2009-08

Extracting Takagi-Sugeno Fuzzy Rules with Interpretable Submodels via Regularization of Linguistic Modifiers

Shang-Ming Zhou,

10.1109/tkde.2008.208
IEEE Transactions on Knowledge and Data Engineering
Institute of Electrical and Electronics Engineers (IEEE)

Extracting Takagi-Sugeno Fuzzy Rules with Interpretable Submodels via Regularization of Linguistic Modifiers

Shang-Ming Zhou, *Member, IEEE* and John Q. Gan, *Senior Member, IEEE*

Abstract— In this paper, a method for constructing Takagi-Sugeno (TS) fuzzy system from data is proposed with the objective of preserving TS submodel comprehensibility, in which linguistic modifiers are suggested to characterise the fuzzy sets. A good property held by the proposed linguistic modifiers is that they can broaden the cores of fuzzy sets while contracting the overlaps of adjoining membership functions during identification of fuzzy systems from data. As a result, the TS submodels identified tend to dominate the system behaviors by automatically matching the global model in corresponding sub-areas, which leads to good TS model interpretability while producing distinguishable input space partitioning. However, the global model accuracy and model interpretability are two conflicting modelling objectives, improving interpretability of fuzzy models generally degrades the global model performance of fuzzy models, and vice versa. Hence one challenging problem is how to construct a TS fuzzy model with not only good global performance but also good submodel interpretability. In order to achieve a good trade-off between global model performance and submodel interpretability, a regularization learning algorithm is presented in which the global model objective function is combined with a local model objective function defined in terms of an extended index of fuzziness of identified membership functions. Moreover, a parsimonious rule-base is obtained by adopting a QR decomposition method to select the important fuzzy rules and reduce the redundant ones. Experimental studies have shown that the TS models identified by the suggested method possess good submodel interpretability and satisfactory global model performance with parsimonious rule-bases.

Index Terms— Interpretability, Distinguishability, Knowledge extraction, Local models, Submodels, Takagi-Sugeno fuzzy models, Regularization, Fuzziness.

1 Introduction

Fuzzy models have been widely and successfully used in many areas such as system modeling and control, data analysis, and pattern recognition. Traditionally, fuzzy rules are generated from human expert knowledge or heuristics, which results in good high-level semantic generalization capability. On the other hand, more and more researchers have made efforts to build fuzzy models from observational data with many successful applications [1]-[4]. Compared to heuristic fuzzy rules, fuzzy rules generated from data are able to extract more specific information about unknown complex systems or processes, however, the wide investigation on data-driven models mainly focuses on the issues of high accuracy, completeness and efficiency. Recently, more and more efforts have been made to approach the problem of interpretability of fuzzy systems [5]-[21], because one of the important incentives of introducing fuzzy methods into complex system modeling is to improve the model interpretability and thus gain deep insights into the complex systems to be modeled. As a matter of fact, comprehensibility preservation during data-driven adaptation and knowledge extraction has been regarded as one of the most important issues in data-driven fuzzy modeling [6][7] [8][9][10][11][12][13].

The first aspect of fuzzy model interpretability is the transparency of input space partitioning, that is, the generated fuzzy sets should be distinguishable and interpretable. Although there exists no unified standard for selecting membership functions (MFs), some researchers have proposed semantic criteria or heuristic criteria to guide the generation of MFs in the interests of preserving or enhancing model interpretability. For instance, several semantic criteria for designing MFs (such as distinguishability of MFs, normalization of MFs, moderate number of linguistic terms per variable, and coverage of the universe of discourse) have been shown to be very helpful in improving fuzzy model interpretability [15][16][20][21]. Particularly, some semantic criteria can be formalised for enhancing fuzzy model interpretability by combining these expressions with global model accuracy measure [16].

Another interesting criterion states that "good" clusters are actually not very fuzzy [22][23][24]. Although some fuzzy algorithms are used in data clustering, the aim of the clustering is to generate a "harder" partitioning of the data sets [24], by which a better interpretation of input space partitioning can be achieved. The requirements directly related to this interpretability in fuzzy modeling are that under the condition of preserving the global accuracy

- *Shang-Ming Zhou is with Centre for Computational Intelligence, Department of Informatics, De Montfort University, Leicester, LE1 9BH, UK. E-mail: smzhou@ ieee.org.*
- *John Q. Gan is with Department of Computing and Electronic Systems, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK. E-mail: jqgan@essex.ac.uk.*

Manuscript received (18 March 2008).

at a satisfactory level, the fuzzy sets should have large core regions and adjacent fuzzy sets should be less overlapped. In order to obtain MFs with large bounded core, Hoppner and Klawonn [25] proposed a novel clustering algorithm by using the distance to the Voronoi cell of a cluster rather than to the cluster prototype, and furthermore they assigned a "reward" to membership degrees that are near to 0 and 1. However, traditional data-driven algorithms for rule generation, such as neuro-fuzzy algorithms [2][3], usually generate fuzzy sets with "too much" overlap due to their accuracy-oriented nature. By using fuzzy sets with "too much" overlap, the distinguishability of input space partitioning is lost so that it is difficult to assign distinct linguistic labels and semantic meanings to these fuzzy sets, which leads to poor model interpretability [12].

