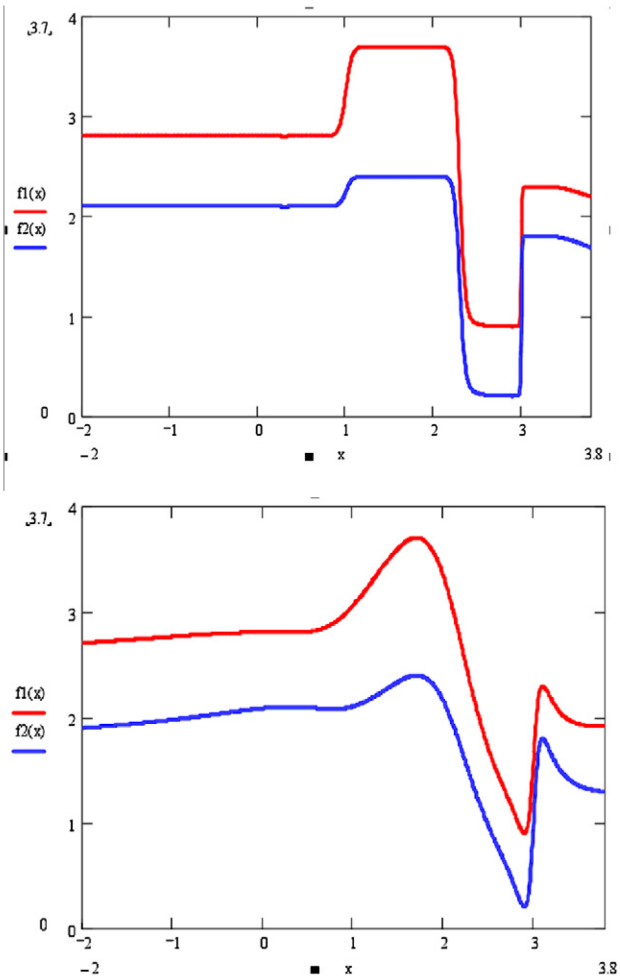

Fig. 6. Plots of the granular model for selected values of m; $m = 1.2$ (blue color), 2.0 (red color). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

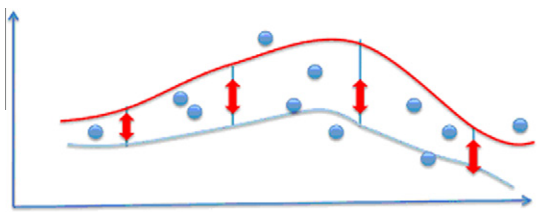

Fig. 7. Illustration of the two objectives to be optimized in the construction of the granular model.

(b) level of coverage of the data expressed as a ratio

$$I = \frac{\text{no. of data covered}}{\text{all data}} \quad (10)$$

The higher the ratio, the better the coverage of the data. Ideally, we may like to see I being very close to 1 however this might result in an unacceptable lack of specificity of the results provided by the granular fuzzy model. These two characteristics can be shown graphically. Accepting a certain value of α in the construction of the information granules through running the principle of justifiable granularity, we calculate the resulting values of I as well as the cumulative length of intervals of the granular outputs of the model.

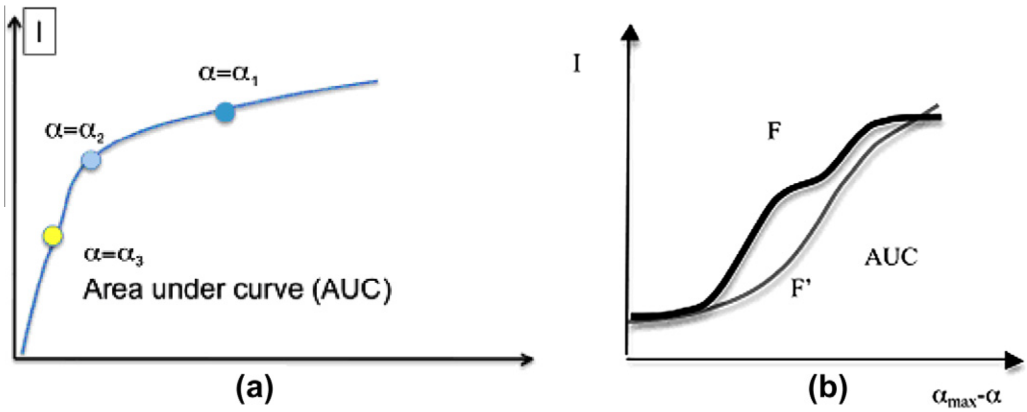

Fig. 8. (a) Plot of I versus accumulated length L for selected values of α, $\alpha_1 > \alpha_2 > \alpha_3$ and (b) Representing characteristics and quality of the granular fuzzy model through computations of the area under curve (AUC) in the $\alpha_{max} - \alpha$ – I coordinates. Note the model F is better than F' which is quantified by the values of the AUC, $AUC(F) > AUC(F')$.

An example plot in the I-L coordinates is shown in Fig. 8(a). An important characterization of the quality of the granular fuzzy model is obtained by computing an area under curve (AUC). The higher the value of the AUC, the better the obtained model is.

One can introduce a certain modification to the I-L plot by replacing the x-coordinate by the one formed as $\alpha_{max} - \alpha$, see Fig. 8 (b). This helps us compare various granular models in terms of their quality. The values of the AUC computed in the $\alpha_{max} - \alpha$ – I coordinates are comparable as we are concerned with the values of α not the lengths of the intervals. As the length of the intervals is not affected when moving to higher values of α (a significant saturation effect is present), the choice of the limit value α_{max} is not critical.

