Granular Fuzzy Rule-Based Models: A Study in a Comprehensive Evaluation and Construction of Fuzzy Models

Xingchen Hu, Witold Pedrycz, *Fellow, IEEE*, and Xianmin Wang

***Abstract*—Fuzzy models are regarded as numeric constructs and as such are optimized and evaluated at the numeric level. In this study, we depart from this commonly accepted position and propose a granular evaluation of fuzzy models and present an augmentation of fuzzy models by forming information granules around numeric values of the parameters and constructions of the models. The concepts and algorithms of granular fuzzy models are discussed in the setting of Takagi–Sugeno rule-based architectures. We show how different protocols of forming and allocating information granules lead to the improvement of the granular performance of the models. Different from the standard numeric performance measure of fuzzy models coming in the form of the root mean squared error index, two performance measures are introduced that are pertinent to granular constructs, namely coverage and specificity. Furthermore, we propose a global indicator implied by these two measures, called an area under the curve, being computed for the characteristics of the granular model expressed in the coverage-specificity coordinates. A series of experimental studies is reported, which offers a comprehensive overview of the introduced performance measure criteria as well as the underlying realization of the granular fuzzy models.**

***Index Terms*—Allocation of information granularity, coverage and specificity performance measures, information granules, granular computing, granular fuzzy models, type-2 information granules.**

I. Introduction

SURPRISINGLY, fuzzy models [1]–[6] exploit information granules—fuzzy sets as their generic building blocks; however, the design of the models is carried out at the numeric level being guided by a performance index that is typical for numeric models. In the overall design, fuzzy models manifest as numeric constructs and as such are evaluated by performance indexes, such as the root mean squared error (RMSE) or alike. They compete with other numeric models when using the same number-driven performance indexes. Interestingly, conceptually more advanced fuzzy models, such as those exploiting type-2 or interval valued fuzzy sets still fall under the same trap. The development process is "standard" to a significant extent. One designs the model, realizes order reduction, carries out defuzzification, (better to say, decoding), arrives at numeric results and evaluates them using the RMSE measure. Type-2 fuzzy sets deliver higher flexibility, (considering a substantially larger number of parameters they bring when constructing fuzzy models); however, their design is still guided by some inherently numeric performance index. Ironically, all of excessive design effort is wasted as at the end, the model is assessed as a purely numeric construct.

The objective of this study is to propose and thoroughly investigate a novel way of an augmentation of fuzzy models by introducing a concept of their granular generalization and a way of its detailed realization. There are no ideal fuzzy models that fully capture, (coincide with) all numeric experimental data. To embrace the data, we make the fuzzy model granular in the sense that its original numeric constructions, (parameters and prototypes) are generalized to become granular parameters and prototypes. These parameters and prototypes are made granular in the sequel resulting in a granular fuzzy model. With the use of information granulation of the parameters and prototypes of fuzzy models, the resulting granular fuzzy model offers higher level tolerance to noise and modeling errors and helps produce results of practical relevance. It is evident that numeric results delivered by any model, no matter how advanced this model could be, do not coincide with the outcomes produced by real-world systems. Granular model augments the existing numeric results by generating prediction intervals so that one can expect where the real-world outcome is going to be located. The practical relevance of non-numeric prediction has been recognized and emphasized in system modeling in the past; one can refer here to the use of prediction intervals in power systems reported in [35] or allude to interval-like prediction in models of linear regression. From the methodological point of view, one can stress that granular modeling delivers a successive layer of system modeling: The approach advocated here builds upon an already constructed numeric model, and involving concepts of

Manuscript received February 5, 2016; revised June 21, 2016; accepted August 29, 2016. Date of publication September 21, 2016; date of current version October 4, 2017. This work was supported by the Canada Research Chair Program and Natural Sciences and Engineering Research Council of Canada. This work was also supported in part by the National Centre for Research under Grant UMO-2012/05/B/ST6/03068 and the National Natural Science Foundation of China under Grant 41372341. The work of Xingchen Hu was supported by the China Scholarship Council under Grant 201306110018.

X. C. Hu is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: xhu4@ualberta.ca).

W. Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada, the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia, and also with the Systems Research Institute, Polish Academy of Sciences, Warsaw 01447, Poland (e-mail: wpedrycz@ualberta.ca).

X. M. Wang is with the Hubei Subsurface Multi-scale Imaging Key Laboratory, Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China (e-mail: wangxianmin781029@hotmail.com).

Digital Object Identifier 10.1109/TFUZZ.2016.2612300

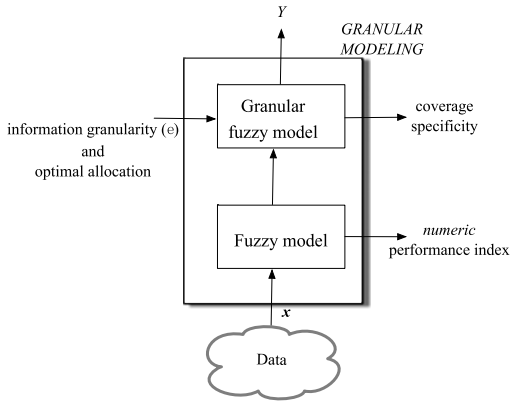

Fig. 1. From fuzzy models to granular fuzzy models: design and evaluation framework.

information granules makes it more aligned with the system under consideration. Information granularity along with its level of granularity is regarded as a certain design asset [25]–[27] facilitating the evaluation of the fuzzy model. A way in which information granules are formed is guided by two performance indexes pertinent to the evaluation of the granular model vis-a-vis numeric data, namely a coverage criterion and a specificity measure. The optimization of allocation of information granularity used in the formation of granular models is supported by the mechanism of population-based optimization. Along with the two measures introduced is a global measure of the granular quality of the fuzzy model, which is independent from a level of information granularity and expressed as an area under the curve (AUC) when assessing the performance of the fuzzy model for various levels of information granularity and determining the curve in the coverage-specificity coordinates. The essence of the study and a way in which granular fuzzy rule-based models emerge are displayed in Fig. 1.

In the study, we propose a two-phase development process of granular fuzzy models, which, in our opinion, is both legitimate and sound from the perspective of the introduced optimization criteria as well as the underlying optimization. The first phase, depicted in Fig. 1, concerns the development of a fuzzy model. Here, we resort ourselves to the spectrum of existing well-established design practices of fuzzy models. No change is being made to the design. We take full advantage of what has been fully documented in the literature and successfully used so far. The second design phase is the crux of the overall construct: here, we augmented the already constructed fuzzy model by making its parameters and prototypes granular, (represented by information granules). The optimization is guided by the aggregate criterion of coverage and specificity, (which is fully in line with a way in which the quality of information granules is quantified vis-à-vis the available numeric evidence).

There are several facets of originality of this study. By bringing the concept, performance measures, and ensuing algorithms of the granular evaluation, we embark on a new and uncharted territory of building and expressing the performance of granular fuzzy models in a holistic way. The way of an effective building granular fuzzy models realized on a basis of the existing model [7], [8] and an assessment of its performance brings another aspect of originality.

The two criteria, such as coverage and specificity are suitable descriptors of the granular output of the granular fuzzy model. The accuracy of the model is considered as the criterion in the design of the fuzzy rule-based model, refer to Fig. 1 highlighting the point of the criteria optimized at the two levels of optimization, (realized for the granular fuzzy model and the fuzzy model). The higher level of modeling completed with the granular counterpart of the model becomes optimized.

The study dwells upon the principles of granular computing [25]–[30], including an allocation of information granularity and hierarchical processing of information granules yielding information granules of higher type and in this way it can exhibit long-ranging implications in terms of the formal settings of information granules, (not necessarily fuzzy sets) [9]–[12]. In this study, we concentrate on fuzzy Takagi–Sugeno (T–S) rule-based models [1], [2], [34]. The rationale behind this choice is as follows: rule-based models of this form are commonly encountered in fuzzy modeling and come with a great deal of well-established design practices, (quite commonly engaging techniques of evolutionary optimization) and applications demonstrating their usefulness and relevance [13]–[20].

The paper is structured as follows. In Section II, we briefly elaborate on the essentials of rule-based fuzzy models. Granular fuzzy models built on a basis of fuzzy rule-based models are discussed in Section III where we cover the underlying concept, discuss on how to make the fuzzy models granular, and quantify the performance of the granular model by invoking the measures of coverage and specificity. After that, in Section IV, we elaborate on different protocols of allocation of information granularity and discuss the use of particle swarm optimization (PSO) as an optimization vehicle. Experimental studies are reported in Section V, while conclusions are drawn in Section VI. To focus the entire discussion and emphasize the underlying concept, we confine the study on information granules formalized as intervals.