However, in the first-order Takagi-Sugeno (TS) fuzzy model [1], a most widely investigated paradigm in fuzzy modeling, the consequents of fuzzy rules are local linear models. As a result, there is another type of model interpretation regarding the interaction between global model and its local linear models (or linear submodels). The purpose of this paper is to provide a new scheme for extracting TS fuzzy rules from data with good local model interpretability by a new type of membership function and learning algorithm. In this approach, the TS local linear models are forced to fit the global model locally and separately during the learning process, and at the same time, the distinguishability of the input space partition can also be improved.

The remainder of the paper is organized as follows. Section 2 addresses the issues arising about local model intepretabilty in TS fuzzy inference systems. In section 3, a linguistic modifier is defined as an MF. Section 4 describes a TS model using linguistic modifiers as MFs, and its local model behaviors are analyzed in detail. In section 5, a hybrid learning scheme is proposed to update both the consequent and premise parameters in the TS model by regularizing the fuzziness of linguistic modifiers, and a pivoted QR decomposition algorithm is used to identify the most influential fuzzy rules. Section 6 includes experimental results to evaluate the performance of the proposed method in terms of global model accuracy and local model interpretability. Section 7 concludes the paper with discussions. In this paper, the first-order TS model is considered, unless otherwise stated.

2 Issues about Interpretability of Local Linear Models in Takagi-Sugeno Fuzzy Systems

It is known that a challenge for the real-time predictive control of nonlinear systems is that a nonlinear (and usually non-convex) optimization problem must be solved at each sampling period by model predictive control, and the non-convex optimization usually involves high computational overhead. As a result, the application for fast systems is hampered where iterative optimization techniques cannot be properly used due to short sampling time intervals. The TS fuzzy model prevails in representing nonlinear systems in the fuzzy control community in that the global TS model interpolates between some local linear models in nature, and these control-relevant local linear models rather than a single nonlinear plant model possess great potential of being effectively used in model predictive control.

The interpretation of fuzzy models including TS fuzzy models heavily depends on human's prior knowledge, which is a subjective issue sometimes. However, if there is no prior knowledge available, such as in data-driven system modelling, a criterion for interpreting TS local linear models, as indicated in [9][10][26], is sensible and applicable and should be adopted in fuzzy modelling [27]. This criterion is summarised as follow:

***Definition 1*: The local models of a TS model are considered to be interpretable if they fit the global model well in their local regions, and result in fuzzy rule consequents that are local linearizations of the non1inear system.**

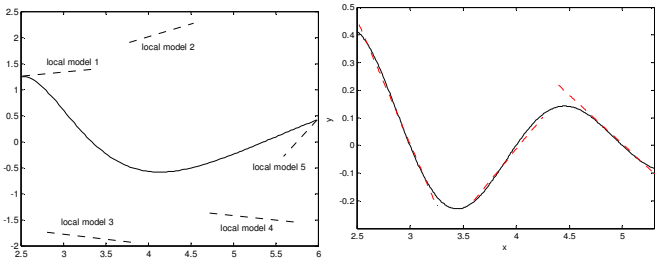

Fig. 1 TS models with (left) non-interpretable local models and (right) interpretable local models: solid line represents the global model, dotted lines represent the local models

According to this definition, interpretable local models of a TS model should dominate the system behaviors separately in their local regions. The submodels shown in *Fig.1(left)* do not match the global model well in the corresponding local areas, so they exhibit poor interpretation of interaction with global model, whilst the submodels shown in *Fig.1(right)* are local linearizations of the global system in local regions, which shows good interpretable interaction with global model. Hence the TS model with interpretable submodels like the ones depicted in *Fig.1(right)* is preferred in fuzzy system modeling. Interestingly, the local error function defined as follows can be used to evaluate the degree of TS local model dominating the behaviors of the global system, and thus work as a measure of the interaction between TS local models and global model [9]:

$$J_L = \sum_{i=1}^{L}\sum_{k}^{N} w_i(k)\left[d(k) - y_i(k)\right]^2 \qquad (1)$$

where given the kth sample, $w_i(k)$ is the normalized firing strength of the ith rule, $y_i(k)$ is the output of the ith submodel, and $d(k)$ is the desired global system output. Because $w_i(k)$ has nonzero values only in a small region of the input space, as a result each fuzzy rule acts like an independent sub-model that is only related to a subset of training data. A smaller J_L value implies that

the TS local models fit the global model better in the corresponding local regions, which indicates that better local model interpretation can be obtained in the sense of Definition 1.

Zhou and Gan recently proposed a unified view of data-driven interpretable fuzzy models in terms of *low-level interpretability* and *high-level interpretability*, while the TS local model interpretability in the sense of Definition 1 works as a criterion for transparency of rule structure in constructing TS fuzzy model with high-level interpretability [10]. In [28], Gan and Harris analyzed the relationship between fuzzy local linearization and local basis function expansion and the role of the local models in global model approximation. It is a rather challenging task to build a locally interpretable TS fuzzy model with good global approximation performance. This is because in the TS fuzzy system modeling, local linear model interpretability and global model ability are two conflicting modeling objectives. As a result, interpretability improvement of local linear models in the sense of Definition 1 usually degrades global model approximation ability. One possibility of attacking this problem, proposed in [26], is to constrain the candidate model parameters of the rules in the TS fuzzy model based on prior knowledge about the modeled process such as stability, minimal or maximal gain, or the setting time. For instance, instead of the identification from the measured input-output data, a TS model with good local model interpretability in the sense of Definition 1 was generated directly from a polynomial Hammerstein system that is assumed to be known [26]. However, in most data-driven system modeling tasks, no prior knowledge about the modeled process is available.