6. Experimental studies

In a series of numeric experiments, we present a design and performance of granular fuzzy models and highlight some essential features of the models along with a role played by the key design parameters of the granular constructs formed at the upper level of the hierarchy. In all experiments, the value of the fuzzification coefficient was set to 2.0 or was subject to some optimization.

6.1. Concrete compressive strength data

This data set coming from the Machine Learning repository (http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength) concerns a compressive strength of concrete with association with its characteristics such as blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate. It consists of 1,030 pairs of input–output data. We form 10 fuzzy rule-based models where each of them is constructed on a basis of 103 randomly selected data. The number of rules in each model is equal to the number of fuzzy clusters determined for each locally available data set. The performance index Q, which is used to evaluate the quality of the model is expressed in the following form

$$Q = \frac{1}{N}\sum_{k=1}^{N}(\hat{y}_k - y_k)^2 \tag{11}$$

where N is the number of data present in the individual data set. Once the information granules have been formed by using the Fuzzy C-Means (FCM) clustering algorithm, the parameters of the linear functions present in the conclusion parts of the rules were estimated by running a standard Least Square Error (LSE) method. The number of rules itself is subject to optimization and here we resort ourselves to a successive enumeration by designing the fuzzy model for increasing the number of rules while monitoring the values of the corresponding performance index (11). The values of Q treated as a function of the number of rules (clusters) for all the models are shown in Fig. 9.

One can note (as it have been expected) that when the number of cluster is increased, the values of the performance index decrease. Essentially all the local models follow the same tendency: there is a cutoff point at $c = 5$ or 6 beyond which the values of the performance index are not significantly affected by further increasing the number of rules. To visualize the performance of the fuzzy models formed here, a collection of plots of output data versus the corresponding outputs of the models is included in Fig. 10.

Once the 10 local models have been constructed, the prototypes present there are clustered at the higher level to create a backbone of the global fuzzy model. Following the principle of justifiable granularity, for each of the prototypes obtained in

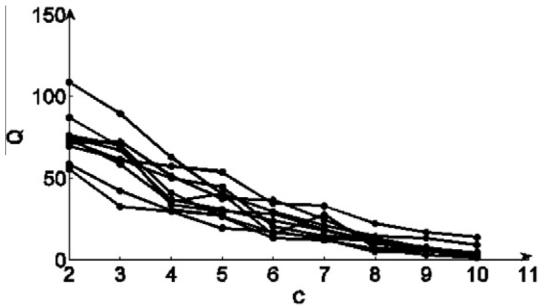

Fig. 9. The performance index of the models versus the number of rules (clusters); shown are the values for all 10 models.

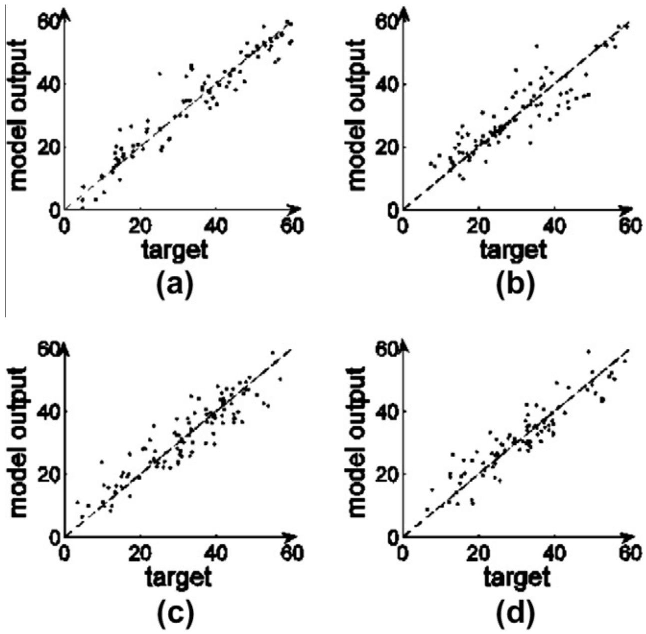

Fig. 10. The performance of selected fuzzy models (data versus model output): (a) $Q = 21.94$ (b) $Q = 31.16$ (c) $Q = 29.25$ (d) $Q = 24.01$.

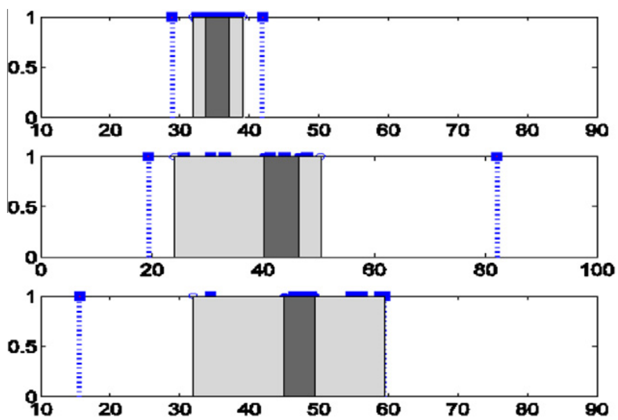

Fig. 11. Prototypes and the associated intervals for two values of α (0.01- grey and 0.76 – black), the number of clusters is set to 3, $c = 3$.