II. Fuzzy Modeling Based on Clustering

In this study, we consider a T–S architecture of the fuzzy model [1], [2] of a multi-input single-output system. The rules describing this model are structured as follows:

$$\text{rule } i^{\text{th}} : \text{If } x \text{ is } B_i\,(\boldsymbol{x}) \text{ then } \tilde{y}_i \text{ is } f_i\,(\boldsymbol{x})\,, i = 1, 2, \ldots, c \quad (1)$$

where $\boldsymbol{x}$ is a vector of *n* input variables, $\tilde{y}_i$ is the output of each rule, and $B_i(\boldsymbol{x})$ is the membership function of the *i*th multivariable fuzzy set defined in the space of input variables. *c* stands for the number of the rules while $f_i(\boldsymbol{x})$ is a local function described as follows [5]:

$$f_i(\boldsymbol{x}) = w_i + \boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{v}_i) \quad (2)$$

$i = 1, 2, \ldots, c$. Along the slope parameters of the hyperplane of the linear model f_i, $\boldsymbol{v}_i$ is a center, (prototype) of the location of the rule specified in the input space $\boldsymbol{R}^n$. As emphasized in [5], this form of the linear function is easily interpretable as a hyperplane, which passes through the point $\boldsymbol{v}_i$ and assumes the corresponding value in the output space equal to w_i. When arranging all the rules together involving their condition parts,

the model reads as follows:

$$\hat{\boldsymbol{y}} = \frac{\sum_{i=1}^{c} B_i(\boldsymbol{x}) f_i(\boldsymbol{x})}{\sum_{i=1}^{c} B_i(\boldsymbol{x})}. \tag{3}$$

When the relationship $\sum_{i=1}^{c} B_i(\boldsymbol{x}) = 1$, holds then

$$\hat{\boldsymbol{y}} = \sum_{i=1}^{c} B_i(\boldsymbol{x}) \left(w_i + \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{v}_i) \right). \tag{4}$$

Let us introduce some auxiliary notation as shown below,

$$\boldsymbol{z}_i = B_i(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{v}_i) \tag{5}$$

$$\boldsymbol{q} = \sum_{i=1}^{c} B_i(\boldsymbol{x}) w_i. \tag{6}$$

Then, the above model is concisely described in the form

$$\hat{\boldsymbol{y}} = \boldsymbol{q} + \sum_{i=1}^{c} \boldsymbol{a}_i^T \boldsymbol{z}_i. \tag{7}$$

When it comes to the design of the model, there are two well-delineated development phases: 1) structure determination, and 2) estimation of parameters of the local linear models. In the design, we use a finite collection of input–output pars of data composed of N pairs in the form $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)$. The first phase of the development process is predominantly concerned with the formation of the condition parts of the model and is typically realized through fuzzy clustering and fuzzy C-means (FCM) [21]–[24], in particular. The number of clusters is equal to the number of the rules. Here, the FCM algorithm is run in the combined input–output space $\boldsymbol{R}^{n+1}$ producing the prototypes $[\boldsymbol{v}_i \; w_i]$, which position the hyperplane passing through the well-defined point in the input space $(\boldsymbol{v}_i)$ and the output space (w_i). The activation level of the ith rule by the input $\boldsymbol{x}$ is computed following the basic calculations used in the FCM algorithm, viz.

$$B_i(\boldsymbol{x}) = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|\boldsymbol{x} - \boldsymbol{v}_i\|}{\|\boldsymbol{x} - \boldsymbol{v}_j\|} \right)^{2/(m-1)}} \tag{8}$$

where $m > 1$ is a fuzzification coefficient. In the second phase, the slopes of the linear models are optimized on the basis of the provided experimental data. More specifically, let us arrange the outputs of the model and the corresponding data in a vector format, namely $\hat{\boldsymbol{y}} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N]^T$, (model) and $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^T$, (data). We define the objective function, (performance index) of the model coming as a sum of squared errors reads as follows:

$$Q = \sum_{k=1}^{N} (y_k - \hat{y}_k)^2 \tag{9}$$

where

$$\hat{y}_k = q_k + \sum_{i=1}^{c} \boldsymbol{a}_i^T z_{ki} \tag{10}$$

$$z_{ki} = B_i(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{v}_i). \tag{11}$$

We introduce the following concise notation:

$$\boldsymbol{p} = [y_1 - q_1, y_2 - q_2, \ldots, y_N - q_N]^T. \tag{12}$$

In the sequel, the parameters of the model are arranged into the cn-dimensional vector

$$\boldsymbol{a} = [a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, a_{22}, \ldots, a_{2n}, \ldots, a_{c1}, a_{c2}, \ldots, a_{cn}]^T. \tag{13}$$

Furthermore, the data are structured in the matrix format

$$\tilde{Z} = \begin{bmatrix} \boldsymbol{z}_{11} & \boldsymbol{z}_{12} & \cdots & \boldsymbol{z}_{1c} \\ \boldsymbol{z}_{21} & \boldsymbol{z}_{22} & \cdots & \boldsymbol{z}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{z}_{N1} & \boldsymbol{z}_{N2} & \cdots & \boldsymbol{z}_{Nc} \end{bmatrix}. \tag{14}$$

Then

$$\begin{aligned} Q &= \sum_{k=1}^{N} \left(y_k - q_k - \sum_{i=1}^{c} \boldsymbol{a}_i^T z_{ki} \right)^2 \\ &= \left(\boldsymbol{p} - \tilde{Z} a \right)^T \left(\boldsymbol{p} - \tilde{Z} a \right). \end{aligned} \tag{15}$$

Minimizing, (15), we estimate the parameters a by following the well-known expression:

$$\boldsymbol{a}_{\text{opt}} = \left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T \boldsymbol{p}. \tag{16}$$

Commonly, along with the above performance index, one considers the following RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y_k - \hat{y}_k)^2}. \tag{17}$$

III. Granular Fuzzy Model: An Underlying Concept and Design Aspects

In what follows, we introduce a fundamental concept of a granular fuzzy model and, then, highlight the required design pursuits along with associated optimization tasks.

A. Concept

In spite of the presence of fuzzy sets used in the development of fuzzy models, these models manifest as numeric constructs, (viz., the output of the fuzzy model is numeric). The radical design departure leading to an evident enhancement of the ensuing model is to admit granular constructions of the model. The original numeric values of the construction of the fuzzy model, (say, prototypes of the clusters $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c, w_1, w_2, \ldots, w_c$ and parameters of the local models $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_c$) are generalized to the form of information granules $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_c, W_1, W_2, \ldots, W_c, \boldsymbol{A}_1, \boldsymbol{A}_2, .., \boldsymbol{A}_c$ so that they are distributed around the original numeric values. To form granular parameters, we admit a certain level of information granularity, (ε), (the details will be discussed in the consecutive sections) and allocate it across the parameters of the model so that a certain optimization criterion becomes optimized.

Such information granules are formed around the original numeric values of the constructions. Alluding to the generic structure of the rule-based model governed by (1) instead of numeric values of the constructions, we consider information

granules located around the numeric values of the constructions. Formally, as a follow-up of the original formulas, we describe the resulting granular model as follows:

$$\text{if } \boldsymbol{x} \text{ is } G(B_i(\boldsymbol{x})) \text{ then } \hat{Y}_i = W_i \oplus \boldsymbol{A}_i^T \otimes (\boldsymbol{x} - \boldsymbol{V}_i) \quad (18)$$

the capital letters emphasize that the corresponding components are information granules and $G(B_i(\boldsymbol{x}))$ denotes an interval of levels of activation, (firing) of the rule generated on a basis of the granular prototypes, viz., $G(B_i(\boldsymbol{x})) = [B_i^-(\boldsymbol{x}), B_i^+(\boldsymbol{x})]$. The operations $\oplus$ and $\otimes$ are carried out for information granules as the calculations in interval arithmetic. Next, the output of the model comes in the form

$$\hat{Y} = \sum_{\substack{i=1 \\ \oplus}}^{c} G(B_i(\boldsymbol{x})) \otimes (W_i \oplus \boldsymbol{A}_i^T \otimes (\boldsymbol{x} - \boldsymbol{V}_i)). \quad (19)$$

To admit granular parameters and prototypes, and optimize an allocation of information granularity, two fundamental issues have to be studied in depth:

1) *A way of characterizing the quality of the granular models:* This problem is more advanced than the one encountered in, (numeric) fuzzy models. In any possible evaluation of the granular model, one has to take this into consideration a fact that a numeric data is confronted with an information granule produced by the granular model. This calls for a prudent quantification of the result.
2) *Optimization of allocation of information granularity across the fuzzy models:* How the level of information granularity (being treated as an essential design asset) can be distributed across the parameters and prototypes of the fuzzy rule-based model to yield the best performance of the granular model is the crux of the design problem.

We note that way of proceeding with a granular fuzzy model exhibits a significant level of generality. We have not committed to any particular formalism of realization of information granules. In the ensuing detailed investigations, we confine ourselves to intervals as this helps focus on the essence of the approach and avoid venturing into computational details.

B. Quantifying Performance of Granular Fuzzy Models

In what follows, we introduce two essential criteria [34] assessing the performance of the granular model. As the results are not numeric any longer, we are required to consider two complementary characterizations of information granules:

coverage: A fundamental criterion used to assess the performance of the granular model concerns coverage. In essence, coverage expresses an extent to which information granule produced by the model Y_k "covers" y_k, viz., the experimental datum is represented by the result produced by the model. Considering a collection of data, the overall coverage is expressed as the following sum:

$$cov = \frac{1}{N} \sum_{k=1}^{N} \text{incl}\,(y_k, Y_k). \quad (20)$$

Evidently, the higher the coverage, the better the model with respect of its modeling capabilities. The inclusion predicate, (incl) has to be specified depending upon the formal way in which information granule Y_k has been formalized. If Y_k comes in the form of a certain interval, the inclusion predicate, returns 1 as y_k is included in the interval $Y_k = [y_k^-, y_k^+]$ and generates 0 otherwise. In a nutshell, by using the above performance measure, one counts the number of instances of inclusion of data y_k in the granular output of the model and returns average value computed over all data. In an ideal situation, *cov* returns 1, viz., all data are "covered" by the granular model. In case of Y_k being fuzzy sets, the inclusion operation returns a membership value of y_k in Y_k, namely $Y_k(y_k)$. The coverage plays an important role; however, one has to take into consideration the quality of the granular output. It is expressed in the form of information specificity.

specificity: This measure evaluates how specific, (detailed) a certain information granule Y_k is. In general, by specificity of Y_k , $spec(Y_k)$, we view a function defined over Y_k satisfying the condition of monotonicity: if $Y_k \subset Y_k'$ then $spec(Y_k) > spec(Y_k')$, and the boundary condition $spec(\{y\}) = 1$.