Interestingly and promisingly, in order to improve the TS local model interpretability, a scheme that combines global learning and local learning (ComGLL) was proposed to train a TS fuzzy model from data [9], in which in addition to the commonly used global error objective, local error objectives measuring the deviations of the outputs of individual local models from the desired outputs are integrated into a learning index as well. In [27], the ComGLL scheme was treated as a multi-objective identification problem and the Pareto-optimal solutions were used to identify the model parameters. The advantage of the ComGLL scheme lies in that as an interpretability-oriented modeling approach, it improves the interpretation of TS local linear models in the sense of Definition 1, that is, these TS local linear models tends to match the global model in local regions. However, one drawback of this scheme is that it does not simultaneously consider improving the transparency of input space partitioning, such as the distinguishability of the generated fuzzy sets, as a result, by experiments it was found that some local models still exhibit some eccentric behaviors.

Another potential strategy of improving TS local model interactions with global model is via the scheme of merging similar fuzzy sets in input space partitions. One popular scheme of merging similar fuzzy sets for improving fuzzy model interpretability is performed in terms of the similarity measure of fuzzy sets [12][31][32]. In order to overcome the computational inefficiency of the similarity measure, some researchers recently proposed the possibility measure of fuzzy sets for producing distinguishable fuzzy sets [33][34]. The difference between the merging scheme (including similarity based merging scheme and possibility based merging (PBM) scheme) and the ComGLL lies in that the merging scheme can be used to improve the interpretability of both Mamdani fuzzy model and TS fuzzy model, whilst the ComGLL scheme specially aims at improving TS local linear model interpretability. The main advantage of the merging scheme in terms of the similarity or possibility of fuzzy sets lies in its ability to improve the distinguishability of input space partitions. However, for TS models, distinguishable fuzzy sets are helpful in lessening the eccentric behaviours of local linear models in an indirect way, but the interactions of these TS local linear models with global model can not be much improved in some cases where the used modeling method does not aim at making the TS local models match the global model in their corresponding regions. It should be noted that there are other senses of TS fuzzy model interpretability [29][30]. But the methods developed in [29][30] for improving TS fuzzy model interpretability did not consider the TS local model interpretation in the sense of Definition 1.

The aim of this paper is to improve the interpretability of TS local models with regard to the interactions between global model and local models as addressed in Definition 1. Specifically speaking, in order to generate distinguishable fuzzy sets and obtain good local model interpretability for a TS model, a linguistic modifier is proposed to characterize MFs whose centers and shapes can be updated automatically. The linguistic modifier has the ability to enlarge ε-insensitive core of a fuzzy set and at the same time lessen the overlap of adjacent MFs. As MFs become less overlapped and possess larger ε-insensitive core regions, a desirable situation for local model interpretation would emerge: there is only one rule that dominates in a local region and the consequents of fuzzy rules (local models) are forced to represent the global model behaviors in the corresponding local areas. Thus, the eccentric behaviors of local models would be remedied greatly. However, local model interpretability improvement could have a side effect on global model accuracy. In order to control the degree of linguistic modification, as an extension of the fuzziness measure proposed in [35], this paper proposes an index of fuzziness to evaluate the performance of linguistic modification of MFs with adjustable crossover points. This index of fuzziness is then regularized with the global model accuracy in a hybrid objective function, and a tradeoff between global approximation ability and local model interpretation can be achieved by minimizing this hybrid objective function. To further conduct rule base reduction, a pivoted QR decomposition algorithm [36][37] is used to identify the most influential fuzzy rules and remove the redundant ones, which leads to a more parsimonious TS fuzzy model.

3 Linguistic Modifiers as Fuzzy Membership Functions

The core of a fuzzy set is a set of points whose membership degrees are one. However, the sizes of the cores of fuzzy sets usually remain unchanged during adaptiation of membersip functions. In the following, we define an ε-insensitive core of a fuzzy set, which changes along with the process of parameter learning.

The ε-insensitive core of a fuzzy set *A* is defined as

$$VCore_{\varepsilon}(A) = \{x \mid 1 \geq A(x) \geq 1-\varepsilon\} \tag{2}$$

where ε is a small positive real number and $A(x)$ is the MF of *A*.