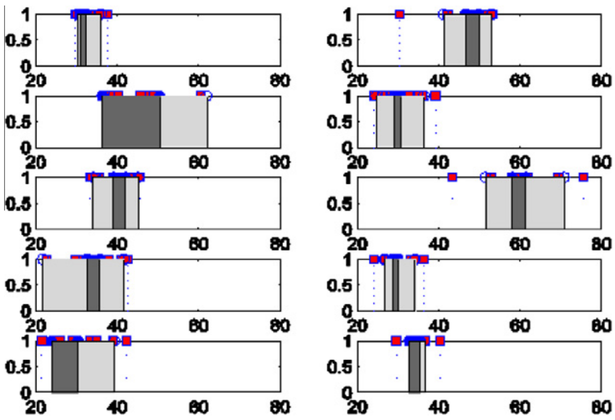

Fig. 12. Prototypes and the associated intervals for two values of α (0.01- grey and 0.71- black), the number of clusters is set to 10.

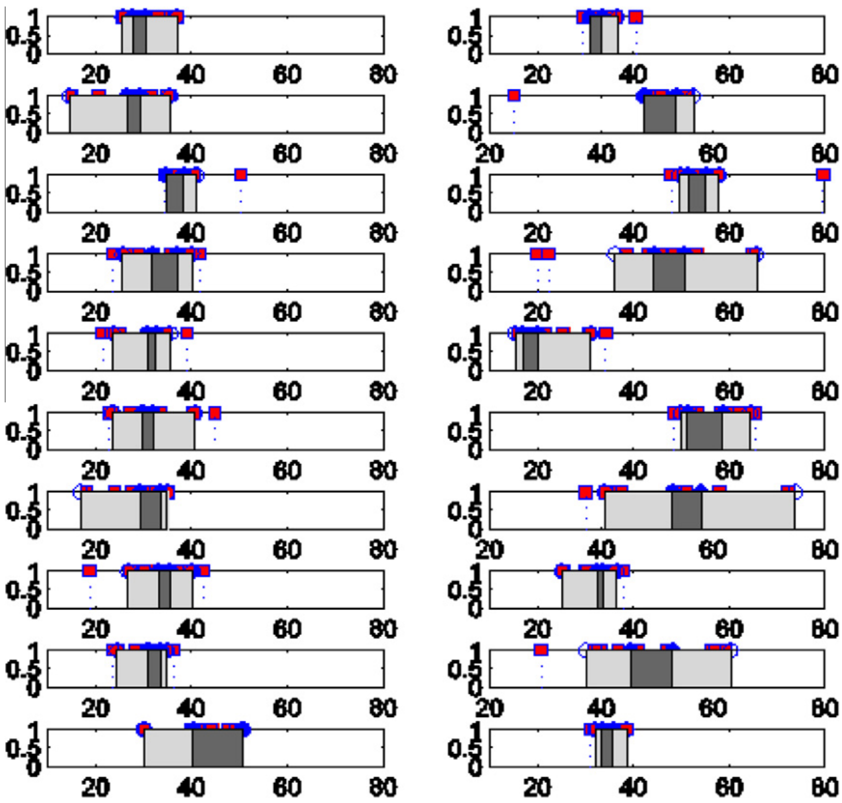

Fig. 13. Prototypes and the associated intervals for two values of α (0.01– grey and 0.54 – black), the number of clusters is set to 20.

this way we form the intervals of the corresponding output values. Several selected information granules are illustrated in Figs. 11–13 where the number of clusters built at the higher level is set to $c = 3$, 10, and 20. The values of the granularity level α are set to quite low and quite high values; this helps contrast the specificity of the resulting intervals (indicated in the figure by different colors of shading).

The most concise and informative characterization of the quality of the overall model can be conveyed by plotting the values of the coverage of data I versus the cumulative length of the intervals produced by the granular fuzzy model for the inputs formed as the prototypes of the information granules obtained for the individual fuzzy models. The resulting relationships are shown in Fig. 14.

It becomes apparent that I is a non-decreasing function of the cumulative length of the intervals. Furthermore the index I assumes higher values for increasing values of "c". The values of AUC are determined by computing the area of the plots presented in the (3-α, I) coordinates. As noted earlier, we use another x-coordinate (that is 3-α) to produce consistent results

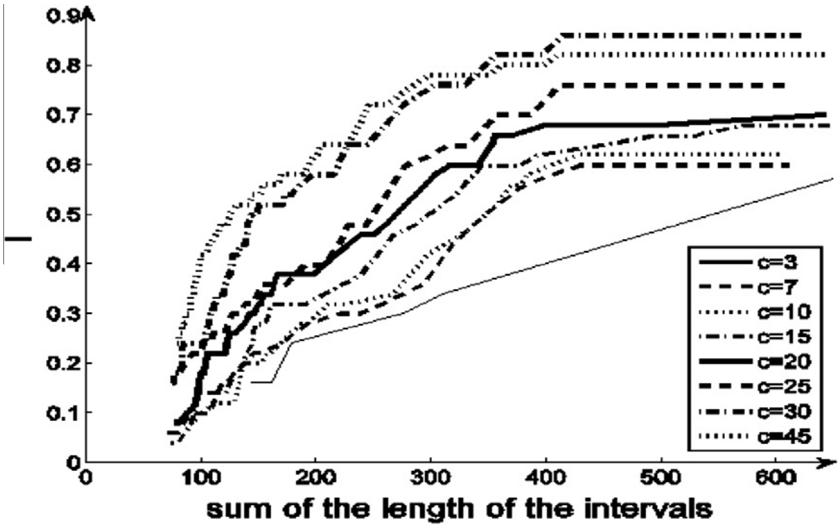

Fig. 14. Coverage of data I versus the cumulative length of intervals L of the output for selected number of clusters.

for all combinations of the parameters. This is not always possible when using the cumulative length of the intervals. For instance, we observe this effect in Fig. 14 where the curve for $c = 3$ starts at higher values of L. The choice of the new variable 3-α is motivated by the fact that for values of higher than 3 there is a saturation effect in the sense no substantial changes in the values of I are reported. The corresponding series of plots is shown in Fig. 15. We report the values of I for $m = 2$ (which is the typical value of the fuzzification coefficient in the FCM method) as well as for the optimal values of m. In all cases reported here, there is a visible difference (improvement) when using the optimized value of "m". As a matter of fact, the optimal values of the fuzzification coefficient vary in-between 1.5 and 6.0. There is also a similar relationship between the values 3-α and the coverage levels which increase quickly when α assumes low values (and the intervals become wider).