When considering an interval form of Y_k, the shorter the interval, the higher its specificity becomes. In a limit case, once Y_k reduces to a single point, the specificity attains its maximal value of 1.

One among possible alternatives using which the specificity can be expressed comes in the following form:

$$spec = \frac{1}{N} \sum_{k=1}^{N} \exp\left(-|y_k^+ - y_k^-|\right). \quad (21)$$

Obviously, instead of the exponential function used above, one could consider any continuous decreasing function of the length of the interval. It is also worth stressing that the coverage and specificity are functions of ε. This fact could be stressed by using the alternative notation *cov*(ε) and *spec*(ε).

The performance of the granular model is assessed by considering the criteria of coverage and specificity. We strive to simultaneously maximize the coverage and the specificity. Apparently, these two measures are in conflict. Higher coverage values imply lower specificity values. To accommodate this request, we consider a product of these two criteria:

$$V = cov\,(\varepsilon) \cdot spec\,(\varepsilon)\,. \quad (22)$$

Interval information granules are related with some constructs known in the literature and referred therein as prediction intervals [35].

For some given ε, the optimization of (22) results in an optimal allocation of information granularity (the problem will be formulated in detail later on).

The characteristics of the obtained granular model can be displayed in the coverage and specificity coordinates. They deliver a comprehensive insight into the performance of the model and their dependence upon the changes in the values of ε. Several plots, see Fig. 2, are displayed showing various ways in which increasing values of ε impact the coverage and specificity. For $\varepsilon = 0$, the specificity is 1 while the coverage is practically equal to 0. With the increase of ε, the coverage increases but we pay a price of specificity reduction as the values of this measure are

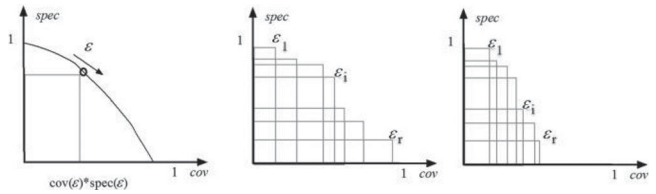

Fig. 2. Performance of the granular model expressed in the coverage—specificity coordinates: (a) monotonically decreasing values of specificity with the increase of coverage (b) significant drop in the specificity with some limited increase in coverage at $\varepsilon = \varepsilon_i$ (c) granular model characterized by low AUC value.

reduced. There could be segments of the curve where coverage still increases not impacting the specificity in a significant measure.

The optimization criterion (22) can also be modified by focusing on one criterion and requesting that another one satisfies some constraint. For instance, we may optimize coverage and the same time requesting that the specificity does not go beyond some threshold γ, which leads to the problem in the form

$$\begin{aligned} &\text{Maximize}_{\text{ allocation of information granularity}} cov(\varepsilon) \\ &\text{subject to } spec(\varepsilon) > \gamma \end{aligned} \tag{23}$$

where the maximization is expressed in a general fashion, (the details will be discussed later in one of the subsequent sections).

To develop a global measure of performance being independent from this level and produce a global characterization of the model, in the coverage-specificity coordinates, we form evaluation for the model for various values of ε and, subsequently, estimation of an AUC as following expression. Higher values of AUC indicate that for some values of ε, the granular model "covers" more data yet producing results of higher specificity implying a high quality of the model

$$\text{AUC} = \frac{1}{r}\sum_{\varepsilon}[cov(\varepsilon)spec(\varepsilon)] \tag{24}$$

where r is the number of values assumed by ε. In general, if an infinite number of the levels of information granularity is sought, the above expression is replaced by an integral over ε, namely

$$\text{AUC} = \int_0^1 [cov(\varepsilon) \cdot spec(\varepsilon)]d\varepsilon. \tag{25}$$

Refer to Fig. 2, in which the performance of the granular model, (expressed in the AUC value) in Fig. 2(c) is far lower than the one in Fig. 2(b).

Although the result—a granular fuzzy model, may, on surface, exhibit some close resemblance with type-2 fuzzy models present quite commonly in the literature, there are two important differences. First, the design promoted here exhibits two well-delineated phases whereas type-2 fuzzy models are built in a single-step process, which inevitably engages a huge search space, (and what implies a huge computing overhead and eventual inefficiency). Second, what is even more important, the evaluation of such models is carried out in a "traditional" manner and this entails the use of the mechanisms of order reduction and a conversion, (decoding) of the result into a numeric outcome so that the standard RMSE, (or any other number-oriented performance index) can be used. In other words, while the enhanced flexibility has been brought to the picture by type-2 fuzzy sets, their potential in system modeling has not been taken advantage of. Type-2 fuzzy model is constructed by being built by the numerically navigated optimization criterion, (which involves a numeric manifestation of the model). In contrast, the granular fuzzy model is constructed through the guidance offered by the two measures of performance discussed *en block*, namely coverage and specificity.

As visualized in Fig. 1, there is a striking advantage of granular fuzzy model in terms of the richness of the produced result. Fuzzy models deliver numeric result. Granular fuzzy model returns a profile, (say, an interval) of possible values of the output and in this sense offers a far more comprehensive modeling insight.

There is a visible and compelling parallel one can point to here as well. Linear regression is a numeric model and the results are numbers. If we take advantage of some statistical assumptions located behind the model, (viz., the nature of noise, linearity of the model), we endow the model with prediction intervals associated with the parameters of the model and confidence curves produced by the regression models [35], [36]. In a nutshell, the results are no longer numeric but become interval-valued. In our investigations, the line of thought is similar although no particular assumptions are required to be satisfied. Furthermore, no restriction on the type of the fuzzy model is being made. As a matter of fact, the term granular model applies equally well to any model, (say, neural networks, cognitive maps, classifiers, polynomials, etc.).

The terms prediction levels, (or prediction intervals) are inherently associated with linear regression and those are derived under some assumptions. The coverage measure used here is far more general, applies to various formalisms of information granules, and does not call for making some prior assumptions. The associated precision of the result is quantified by the specificity measure forming a part of the optimized objective function.

In addition to the granular characterization, (and optimization) of the output of the model, one is provided by an interesting alternative to assess the granularity of the optimized granular parameters of the model, thus resulting in a generalized sensitivity analysis.

C. Allocation of Information Granularity

Here, we elaborate on a way in which the level of information granularity, (ε) can be distributed, (allocated) across the fuzzy model. Here are two alternatives to be considered by introducing granularity to the parameters of the linear models of the conclusion, making prototypes granular and allocating information granularity across the condition and conclusion parts.

1) Granulation of Parameters of Local Linear Functions: Given the numeric value of the parameter of the local model, say

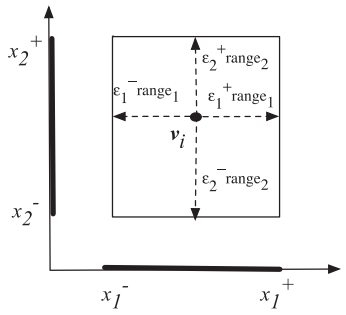

Fig. 3. Constructing a granular prototype; a two-dimensional example. range$_1$ and range$_2$ are the ranges of values assumed by the corresponding variables x_1 and x_2.

a_{ij}, $i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, n$, we make it granular by admitting a certain interval $[a_{ij}^-, a_{ij}^+]$ formed around the original numeric value. Obviously, because of the granular form of the parameters $\boldsymbol{A} = [\boldsymbol{a}^-, \boldsymbol{a}^+]$, the numeric outputs become effectively granular, (more specifically, intervals) $\boldsymbol{Y} = [\boldsymbol{y}^-, \boldsymbol{y}^+]$ whose bounds are described in the following form:

$$\left[y_k^-, y_k^+\right] = \sum_{i=1}^{c} B_i(\boldsymbol{x}) \otimes \left[w_i \oplus \left[\boldsymbol{a}_i^{-T}, \boldsymbol{a}_i^{+T}\right] \otimes (\boldsymbol{x}_k - \boldsymbol{v}_i)\right]. \tag{26}$$

2) Granulation of Prototypes: The prototype produced by the FCM algorithm is numeric. To generalize it to a granular construct, we build a hyper-rectangle prototype $\boldsymbol{V}_i$. A two-dimensional ($n = 2$) illustration is displayed in Fig. 3. In general, we admit the sides of the hypercube to be of different lengths and distributed asymmetrically around the prototypes. The resulting granular prototype is fully described by 2^n parameters, (4 in the two-dimensional case).

