From Fuzzy Rule-Based Models to Granular Models

Ye Cui, Hanyu E, Witold Pedrycz, *Life Fellow, IEEE*, Zhiwu Li, *Fellow, IEEE*, and Xianmin Wang

***Abstract*—Fuzzy rule-based models constructed in the presence of numeric data are nonlinear numeric models producing for any input some numeric output. There are no ideal models so the obtained numeric output could create a false illusion of achieved accuracy. A desirable approach is to augment the results with some measure of confidence (credibility) by admitting a granular rather than numeric format of the produced output values of the model. Our focus of this study is on fuzzy Takagi–Sugeno rule-based models whose conclusions are constant. The ultimate objective is to extend such models to the generalized granular structure with the conclusions formed as information granules. We study information granules described by intervals and fuzzy sets as well as probabilistic Gaussian information granules. The original design of the granular model is realized by involving the principle of justifiable granularity. Using this principle, we also show how to determine the equivalence between information granules. The construction of probabilistic information granules of the model is completed with the aid of optimized Gaussian process models. The granular models built in this way constitute a substantial and application-oriented departure from the numeric fuzzy models by offering a comprehensive insight into the quality of the produced results. The experimental studies based on synthetic and publicly available data demonstrate the design process and discuss the quality of the obtained results.**

***Index Terms*—Gaussian process, granular model, information granule, rule-based model, the principle of justifiable granularity.**

I. Introduction

FUZZY rule-based models are constructed on a basis of fuzzy rules. Two main categories have been studied, namely Takagi–Sugeno (TS) (functional) models and Mamdani (relational) models [1]. In the TS models, the conclusions are (numeric) local functions (f_i). The rules realize a nonlinear mapping from an n-dimensional input space to a 1-D output space, namely "if $\boldsymbol{x}$ is A_i, then y is f_i". In Mamdani models, the conclusions are fuzzy sets with the rules in the form "if $\boldsymbol{x}$ is A_i, then y is B_i". The TS models started to gain popularity due to a number of compelling reasons: 1) TS models realize numeric mappings owing to the nonlinear membership functions in the conditions and the nonlinearity of the local functions. 2) The data-driven design process is well established and well documented in the literature [2], [3], [4]. The models are designed in the presence of multivariable data. 3) The accuracy is achieved through the minimization of the loss function; the number of rules could be increased to reduce approximation error.

Relational models, although historically emerged earlier than TS models, have not gained the same level of visibility, interest, and applicability. The models are predominantly expert-designed – the rules are hand-crafted. The dimensionality of systems being modeled in this way is low. Usually, the models are built for only a few input variables. When proceeding with problems of higher dimensionality, we encounter an evident knowledge acquisition bottleneck as clearly identified in numerous studies on expert systems [5], [6], [7], [8]. The inference mechanisms are based on relational calculus and subsequently the obtained result is a fuzzy set. As the design objective of fuzzy models has been driven by the accuracy criterion, fuzzy sets are decoded to a single numeric value in the process one commonly refers to as defuzzification. Note that the (numeric) accuracy of relational models is usually lower than the TS ones [9], [10]. Surprisingly, the advantage of fuzzy models of providing outputs in the form of fuzzy sets (or information granules, in general) have not been explored. Having this in mind, our objective is to deliver granular results of TS with the granularity of the results where information granularity serves as a tangible measure expressing their confidence (credibility).

We advocate here an inherent phenomenon of type elevation of information granules. Numeric data (regarded as type-0 information granule) used to build a (numeric) model come with the usage and evaluation of their results realized in the form of information granules of higher type, at least type-1. This phenomenon could be referred to in the discipline of Granular computing as an elevation principle of type of information granularity [11], [12], [13]. For instance, in this way we talk about type-1 and type-2 fuzzy sets. Type-1 fuzzy sets define varying degrees of importance of information depending on the numerical location of the data, while type-2 fuzzy sets

Received 6 April 2024; revised 27 July 2024, 2 September 2024, and 10 October 2024; accepted 16 October 2024. Date of publication 30 October 2024; date of current version 5 February 2025. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2022-03045, in part by Alberta Innovates under Grant 232403216, in part by Canada First Research Excellence Fund under Grant FES-T11-P01, in part by the National Key R&D Program of China under Grant 2018YFB1700104, and in part by the Science Technology Development Fund, MSAR, under Grant 0029/2023/RIA1. Recommended by Associate Editor P. Sussner. *(Ye Cui and Hanyu E contributed equally to this work.) (Corresponding author: Hanyu E.)*

Ye Cui and Hanyu E are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada (e-mail: ycui7@ualberta.ca; hanyu6@ualberta.ca).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada, also with the Systems Research Institute, Polish Academy of Sciences, 00-901 Warsaw, Poland, and also with the Research Center of Performance and Productivity Analysis, Istinye University, 34396 Istanbul, Türkiye (e-mail: wpedrycz@ualberta.ca).

Zhiwu Li is with the Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China, and also with the School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China (e-mail: systemscontrol@gmail.com).

Xianmin Wang is with the Hubei Subsurface Multiscale 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.2024.3483817

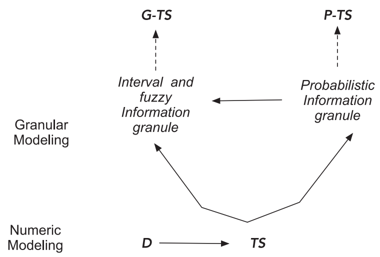

Fig. 1. From numeric to granular modeling: the design of granular rule-based models.

introduce a defined level of granularity (uncertainty). This specification of uncertainty enables the formation of higher-type information granules [14], [15], [16].

Capitalizing on the well-established methodology of TS models, we develop granular augmentation of the numeric model producing results in the form of information granules. The resulting constructs emerge as granular rule-based models. This methodology offers an original and far-reaching avenue of fuzzy modeling making the rule-based models more in rapport with application environments and becomes instrumental in a prudent deployment of fuzzy models.

An overall design framework of the study is portrayed in Fig. 1. The figure underlines two layers of design. Starting with a well-established numeric TS model, two alternative information granularity elevation strategies leading to Granular modeling are developed by building granular (interval or fuzzy set based) TS, G-TS and probabilistic TS, P-TS by involving Gaussian process (GP) mechanism. Furthermore, as visualized in this figure, the granular equivalence between types of information granules is established.

The main objectives of the study are outlined as follows.

1) We deliver a critical analysis of design practices of rule-based models vis-à-vis the character of their numeric results and an associated lack of quantification of their quality/confidence.
2) We form a proposal of forming ways of elevation of type of information granularity of results in rule-based models by involving interval, fuzzy and probabilistic information granules.
3) We explore granular equivalence between information granules through the use of the principle of justifiable granularity encountered as a part of the methodological framework of Granular computing.

These objectives exhibit some aspects of originality by bringing novel ways of emphasizing and constructing granular rule-based models.

1) As for the methodology and design methods, these have not been raised and studied systematically. Some preliminary studies have been carried out in [17] and [18], however, they do not tackle the spectrum of design practices established in this article.
2) Additionally, another novel aspect of our work is the exploration of methods for transforming between different types of information granules.

The exposure of the material is structured in a top-down manner. The rest of this article is organized as follows. In Section II, the literature review for the improvement of the regression model is depicted. In Section III, the basic knowledge is introduced, namely a design of TS rule-based model, construction of information granules by the principle of justifiable granularity and GP method. Section IV discusses the generation of granular models, granular TS model, and probabilistic TS model. The experimental results covered in Section V. Finally, Section VI concludes this article.

In the study, we adhere to the standard notation. Vectors are shown in boldface and information granules (intervals, fuzzy sets and probabilistic information granules) are shown in capital letters. The data come in the form of input–output pairs ($\boldsymbol{x}_k$, target$_k$), $k = 12, \ldots, N$, where $\boldsymbol{x}_k$ is defined in the n-dimensional input space $\boldsymbol{R}^n$.

II. Literature Review

The landscape of regression modeling has been undergoing a significant transformation with the integration of granular computing concepts, reflecting a shift from purely numeric predictions to models that incorporate measures of uncertainty and granularity. This brief and focused review delves into the progression of traditional regression models, such as neural networks (NN) and linear models, towards incorporating information granules like intervals. NN are combined with prediction intervals (PI) and confidence intervals (CI) through several methods to improve the reliability of their predictions [19], [20]. Khosravi et al. [21] proposed some combination techniques, such as delta and Bayesian. Delta method involves calculating the gradient and Jacobian matrices, the first derivative of output with respect to its input, which estimate the total variance and form the PI around the numeric predictions. The Bayesian method integrates Bayesian inference into the NN training process, treating the model parameters as random variables with prior distributions. As data are observed, update the prior distributions to posterior distributions using Bayes' theorem. Then draw samples from the posterior distributions of the NN parameters. These samples represent different plausible parameter configurations given the observed data. In this case, we can use the sampled parameters to generate multiple predictions for each input, then the prediction intervals can be constructed based on the distribution of these multiple predictions. In addition, building on the concepts of NN and granularity, Granular neural networks (GNNs) represent a fusion of neural computation with granular computing [22], [23], [24], [25]. GNN adapts the learning mechanisms of neural networks to work with granulated data or generates the granular weights instead of the numeric weights to produce the interval-valued outputs. In linear regression models, the concepts of confidence and prediction intervals are widely used and can be easily determined with the formulas available in the published studies [26]. Cui et al. [17] proposed a method to augment the rule-based model to produce the prediction intervals. Based on the derived numeric results, the prediction intervals are generated with the formula in [27]. Sanz et al. [28] developed a compact evolutionary interval-valued fuzzy rule-based classification system specifically for real-world financial applications. In this system, interval-valued fuzzy sets (IVFSs) were constructed. This extended the classical fuzzy reasoning method to work with intervals instead of numeric values throughout the inference process. Sanz et al. [29] also outlined the use of an interval-valued fuzzy rule-based classification system specifically designed for medical diagnosis of cardiovascular diseases. The methodology integrated fuzzy rule-based classification systems with interval-valued fuzzy sets in three steps: The linguistic labels of the classifier were modeled using IVFSs; during the inference process, the K_α operator was employed to handle the interval data more effectively; using genetic tuning to adjust the degree of ignorance that each IVFS represents and K_α parameters. In summary, while previous studies have achieved some enhancements in regression models, these improvements have often focused on interval information granules. It becomes also beneficial to explore other types of information granules.