The plots of the outputs from the data versus the outputs (intervals) of the granular fuzzy model are presented in Figs. 16 and 17. Here we contrast the lengths of the intervals and their position when using a different number of clusters (c was set to 3 and 40, respectively).

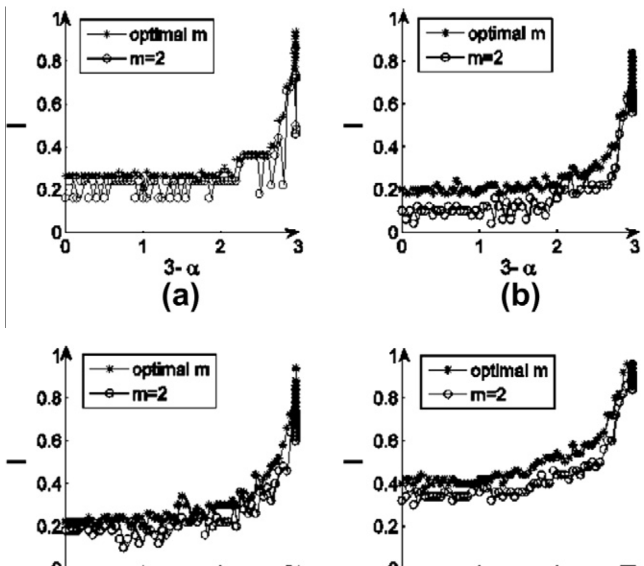

Fig. 15. I versus the level of granularity, 3-α and selected values of c: (a) $c = 3$, (b) $c = 10$, (c) $c = 20$, (d) $c = 40$.

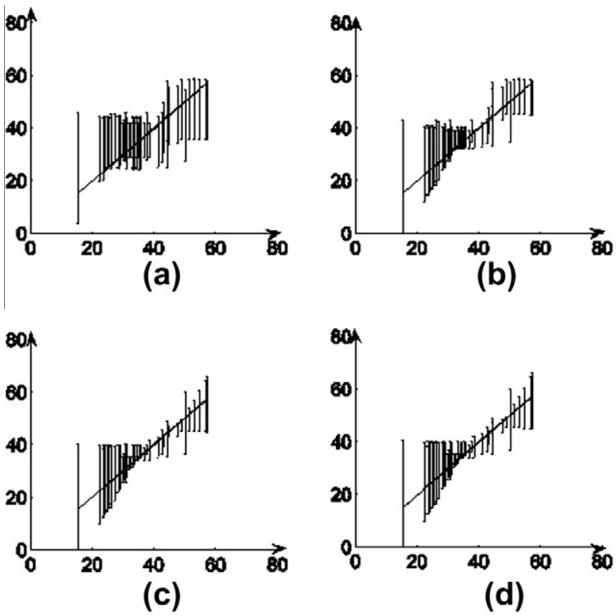

Fig. 16. Scatter plot of data versus the intervals generated by the corresponding inputs obtained for different values of alpha: (a) $\alpha = 0.01$, (b) $\alpha = 0.1$, (c) $\alpha = 1$, (d) $\alpha = 3$. The number of clusters is equal to 3.

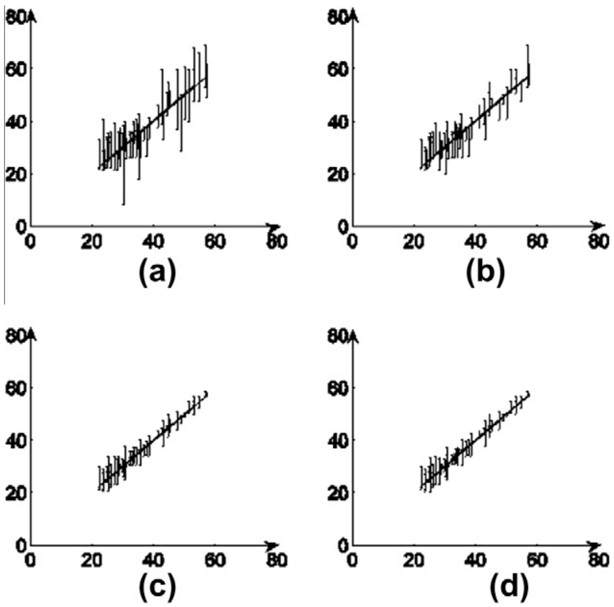

Fig. 17. Scatter plot of data versus the intervals generated by the corresponding inputs obtained for different values of α: (a) $\alpha = 0.01$, (b) $\alpha = 0.1$, (c) $\alpha = 1$, (d) $\alpha = 3$. The number of clusters is set to 40.

Table 2
Values of the AUC for selected numbers of clusters and two strategies of the determination of the prototypes of the granular fuzzy model.