In order to remedy the eccentric behaviors of local models in a TS fuzzy model, a special MF called linguistic modifier is introduced to simultaneously adjust the overlapping degree of adjacent MFs and the ε-insensitive core of a fuzzy set. Given an initial fuzzy set $A^0(x)$, the modifier produces a new fuzzy set A in a relaxation way as follows:

$$A(x) = \begin{cases} \dfrac{1}{(\mu_{C_1})^{p-1}}\left(A^0(x)\right)^p & x < C_1 \\ 1-\dfrac{1}{\left(1-\mu_{C_1}\right)^{p-1}}\left(1-A^0(x)\right)^p & C_1 \leq x < \beta \\ 1-\dfrac{1}{\left(1-\mu_{C_2}\right)^{p-1}}\left(1-A^0(x)\right)^p & \beta \leq x < C_2 \\ \dfrac{1}{(\mu_{C_2})^{p-1}}\left(A^0(x)\right)^p & C_2 \leq x \end{cases} \tag{3}$$

where $p \geq 1$ is the linguistic modifier parameter to control the fuzziness, C_1 and C_2 are called left and right crossover points of A respectively and their membership degrees are evaluated by the following equations:

$$\mu_{C_1} = A^0(C_1) \text{ and } \mu_{C_2} = A^0(C_2) \tag{4}$$

and β is the core center of A and defined by

$$\beta = \frac{1}{2}\left(\beta_1^0 + \beta_2^0\right) \tag{5}$$

where β_1^0, β_2^0 are the lower and upper bounds of the core of the set $normA^0$ respectively, and $normA^0$ is the norm set of A^0 defined as

$$normA^0(x) = A^0(x) / \sup_x(A^0(x)) \tag{6}$$

The examples of linguistic modifiers are illustrated in *Fig. 2*. It can be proved that for linguistic modifier (3), as p increases, the membership degrees of the points belonging to $(-\infty, C_1)$ or $(C_2, +\infty)$ will decrease, while the membership degrees of the points falling into (C_1, β) or (β, C_2) will increase and approach to 1. Therefore, this relaxation linguistic modifier can adjust the MF's shape by enlarging the ε-insensitive core of the fuzzy set and at the same time reducing the overlap of adjacent MFs, which is a useful property in improving TS fuzzy model interpretability. In this paper, the linguistic modifier will be optimally adjusted by regularizing its fuzziness with global model accuracy.

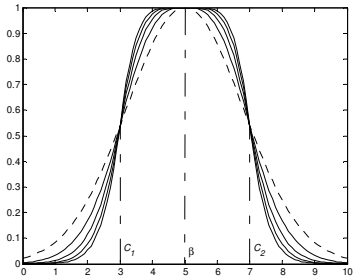

Fig. 2 Linguistic modifiers

It is noteworthy that trapezoidal MFs have the potential of improving local model interpretability by enlarging core regions and narrowing the overlap of adjacent MFs, but by experiments it is found that by using trapezoidal MFs in a TS system, its local models would be easily over-fitted. Although linguistic modifiers with large *p* values seem to approach to a kind of trapezoidal MFs, parameter p provides a flexible way for updating the overlapping of adjacent MFs and the ε-insensitive core regions through a relaxation process. In the extreme case of $p \to \infty$, the fuzzy sets characterized by linguistic modifiers become classic interval sets. Thus, during the interpretation improvement, local models could be restrained from being over-fitted by applying a fuzziness regularization scheme in the learning process.

4. Behavioral Analysis of Submodels in Takagi-Sugeno Fuzzy System Using Linguistic Modifiers as Membership Functions

4.1 Takagi-Sugeno Model Using Linguistic Modifiers as Membership Fuctions

In this paper, the TS model based on the following rules will be addressed:

$$Rule_i: \text{if } x_1 \text{ is } A_{i_1,1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{i_n,n}, \text{ then}$$
$$y_i = a_{i0} + a_{i1}x_1 + \cdots + a_{in}x_n \tag{7}$$

where x_j are input variables, y_i is the output of the *i*th local model, $A_{i_j,j}$ are fuzzy sets on domain x_j, a_{ij} are the consequent parameters that are to be identified based on given data, and $Rule_i$ is the *i*th rule of the TS model. If the number of fuzzy sets on domain x_j is L_j, then $1 \leq i_1 \leq L_1$, ..., $1 \leq i_n \leq L_n$, and $1 \leq i \leq L = \prod_{j=1}^{n} L_j$.

The global output of the TS model is calculated by

$$y = \sum_{i=1}^{L} w_i y_i \tag{8}$$

where w_i is the normalized firing strength of rule $Rule_i$: $w_i = \tau_i / \sum_{l=1}^{L} \tau_l$, and τ_i is called the firing strength of rule $Rule_i$, which is defined by

$$\tau_i = \prod_{j=1}^{n} A_{i_j,j}(x_j) \tag{9}$$

It can be seen that in this TS model, given the fuzzy sets about every variable on its domain of discourse, the rule base includes all the possible combinations of these fuzzy sets to cover the whole input space. To represent the rules clearly, we sort the rules as follows: related to a combination of fuzzy sets $A_{i_1,1}$, …, $A_{i_n,n}$, the rule is indexed as the ith rule, where $i = \sum_{j=1}^{n-1}[(i_j - 1)\cdot \prod_{q=j+1}^{n} L_q] + i_n$.