III. Rule-Based Models and Their Developments

The generic TS rule-based model [30] dwells upon a collection of rules

$$-\text{if } \boldsymbol{x} \text{ is } A_i(\boldsymbol{x}), \text{ then } y_i = b_i \tag{1}$$

where $A_i(\boldsymbol{x})$s, $i = 12,\ldots,c$, are fuzzy sets defined in $\boldsymbol{R}^n$ and the conclusions of the rules are constants (b_i). In general, the conclusions could be considered as linear functions or polynomials. For any input $\boldsymbol{x}_k$, the resulting output (conclusion) is determined as a linear combination of local constant functions weighted by the corresponding activation levels of the rules $A_i(\boldsymbol{x}_k)$, namely $\hat{y}_k = \sum_{i=1}^{c} A_i(\boldsymbol{x}_k)\, b_i$.

The construction of rule-based model is well-documented and supported by numerous detailed analyses [31], [32], [33]. The design is completed on a basis of a collection of input–output pairs of data ($\boldsymbol{x}_k$, target$_k$). The fuzzy sets $A_i(\boldsymbol{x})$ standing in the condition parts of the rules are developed by clustering data $\boldsymbol{x}_k$ located in the input space. Commonly, fuzzy C-means is used, which leads to a collection of prototypes (v_i) and subsequently the fuzzy sets A_i, $i = 12,\ldots,c$. The conclusion parts are optimized by minimizing a sum of squared errors regarded as a loss function L

$$L = \frac{1}{N} \sum_{k=1}^{N} \left(\hat{y}_k - \text{target}_k\right)^2. \tag{2}$$

The analytical solution to the above estimation problem, namely a vector of constant conclusions $\boldsymbol{b} = [b_1, b_2, \ldots, b_c]^T$ is obtained in the following way:

$$\boldsymbol{b} = \left(A^T A\right)^{-1} A^T \textbf{target} \tag{3}$$

where

$$\mathrm{A} = \begin{bmatrix} \mathrm{A}_1(\mathbf{x}_1) & \cdots & \mathrm{A}_c(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \mathrm{A}_1(\mathbf{x}_\mathrm{N}) & \cdots & \mathrm{A}_c(\mathbf{x}_\mathrm{N}) \end{bmatrix}, \quad \textbf{target} = \begin{bmatrix} \text{target}_1 \\ \vdots \\ \text{target}_\mathrm{N} \end{bmatrix}.$$

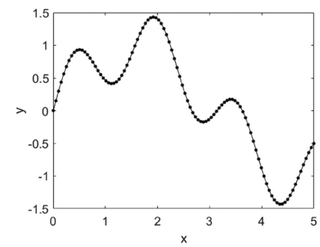

Fig. 2. Nonlinear function along with uniformly distributed data (data shown by dots).

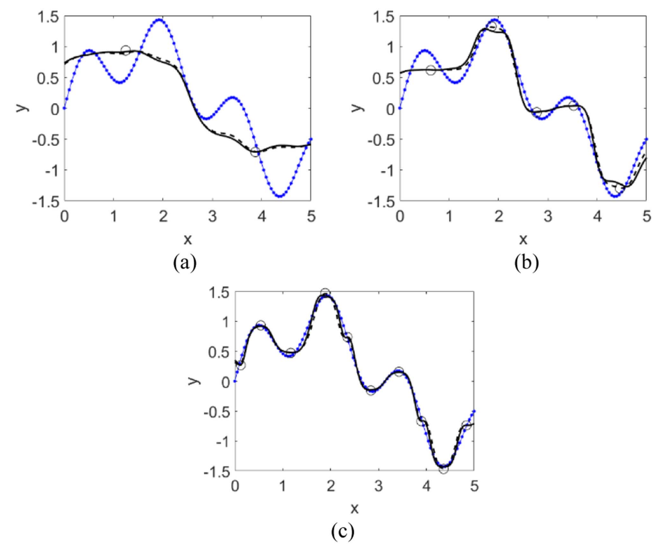

Fig. 3. Resulting rule-based models. (a) $c = 2$. (b) $c = 5$. (c) $c = 10$.

As an illustrative example being used throughout the study, we consider a 1-D nonlinear function described as

$$f\ (x) = \sin(x)\ + 0.5\, \sin(4x) \tag{4}$$

defined over the input space [0, 5]. The plot of the function is displayed in Fig. 2.

The training data consists of 400 input–output pairs, $D = \{(x_k,\ \text{target}_k)\}$, $k = 12,\ldots,400$. Fig. 3 shows the location of the obtained pairs $(v_i,\ b_i)$, $i = 12,\ldots,c$, with $c = 2, 5, 10$ clusters. The prediction results for both the training and testing datasets are depicted as solid and dashed lines, respectively. The values of the root-mean-squared error (RMSE), i.e., $\sqrt{L}$, computed over the testing data are 0.41, 0.16, and 0.09, respectively. As visualized, higher numbers of rules (5 or 10) produce a good fit to the original function.

A. Construction of Information Granules

We discuss two key ways of building information granules: 1) generating type-2 fuzzy sets (intervals and fuzzy sets) by engaging the principle of justifiable granularity. 2) Forming probabilistic information granules produced by the GP model.

1) Principle of Justifiable Granularity: The principle of justifiable granularity [34] refers to the way of representing a collection of numeric data by a single information granule (for instance, an interval, or a fuzzy set) by assuring that the constructed information granule is supported by the data and at the same time it is specific enough (viz. carries a well-defined semantics). In other words, information granules deliver a certain level of abstraction at which the essence of numerous data is represented as an interpretable piece of knowledge. For instance, by eyeballing a list of temperature readings over the last month, the information granule comes as *moderate* temperature described as a certain fuzzy set. This fuzzy set conveys a meaningful and actionable piece of knowledge. In what follows, we outline a process of building information granules.

We have a set weighted data x_k, $k = 12, \ldots, N$, along with their corresponding weights ω_k, assuming values on [0, 1]. The optimization procedure behind the construction of the information granule *T* consists of two main phases. First, a numeric representative of data is formed; it could be a mean, weighted mean or modal value, among other options. Next the parameters of the information granule for the increasing and decreasing parts of the granule are determined. For the interval $T = [lb, ub]$, we separately determine the lower and upper bound. For *T* being a fuzzy set with some type of membership function specified in advance, the parameters of the increasing and decreasing portions of membership function are estimated. Following the above process, we determine an optimal interval granule distributed around the mean *m* of the available data.

In the optimization process, the two essential characteristics of information granules, namely coverage and specificity, are considered.

1) Coverage (*cov*) describes the extent to which *T* is supported by the data.
2) Specificity (*sp*) states how specific the constructed granule is.

The detailed calculations proceed as follows.

For the lower bound of the interval granule, we determine the coverage and specificity as follows:

$$cov = \frac{1}{N} \sum_{k:\ x_k \in [lb,\ m]} \omega_k$$
$$sp = \frac{1}{N} \sum_{k=1}^{N} \max\left(0, \left(1 - \frac{|m - lb|}{\text{range}}\right)\right) \tag{5}$$

where range is the distance between minimal value of *x* in the data and mean value *m*. (Instead of the mean value, one can consider some other numeric representative of data such as median.)

For the upper bound, we compute coverage and specificity as follows:

$$cov = \frac{1}{N} \sum_{k:\ x_k \in [m,\ ub]} \omega_k$$
$$sp = \frac{1}{N} \sum_{k=1}^{N} \max\left(0, \left(1 - \frac{|\text{ub} - \text{m}|}{\text{range}}\right)\right) \tag{6}$$

here range is the distance between maximal value of *x* in data and mean value *m*.

From the abovementioned formulas, it is visible that the higher value of coverage comes with lower values of specificity and vice versa. Given this inherent conflicting nature of coverage and specificity, our objective is to produce the highest values of these two measures; we aim at the maximization of their product, $V = cov{*}sp$. This optimization is carried out for the lower bound and the upper bound independently. The optimization of the bounds *lb* and *ub* is achieved by maximizing the *V*.

Consider triangular fuzzy set *T*(*x*). It is characterized by three key parameters: the lower bound (*lb*), the upper bound (*ub*), and the central point whose value is set as *m*.

In this case, the coverage of the lower bound is described as

$$cov = \frac{1}{N} \sum_{k:\ x_k \in [lb,\ m]} \min(T(x_k), \omega_k). \tag{7}$$

However, it should be noted that the computation of specificity for the triangular fuzzy set concerns a family of α-cuts where $\alpha \in [01]$. Using lower bound as the example (viz. focusing on the increasing part of the membership function), for a fixed value of α, the interval granule is $[lb+(m-lb)\alpha, m]$, the corresponding specificity is $sp_\alpha = 1 - \frac{|m-(lb+(m-lb)\alpha)|}{range}$. Thus, the specificity is computed as the integral taken over α-cuts, namely

$$sp = \int_0^1 sp_\alpha \, d\alpha \tag{8}$$

where range is distance between *m* and minimal value in data set.

For computing the upper bound, the detailed formulas of coverage and specificity read as follows:

$$cov = \frac{1}{N} \sum_{k:\ x_k \in [m,\ ub]} \min(\mathrm{T}(x_k), \omega_k)$$
$$sp = \int_0^1 sp_\alpha \, d\alpha \tag{9}$$

where $sp_\alpha = 1 - \frac{|(ub-(ub-m)\alpha)-m|}{\text{range}}$ and range $= \max(x) - m$.

Therefore, the objective function for determining the optimal lower or upper bounds is defined as the product of two components $V = cov{*}sp$.