Strategy of prototype formation	$c = 3$		$c = 10$		$c = 20$		$c = 40$	
	Optimal m	$m = 2$	Optimal m	$m = 2$	Optimal m	$m = 2$	Optimal m	$m = 2$
Clustering	0.9271	0.7682	0.7597	0.4902	0.9184	0.7271	1.5105	1.2774
Selection	0.7680	0.2579	0.7846	0.4809	0.8326	0.5736	0.9558	0.6450

Table 3
Values of the AUC for selected numbers of clusters and two strategies of the determination of the prototypes of the granular fuzzy model.

Strategy of prototype formation	$c=3$		$c=10$		$c=20$		$c=40$	
	Optimal m	$m=2$	Optimal m	$m=2$	Optimal m	$m=2$	Optimal m	$m=2$
Clustering	1.6741	1.3099	1.9393	1.6502	2.0171	1.6353	2.0832	1.7697
Selection	1.9867	1.3678	1.9668	1.5696	2.0300	1.7174	1.7331	1.5247

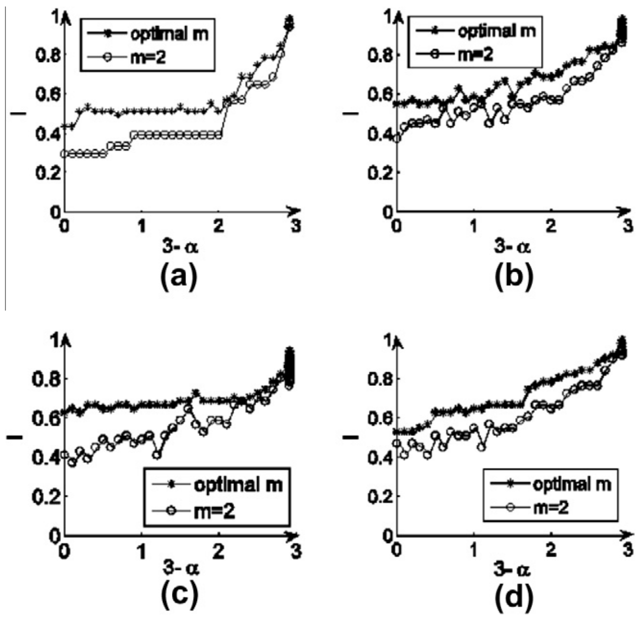

Fig. 18. I versus the level of granularity, 3-α and selected values of c: (a) $c=3$, (b) $c=10$, (c) $c=20$, (d) $c=40$. Here use is the strategy of prototype clustering.

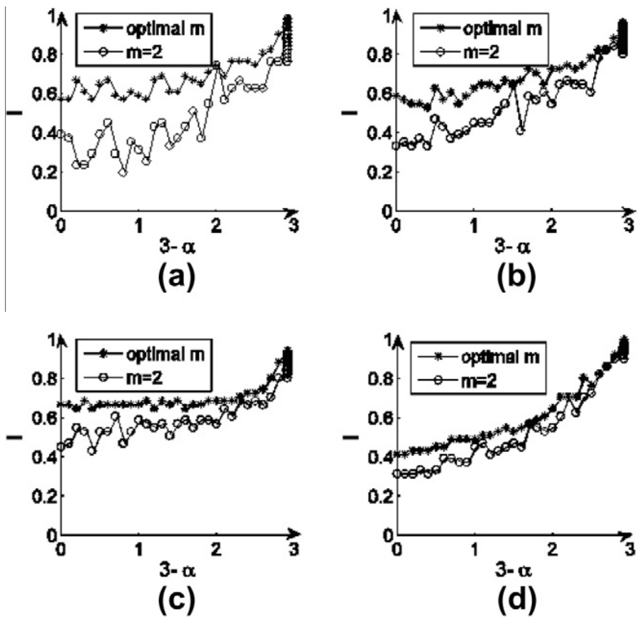

Fig. 19. I versus the level of granularity, 3-α and selected values of c: (a) $c=3$, (b) $c=10$, (c) $c=20$, (d) $c=40$.

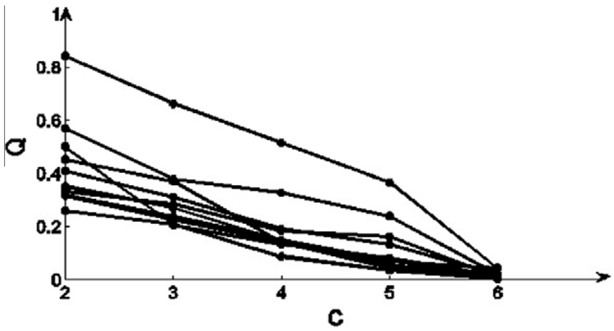

Fig. 20. The performance index Q versus the number of rules (clusters); $m = 2.0$.

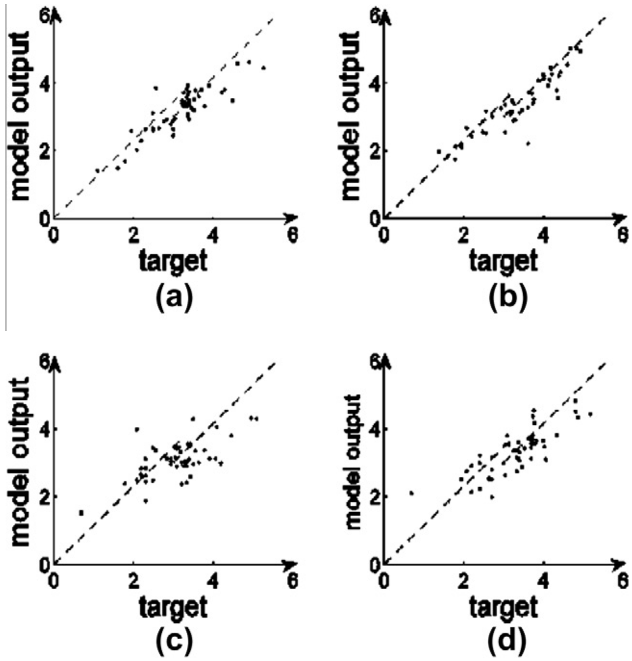

Fig. 21. The performance of four lower models (data output versus model output): (a) $Q = 0.15$ (b) $Q = 0.13$ (c) $Q = 0.14$ (d) $Q = 0.14$.