In this paper, the MFs of fuzzy sets $A_{i_j,j}$ are chosen to be the linguistic modifiers defined by (3), i.e.,

$$A_{i_j,j}(x_j) = A_{i_j,j}(x_j; C^{(1)}_{i_j,j}, \beta_{i_j,j}, C^{(2)}_{i_j,j}, p_{i_j,j}) \tag{10}$$

In the following subsection, we will analyze the local model behaviors and discuss why linguistic modifiers have the potential of improving local model interpretability.

4.2 Local Model Behaviors and Model Interpretability

The behaviors of local models can be characterized by the consequent parameters, while the behaviors of global model can be described by the derivative of the model output with respect to (w.r.t.) model input. This subsection will analyze the relationship between local behaviors and global behaviors. The derivative of the TS model output w.r.t. model input is as follows:

$$\frac{\partial y}{\partial x} = \sum_{i=1}^{L} w_i a_i + \sum_{i=1}^{L} a_i^T x \frac{\partial w_i}{\partial x} \tag{11}$$

where $a_i = (a_{i0}, a_{i1}, \cdots, a_{in})^T$ and $x = (1, x_1, \cdots, x_n)^T$. It can be seen from (11) that the global model behaviors depend on both local model behaviors characterized by vectors a_i and the variation of firing strength. Hence, it is possible to achieve good local model interpretability, i.e., the match of local model behaviors with global model behaviors in specific local regions, by controlling the variation of firing strength. The factors directly affecting w_i and $\partial w_i / \partial x$ are the size of core regions and the size of overlapping regions of adjacent fuzzy sets. The following theorem about partition of unity is useful to the analysis of the influence of local model behaviors on the global model behaviors.

Theorem 1: *In the input space partitioning by* $A_{i_j,j}$ $(j = 1, \cdots, n;\ 1 \le i_1 \le L_1;\ \ldots,\ 1 \le i_n \le L_n)$, *for any given sample input* $x^0 = (x_1^0, \cdots, x_n^0)^T$, *if there exist two adjacent fuzzy sets* $A_{i_j',j}$ *and* $A_{i_j'',j}$ *on each domain* x_j *such that* $A_{i_j',j}(x_j^0) + A_{i_j'',j}(x_j^0) = 1$ *and* $A_{i_j,j}(x_j^0) = 0$ $(i_j \ne i_j', i_j'')$, $j = 1, \cdots, n$, *then*

$$\sum_{i=1}^{L} \prod_{j=1}^{n} A_{i_j,j}(x^0) = 1 \tag{12}$$

where $L = \prod_{j=1}^{n} L_j$, $i_j'' = i_j' + 1$.

Proof. Please see the Supplemental Material. □

In this paper, the MFs of fuzzy sets $A_{i_j,j}$ are obtained in terms of (3) with $C^{(1)}_{i_j,j}, \beta_{i_j,j}, C^{(2)}_{i_j,j}, p_{i_j,j}$, in which $A^0_{i_j,j}(x_j) = \exp\left(-\left(x_j - \beta_{i_j,j}\right)^2 / \left(2\sigma^2_{i_j,j}\right)\right)$, the left and right crossover points, $C^{(1)}_{i_j,j}$ and $C^{(2)}_{i_j,j}$, are defined as $C^{(1)}_{i_j,j} = \left(\sigma_{i_j-1,j}\beta_{i_j,j} + \sigma_{i_j,j}\beta_{i_j-1,j}\right) / \left(\sigma_{i_j-1,j} + \sigma_{i_j,j}\right)$ and $C^{(2)}_{i_j,j} = \left(\sigma_{i_j,j}\beta_{i_j+1,j} + \sigma_{i_j+1,j}\beta_{i_j,j}\right) / \left(\sigma_{i_j,j} + \sigma_{i_j+1,j}\right)$ separately, while $\mu_{C^{(1)}_{i_j,j}}$ and $\mu_{C^{(2)}_{i_j,j}}$ are calculated by $\mu_{C^{(1)}_{i_j,j}} = A^0_{i_j,j}(C^{(1)}_{i_j,j})$ and $\mu_{C^{(2)}_{i_j,j}} = A^0_{i_j,j}(C^{(2)}_{i_j,j})$. Hence, in $A_{i_j,j}$, only linguistic parameter $p_{i_j,j}$ and the centre $\beta_{i_j,j}$ are updated by a training process.

Let us move on to the analysis of local model behaviors. First consider w_i and the first term on the right side of (11). By using the linguistic modifiers, for any input $x = (x_1, \cdots, x_n)^T$, there always exists one fuzzy set $A_{i_j^0,j}$ on domain x_j such that $C^{(1)}_{i_j^0,j} \le x_j < C^{(2)}_{i_j^0,j}$, thus $\lim_{p_{i_j^0,j} \to \infty} A_{i_j^0,j}(x_j) = 1$ and $\lim_{p_{i_j^1,j} \to \infty} A_{i_j^1,j}(x_j) = 0$ $(i_j^1 \ne i_j^0)$. Therefore, in terms of (9) the firing strength of the $i_0 th$ rule will approach to 1 as $p_{i_j^0,j}$ increases, i.e.,

$$\lim_{\substack{p_{i_j^0,j} \to \infty \\ (j=1,\cdots,n)}} \tau_{i_0} = 1 \tag{13}$$

where $i_0 = \sum_{j=1}^{n-1}[(i_j^0 - 1)\cdot \prod_{q=j+1}^{n} L_q] + i_n^0$. Because $\lim_{p_{i_j^0,j} \to \infty, p_{i_j^1,j} \to \infty} \left(A_{i_j^0,j}(x_j) + A_{i_j^1,j}(x_j)\right) = 1$ $(i_j^1 \ne i_j^0)$ $(j = 1,$