Given a certain data $\boldsymbol{x}$, a level of activation of the *i*th rule comes in the form of an interval. Alluding to the way in which this is done in fuzzy models, here we have to express on how to determine a distance between a numeric vector and the hyper rectangular information granule $\boldsymbol{V}_i$. Intuitively, the result should be non-numeric and we should take into consideration a range of possible values assumed by the distance involving their extreme values. An intuitively sound option would be to take the most distant and the closest vertices of $\boldsymbol{V}_i$. This, however, becomes computationally questionable when *n* attains higher values as we have to consider 2^n vertices of the hyper rectangle to determine the bounds of these distance values. An alternative is to project the hyper rectangle on the corresponding axes and determine the distance between the projected $\boldsymbol{x}$, say $\boldsymbol{x}_j$ and the most distant and the closest point of the interval resulting from the projection of the hyper rectangle on the same *j*th input variable, for details refer to Fig. 3.

We arrive at the detailed formulas

$$B_i^1(\boldsymbol{x}_k) = \begin{cases} \dfrac{1}{\sum_{j=1}^{c}\left(\dfrac{\left\|\boldsymbol{x}_k - \boldsymbol{v}_i^-\right\|}{\left\|\boldsymbol{x}_k - \boldsymbol{v}_j^-\right\|}\right)^{2/(m-1)}} & \text{if } \boldsymbol{x}_k \notin \boldsymbol{V}_i \\ 1 & \text{if } \boldsymbol{x}_k \in \boldsymbol{V}_i \end{cases} \tag{27}$$

$$B_i^2(\boldsymbol{x}_k) = \begin{cases} \dfrac{1}{\sum_{j=1}^{c}\left(\dfrac{\left\|\boldsymbol{x}_k - \boldsymbol{v}_i^+\right\|}{\left\|\boldsymbol{x}_k - \boldsymbol{v}_j^+\right\|}\right)^{2/(m-1)}} & \text{if } \boldsymbol{x}_k \notin \boldsymbol{V}_i \\ 1 & \text{if } \boldsymbol{x}_k \in \boldsymbol{V}_i \end{cases} \tag{28}$$

$$\left\|\boldsymbol{x}_k - \boldsymbol{v}_i^-\right\| = \sqrt{\sum_{j=1}^{n} \frac{\left(x_{kj} - b_{ij}^-\right)^2}{\sigma_j^2}} \tag{29}$$

$$\left\|\boldsymbol{x}_k - \boldsymbol{v}_i^+\right\| = \sqrt{\sum_{j=1}^{n} \frac{\left(x_{kj} - b_{ij}^+\right)^2}{\sigma_j^2}} \tag{30}$$

where $m > 1$, and σ_j is a standard deviation of the *j*th variable while b_{ij}^- and b_{ij}^+ are defined as follows:

$$b_{ij}^- = \begin{cases} x_j^- & \text{if } x_{kj} < x_j^- \\ x_{kj} & \text{if } x_j^- \le x_{kj} \le x_j^+ \\ x_j^+ & \text{if } x_{kj} > x_j^+ \end{cases} \tag{31}$$

$$b_{ij}^+ = \begin{cases} x_j^+ & \text{if } x_{kj} < x_j^- \\ x_{kj} & \text{if } x_j^- \le x_{kj} \le x_j^+ \\ x_j^- & \text{if } x_{kj} > x_j^+ \end{cases} \tag{32}$$

$$\boldsymbol{v}_i^- = \left[b_{i1}^-, b_{i2}^-, \ldots, b_{in}^-\right] \tag{33}$$

$$\boldsymbol{v}_i^+ = \left[b_{i1}^+, b_{i2}^+, \ldots, b_{in}^+\right]. \tag{34}$$

Then

$$B_i^-(\boldsymbol{x}_k) = \min\left(B_i^1(\boldsymbol{x}_k), B_i^2(\boldsymbol{x}_k)\right) \tag{35}$$

$$B_i^+(\boldsymbol{x}_k) = \max\left(B_i^1(\boldsymbol{x}_k), B_i^2(\boldsymbol{x}_k)\right). \tag{36}$$

The output interval reads as follows:

$$\left[y_k^-, y_k^+\right] = \sum_{i=1}^{c} \left[B_i^-(\boldsymbol{x}_k), B_i^+(\boldsymbol{x}_k)\right] \otimes \left[\left[w_i^-, w_i^+\right] \oplus \boldsymbol{a}_i^T \otimes \left[\boldsymbol{x}_k - \boldsymbol{v}_i^-, \boldsymbol{x}_k - \boldsymbol{v}_i^+\right]\right]. \tag{37}$$

3) Granulation of Parameters and Prototypes of Fuzzy Models: Based on the above two strategies, we consider to combine the granulation of both parameters and prototypes. The components of granulation, for instance, the granulating parameters, distance, and partitions calculation, are same as previous description. In contrast, the calculation of granular output is changed as following expression, which contains all granular components of the rule-based fuzzy model:

$$\left[y_k^-, y_k^+\right] = \sum_{\mathrm{i}=1}^{c} \left[B_i^-(\boldsymbol{x}_k), B_i^+(\boldsymbol{x}_k)\right] \otimes \left[\left[w_i^-, w_i^+\right] \oplus \left[\boldsymbol{a}_i^{-T}, \boldsymbol{a}_i^{+T}\right] \otimes \left[\boldsymbol{x}_k - \boldsymbol{v}_i^-, \boldsymbol{x}_k - \boldsymbol{v}_i^+\right]\right]. \tag{38}$$

In the sequel, we carefully look at the ways of allocating information granularity and the quality of the granular model.

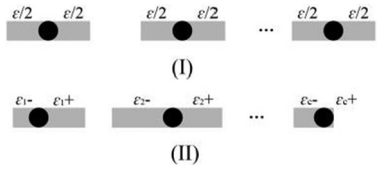

Fig. 4. Two scenarios of allocation of information granularity to parameters: (I) uniform, and (II) nonuniform.

IV. Optimal Allocation of Information Granularity: A Detailed Algorithmic Setting

Once the (numeric) rule-based model has been constructed, (and we are not particularly concerned how it has been formed), on its basis, we form a granular fuzzy model. This is accomplished by three strategies:

1) granulation the parameters of the fuzzy model, which means making the original numeric parameters $\boldsymbol{a}_i$ granular, or being more specific, interval-valued, as illustrated in Fig. 4 outlining the overall process of forming granules;
2) granulation the prototypes of the fuzzy model, which means making the prototypes $\boldsymbol{v}_i$ granular, in particular, making the single point into a hypercube, as shown a two-dimensional example forming a prototype granule in Fig. 3;
3) granulation both parameters and prototypes of the fuzzy model.

There are several essential components contributing to the entire construction:

1) a level of information granularity ε is provided in advance. It can be regarded as a supplied design asset so that the model is made granular. The values of ε are confined to the unit interval. The higher the values of ε, the more design flexibility is being offered. In the limit when $\varepsilon = 0$ the granular model becomes the original one, (numeric model);
2) a way of allocation information granularity, viz., a method of making the original numeric parameters granular;
3) a way of expressing the quality of the produced granular model.

A. Allocation of Information Granularity Protocols

For the given level of information granularity ε, we consider the two generic scenarios, (uniform and nonuniform) using which information granules are being formed in above three strategies.

1) Allocation of Granularity of Parameters

a) *Uniform and symmetric allocation of granularity*: The first scenario realizes a symmetric and uniform allocation of the level of information granularity by building the bounds of the intervals as follows:

$$a_{ij}^{-} = \begin{cases} \min\left(a_{ij}\left(1-\varepsilon_0\right), a_{ij}\left(1+\varepsilon_0\right)\right) & \text{if } a_{ij} \neq 0 \\ -\varepsilon_0 & \text{if } a_{ij} = 0 \end{cases} \tag{39}$$

$$a_{ij}^{+} = \begin{cases} \max\left(a_{ij}\left(1-\varepsilon_0\right), a_{ij}\left(1+\varepsilon_0\right)\right) & \text{if } a_{ij} \neq 0 \\ \varepsilon_0 & \text{if } a_{ij} = 0 \end{cases} \tag{40}$$

where $\varepsilon_0 = \varepsilon/2n$.

b) *Nonuniform and asymmetric allocation of granularity*: Here, we consider the length of each interval around the numeric values of the parameters is not equal:

$$a_{ij}^{-} = \begin{cases} \min\left(a_{ij}\left(1-\varepsilon_{ij}^{-}\right), a_{ij}\left(1+\varepsilon_{ij}^{+}\right)\right) & \text{if } a_{ij} \neq 0 \\ -\varepsilon_{ij}^{-} & \text{if } a_{ij} = 0 \end{cases} \tag{41}$$

$$a_{ij}^{+} = \begin{cases} \max\left(a_{ij}\left(1-\varepsilon_{ij}^{-}\right), a_{ij}\left(1+\varepsilon_{ij}^{+}\right)\right) & \text{if } a_{ij} \neq 0 \\ \varepsilon_{ij}^{+} & \text{if } a_{ij} = 0 \end{cases}. \tag{42}$$

The variables ε_{ij}^{-} and ε_{ij}^{+} are satisfied the following conditions:

$$0 \leq \varepsilon_{ij}^{-}, \varepsilon_{ij}^{+} \leq 1 \tag{43}$$

$$\sum_{i=1}^{c}\left(\sum_{j=1}^{n}\varepsilon_{ij}^{-} + \sum_{j=1}^{n}\varepsilon_{ij}^{+}\right) = \varepsilon. \tag{44}$$