The principle of justifiable granularity becomes instrumental in the transformation between information granules. Given is an information granule *A*, namely an interval built on a basis of some data D_0. An information granule *B* is equivalent to *A* if it satisfies the equality: $cov(A){*}sp(A) = cov(B){*}sp(B)$. Where the coverage measures, *cov*(*A*) and *cov*(*B*) are computed with respect to some data D_0. Thus, in this manner, we can transform an interval *A* to a trapezoidal fuzzy set or transform between two fuzzy sets having different membership functions.

2) GP Model: GP model [35] generates a probabilistic information granule described by some Gaussian distribution. It is a Bayesian model which in the presence of data *D*, for any input $\boldsymbol{x}*$ produces the corresponding output $N(m(\boldsymbol{x}*), \sigma(\boldsymbol{x}*))$. $m(\boldsymbol{x}*)$ can be used as the numeric output of the model as it represents the best estimate for input $\boldsymbol{x}*$, and $\sigma(\boldsymbol{x}*)$ represents the uncertainty associated with this prediction, where

$$m(\boldsymbol{x}^*) = k(\boldsymbol{x}, \boldsymbol{x}^*)^T \, k(X, X)^{-1} \mathbf{target} \tag{10}$$

and

$$\sigma(\boldsymbol{x}^*) = k(\boldsymbol{x}^*, \boldsymbol{x}^*) - k(X, \boldsymbol{x}^*)^T k(X, X)^{-1} k(X, \boldsymbol{x}^*) \quad (11)$$

where k stands for the kernel function and

$$\mathrm{k}\ (\mathrm{X}, \boldsymbol{x}^*) = \begin{bmatrix} \mathrm{k}\,(\boldsymbol{x}_1, \boldsymbol{x}^*) \\ \vdots \\ \mathrm{k}\,(\boldsymbol{x}_\mathrm{N}, \boldsymbol{x}^*) \end{bmatrix}; \ \ \textbf{target} = \begin{bmatrix} \mathrm{target}_1 \\ \mathrm{target}_2 \\ \vdots \\ \mathrm{target}_\mathrm{N} \end{bmatrix}$$

$$k(X, X = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{bmatrix}. \quad (12)$$

The GP model comes with a high degree of flexibility as there are a number of possible kernels among which one can include RBF kernels, Matern kernels, and rational quadratic kernels as representative examples

RBF kernels

$$k\,(\boldsymbol{x}_i, x_j) = a^2 \exp\left(-\frac{1}{2l^2}\|\boldsymbol{x}_i - x_j\|^2\right) \quad (13)$$

where $l > 0$ is the length-scale (larger l means more slowly varying kernel) and $a > 0$ is a scaling factor.

Matern kernels

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{2^{1-v}}{\Gamma(v)}\left(\frac{\sqrt{2v}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{l}\right)^v K_v\left(\frac{\sqrt{2v}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{l}\right) \quad (14)$$

where l is the length scale, v is a positive shape parameter that determines the smoothness of the kernel. Γ denotes the gamma function, K is the modified Bessel function of the second kind.

Rational quadratic kernel comes in the form

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left(1 + \frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2al}\right)^{-a} \quad (15)$$

where $a > 0$ specifies the degree of smoothness, and $l > 0$ controls the scale.

The symbol $\|\cdot\|$ stands for Euclidean distance. The parameters of the above kernel functions, organized in a single vector θ, are optimized by maximizing the log-likelihood estimation criterion (*LL*), essentially equating to minimize negative log-likelihood (NLL), i.e., $\theta_{\mathrm{opt}} = \mathrm{argmin}_\theta$ (NLL), where

$$LL = -\frac{1}{2}\textbf{target}^T k(X, X)^{-1}\textbf{target} - \frac{1}{2}\log|k(X, X)| - \frac{N}{2}\log(2\pi) \quad (16)$$

where NLL $= -LL$, $|k(X, X)|$ denotes the determinant of $k(X, X)$.

The optimization process involves the limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bounds algorithm [36], which is a type of Quasi–Newton methods. It uses the gradient of the objective function NLL and the estimated value of the Hessian matrix H to guide the updates of the parameters to converge to the extreme value faster. The parameters are updated by

$$\boldsymbol{\theta}^{iter+1} = \boldsymbol{\theta}^{iter} - \left[H(\mathrm{NLL}^{iter})\right]^{-1}\frac{\partial NLL^{iter}}{\partial \boldsymbol{\theta}^{iter}}. \quad (17)$$

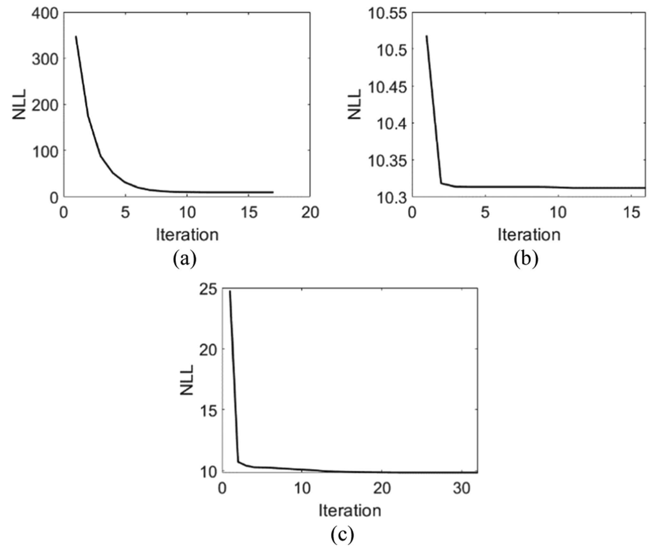

Fig. 4. Optimization of *NLL* for different kernels. (a) RBF kernel. (b) Matern kernel. (c) Rational quadratic kernel.

Returning to the 1-D data presented in (4), the GP model is constructed by using the pairs (v_i, b_i), $i = 12, \ldots, c$, being treated as the observed data. Considering $c = 10$, the optimization of the NLL leads to the results depicted in Fig. 4 and 5, respectively. In Fig. 4, the initial settings for the parameters in different kernel functions are as follows: For RBF kernels, 1.0 and 1.0 for l and a, respectively, for Matern kernels, both l and v start at 1.0, the Rational quadratic kernels similarly start with parameters l and a each at 1.0.

From above Fig. 5, the length scale parameter l in RBF, Matern, and Rational quadratic kernel function are 0.49, 0.56, and 0.54, respectively, which indicates the importance of capturing key information within a relatively compact feature space. The smaller length scale means that data points, even those closely spaced, might be perceived as distinct, rendering the model particularly sensitive to data variations. The vertical scale or the amplitude of the RBF function is controlled by parameter a, it is 0.69. This smaller value suggests a limited scope of fluctuation in the model's output. In Matern kernel, the optimal parameter v is 9.5 so that it generates a smooth function. As for the parameter a, larger a indicates a preference for capturing variations across broader scales within the data. When a is larger, the kernel behaves more similarly to the RBF kernel, suggesting that the model perceives the data variations as more uniform across different scales, rather than distinctly variable, leading to a model less sensitive to small-scale fluctuations.

Continuing with examples, we consider the RBF kernel as an example, for new data $\boldsymbol{x}^*$, $m(\boldsymbol{x}^*)$ is considered as the predicted numeric output following (10) for different number of rules and shown with dash line in Fig. 6. The corresponding optimal

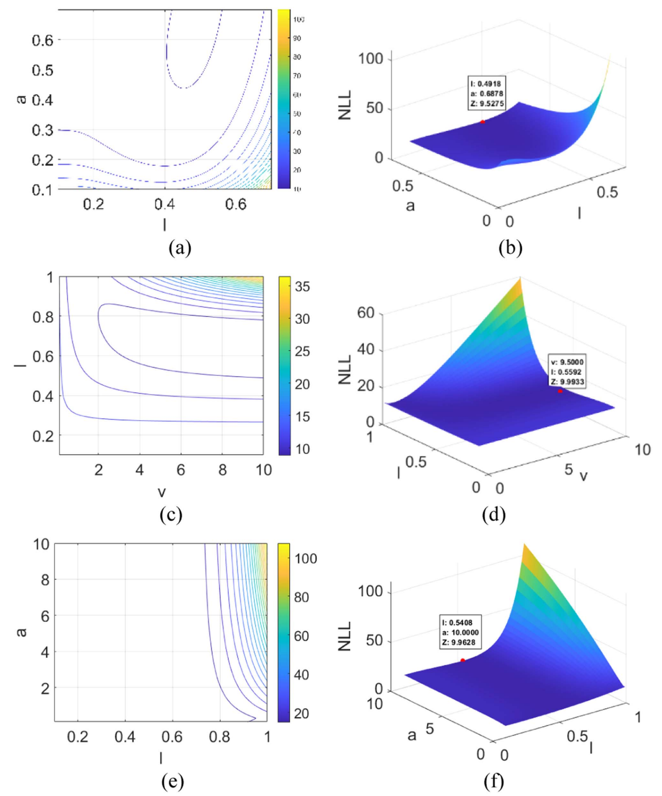

Fig. 5. Contour and 3-D plots of NLL as a function of parameters. (a) and (b) RBF kernel. (c) and (d) Matern kernel. (e) and (f) Rational quadratic kernel.

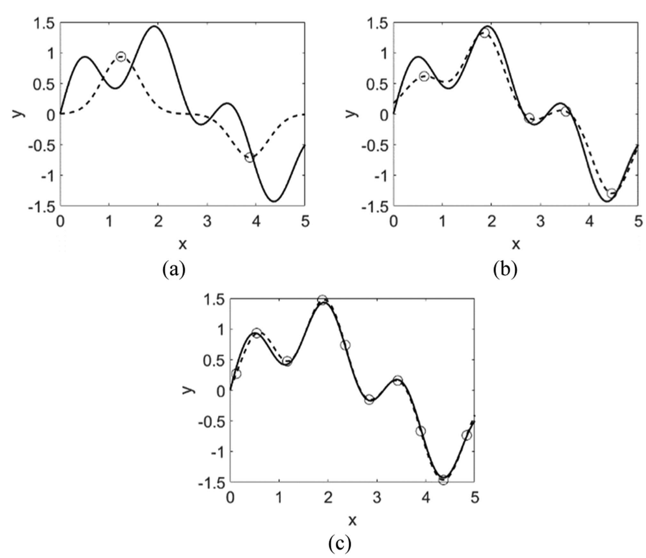

Fig. 6. Prediction results for GP models. (a) $c = 2$. (b) $c = 5$. (c) $c = 10$.

parameters by minimizing NLL (16) and resulting RMSE values are shown in Table I.