As could have been expected, the specificity of output information granules increases with the larger number of rules (clusters) used in the granular fuzzy model. Table 2 summarizes the values of the AUC for different number of "c" and the two strategies of forming information granules at the higher level, which is clustering the existing prototypes and the selection of their subsets. The fuzzification coefficient is also optimized. (see Table 3).

Two observations are well supported by the obtained results. In general, the quality of the granular model improves with the increase in the number of clusters. This becomes for both selection strategies (clustering of prototypes or their selection). The clustering of the prototypes is a better strategy leading to the higher values of the AUC. There is also a beneficial effect of the optimization of the fuzzification coefficient – in all cases we report the improvement, which is more visible when associated with the selection strategy.

When we use different number of clusters (rules) in the models formed at the lower (more specifically, 6,5,6,5,4,6,4,4,5,6), the performance of the granular fuzzy model is illustrated in Fig. 18.

In case of using the strategy of selecting prototypes (out of the prototypes available at the lower level), the obtained results are displayed in Fig. 19.

Here the optimal values of "m" are reported to be located in the range of 1.5–6.

6.2. PM10 data set

This data set comes from StatLib (http://lib.stat.cmu.edu/datasets/) and comprises 500 observations originating in a study where air pollution at a road is related to traffic volume and meteorological variables, collected by the Norwegian Public

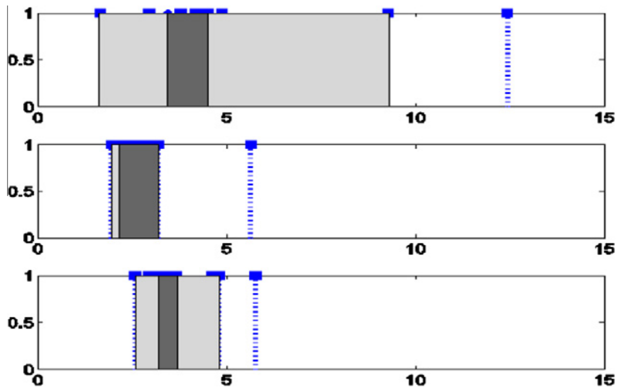

Fig. 22. Intervals of prototypes for α equal to 0.01 (grey) and 1.6 (black); $c = 3$.

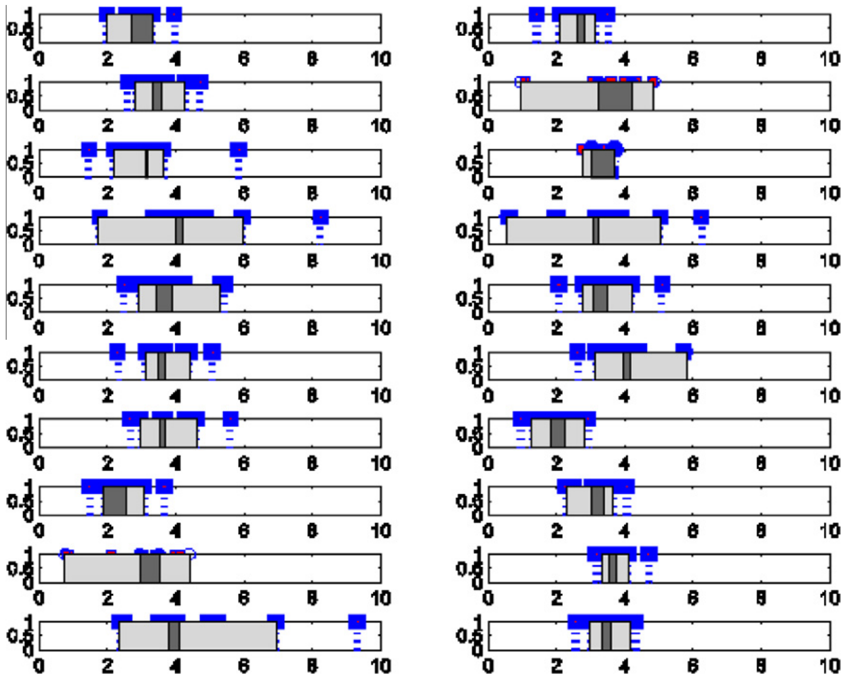

Fig. 23. Intervals of prototypes for the values of α equal to 0.01 (grey) and 2.96 (black); $c = 20$.

Roads Administration. The number of input variables is equal to 7. The overall data set is divided into 10 equal subsets, which were used to construct 10 local fuzzy models at the lower level. The performance of these models reported for selected values of "c" ranging from 2 to 6 is illustrated in Fig. 20. The number of clusters used in each local model is set to 4.

The performance of the local models is quantified by plotting the outputs of the corresponding models versus the data; the series of plots is provided in Fig. 21.

Once the models have been constructed, the prototypes coming from all of them are clustered to form the granular fuzzy model. Following the principle of justifiable granularity, for each of the prototypes obtained here we form the intervals at the output. Several selected results are illustrated in Figs. 22 and 23 for $c = 3$ and 20.