2) Allocation of Granularity of Prototypes

a) *Uniform and symmetric allocation of granularity*: The first scenario realizes a symmetric and uniform allocation of the level of information granularity by building the bounds of the intervals as follows:

$$\boldsymbol{v}_i \rightarrow \boldsymbol{V}_i\left(\left[\varepsilon_0 range_i^x, \varepsilon_0 range_i^x, \ldots, \varepsilon_0 range_i^x\right]_{1\times n}\right) \tag{45}$$

where $range_i^x = x_{\max} - x_{\min}$.

$$w_i^{-} = \min\left(w_i - \varepsilon_0 range_i^y, w_i + \varepsilon_0 range_i^y\right) \tag{46}$$

$$w_i^{+} = \max\left(w_i - \varepsilon_0 range_i^y, w_i + \varepsilon_0 range_i^y\right) \tag{47}$$

where $\varepsilon_0 = \varepsilon/2c(n+1)$ and $range_i^y = y_{\max} - y_{\min}$.

b) *Nonuniform and asymmetric allocation of granularity*: Here, we consider the length of each interval around the numeric values of the parameters is not equal:

$$\begin{aligned} \boldsymbol{v}_i \rightarrow \boldsymbol{V}_i\left(\varepsilon_{i1}^{-} range_i^x, \varepsilon_{i1}^{+} range_i^x, \varepsilon_{i2}^{-} range_i^x,\right. \\ \left. \times\ \varepsilon_{i2}^{+} range_i^x, \ldots, \varepsilon_{in}^{-} range_i^x, \varepsilon_{in}^{+} range_i^x\right) \end{aligned} \tag{48}$$

$$w_i^{-} = \min\left(w_i - \varepsilon_i^{w-} range_i^y, w_i + \varepsilon_i^{w+} range_i^y\right) \tag{49}$$

$$w_i^{+} = \max\left(w_i - \varepsilon_i^{w-} range_i^y, w_i + \varepsilon_i^{w+} range_i^y\right). \tag{50}$$

The variables ε_{ij}^{v-}, ε_{ij}^{v+}, ε_{i}^{w-}, and ε_{i}^{w+} satisfy the following conditions:

$$0 \le \varepsilon_{ij}^{v-}, \varepsilon_{ij}^{v+}, \varepsilon_{i}^{w-}, \varepsilon_{i}^{w+} \le 1 \tag{51}$$

$$\sum_{i=1}^{c}\left(\sum_{j=1}^{n}\varepsilon_{ij}^{v-}+\sum_{j=1}^{n}\varepsilon_{ij}^{v+}+\varepsilon_{i}^{w-}+\varepsilon_{i}^{w+}\right)=\varepsilon. \tag{52}$$

3) Allocation of Granularity of Parameters and Prototypes

a) *Uniform and symmetric allocation of granularity*: Here the parameters and prototypes are made into granular intervals simultaneously by using the same formulas as in (39), (40), and (45), where $\varepsilon_0 = \varepsilon/2(cn + c(n + 1))$.

b) *Nonuniform and asymmetric allocation of granularity*: Here we consider the length of each interval around the numeric values of the parameters is not equal.

In this case, the information granularity is allocated around both parameters and prototypes asymmetrically by using the same formulas as in (41)–(42) and (45)–(47). The variables ε_{ij}^{-}, ε_{ij}^{+}, ε_{ij}^{v-}, ε_{ij}^{v+}, ε_{i}^{w-}, and ε_{i}^{w+} satisfy the following conditions:

$$0 \le \varepsilon_{ij}^{-}, \varepsilon_{ij}^{+}, \varepsilon_{ij}^{v-}, \varepsilon_{ij}^{v+}, \varepsilon_{i}^{w-}, \varepsilon_{i}^{w+} \le 1 \tag{53}$$

$$\sum_{i=1}^{c}\left(\sum_{j=1}^{n}\varepsilon_{ij}^{-}+\sum_{j=1}^{n}\varepsilon_{ij}^{+}+\sum_{j=1}^{n}\varepsilon_{ij}^{v-}+\sum_{j=1}^{n}\varepsilon_{ij}^{v+}+\varepsilon_{i}^{w-}+\varepsilon_{i}^{w+}\right)=\varepsilon. \tag{54}$$

In case b), given the large number of values to be determined, we can consider resorting to some evolutionary optimization techniques, such as PSO, genetic algorithm (GA), and differential evolution (DE).

B. PSO in the Optimal Allocation of Information Granularity

In the determination of the values of ε_{ij} we consider a population-based optimization technique PSO method, in this study. PSO algorithm emulates the behavior of insect or bird flocking to search the solutions by accommodating its own past experience or experience of the individuals in the flock [30]–[32], comparing with GA and DE methods, PSO has advantage of encouraging the cooperation and sharing information between the particles. Therefore, PSO comes as a sound optimization alternative. Each particle moves in the search space and its movement involves its previous direction and depends upon a location of the best particle in the population as well as the best location of the particle obtained so far. The position ***pos*** and velocity ***vel*** of the particles of each generation are updated as follows:

$$\boldsymbol{pos}_t = \boldsymbol{pos}_t + \boldsymbol{vel}_t \tag{55}$$

$$\boldsymbol{vel}_t = \lambda \cdot \boldsymbol{vel}_t + \xi_1 \cdot \boldsymbol{rand}_1(\boldsymbol{p}_{\text{best}} - \boldsymbol{pos}_t) + \xi_2 \cdot \boldsymbol{rand}_2(\boldsymbol{g}_{\text{best}} - \boldsymbol{pos}_t) \tag{56}$$

where the subscript *t* stands for index of the individual particle and λ is an inertia weight. $\boldsymbol{p}_{\text{best}}$ stands for the personal best solution found so far, whereas $\boldsymbol{g}_{\text{best}}$ is the best position reported among all particles in the population. ξ_1 and ξ_2 are constants while $\boldsymbol{rand}_1$ and $\boldsymbol{rand}_2$ are the vectors of random numbers within the range [0, 1]. The search space of each variable in the PSO is set within the range [0, 1]. Meanwhile, to keep the balance of information granularity, [described by (44), (52), and (54)], we employ the normalization condition

$$\boldsymbol{pos}_t^{*} = \frac{\boldsymbol{pos}_t \cdot \varepsilon}{\sum_{i=1}^{M} pos_{ti}} \tag{57}$$

where *M* denotes the dimensionality of the search space. AUC is used as the fitness function.

V. Experimental Studies

In this section, we present a series of experiments, including both synthetic and publicly available data, to illustrate the design process and show the performance of the resulting granular rule-based fuzzy models. First, the numeric T–S model is constructed, (following the process described in Section II), which is followed by the formation of its granular counterpart. With regard to the PSO algorithm, the number of generations and the size of the population were set as 500 and 20, respectively. These values were selected on a trial-and-error basis: it has been found that they are sufficient to achieve convergence of the method and going beyond these values does not lead to a visible improvement of the performance.

A. Synthetic Data

A two-variable nonlinear function [33] comes in the following form:

$$y = \left(1 + x_1^{-2} + x_2^{-1.5}\right)^2 \tag{58}$$

where x_1 and x_2 are two independent variables assuming values within the range [1], [5]. The data set is composed of 900, (30×30) input–output data pairs where each input is distributed according to the uniform random distribution in [1], [5]. The dataset is split into a training set (70%) and testing set (30%). The initial values of ε are used in the range [0, 1] with a step 0.1.

Fig. 5 displays the relationships between coverage and specificity obtained for different granular models when allocating information granularity. The lines with circles represent a uniform allocation of information granularity while the lines with pentagram correspond to the case where information granularity has been optimized. As expected, considering the conflicting nature of the coverage and specificity criteria, the increase in the specificity comes at the expense of the decreasing values of coverage. Table I includes AUC values corresponding to the two scenarios of allocation of information granularity. The reason that the performance of the model being associated with both the parameters of the linear functions and the prototypes is no better than the performance obtained when information granularity associates only with the condition part, (viz., granular prototypes) relates to the constraint imposed on the predetermined level of information granularity ε. Granularity allocation associated with the parameters of the linear functions forming the condition parts of the rules shares a portion of the overall

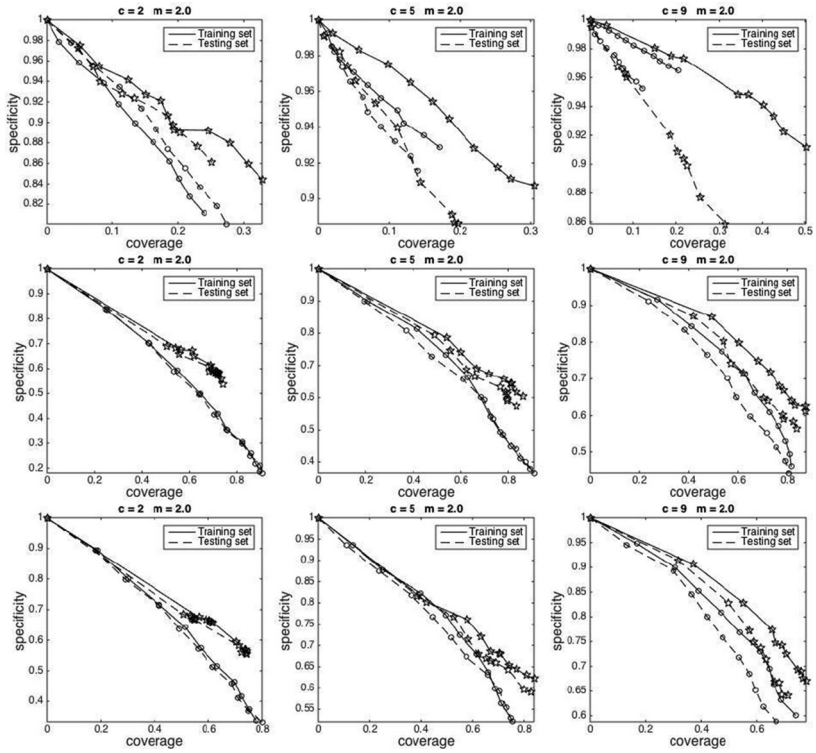

Fig. 5. Coverage and specificity obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part—first row, granulation of prototypes—second row, and granulation of parameters and prototypes—third row.