Compared with the output observed from the rule-based model in Fig. 3, it is noted that with smaller rules ($c = 2, 5$), the rule-based model surpasses the GP model in performance. However, this trend reverses with an increase in the number

TABLE I
PARAMETERS OF KERNELS AND THE CORRESPONDING RMSE VALUES FOR DIFFERENT NUMBER OF C AND SELECTED KERNELS

kernel	c=2	c=5	c=10
RBF	l:0.34 a:0.68 $RMSE$:0.670	l:0.39 a:0.76 $RMSE$:0.173	l:0.49 a:0.69 $RMSE$:0.052
Matern	l:0.12 v:1.11 $RMSE$:0.784	l:0.99 v:0.48 $RMSE$:0.221	l:0.56 v:9.5 $RMSE$:0.051
Rational quadratic	l:0.01 a:1.33 $RMSE$:0.822	l:0.67 a:1.65 $RMSE$:0.325	l:0.54 a:10.00 $RMSE$:0.050

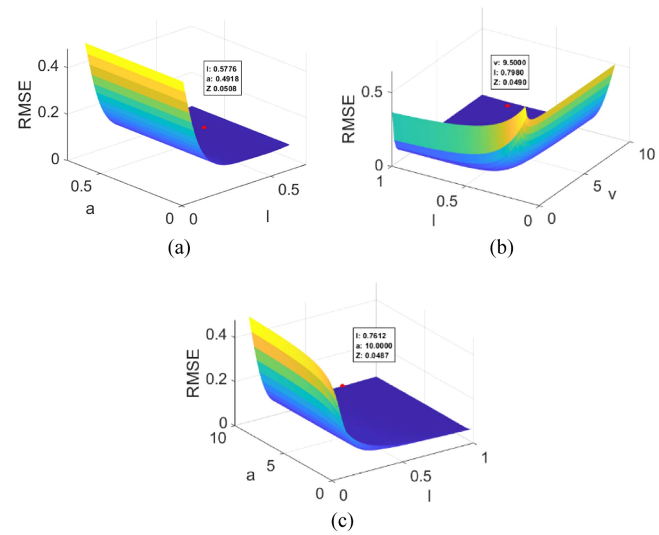

Fig. 7. RMSE as a function of parameters. (a) RBF kernel. (b) Matern kernel. (c) Rational quadratic kernel.

of rules ($c = 10$), where the GP model then exhibits better performance as the number of observed data becomes larger.

As the mean value $m(\boldsymbol{x}^*)$ is used as the numeric output of the model, the RMSE could also be considered as the objective function to optimize the corresponding parameters of the kernels. In what follows, we present how the RMSE varies with the parameters of some kernels.

Comparing Fig. 5–7, it is observed that the trends of the objective functions, RMSE and NLL, are inversely related with respect to the parameters. Furthermore, when analyzing the outcomes ($c = 10$) presented in Table I alongside those depicted in Fig. 7, it is evident that the RMSE values from Fig. 7 are smaller. This implies that employing RMSE as the objective function yields more precise results compared to utilizing NLL as the target function. To further confirm this, we employ a GP model with an RBF kernel function for 2-D synthetic data depicted in Fig. 8(a). We achieve parameter optimization by minimizing the objective functions RMSE and NLL. Subsequently, the optimal parameter values are used to derive the predicted numerical results, which are illustrated in Fig. 8(b) and (c), respectively, where circles show a location of the prototypes. The RMSE values obtained for the two methods are 0.425 and 0.503, respectively. Thus, in what follows, we use RMSE as the objective function to optimize the parameters.

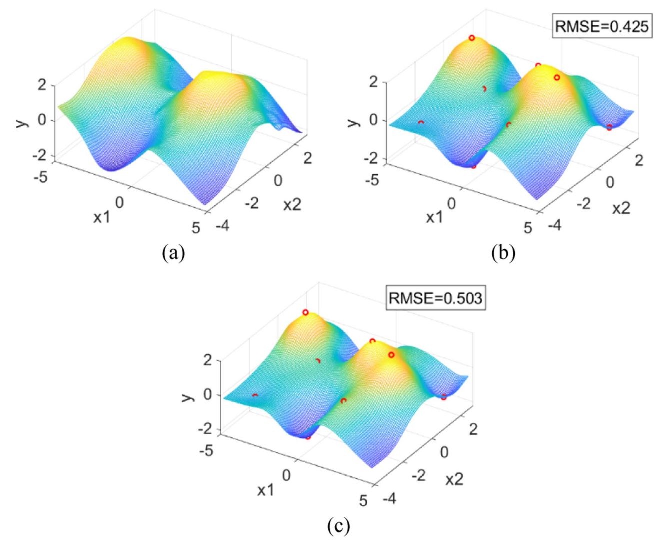

Fig. 8. Two-dimensional data and model outputs. (a) Data. (b) Outputs with RMSE as objective function. (c) Outputs with NLL as objective function.

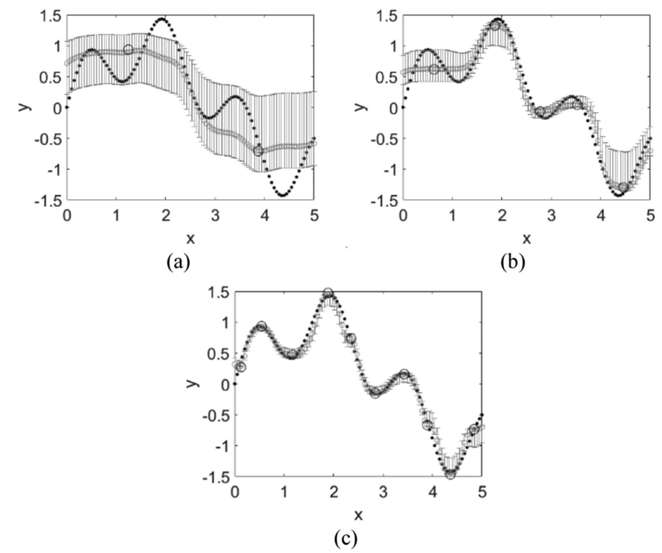

Fig. 9. Predicted interval output of rule-based models. (a) $c = 2$. (b) $c = 5$. (c) $c = 10$.

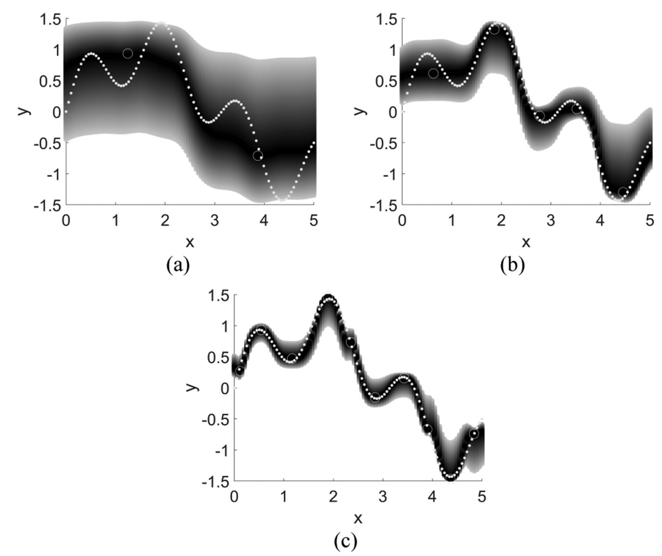

Fig. 10. Predicted triangular fuzzy set output of rule-based models. (a) $c = 2$. (b) $c = 5$. (c) $c = 10$.

IV. Granular Fuzzy Models

The numeric value of conclusion (b_i) can be considered as the optimal constant in the output space. The model can be elevated to the granular counterpart by building granular conclusions B_i distributed around b_i. This leads to the granular rules, which in contrast to the TS model, produce granular outputs. The structure of the ith granular rule now reads as

$$-\text{If } \boldsymbol{x} \text{ is } A_i(\boldsymbol{x}), \text{ then } y_i = B_i, i = 12, \ldots, c. \quad (18)$$

A. Granular Rule-Based Model

Structurally, the rules (18) can be referred to relational rules. They resemble rules originally introduced by Mamdani. The granular B_i can be represented either as an interval or a fuzzy set. For the interval output, we have $B_i = [lb_i, ub_i]$. B_i can be also treated as fuzzy sets, in particular described by a triangular fuzzy set, denoted as $B_i = T(\boldsymbol{x})$ with the parameters lower bound lb_i, constant conclusion b_i and upper bound ub_i. The information granules are constructed by invoking the principle of justifiable granularity, as discussed in Section III. Let us focus on the construction of B_i standing in the ith rule. As before, we have a collection of weighted data (target$_k$, ω_k), $k = 12, \ldots, N$, where ω_k represents the associated weight of corresponding target$_k$ that is equal to the corresponding membership grade $A_i(\boldsymbol{x}_k)$ standing in the condition part of rule. In the calculations of coverage and specificity, we confine to the range that spreads across the interval between the constants $b_{i\text{-}1}$ and b_{i+1} (assuming that the constants have been organized in increasing order).