Considering a fixed value of m equal to 2, the corresponding plots of "I" versus the coordinate 6-α are shown in Fig. 24. We observe that for higher values of α the values of I are significantly reduced. Some cutoff points can be identified where the values of α are relatively high (so the bounds are quite tight) yet the granular model still captures a significant fraction of data.

The plots analogous to those reported in Fig. 25 where several prototypes were selected out of all prototypes used by the local models.

The values of the AUC obtained for all the models are summarized in Table 4.

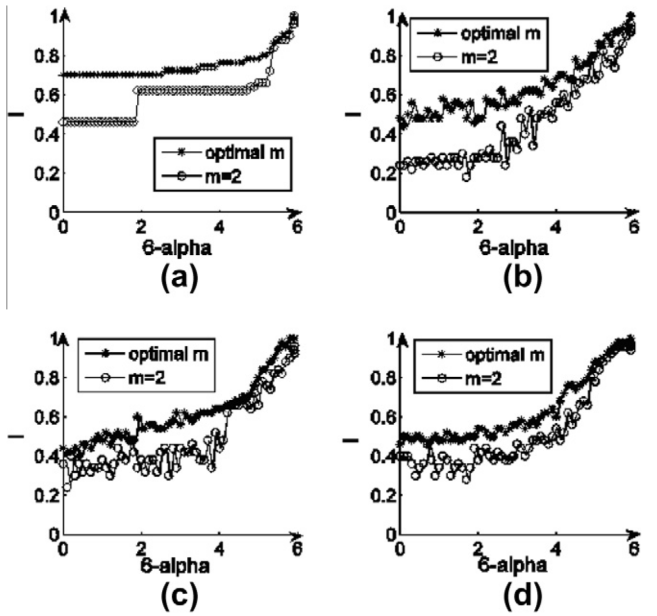

Fig. 24. Values of I versus the level of granularity, 6-α and selected values of c: (a) $c = 3$, (b) $c = 10$, (c) $c = 20$, (d) $c = 40$.

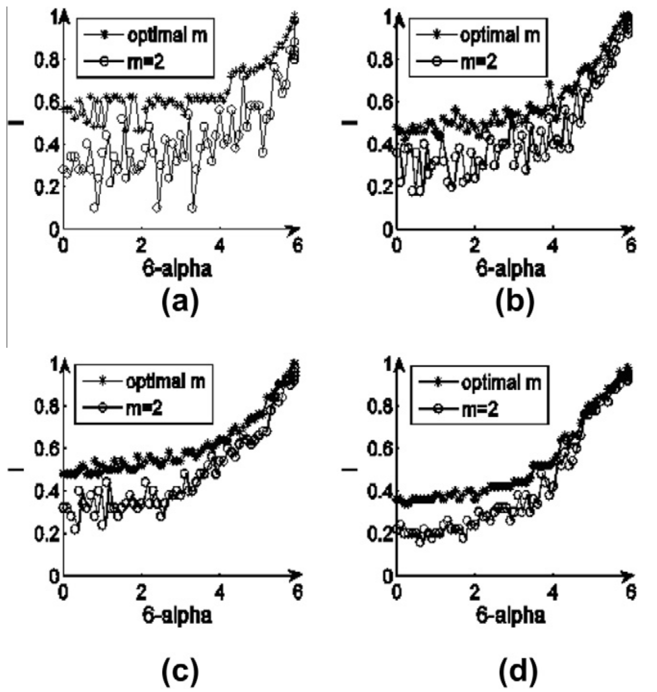

Fig. 25. Plots of I versus the level of granularity, 6-α, and selected values of c: (a) $c = 3$, (b) $c = 10$, (c) $c = 20$, (d) $c = 40$.

7. Conclusions

The methodological underpinnings of granular fuzzy models stem from the need of an effective conceptual and algorithmic representation and quantification of diversity of locally available sources of knowledge – fuzzy models. The diversity of views at the problem/system/ phenomenon is quantified via the granularity of results produced by the global model constructed at the higher level of hierarchy. The granular nature of the results formed there is inherent to the diversity of

Table 4
Values of the AUC for selected numbers of clusters and two strategies of the determination of the prototypes of the granular fuzzy model.

Strategy of prototype formation	$c = 3$		$c = 10$		$c = 20$		$c = 40$	
	Optimal m	$m = 2$	Optimal m	$m = 2$	Optimal m	$m = 2$	Optimal m	$m = 2$
Clustering	4.4040	3.5514	3.7670	2.6678	3.6630	2.9069	3.7247	3.0581
Selection	3.7916	2.4319	3.5070	2.6392	3.6470	2.8421	3.1222	2.4868

the sources of knowledge. The quantification of granularity itself (viz. the multifaceted nature of available models) is a direct result of multiobjective optimization – it is shown that the criteria of coverage of data and specificity of information are conflicting in nature. The choice of a suitable tradeoff in the satisfaction of these two requirements is left to the user. Nevertheless the AUC is helpful here as it can quantify an overall performance of the global model and rank several global models through the use of the AUC values.

Granular fuzzy models subsume the concept of type-2 fuzzy models in the sense they offer compelling, algorithmically well-supported evidence behind the emergence of fuzzy models of higher type. They are more general than type-2 fuzzy models as here we are not confined to any particular architecture of fuzzy models and a way in which type-2 fuzzy sets are incorporated into specific components of the models.