TABLE I
COMPARISON OF AUC VALUES OF GRANULAR MODELS AFTER OPTIMIZATION OF ALLOCATION OF INFORMATION GRANULARITY

Granularity of	$c = 2$		$c = 5$		$c = 9$	
	Training	Testing	Training	Testing	Training	Testing
Parameters	0.1721	0.1361	0.1587	0.0953	0.2904	0.1476
Prototypes	0.4120	0.3981	0.4927	0.4587	0.5104	0.4553
Para. and Prot.	0.4058	0.3809	0.4661	0.4355	0.4905	0.4328

level of available information granularity ε, however, it seems to exhibit less visible contribution to the improvement of the performance of the granular fuzzy model, especially when it comes to the coverage criterion.

Some general tendencies are apparent. The AUC values of the granular models with the optimized granular parameters are higher than those produced when realizing a uniform allocation of information granularity. Furthermore, one can quantify the improvement delivered by the PSO method. The obtained curves display different shapes and in some cases. For example, Fig. 5 for $c = 5$ and $c = 9$, there are relatively flat regions of the curve meaning that one can increase coverage, by increasing the value of (ε) not sacrificing much the specificity of the results. With the optimized model, some of the coverage and specificity values have improved meaning that the model leads to the better coverage with the similar values of specificity or in some cases the coverage is not enhanced too much but the specificity has some improvement. Moreover, in some cases, (for instance, when $c = 2$), after optimization, the specificity has shown a significant improvement while the coverage is reduced, and the points in the coverage-specificity coordinates are positioned very closely. A possible reason is that in those models, the allocation of information granularity has a strong impact on the specificity improvement and a far less visible effect on coverage enhancement. As a result, one improved specificity with sacrificing the coverage.

TABLE II
PUBLICLY AVAILABLE DATASETS USED IN EXPERIMENTS—A BRIEF DESCRIPTION

Name	Instances number	Inputs number	Origin of the data
Auto–MPG	392	7	UCI machine learning repository http://archive.ics.uci.edu/ml/
Boston housing	506	13	UCI machine learning repository https://archive.ics.uci.edu/ml/
Stock	950	9	StatLib repository http://www.dcc.fc.up.pt/∼ltorgo/Regression/
PM10	500	7	CMU StatLib library http://lib.stat.cmu.edu/datasets/
Red wine	·1599	11	UCI machine learning repository http://archive.ics.uci.edu/ml/
Forest fires	517	12	UCI machine learning repository https://archive.ics.uci.edu/ml/

B. Real-World Data

The following series of experiments concerns a collection of publicly available data, Table II.

Figs. 6–11 display the specificity-coverage dependencies. As before, lines with circles represent symmetric allocation of

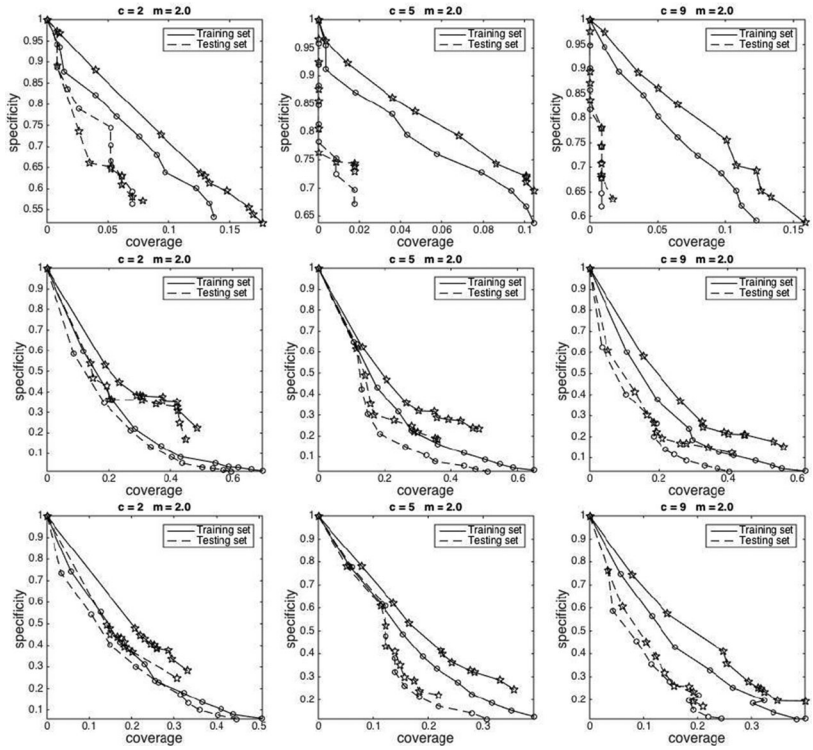

Fig. 6. Coverage and specificity for auto-MPG dataset obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part—first row, granulation of prototypes—second row, and granulation of parameters and prototypes—third row.

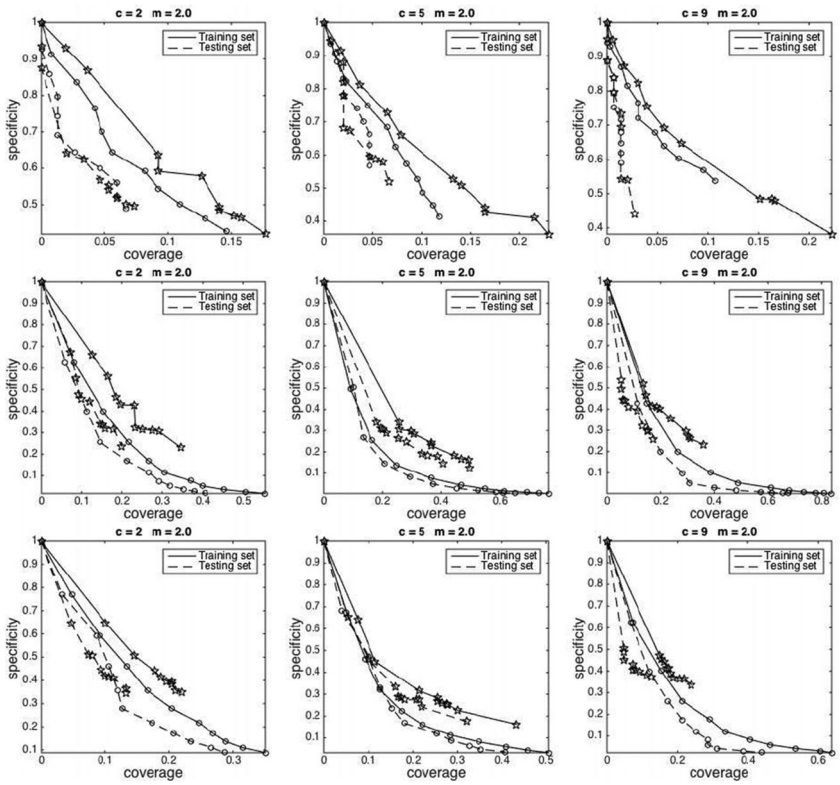

Fig. 7. Coverage and specificity for Boston Housing dataset obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part—first row, granulation of prototypes—second row, and granulation of parameters and prototypes—third row.