Once the optimal lower and upper bound for the conclusion of each rule have been determined, we aggregate the granular rules to produce the final result in the form of interval as follows:

$$Y_k = \sum_{i=1}^{c} Ai(\boldsymbol{x}_k)\ B_i = \sum_{i=1}^{c} Ai(\boldsymbol{x}_k)\ [lb_i,\ ub_i]. \quad (19)$$

Y_k becomes an interval $[y_k^L,\ y_k^R]$, where $y_k^L = \sum_{i=1}^{c} Ai(\boldsymbol{x}_k)\ lb_i$ and $y_k^R = \sum_{i=1}^{c} Ai(\boldsymbol{x}_k)\ ub_i$. For triangular fuzzy set Y_k, it is described as $T(\boldsymbol{x})$ with parameters y_k^L, y_k^M, and y_k^R, where $y_k^M = \sum_{i=1}^{c} Ai(\boldsymbol{x}_k)\ b_i$, $y_k^L = \sum_{i=1}^{c} Ai(\boldsymbol{x}_k)\ lb_i$, and $y_k^R = \sum_{i=1}^{c} Ai(\boldsymbol{x}_k)\ ub_i$.

For the rule-based model for data shown in Fig. 2, the corresponding granular results (interval and triangular fuzzy set) for c being 2, 5, and 10 are shown in Figs. 9 and 10, respectively. The values of V obtained for the increasing number of rules for interval conclusions are 0.47, 0.48, and 0.63, respectively. The fuzzy sets displayed come with the values of V equal to 0.28, 0.44, and 0.57, respectively.

There is a visible tendency: with the increase of the number of rules, the specificity increases and subsequently the values

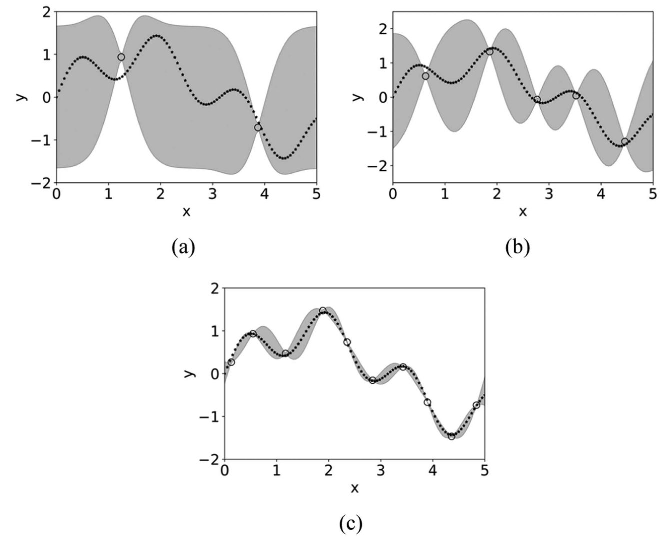

Fig. 11. Probabilistic granular results of GP models. (a) $c = 2$. (b) $c = 5$. (c) $c = 10$.

of V become higher. This is visible both for the intervals and triangular fuzzy sets.

B. Probabilistic GP-Based Fuzzy Rule-Based Model

To form the GP model, we consider the prototypes pairs $(v_i,\ b_i)$, $i = 12,\ldots,c$, that could be sought as the numeric landmarks of the fuzzy rule-based model. Then, the probability distribution of the prediction results for the new data is obtained by (10) and (11). The optimization of the parameters standing in different kernel functions has been studied in Section III. Using RBF kernel as an illustrative example, based on these optimized parameters, the Gaussian probabilistic information granules (prediction with 95% confidence interval) are shown in Fig. 11 with c being 2, 5, and 10. The circles denote prototypes and data are represented by dots.

The predicted outcomes, expressed in terms of probabilistic granularity, exhibit a notable specificity for data near the observed points. A larger number of rules makes more specific prediction information granules. Compared with the interval and fuzzy information granules in Figs. 8 and 9, it is observed that probabilistic granules seem to be broad (low specificity) when there are few rules. However, with an increase in the number of rules, the granularity of the results from the GP model becomes finer.

Probabilistic information granules can be converted into intervals or fuzzy sets by using the principle of justifiable granularity. As the probabilistic information granules produced by GP are governed by normal distribution, the calculations are carried out as follows. Let $N(0, \sigma)$ denote a Gaussian probability function $p(x)$ with the zero mean and standard deviation σ. Assume T is an interval distributed around zero with a spread d to be optimized. The coverage is computed in the form $cov(T) = \int_{-d}^{d} p(x)dx$, whereas the specificity $sp(T)$ is $1 - 2d/6\sigma = 1 - d/3\sigma$. d_{opt} is produced by maximizing the product of coverage and specificity; $d_{\text{opt}} = \max_d \left[cov(T)sp(T)\right]$. In the same way, one obtains the parameters of the fuzzy set. The corresponding coverage and specificity are $cov(T) = \int_{-d}^{d} T(x)p(x)dx$ and $sp(T) = \int_0^1 1 - \frac{|b_\alpha - a_\alpha|}{6\sigma} dx$, where a_α and b_α are the lower and upper bound of the α-cut for fuzzy set $T(x)$. The optimal spread d for interval and triangular fuzzy set could be determined by maximizing the product of coverage and specificity. Through computation, the relationship between the spread d and standard deviation σ for interval granule is $d = 1.15715\sigma$ and for triangular fuzzy set, one has $d = 1.64395\sigma$.

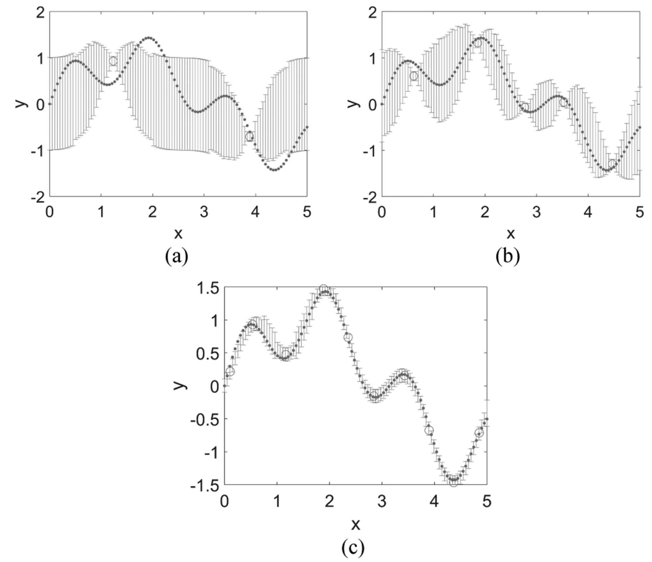

Fig. 12. Interval granules of the GP model. (a) $c = 2$. (b) $c = 5$. (c) $c = 10$.

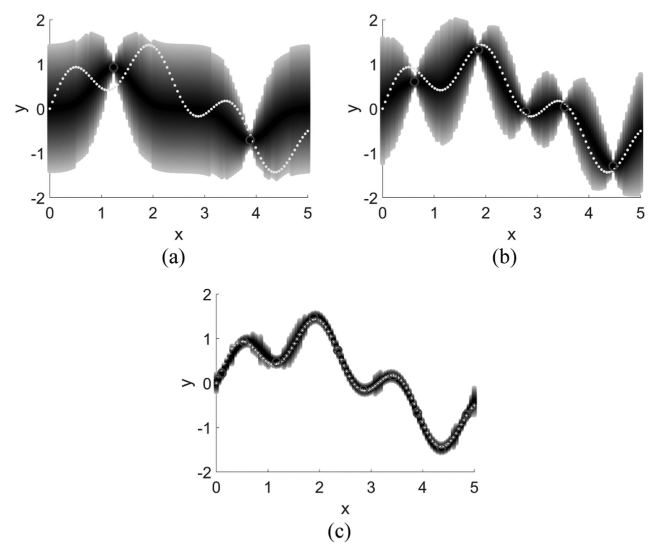

Fig. 13. Triangular granules of the GP model. (a) $c = 2$. (b) $c = 5$. (c) $c = 10$.

The probabilistic granular results converted to interval and triangular information granules are shown in Fig. 12 and 13, respectively.

The probabilistic information granules resulted from other kernels could also be converted to intervals or fuzzy sets. In

TABLE II
VALUES OF V OBTAINED FOR DIFFERENT NUMBER OF RULES AND SOME KERNELS

Kernel function	*c*=2	*c*=5	*c*=10
RBF kernel	V_{interval}:0.29 $V_{\text{triangular}}$:0.25	V_{interval}:0.59 $V_{\text{triangular}}$:0.53	V_{interval}:0.85 $V_{\text{triangular}}$:0.61
Matern kernel	V_{interval}:0.28 $V_{\text{triangular}}$:0.17	V_{interval}:0.60 $V_{\text{triangular}}$:0.40	V_{interval}:0.85 $V_{\text{triangular}}$:0.58
Rational quadratic	V_{interval}:0.26 $V_{\text{triangular}}$:0.13	V_{interval}:0.61 $V_{\text{triangular}}$:0.49	V_{interval}:0.86 $V_{\text{triangular}}$:0.62

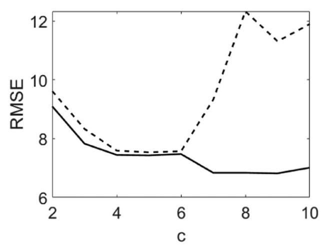

Fig. 14. Performance index as a function of the number of clusters.

Table II, the values of *V* are recorded for different kernel functions.

As the number of rules grows, the model gains the ability to capture more detailed patterns, thereby reducing the discrepancy between the granular and actual values. The performance obtained for different kernel functions does not vary visibly.

V. EXPERIMENTAL STUDIES

In this section, detailed experimental studies on machine learning dataset are conducted, focusing on evaluating the performance of various models. To ensure a comprehensive analysis, the dataset is randomly partitioned (uniform distribution) into the training and testing sets at the rate 80-20%. The experimental setup is as follows.

Rule-Based Model: It involves creating *c* rules, $c = 2, 3,\ldots, 10$. The optimal number of rules c_{opt} is determined when further increase of the value of *c* does not lead to noticeable improvements in performance, i.e., $L(c_{opt} + 1)/L(c_{opt}) < 0.95$. Then, combine these rules to make predictions in the numeric level. The principle of justifiable granularity is used to generate granular rules, intervals, or fuzzy sets.

GP-Based Model: A probabilistic model used in regression and classification in which predictions are made based on Gaussian distribution. The mean value can be considered as the numeric output and the corresponding Gaussian distribution could be converted to the interval or triangular information granules.