$\cdots, n)$, in terms of Theorem 1, we have

$$\lim_{\substack{p_{i_j,j}\to\infty \\ (j=1,\cdots,n)}} \left(\sum_{i=1}^{L}\tau_i(x)\right)=1 \tag{14}$$

and

$$\lim_{\substack{p_{i_0 j,j}\to\infty \\ (j=1,\cdots,n)}} w_{i_0}=1\ ,\ \lim_{\substack{p_{i_j,j}\to\infty \\ (j=1,\cdots,n)}} w_i=0, i\neq i_0 \tag{15}$$

From (11), it is clear that when parameters $p_{i_j,j}$ increase to some extent, there is only one local model, characterized by a_{i_o}, dominating the first term on the right side of (11).

Now consider $\partial w_i/\partial x$ and the second term on the right side of (11). $\partial w_i/\partial x$ can be derived as follows:

$$\partial w_i/\partial x_j=\left(\sum_{l=1}^{L}\tau_l\frac{\partial\tau_i}{\partial x_j}-\tau_i\sum_{l=1}^{L}\frac{\partial\tau_l}{\partial x_j}\right)\Big/\left(\sum_{l=1}^{L}\tau_l\right)^2 \tag{16}$$

$$\partial\tau_i/\partial x_j=\prod_{q\neq j}A_{i_q,q}(x_q)\partial A_{i_j,j}(x_j)/\partial x_j \tag{17}$$

$$\partial w_i/\partial x_j=\Omega_{ij}\Big/\left(\sum_{l=1}^{L}\tau_l\right)^2 \tag{18}$$

where

$$\begin{aligned}\Omega_{ij}=&\left(\prod_{q\neq j}A_{i_q,q}(x_q)\partial A_{i_j,j}(x_j)/\partial x_j\right)\left(\sum_{l=1}^{L}\tau_l\right)-\\ &\tau_i\sum_{i_1=1}^{L_1}\cdots\sum_{i_n=1}^{L_n}\left(\prod_{q\neq j}A_{i_q,q}(x_q)\partial A_{i_j,j}(x_j)/\partial x_j\right)\end{aligned} \tag{19}$$

and $\partial A_{i_j,j}(x_j)/\partial x_j$ will take different values in different regions. For $x_j<C_{i_j,j}^{(1)}$,

$$\partial A_{i_j,j}(x_j)/\partial x_j=p_{i_j,j}\left(\frac{A_{i_j,j}^0(x_j)}{\mu_{C_{i_j,j}^{(1)}}}\right)^{p_{i_j,j}-1}\partial A_{i_j,j}^0(x_j)/\partial x_j \tag{20}$$

For $C_{i_j,j}^{(1)}\le x_j<\beta_{i_j,j}$,

$$\partial A_{i_j,j}(x_j)/\partial x_j=p_{i_j,j}\left(\frac{1-A_{i_j,j}^0(x_j)}{1-\mu_{C_{i_j,j}^{(1)}}}\right)^{p_{i_j,j}-1}\frac{\partial A_{i_j,j}^0(x_j)}{\partial x_j} \tag{21}$$

For $\beta_{i_j,j}<x_j<C_{i_j,j}^{(2)}$,

$$\partial A_{i_j,j}(x_j)/\partial x_j=p_{i_j,j}\left(\frac{1-A_{i_j,j}^0(x_j)}{1-\mu_{C_{i_j,j}^{(2)}}}\right)^{p_{i_j,j}-1}\frac{\partial A_{i_j,j}^0(x_j)}{\partial x_j} \tag{22}$$

For $C_{i_j,j}^{(2)}\le x_j$,

$$\partial A_{i_j,j}(x_j)/\partial x_j=p_{i_j,j}\left(\frac{A_{i_j,j}^0(x_j)}{\mu_{C_{i_j,j}^{(2)}}}\right)^{p_{i_j,j}-1}\frac{\partial A_{i_j,j}^0(x_j)}{\partial x_j} \tag{23}$$

For $x_j=\beta_{i_j,j}$,

$$\partial A_{i_j,j}(x_j)/\partial x_j=0 \tag{24}$$

Because $\lim_{p_{ij,j}\to\infty}p_{i_j,j}\alpha^{p_{ij,j}-1}=0$ if $|\alpha|<1$, and the absolute values of all the terms in the brackets of (21)-(24) are less than 1, so $\lim_{p_{i_j,j}\to\infty}\partial A_{i_j,j}(x_j)/\partial x_j=0$ holds. Then, $\lim_{p_{i_j,j}\to\infty}\partial w_i/\partial x_j=0$ is true. In other words, as parameters $p_{i_j,j}$ increase to some extent, the influence of the second term on the right side of (11) will become weak, and the global model behaviors will tend to be dominated by a single local model characterized by a_{i_o}. This is a good local model interpretation expected in the subsequent model applications such as state estimation and control.