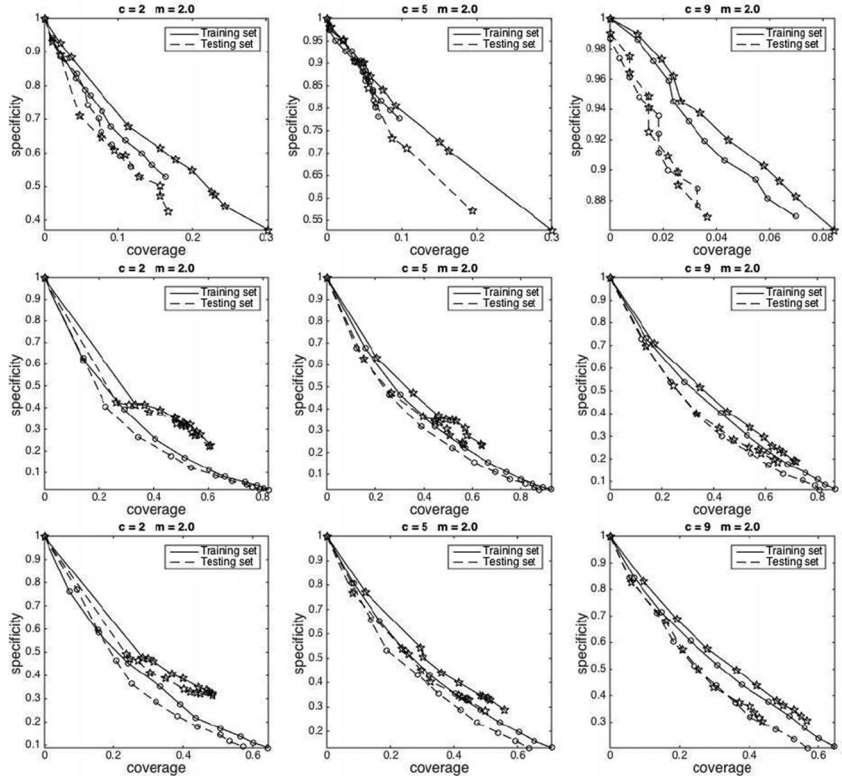

Fig. 8. Coverage and specificity for Stock dataset obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part—first row, granulation of prototypes—second row, and granulation of parameters and prototypes—third row.

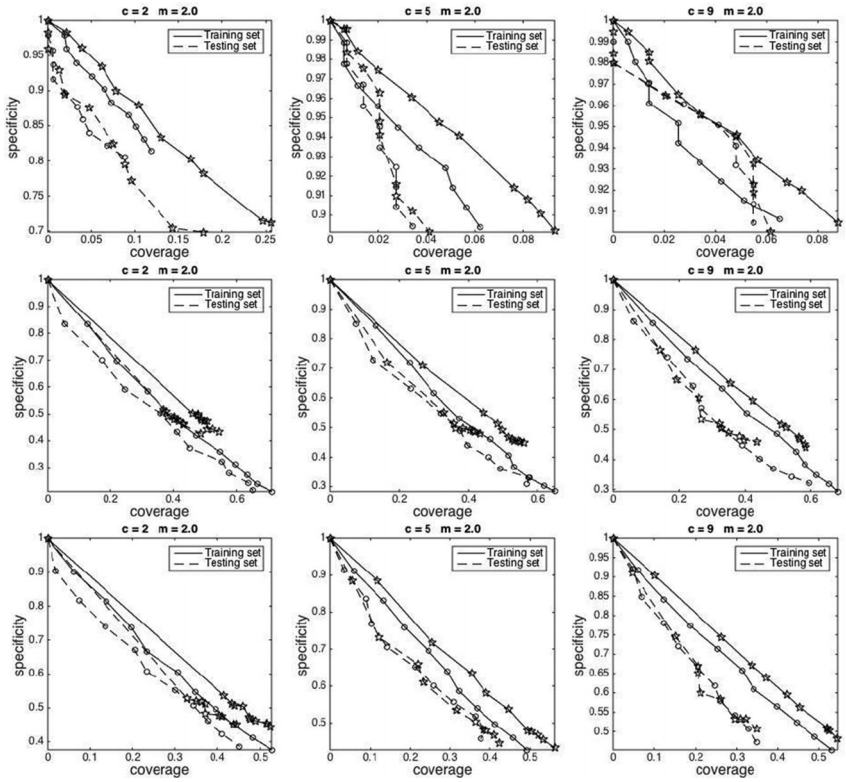

Fig. 9. Coverage and specificity for PM10 dataset obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part—first row, granulation of prototypes—second row, and granulation of parameters and prototypes—third row.

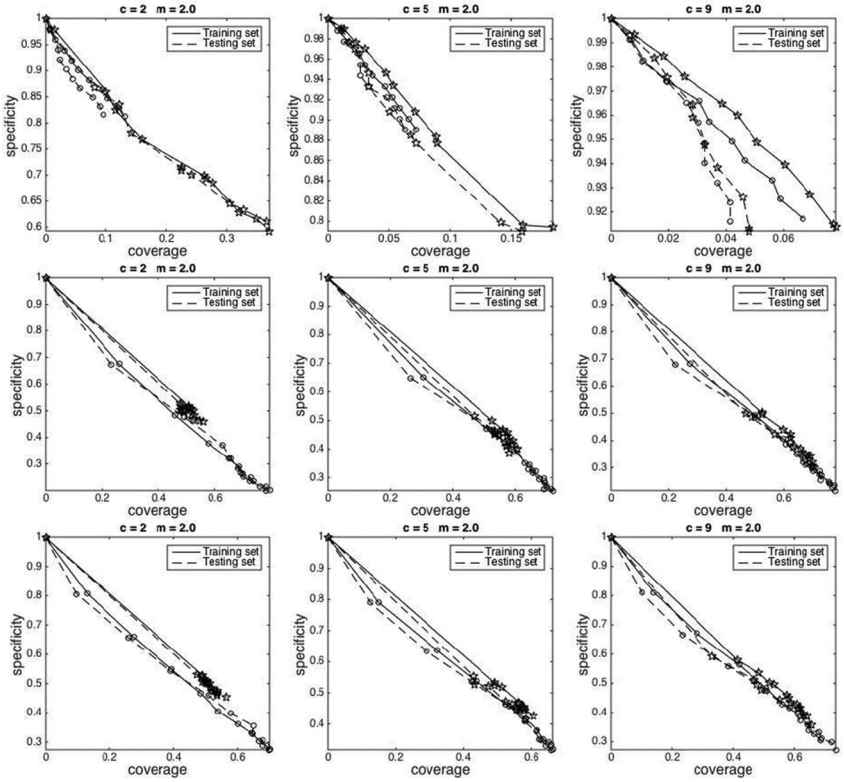

Fig. 10. Coverage and specificity for Red wine dataset obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part—first row, granulation of prototypes—second row, and granulation of parameters and prototypes—third row.

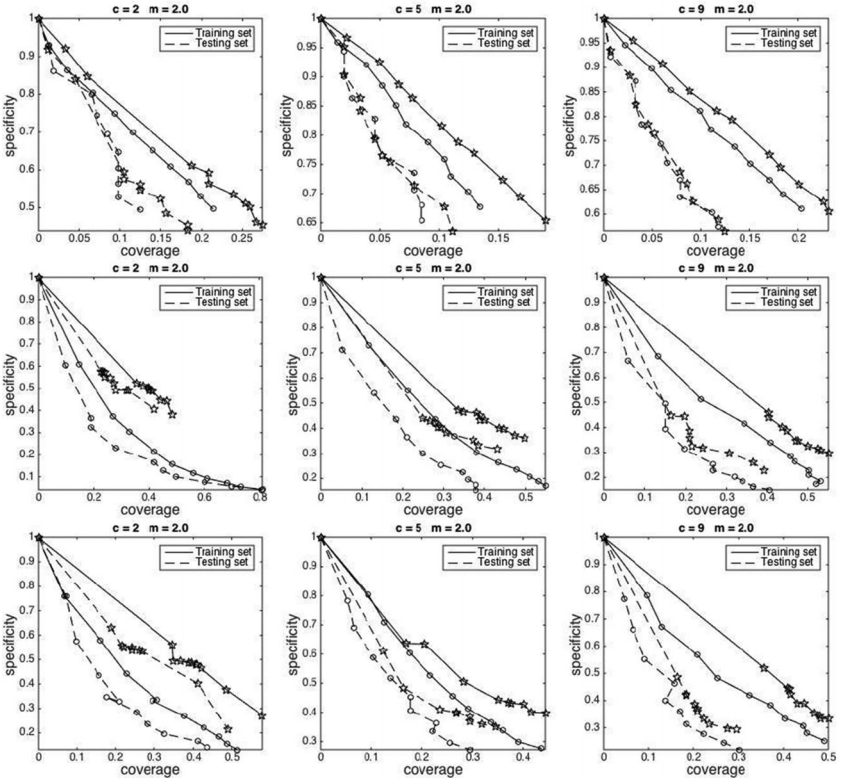

Fig. 11. Coverage and specificity for Forest fire dataset obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part—first row, granulation of prototypes—second row, granulation of parameters and prototypes—third row.

TABLE III
AUC VALUES OF GRANULAR FUZZY MODELS AFTER OPTIMIZATION OF ALLOCATION OF INFORMATION GRANULARITY

Data sets	Granularity of	$c = 2$		$c = 5$		$c = 9$	
		Training	Testing	Training	Testing	Training	Testing
Auto-MPG	Parameters	0.0719	0.0290	0.0494	0.0045	0.0639	0.0043
	Prototypes	0.1217	0.0944	0.1034	0.0642	0.0895	0.0464
	Para. and Prot.	0.1001	0.0733	0.0871	0.0529	0.0798	0.0414
Boston Housing	Parameters	0.0593	0.0220	0.0602	0.0217	0.0478	0.0067
	Prototypes	0.0873	0.0540	0.0819	0.0641	0.0780	0.0372
	Para. and Prot.	0.0759	0.0421	0.0558	0.0507	0.0729	0.0329
Stock	Parameters	0.0879	0.0529	0.0681	0.0488	0.0392	0.0153
	Prototypes	0.1589	0.1476	0.1646	0.1395	0.1589	0.1284
	Para. and Prot.	0.1507	0.1333	0.1574	0.1312	0.1639	0.1227
PM10	Parameters	0.1019	0.0506	0.0468	0.0203	0.0698	0.0305
	Prototypes	0.2340	0.1962	0.2401	0.1868	0.2519	0.1611
	Para. and Prot.	0.2278	0.1895	0.2184	0.1530	0.2312	0.1377
Red wine	Parameters	0.1513	0.1407	0.0654	0.0518	0.0442	0.0291
	Prototypes	0.2590	0.2457	0.2566	0.2405	0.2497	0.2285
	Para. and Prot.	0.2547	0.2504	0.2617	0.2457	0.2574	0.2392
Forest fire	Parameters	0.1072	0.0631	0.0817	0.0410	0.1019	0.0439
	Prototypes	0.1967	0.1451	0.1724	0.1210	0.1689	0.0836
	Para. and Prot.	0.1855	0.1304	0.1554	0.1000	0.1725	0.0783

information granularity and the line with pentagram represents optimized allocation of granularity. There is some diversity between the datasets; nevertheless the drop in the specificity values with the increase of the coverage exhibits a common monotonic relationship. The optimal allocation of information granularity gives rise to the better granular models, (expressed in terms of the reported AUC values); however, the improvement varies across the data.