Using the dataset Boston Housing as an example, we run the rule-based model and GP-based model and then express their results as follows. Fig. 14 shows the performance index RMSE regarded as a function of the number of clusters for training data with solid line and testing data with a dash line. As the number of clusters increases, there is a noticeable decrease in training error; however, the testing error begins to increase once the number of clusters exceeds six. The primary reason for the rise is likely overfitting. With more clusters, the model tends to fit

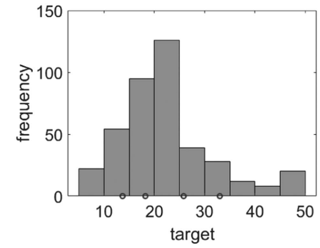

Fig. 15. Distribution of prototypes versus the spread of the data targets.

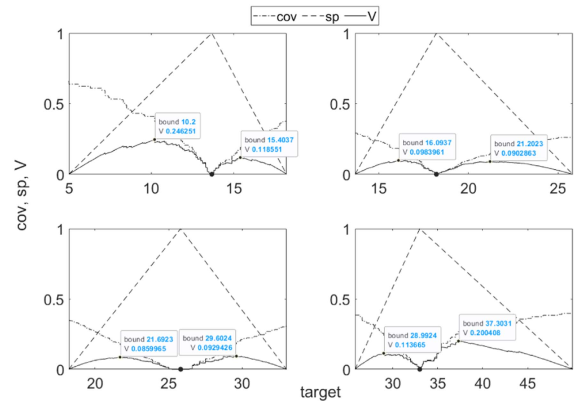

Fig. 16. Granular criteria as a function of the corresponding bounds for each b_i.

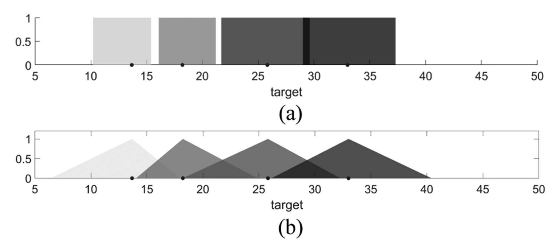

Fig. 17. Conclusion part of granular rules. (a) Interval. (b) Triangular fuzzy set.

the training data more closely, capturing noise and idiosyncrasies in the training set that fail to generalize to the testing set.

Considering $c = 4$ for illustrative purposes, the final optimized b_i are shown with circles in the Fig. 15. It shows the position of the prototypes related to the original data output.

The number of clusters affects the performance of a rule-based model. The performance RMSE of the model at the numeric level is 7.39. Considering the optimal prototypes, the principle of justifiable granularity is used to determine the information granules, i.e., their lower bounds and upper bounds. Using interval-valued granules as illustration, Fig. 16 presents the characteristics of the granules, *cov*, *sp*, and *V*, as a function of the bound range around the corresponding prototypes. The optimal values for lower and upper bounds are determined by maximizing the corresponding index *V* and they are visualized in Fig. 16.

Then, the granular results for each rule (each consequent b_i) are shown in Fig. 16. Fig. 17(a) shows the interval outputs and (b) depicts the triangular outputs for the corresponding prototypes marked with circles.

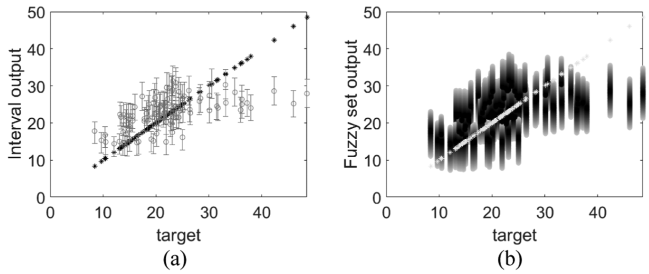

Fig. 18. Granular results of rule-based model. (a) Interval. (b) Triangular fuzzy set.

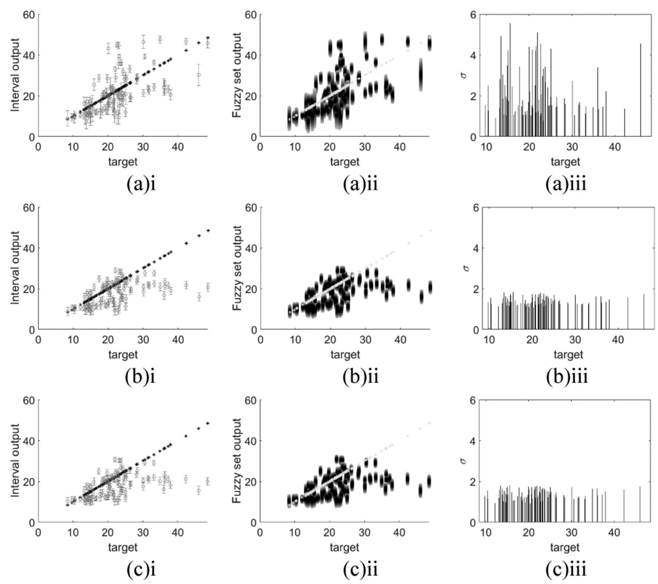

Fig. 19. Granular results for GP-based model with. (a) RBF. (b) Matern. (c) Rational quadratic.

Then, we can aggregate all granular outputs following (15). In Fig. 18, we show the predicted interval and triangular output as a function of data target when the number of clusters is 4. The evaluation criterion V for interval and triangular information granules is 0.37 and 0.32, respectively.

As for the GP method, we use the optimal four prototypes as the observed data. Based on the observations, we could obtain the predictions for the new data. In what follows, we show the final granular results converted by the probabilistic results with different kernel functions used in GP method. The performance index V for the granular model with RBF kernel function is 0.21 and 0.11 in terms of the interval granule and fuzzy sets, respectively. As for the Matern kernel, the index is 0.19, 0.08, respectively, for Rational quadratic kernel, they are 0.18, 0.08 for interval and fuzzy set information granules, respectively. The granules and the corresponding standard deviation σ of the probabilistic results are depicted in the following Fig. 19.

In addition, the experiments are also implemented based on other datasets, with each dataset being tested 50 times to ensure robust results. The optimal number of clusters c_{opt} that perform well in most trials is selected. The optimal parameters of kernels are determined by minimizing the RMSE values. The quality of

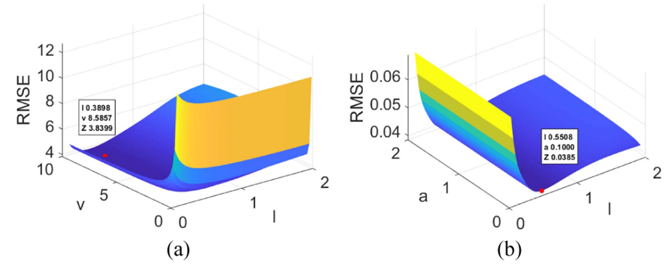

Fig. 20. RMSE as a function of parameters for GP-based model with optimal kernels. (a) House price. (b) Apple Final.

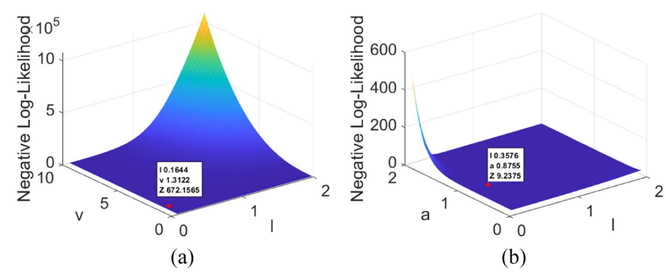

Fig. 21. NLL as a function of parameters for dataset. (a) House price. (b) Apple Final.

the model is assessed and recorded in the numeric and granular level for testing data, it is RMSE and product (V) of coverage cov and specificity sp, respectively, refer to Table III.

Based on the data presented in Table III, it is evident that the GP model yields varying outcomes when applied with distinct kernel functions. This variation underscores the significance of selecting an appropriate kernel function tailored to the specific dataset. In essence, this highlights the importance of understanding the characteristics of the dataset and the behavior of different kernels to achieve the best performance from the GP model.

Using the dataset House price, and Apple Final with the optimal kernels marked in Table III as illustrative examples, the RMSE as a function of parameters and NLL as a function of parameters are shown in Figs. 20 and 21, respectively.

The findings indicate divergent trends in RMSE and NLL as parameter values change. By minimizing RMSE or NLL to optimize parameters, the model with optimal parameters achieves performance index RMSE of 3.84 and 4.65, respectively, for dataset House price. For dataset Apple Final, they are 0.038 and 0.042, respectively. The evidence presented further solidifies the argument that employing RMSE as the metric for parameter optimization is a judicious decision.

It is worth noting that the GP model, being more flexible, outperforms the rule-based model in terms of accuracy and generalizability (granularity) for lower (one) dimensional data. However, in terms of machine learning dataset, mostly the rule-based model might be simpler to understand with high accuracy. The reason is that the GP model performs badly for high-dimensional data as the distance in the kernel functions could be challenges or problematic in high dimensional spaces [37].