However, it should be noted that when the MFs become less overlapped and have large core regions, the global approximation ability of the TS model would be degraded. In the next section a learning scheme is proposed to balance the model accuracy and interpretability based on a combined performance measure, so that the linguistic modifier parameters can be optimally adjusted.

5 A Learning Algorithm Based on Fuzziness Regularization

5.1 Fuzziness Measure of a Fuzzy Set

The proposed fuzziness measure is based on the distance between a fuzzy set *A* and an ordinary (crisp) set $\underline{A}$ that is near to *A* and defined as follows:

$$\underline{A}(x)=\begin{cases}1 & if\ C^{(1)}\le x\le C^{(2)}\\ 0 & otherwise\end{cases} \tag{25}$$

where $C^{(1)}$ and $C^{(2)}$ are the left and right crossover points of fuzzy set *A*, respectively. Given a data set $\{x(k)\}_{k=1}^{N}$ on domain *x*, the index of fuzziness of *A* is defined based on the distance between *A* and $\underline{A}$ as follows:

$$F(A)=\frac{2}{N^{1/r}}\left\|A(x)-\underline{A}(x)\right\|_r \tag{26}$$

where *N* is the size of the data set and *r* is the order of the distance. It can be easily proved that $F(A)\ge F(A^*)$, where A^* is a sharper version of *A* in the sense that

$$\begin{aligned} &A^*(x) \geq A(x), \quad if\ C^{(1)} \leq x \leq C^{(2)} \\ &A^*(x) \leq A(x), \qquad otherwise \end{aligned} \tag{27}$$

Obviously, in case $A(C^{(1)}) = A(C^{(2)}) = 0.5$, $F(A)$ becomes the classic fuzziness measure proposed in [35] and (27) is reduced to one of the properties proposed by De Luca and Termini for fuzziness measures of fuzzy sets [40]. In particular, when *r*=2 (Euclidean distance is used), (26) defines a *quadratic index of fuzziness*:

$$F_q(A) = \frac{2}{\sqrt{N}} \sqrt{\sum_{k=1}^{N} \left(A(x(k)) - \underline{A}(x(k))\right)^2} \tag{28}$$

which will be used in the proposed learning algorithm in this paper.

5.2 Fuzziness Index Regularization and the Learning Algorithm

For a given data set $\{(x(k), d(k))\}_{k=1}^{N}$, a hybrid learning scheme is employed to update both the consequent parameters a_{ij} and the premise parameters $\beta_{i_j,j}$ and $p_{i_j,j}$. In the first pass, the premise parameters $\beta_{i_j,j}$ and $p_{i_j,j}$ are fixed, and the consequent parameters a_{ij} are identified by least square estimation in terms of the global model accuracy measure. In the second pass, the newly obtained consequent parameters a_{ij} are fixed, and the premise parameters $\beta_{i_j,j}$ and $p_{i_j,j}$ are updated by a gradient descent algorithm in terms of the following combined performance measure:

$$J = J_G + \lambda J_F \tag{29}$$

where $0 \leq \lambda$ is the regularization coefficient, J_G is the global accuracy measure defined as

$$J_G = \frac{1}{2} \sum_{k=1}^{N} \left\| d(k) - y(k) \right\|^2 \tag{30}$$

and J_F is the index of fuzziness of the TS fuzzy model, defined by

$$J_F = \sum_{i_1=1}^{L_1} \cdots \sum_{i_n=1}^{L_n} \sum_{j=1}^{n} F(A_{i_j,j}) \tag{31}$$

where $F(A_{i_j,j})$ is the *quadratic index of fuzziness* defined by (28). It can be seen that the updating of the parameters depends not only on the global model accuracy, but also on how much the degree of fuzziness of the linguistic modifiers is.

1) To update consequent parameters: In order to identify the consequent parameters in the TS model, we reformulate some expressions first. A base matrix *M* is defined as follows:

$$M = \begin{bmatrix} M_1^T(1) & \cdots & M_L^T(1) \\ & & \\ M_1^T(N) & \cdots & M_L^T(N) \end{bmatrix}_{N \times L(n+1)} \tag{32}$$

where $M_i^T = (w_i \ w_i x_1 \cdots w_i x_n)$. Let the *k*th row vector of matrix M be $M^T(k) = (M_1^T(k) \cdots M_L^T(k))$, then $M^T = [M(1) \cdots M(N)]$. Let $a = (a_1^T \ a_2^T \cdots a_L^T)^T$ denote the consequent parameters, where $a_i = (a_{i0} \ a_{i1} \cdots a_{in})^T$. Also let $d = (d(1) \cdots d(N))^T$ be the desired output vector. Because the consequent parameters in a do not make any contribution to the index of fuzziness of the TS model, they can be identified practically based on the global approximation accuracy measure J_G. In terms of (8) we have

$$M \cdot a = d \tag{33}$$

where the dimensions of M, a and d are $N \times L \cdot (n+1)$, $L \cdot (n+1) \times 1$ and $N \times 1$ respectively. Since the number of training pattern pairs is usually greater than $L \cdot (n+1)$, (33) defines a typical ill-posed problem and generally there does not exist an exact solution for vector a if there is no regularization information about a added to the global approximation accuracy J_G. Therefore, we usually seek a least square estimate of a to minimize J_G. The optimal estimate a^* can be obtained by

$$a^* = M^+ d \tag{34}$$

where M^+ is the Moore-Penrose inverse of matrix M. When M is of column full rank, the Moore-Penrose inverse of matrix M can be expressed as $M^+ = (M^T M)^{-1} M^T$. In case of singularity of $M^T M$, $M^+ = (M^T M + I)^{-1} M^T$ with identity matrix I.