Table III includes the values of AUC obtained for granular models where information granularity has been optimized. The optimization produced an improvement, (in terms of the AUC values) in the broad range 18.6%–286.2%. One observation could be made with regard to the allocation of information granularity across the condition and conclusion parts of the rules.

VI. CONCLUSION

In this study, the fundamentals of granular fuzzy models, their construction and evaluation have been established. We augmented numeric fuzzy models by forming information granules around numeric values of the parameters and prototypes of the models, and evaluated the models, (and it can be any other models, in general) in granular concept brings another more general perspective at the comprehensive evaluation of models and enriches a look at system modeling. We showed that the coverage and specificity measures serve as the two essential measures quantifying a well-rounded way of expressing the quality of the granular model. The AUC comes as a global indicator of the quality of the model.

There are several main and promising directions of further investigations. We highlight the three of them exhibiting some potential and direct practical implications:

1) *Engagement of other formalisms of granular computing*: The principles of granular fuzzy models were outlined with the use of intervals. While this was done for illustrative purposes as being conceptually the simplest and computationally feasible, the fundamentals can be used exploited by involving fuzzy sets, rough sets, and other formal settings. In some cases, in light of the representation principle, (with fuzzy sets described by a family of α-cuts) the results obtained for the interval representation of information granules can be immediately considered in coping with fuzzy sets.
2) *Designing of granular models of higher type*: Following the design process discussed in this study, the numeric parameters of the model are transformed, (generalized) into information granules of type-1. To enhance the coverage of the data, the granular parameters and constructions can be further generalized to build information granules of type-2, (say, granular intervals generalizing intervals with numeric bounds to the intervals whose bounds are information granules themselves.

REFERENCES

[1] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan./Feb. 1985.
[2] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets Syst.* vol. 28, no. 1, pp. 15–33, 1988.
[3] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. Man-Mach. Stud.*, vol. 7, no. 1, pp. 1–13, 1975.
[4] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan./Feb. 1985.
[5] W. Pedrycz and H. Izakian, "Cluster-centric fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1585–1597, Dec. 2014.
[6] W. Pedrycz and M. Reformat, "Rule-based modeling of nonlinear relationships," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 2, pp. 256–269, May 1997.
[7] M. Fazzolari *et al.*, "A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions," *IEEE Trans. Fuzzy Syst*., vol. 21, no. 1, pp. 45–65, Feb. 2013.

[8] O. Cordón, "A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems," *Int. J. Approximate Reasoning*, vol. 52, no. 6, pp. 894–913, 2011.
[9] C. F. Juang, C. M. Hsiao, and C. H. Hsu, "Hierarchical cluster-based multispecies particle-swarm optimization for fuzzy-system optimization," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 1, pp. 14–26, Feb. 2010.
[10] X.-C. Wang, X. D. Liu, and L. S. Zhang, "A rapid fuzzy rule clustering method based on granular computing," *Appl. Soft Comput.*, vol. 24, pp. 534–542, 2014.
[11] B. Hartmann *et al*., "Supervised hierarchical clustering in fuzzy model identification," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 6, pp. 1163–1176, Dec. 2011.
[12] W.-Q. Zhao, K. Li, and G. W. Irwin, "A new gradient descent approach for local learning of fuzzy neural models," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 1, pp. 30–44, Feb. 2013.
[13] M. Chadli and H. R. Karimi, "Robust observer design for unknown inputs Takagi–Sugeno models," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 1, pp. 158–164, Feb. 2013.
[14] X.-P. Xie *et al.*, "Control synthesis of discrete-time T–S fuzzy systems based on a novel non-PDC control scheme," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 1, pp. 147–157, Feb. 2013.
[15] G. Tsekouras *et al.*, "A hierarchical fuzzy-clustering approach to fuzzy modeling," *Fuzzy Sets Syst*., vol. 150, no. 2, pp. 245–266, 2005.
[16] Q. Zhang and M. Mahfouf, "A hierarchical Mamdani-type fuzzy modelling approach with new training data selection and multi-objective optimisation mechanisms: A special application for the prediction of mechanical properties of alloy steels," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2419–2443, 2011.
[17] K.-P. Lin, "A novel evolutionary kernel intuitionistic fuzzy C-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1074–1087, Oct. 2014.
[18] C. S. Li *et al.*, "T–S fuzzy model identification with a gravitational search-based hyperplane clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 305–317, Apr. 2012.
[19] A. Celikyilmaz and I. B. Turksen, "Enhanced fuzzy system models with improved fuzzy clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 3, pp. 779–794, Jun. 2008.
[20] J.-J. Wu *et al.*, "A generalization of distance functions for fuzzy-means clustering with centroids of arithmetic means," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 3, pp. 557–571, Jun. 2012.
[21] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2, pp. 191–203, 1984.
[22] S. R. Kannan, S. Ramathilagam, and P. C. Chung, "Effective fuzzy c-means clustering algorithms for data clustering problems," *Exp. Syst. Appl.*, vol. 39, no. 7, pp. 6292–6300, 2012.
[23] T. C. Havens *et al.*, "Fuzzy c-means algorithms for very large data," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 6, pp. 1130–1146, Dec. 2012.
[24] W. Pedrycz and M. L. Song, "Granular fuzzy models: A study in knowledge management in fuzzy modeling," *Int. J. Approximate Reasoning*, vol. 53, no. 7, pp. 1061–1079, 2012.
[25] W. Pedrycz, "Allocation of information granularity in optimization and decision-making models: towards building the foundations of granular computing," *Eur. J. Oper. Res.*, vol. 232, no. 1, pp. 137–145, 2014.
[26] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst*., vol. 90, no. 2, pp. 111–127, 1997.
[27] W. Pedrycz, A. Skowron, and V. Kreinovich, Eds. *Handbook of Granular Computing*. New York, NY, USA: Wiley, 2008.
[28] A. Bargiela and W. Pedrycz, "Toward a theory of granular computing for human-centered information processing," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 2, pp. 320–330, Apr. 2008.
[29] J.-T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.
[30] J. Kennedy, "Particle swarm optimization," *Encyclopedia of Machine Learning*. New York, NY, USA: Springer, 2010, pp. 760–766.
[31] J. J. Liang *et al.*, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Trans. Evol. Comput.*, vol. 10, no. 3, pp. 281–295, Jun. 2006.
[32] G.-H. Wu *et al.*, "A new granular particle swarm optimization variant for Granular optimization problems," in *Proc. Joint IEEE IFSA World Congr. NAFIPS Annu. Meeting*, 2013, pp. 24–28.
[33] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 1, pp. 7–31, Feb. 1993.
[34] W. Pedrycz and W. Homenda, "Building the fundamentals of granular computing: a principle of justifiable granularity," *Appl. Soft Comput.*, vol. 13, no. 10, pp. 4209–4218, 2013.
[35] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1496–1503, Aug. 2010.
[36] D. L. Shrestha and D. P. Solomatine, "Machine learning approaches for estimation of prediction interval for the model output," *Neural Netw*., vol. 19, pp. 225–235, 2006.

Xingchen Hu received the B.E. degree in spacecraft design and engineering from Beihang University, Beijing, China, in 2011 and the M.E. degree in control science and engineering from the National University of Defense Technology, Changsha, China, in 2013. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada.

He is working under the supervision of Prof. W. Pedrycz, and his research interests include computational intelligence, granular computing, knowledge discovery and data mining, and evolutionary optimization.

Witold Pedrycz (F'99) received the M.Sc., Ph.D., and D.Sc. degrees from the Silesian University of Technology, Gliwice, Poland.

He is a Professor and Canada Research Chair, (CRC - Computational Intelligence) with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. His main research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering. He has published numerous papers in this area. He is intensively involved in editorial activities.

Dr. Pedrycz is an Editor-in-Chief of *Information Sciences* and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—SYSTEMS. He currently serves as an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He is a Member of a number of editorial boards of other international journals.

Xianmin Wang received the B.A. and Ph.D. degrees from Wuhan University in 2001 and 2005, respectively.

She was a Postdoctoral Researcher in Huazhong University of Science and Technology from 2005 to 2007. She is currently a Professor in the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China. Her main research interests include data mining and computational intelligence, remote sensing geology, and planetary science.