TABLE III
PERFORMANCE OF THE GRANULAR MODELS FOR SEVERAL DATASETS; THE BEST RESULTS SHOWN IN BOLDFACE FOR TESTING DATA

Dataset/c_{opt}	Model			
	Granular rule-based model	*GP-based model*		
		RBF kernel	*Matern kernel*	*Rational quadratic*
Boston Housing (13 features and target in [5, 50]) / c_{opt}=4	***RMSE*: 8.15±0.813** cov_{int}: 0.41±0.045 sp_{int}: 0.85±0.008 V_{int}: 0.34±0.037 cov_{fs}: 0.35±0.032 sp_{fs}: 0.86±0.008 V_{fs}: 0.30±0.027	*RMSE*: 9.35±1.296 cov_{int}: 0.23±0.041 sp_{int}: 0.87±0.017 V_{int}: 0.20±0.036 cov_{fs}: 0.12±0.026 sp_{fs}: 0.93±0.008 V_{fs}: 0.11±0.024	*RMSE*: 9.09±1.020 cov_{int}: 0.22±0.044 sp_{int}: 0.92±0.005 V_{int}: 0.20±0.006 cov_{fs}: 0.11±0.028 sp_{fs}: 0.96±0.002 V_{fs}: 0.11±0.027	*RMSE*: 10.13±1.229 cov_{int}: 0.22±0.030 sp_{int}: 0.94±0.012 V_{int}: 0.21±0.014 cov_{fs}: 0.09±0.021 sp_{fs}: 0.94±0.001 V_{fs}: 0.08±0.009
House price (6 features and target in [1.68, 44.64]) / c_{opt}=7	*RMSE*: 4.34±0.099 cov_{int}: 0.29±0.038 sp_{int}: 0.88±0.010 V_{int}: 0.30±0.021 cov_{fs}: 0.26±0.009 sp_{fs}: 0.92±0.025 V_{fs}: 0.25±0.019	*RMSE*: 5.50±0.068 cov_{int}: 0.14±0.019 sp_{int}: 0.93±0.012 V_{int}: 0.14±0.016 cov_{fs}: 0.04±0.011 sp_{fs}: 0.98±0.005 V_{fs}: 0.09±0.010	***RMSE*: 3.91±0.085** cov_{int}: 0.36±0.010 sp_{int}: 0.87±0.018 V_{int}: 0.36±0.012 cov_{fs}: 0.19±0.031 sp_{fs}: 0.97±0.006 V_{fs}: 0.21±0.018	*RMSE*: 4.11±0.088 cov_{int}: 0.33±0.031 sp_{int}: 0.90±0.012 V_{int}: 0.32±0.019 cov_{fs}: 0.17±0.024 sp_{fs}: 0.97±0.007 V_{fs}: 0.21±0.019
Apple Final (8 features and target in [0, 0.345]) / c_{opt}=9	***RMSE*: 0.04±0.004** cov_{int}: 0.56±0.023 sp_{int}: 0.79±0.008 V_{int}: 0.46±0.014 cov_{fs}: 0.47±0.019 sp_{fs}: 0.89±0.002 V_{fs}: 0.43±0.014	*RMSE*: 0.04±0.005 cov_{int}: 0.99±0.014 sp_{int}: 0.39±0.005 V_{int}: 0.40±0.010 cov_{fs}: 0.69±0.018 sp_{fs}: 0.61±0.006 V_{fs}: 0.43±0.007	*RMSE*: 0.05±0.004 cov_{int}: 0.98±0.004 sp_{int}: 0.45±0.009 V_{int}: 0.45±0.007 cov_{fs}: 0.69±0.006 sp_{fs}: 0.62±0.015 V_{fs}: 0.44±0.010	*RMSE*: 0.05±0.003 cov_{int}: 0.99±0.005 sp_{int}: 0.17±0.011 V_{int}: 0.18±0.009 cov_{fs}: 0.90±0.004 sp_{fs}: 0.14±0.016 V_{fs}: 0.14±0.009
Combined cycle power plant (3 features and target in [420.26, 495.76]) / c_{opt}=9	***RMSE*: 8.05±0.485** cov_{int}: 0.33±0.046 sp_{int}: 0.89±0.010 V_{int}: 0.31±0.014 cov_{fs}: 0.31±0.013 sp_{fs}: 0.90±0.042 V_{fs}: 0.29±0.017	*RMSE*: 13.29±2.162 cov_{int}: 0.02±0.033 sp_{int}: 0.99±0.018 V_{int}: 0.04±0.022 cov_{fs}: 0.01±0.040 sp_{fs}: 1.00±0.011 V_{fs}: 0.02±0.018	*RMSE*: 8.77±0.843 cov_{int}: 0.21±0.017 sp_{int}: 0.96±0.010 V_{int}: 0.21±0.013 cov_{fs}: 0.09±0.044 sp_{fs}: 0.98±0.010 V_{fs}: 0.10±0.014	*RMSE*: 8.48±1.089 cov_{int}: 0.09±0.037 sp_{int}: 0.94±0.018 V_{int}: 0.10±0.020 cov_{fs}: 0.04±0.011 sp_{fs}: 0.97±0.018 V_{fs}: 0.06±0.020
Air quality (14 features and target in [-200, 1189]) / c_{opt}=4	***RMSE*: 186.80±13.952** cov_{int}: 0.25±0.043 sp_{int}: 0.88±0.012 V_{int}: 0.23±0.019 cov_{fs}: 0.20±0.012 sp_{fs}: 0.92±0.048 V_{fs}: 0.20±0.018	*RMSE*: 207.76±6.037 cov_{int}: 0.05±0.016 sp_{int}: 0.96±0.033 V_{int}: 0.06±0.016 cov_{fs}: 0.03±0.015 sp_{fs}: 0.99±0.044 V_{fs}: 0.04±0.015	*RMSE*: 210.89±12.259 cov_{int}: 0.13±0.022 sp_{int}: 0.95±0.012 V_{int}: 0.13±0.013 cov_{fs}: 0.07±0.047 sp_{fs}: 0.96±0.012 V_{fs}: 0.07±0.019	*RMSE*: 193.10±8.856 cov_{int}: 0.02±0.016 sp_{int}: 0.98±0.012 V_{int}: 0.03±0.014 cov_{fs}: 0.03±0.013 sp_{fs}: 0.99±0.020 V_{fs}: 0.03±0.017
Abalone (7 features and target in [1, 29]) / c_{opt}=9	***RMSE*: 4.44±0.308** cov_{int}: 0.23±0.012 sp_{int}: 0.97±0.007 V_{int}: 0.23±0.010 cov_{fs}: 0.18±0.015 sp_{fs}: 0.95±0.003 V_{fs}: 0.19±0.015	*RMSE*: 5.06±0.529 cov_{int}: 0.24±0.019 sp_{int}: 0.96±0.006 V_{int}: 0.24±0.011 cov_{fs}: 0.14±0.015 sp_{fs}: 0.99±0.006 V_{fs}: 0.14±0.013	*RMSE*: 5.27±0.326 cov_{int}: 0.20±0.015 sp_{int}: 0.98±0.009 V_{int}: 0.21±0.012 cov_{fs}: 0.11±0.017 sp_{fs}: 0.98±0.001 V_{fs}: 0.12±0.011	*RMSE*: 5.34±0.531 cov_{int}: 0.18±0.015 sp_{int}: 0.96±0.007 V_{int}: 0.18±0.011 cov_{fs}: 0.12±0.016 sp_{fs}: 0.96±0.008 V_{fs}: 0.13±0.012
Concrete Compressive (7 features and target in [2.3, 81.7]) / c_{opt}=5	***RMSE*: 13.07±2.296** cov_{int}: 0.42±0.014 sp_{int}: 0.82±0.011 V_{int}: 0.35±0.012 cov_{fs}: 0.41±0.019 sp_{fs}: 0.77±0.005 V_{fs}: 0.33±0.017	*RMSE*: 16.34±1.332 cov_{int}: 0.34±0.020 sp_{int}: 0.85±0.009 V_{int}: 0.29±0.018 cov_{fs}: 0.18±0.020 sp_{fs}: 0.90±0.012 V_{fs}: 0.17±0.016	*RMSE*: 16.90±1.063 cov_{int}: 0.14±0.015 sp_{int}: 0.90±0.011 V_{int}: 0.14±0.012 cov_{fs}: 0.12±0.010 sp_{fs}: 0.95±0.007 V_{fs}: 0.13±0.011	*RMSE*: 17.21±2.323 cov_{int}: 0.16±0.014 sp_{int}: 0.93±0.006 V_{int}: 0.16±0.013 cov_{fs}: 0.09±0.016 sp_{fs}: 0.95±0.008 V_{fs}: 0.10±0.014
Parkinson's (17 features and target in [0,1]) / c_{opt}=4	***RMSE*: 0.19±0.015** cov_{int}: 0.23±0.010 sp_{int}: 0.87±0.015 V_{int}: 0.21±0.013 cov_{fs}: 0.18±0.017 sp_{fs}: 0.90±0.012 V_{fs}: 0.18±0.013	*RMSE*: 0.28±0.019 cov_{int}: 0.46±0.027 sp_{int}: 0.71±0.011 V_{int}: 0.34±0.018 cov_{fs}: 0.26±0.011 sp_{fs}: 0.84±0.016 V_{fs}: 0.23±0.013	*RMSE*: 0.22±0.017 cov_{int}: 0.88±0.026 sp_{int}: 0.18±0.015 V_{int}: 0.17±0.019 cov_{fs}: 0.59±0.011 sp_{fs}: 0.56±0.016 V_{fs}: 0.34±0.012	*RMSE*: 0.21±0.012 cov_{int}: 0.94±0.016 sp_{int}: 0.12±0.010 V_{int}: 0.12±0.012 cov_{fs}: 0.59±0.010 sp_{fs}: 0.48±0.019 V_{fs}: 0.30±0.017
Magic Gamma Telescope (10 features and target in [15.3,408.5]) / c_{opt}=6	*RMSE*: 66.99±2.280 cov_{int}: 0.37±0.019 sp_{int}: 0.84±0.010 V_{int}: 0.32±0.016 cov_{fs}: 0.28±0.019 sp_{fs}: 0.88±0.007 V_{fs}: 0.25±0.019	*RMSE*: 68.60±1.793 cov_{int}: 0.05±0.018 sp_{int}: 0.98±0.009 V_{int}: 0.06±0.015 cov_{fs}: 0.09±0.017 sp_{fs}: 0.94±0.011 V_{fs}: 0.09±0.013	***RMSE*: 63.45±1.676** cov_{int}: 0.21±0.016 sp_{int}: 0.95±0.010 V_{int}: 0.20±0.011 cov_{fs}: 0.11±0.015 sp_{fs}: 0.92±0.006 V_{fs}: 0.11±0.013	*RMSE*: 64.33±1.490 cov_{int}: 0.25±0.018 sp_{int}: 0.88±0.010 V_{int}: 0.23±0.016 cov_{fs}: 0.12±0.018 sp_{fs}: 0.91±0.011 V_{fs}: 0.11±0.017

* In the above table, 'int' stands for interval; 'fs' means fuzzy set.