It is worth noting that the notion of granular fuzzy modeling stretches beyond fuzzy models. In essence, as we deal with models articulating locally available knowledge, the quantification of the diversity of such sources becomes encapsulated through information granules produced by granular models. In this way, we can talk about *granular* neural networks (in case of local models being formed as neural networks), *granular* regression (when dealing with individual regression models) or *granular* fuzzy cognitive maps (when the local models are fuzzy cognitive maps), etc. One can look into further generalizations in the form of granular fuzzy models of higher type; say type-2 granular fuzzy models or *granular*^{2} fuzzy models. Those are a result of dealing of several hierarchies of sources of knowledge, namely fuzzy models formed in a two-level hierarchical architecture of the knowledge sources.

Acknowledgments

Support from the Natural Sciences and Engineering of Canada (NSERC) and Canada Research Chair (CRC) Program is gratefully acknowledged.

References

[1] R. Alcala, P. Ducange, F. Herrera, B. Lazzerini, F. Marcelloni, A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems, IEEE Transactions on Fuzzy Systems 17 (2009) 1106–1122.
[2] T.H. Chang, T.C. Wang, Using the fuzzy multi-criteria decision making approach for measuring the possibility of successful knowledge management, Information Sciences 179 (4) (2009) 355–370.
[3] M. Delgado, F. Gomez-Skarmeta, F. Martin, A fuzzy clustering based rapid prototyping for fuzzy rule-based modeling, IEEE Transactions on Fuzzy Systems 5 (1997) 223–233.
[4] M. Eftekhari, S.D. Katebi, M. Karimi, A.H. Jahanmiri, Eliciting transparent fuzzy model using differential evolution, Applied Soft Computing 8 (2008) 466–476.
[5] J.L. García-Lapresta, A general class of simple majority decision rules based on linguistic opinions, Information Sciences 176 (4) (2006) 352–365.
[6] A.F. Gobi, W. Pedrycz, Fuzzy modeling through logic optimization, International Journal of Approximate Reasoning 45 (2007) 488–510.
[7] F. Herrera, L. Martínez, P.J. Sánchez, Managing non-homogeneous information in group decision making, European Journal of Operational Research 166 (1) (2005) 115–132.
[8] W.F. Hsiao, H.H. Lin, T.M. Chang, Fuzzy consensus measure on verbal opinions, Expert Systems with Applications 35 (3) (2008) 836–842.
[9] H. Ishibuchi, Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximate Reasoning 44 (2007) 4–31.
[10] J. Kacprzyk, S. Zadrozny, Computing with words in intelligent database querying: standalone and Internet-based applications, Information Sciences 134 (1–4) (2001) 71–1099.
[11] X. Liang, W. Pedrycz, Logic-based fuzzy networks: a study in system modeling with triangular norms and uninorms, Fuzzy Sets and Systems 160 (24) (2009) 3475–3502.
[12] H.S. Park, W. Pedrycz, S.K. Oh, Granular neural networks and their development through context-based clustering and adjustable dimensionality of receptive fields, IEEE Transactiona on Neural Networks 20 (10) (2009) 1604–1616.
[13] G. Pasi, R.R. Yager, Modeling the concept of majority opinion in group decision making, Information Sciences 176 (4) (2006) 390–414.
[14] W. Pedrycz, Knowledge-Based Clustering: From Data to Information Granules, J, Wiley, Hoboken, NJ, 2005.
[15] W. Pedrycz, V. Loia, S. Senatore, Fuzzy clustering with viewpoints, IEEE Transactions on Fuzzy Systems 18 (2) (2010) 274–284.
[16] W. Pedrycz, P. Rai, Collaborative clustering with the use of fuzzy C-means and its quantification, Fuzzy Sets and Systems 159 (18) (2008) 2399–2427.
[17] D.T. Pham, M. Castellani, Evolutionary learning of fuzzy models, Engineering Applications of Artificial Intelligence 19 (2006) 583–592.
[18] H. Roubos, M. Setnes, Compact and transparent fuzzy models and classifiers through iterative complexity reduction, IEEE Transactions on Fuzzy Systems 9 (2001) 516–524.
[19] L.A. Zadeh, Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (1997) 111–117.
[20] L.A. Zadeh, Toward a generalized theory of uncertainty (GTU)—an outline, Information Sciences 172 (2005) 1–40.
[21] S. Zadrozny, J. Kacprzyk, Computing with words for text processing: an approach to the text categorization, Information Sciences 176 (4) (2006) 415–437.
[22] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, IEEE Transactions on Systems Man and Cybernetics 18 (1988) 183–190.
[23] R.R. Yager, Hierarchical aggregation functions generated from belief structures, IEEE Transactions on Fuzzy Systems 8 (5) (2000) 81–490.

[24] D. Wu, J.M. Mendel, Aggregation using the linguistic weighted average and interval type-2 fuzzy sets, IEEE Transactions on Fuzzy Systems 15 (6) (2007) 1145–1161.
[25] H. Chen, P. Tiho, X. Yao, Predictive ensemble pruning by expectation propagation, IEEE Transactions on Knowledge and Data Engineering 21 (7) (2009) 999–1013.
[26] Z.-H. Zhou, Y. Yu, Ensembling local learners through multimodal perturbation, IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics 35 (4) (2005) 725–735.
[27] G. Fumera, F. Roli, A theoretical and experimental analysis of linear combiners for multiple classifier systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (6) (2005) 942–956.
[28] L.I. Kuncheva, J.J. Rodriguez, Classifier ensembles with a random linear oracle, IEEE Transactions on Knowledge and Data Engineering 19 (4) (2007) 500–508.
[29] M.M. Islam, X. Yao, S.M. Shahriar Nirjon, M.A. Islam, K. Murase, Bagging and boosting negatively correlated neural networks, IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics 38 (3) (2008) 771–784.