VI. Conclusion

Rule-based models traditionally produce numeric outcomes, but it is important to acknowledge that these numeric results are approximations rather than precise values. Recognizing the limitations inherent in numeric outputs, this study advocates for a transformative approach within the realm of fuzzy rule-based modeling. The proposal for incorporating granular output into the rule-based model addresses these uncertainties by enhancing the model's ability to express and manage them. To this end, the research introduces two novel model frameworks. The first framework focuses on the generation of granular rules, offering a more nuanced interpretation of data. The second framework utilizes the GP method for a more sophisticated function fitting approach, diverging from conventional rule-based methodologies. The empirical findings from this study strongly support the effectiveness of the granular approach, underlining its practical advantages over traditional methods in the following way: Granular outputs can better represent the ambiguity and imprecision inherent in many real-world systems, offering more nuanced insights into the model's predictions. The proposed models have broad applicability across various domains, particularly in areas such as housing price prediction and energy consumption forecasting. Their ability to handle granular outputs makes them especially valuable in sectors where data is subject to variability and uncertainty.

In future research, active learning could be considered to further improve the quality of the granular models. Active learning potentially refines the granular models by selectively querying the most informative data points for inclusion in the training set. This approach would likely enhance the efficiency and effectiveness of model training, particularly in complex or uncertain data environments.

References

[1] 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, Jan. 1975, doi: 10.1016/S0020-7373(75)80002-2.

[2] E. Hanyu, Y. Cui, W. Pedrycz, and Z. Li, "Enhancements of rule-based models through refinements of Fuzzy C-Means," *Knowl. Based Syst.*, vol. 170, pp. 43–60, Apr. 2019, doi: 10.1016/j.knosys.2019.01.027.

[3] T. W. LIAO, "A review of interval valued Type 2 fuzzy rule-based classifiers," in *Proc. Int. Conf. Mach. Learn. Cybern.*, IEEE, Jul. 2023, pp. 411–416. doi: 10.1109/ICMLC58545.2023.10327955.

[4] P. Baranyi, "Transition between TS fuzzy models and the associated convex hulls by TS fuzzy model transformation," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 4, pp. 2272–2282, Apr. 2024, doi: 10.1109/TFUZZ.2023.3348160.

[5] G. Xue, J. Wang, K. Zhang, and N. R. Pal, "High-dimensional fuzzy inference systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 1, pp. 507–519, Jan. 2024, doi: 10.1109/TSMC.2023.3311475.

[6] H. E, Y. Cui, W. Pedrycz, and Z. Li, "Design of distributed rule-based models in the presence of large data," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 7, pp. 2479–2486, Jul. 2023, doi: 10.1109/TFUZZ.2022.3226250.

[7] Y. Cui, H. E, W. Pedrycz, and Z. Li, "Designing distributed fuzzy rule-based models," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 7, pp. 2047–2053, Jul. 2021, doi: 10.1109/TFUZZ.2020.2984971.

[8] J. Alcala-Fdez, R. Alcala, and F. Herrera, "A fuzzy association Rule-Based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 857–872, Oct. 2011, doi: 10.1109/TFUZZ.2011.2147794.

[9] H. Ying, Y. Ding, S. Li, and S. Shao, "Comparison of necessary conditions for typical Takagi–Sugeno and Mamdani fuzzy systems as universal approximators," *IEEE Trans. Syst., Man, Cybern. - A, Syst. Humans*, vol. 29, no. 5, pp. 508–514, Sep. 1999, doi: 10.1109/3468.784177.

[10] N. Garcia-Diaz, C. Lopez-Martin, and A. Chavoya, "A comparative study of two fuzzy logic models for software development effort estimation," *Procedia Technol.*, vol. 7, pp. 305–314, 2013, doi: 10.1016/j.protcy.2013.04.038.

[11] X. Zhu, W. Pedrycz, and Z. Li, "A development of hierarchically structured granular models realized through allocation of information granularity," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 12, pp. 3845–3858, Dec. 2021, doi: 10.1109/TFUZZ.2020.3028939.

[12] W. Pedrycz, "From data to information granules: An environment of granular computing," in *Proc. IEEE 20th Int. Conf. Cogn. Inform. Cogn. Comput.*, IEEE, Oct. 2021, pp. 2–2, doi: 10.1109/ICCICC53683.2021.9811327.

[13] X. Hu, H. Chen, C. Chen, B. Sun, J. Huang, and K. Huang, "Allocation of information granularity: A multi-objective evolutionary optimization using conflict information," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, IEEE, Jun. 2019, pp. 1–6, doi: 10.1109/FUZZ-IEEE.2019.8858939.

[14] B. Zhang, W. Pedrycz, X. Wang, and A. Gacek, "Design of interval Type-2 information granules based on the principle of justifiable granularity," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 11, pp. 3456–3469, Nov. 2021, doi: 10.1109/TFUZZ.2020.3023758.

[15] J. E. Moreno et al., "Design of an interval Type-2 fuzzy model with justifiable uncertainty," *Inf. Sci. (N Y)*, vol. 513, pp. 206–221, Mar. 2020, doi: 10.1016/j.ins.2019.10.042.

[16] M. A. Sanchez, O. Castillo, and J. R. Castro, "Type-2 fuzzy granular models," 2017, pp. 1–3. doi: 10.1007/978-3-319-41288-7_1.

[17] Y. Cui, E. Hanyu, W. Pedrycz, and Z. Li, "Augmentation of rule-based models with a granular quantification of results," *Soft. Comput.*, vol. 23, no. 23, pp. 12745–12759, Dec. 2019, doi: 10.1007/s00500-019-03825-7.

[18] T. Ouyang, "Structural rule-based modeling with granular computing," *Appl. Soft. Comput.*, vol. 128, Oct. 2022, Art. no. 109519, doi: 10.1016/j.asoc.2022.109519.

[19] Y. Liu, J. Zhao, W. Wang, and W. Pedrycz, "Prediction intervals for granular data streams based on evolving Type-2 fuzzy granular neural network dynamic ensemble," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 4, pp. 874–888, Apr. 2021, doi: 10.1109/TFUZZ.2020.2966172.

[20] J. Hu, Y. Zhou, H. Li, and P. Liang, "An interval forecast model for infectious diseases using fuzzy information granulation and spatial-temporal graph neural network," *J. Intell. Fuzzy Syst.*, pp. 1–15, Apr. 2024, doi: 10.3233/JIFS-236766.

[21] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Trans Neural Netw.*, vol. 22, no. 9, pp. 1341–1356, Sep. 2011, doi: 10.1109/TNN.2011.2162110.

[22] Y. Chen and D. Miao, "Granular regression with a gradient descent method," *Inf. Sci. (N Y)*, vol. 537, pp. 246–260, Oct. 2020, doi: 10.1016/j.ins.2020.05.101.

[23] Y. Q. Zhang, "Granular neural networks," in *Encyclopedia of Complexity and Systems Science*. Berlin, Heidelberg: Springer, 2020, pp. 1–14, doi: 10.1007/978-3-642-27737-5_261-2.

[24] Y. Chen, X. Zhang, Y. Zhuang, B. Yao, and B. Lin, "Granular neural networks with a reference frame," *Knowl. Based Syst.*, vol. 260, Jan. 2023, Art. no. 110147, doi: 10.1016/j.knosys.2022.110147.

[25] M. Song, R. Wang, and Y. Li, "Hybrid time series interval prediction by granular neural network and ARIMA," *Granular Comput.*, vol. 9, no. 1, Mar. 2024, Art. no. 3, doi: 10.1007/s41066-023-00422-w.

[26] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103, New York, NY, USA: Springer, 2013. doi: 10.1007/978-1-4614-7138-7.

[27] Y. S. Lee and S. Scholtes, "Empirical prediction intervals revisited," *Int. J. Forecasting*, vol. 30, no. 2, pp. 217–234, Apr. 2014, doi: 10.1016/j.ijforecast.2013.07.018.

[28] J. Sanz, D. Bernardo, F. Herrera, H. Bustince, and H. Hagras, "A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 973–990, Aug. 2015, doi: 10.1109/TFUZZ.2014.2336263.

[29] J. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, “Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system,” *Appl. Soft. Comput.*, vol. 20, pp. 103–111, Jul. 2014, doi: 10.1016/j.asoc.2013.11.009.

[30] 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, doi: 10.1109/TSMC.1985.6313399.

[31] X. Zhu, D. Wang, W. Pedrycz, and Z. Li, “Fuzzy rule-based local surrogate models for black-box model explanation,” *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 6, pp. 2056–2064, Jun. 2023, doi: 10.1109/TFUZZ.2022.3218426.

[32] W. Zhao, Q. Niu, K. Li, and G. W. Irwin, “A hybrid learning method for constructing compact rule-based fuzzy models,” *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1807–1821, Dec. 2013, doi: 10.1109/TSMCB.2012.2231068.

[33] E. H., Y. Cui, W. Pedrycz, A. Robinson Fayek, Z. Li, and J. Li, “Design of fuzzy rule-based models with fuzzy relational factorization,” *Expert Syst. Appl.*, vol. 206, Nov. 2022, Art. no. 117904, doi: 10.1016/j.eswa.2022.117904.

[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, Oct. 2013, doi: 10.1016/j.asoc.2013.06.017.

[35] M. Seeger, “Gaussian processes for machine learning,” *Int. J. Neural Syst.*, vol. 14, no. 02, pp. 69–106, Apr. 2004, doi: 10.1142/S0129065704001899.

[36] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, “Algorithm 778: L-BFGS-B,” *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, Dec. 1997, doi: 10.1145/279232.279236.

[37] M. Binois and N. Wycoff, “A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization,” *ACM Trans. Evol. Learn. Optim.*, vol. 2, no. 2, pp. 1–26, Jun. 2022, doi: 10.1145/3545611.