2) To update premise parameters: The premise parameters $\{\beta_{i_j,j}\}$ and $\{p_{i_j,j}\}$ are updated in terms of the combined objective function defined in (29), which aims at achieving a good trade-off between global approximation ability and local model interpretability. The equations for updating the premise parameters are developed as follows:

$$\nu_{i_j,j}(t+1) = \nu_{i_j,j}(t) - \rho_{i_j,j} \frac{\partial J}{\partial \nu_{i_j,j}} \tag{35}$$

where *t* is the iteration step, $\rho_{i_j,j}$ is the learning rate, $\nu_{i_j,j} = \beta_{i_j,j}$ or $p_{i_j,j}$ representing the premise parameters[1]. In order to keep $p_{i_j,j} > 1$ during adaptation, the following transformation is used to indirectly update

[1] Please see the Support Material for the computing results of these partial derivatives.

$p_{i_j,j}$ by adjusting $u_{i_j,j}$:

$$u_{i_j,j} = \log(p_{i_j,j} - 1) \tag{36}$$

$$\frac{\partial}{\partial u_{i_j,j}} = (p_{i_j,j} - 1)\frac{\partial}{\partial p_{i_j,j}} \tag{37}$$

To speed up the learning process, the following adaptive learning rates are adopted in our experiments:

$$\rho_{i_j,j} = \kappa / \sqrt{\left(\partial J_G / \partial \nu_{i_j,j}\right)^2 + \left(\partial J_F / \partial \nu_{i_j,j}\right)^2} \tag{38}$$

where $\nu_{i_j,j} = \beta_{i_j,j}$ or $u_{i_j,j}$, and $\kappa(>0)$ is the step size indicating the length of each gradient transition in the parameter space.

5.3 Rule Selection with Pivoted QR Decomposition of Firing Strength Matrix

In fuzzy modeling, it is very important to partition the input space optimally in terms of given criteria. In order to attack the curse of dimensionality in fuzzy modeling, fuzzy rule selection is usually performed [3][8][12][41]. The approach to fuzzy rule selection used in this paper involves the estimation of singular values of the firing strength matrix W. Each column of matrix W corresponds to one fuzzy rule. The important fuzzy rules correspond to the columns of the matrix that are linearly independent of each other [8]. One method to pick up the most influential fuzzy rules is to apply the SVD-QR with column pivoting algorithm to W [36]. As indicated in [8], redundant fuzzy partitions (corresponding to the linear dependent or zero-valued columns) are associated with near zero singular values of W. The smaller are the singular values, the less influential are the associated fuzzy rules. However, the rule ranking result by the SVD-QR with column pivoting algorithm is heavily dependent on the estimation of an effective rank [36]. The problem is that there is usually no clear gap between small singular values and other "large" singular values, and different ranks often produce dramatically different rule ranking results. A method to avoid the estimation of the effective rank is to apply the pivoted QR decomposition [37] directly to matrix W. The pivoted QR decomposition algorithm for ranking fuzzy rules is summarized as follows:

1) Calculate the QR decomposition of W and get the permutation matrix Π via $W\Pi = QR$, where Q is an unitary matrix, R is an upper triangular matrix. The absolute values of the diagonal elements of R, denoted as $|R_{ii}|$, decrease as i increases and are named as R-values;

2) Rank fuzzy rules in terms of the R-values and the permutation matrix Π in which each column has one element taking value 1 and all the other elements taking value 0. Each column of Π corresponds to a fuzzy rule. The numbering of the rule that corresponds to the jth column is the same as the numbering of the row where the "1" element of the jth column is located. The rule corresponding to the first column is the most important, and in descending order the rule corresponding to the last column is the least important.

It is indicated that the R-values tend to track the singular values of W, hence they can be used to identify the influential rules. In this paper we use the R-values of matrix W to perform the rule ranking for TS fuzzy model by applying the pivoted QR decomposition algorithm.

5.4 Learning Scheme Implementation

To summarize, the proposed learning scheme for improving TS local model interpretability is described as follows:

Step 1. Initialize the TS model.

1.1) Initialize the input space partitioning, for example via unsupervised clustering on input-output data set.

1.2) Construct the linguistic modifiers in terms of the initially generated fuzzy sets, and set the initial modifier parameters $p_{i_j,j}$ ($p_{i_j,j} > 1$).

1.3) Set a threshold J_0 for stopping learning process and a threshold fs_0 for stopping rule selection.

1.4) Choose a value of regularization coefficient λ.

1.5) Choose a value of learning step $\kappa(>0)$.

Step 2. Identify the consequent parameters using least square estimation while keeping the premise parameters fixed.

Step 3. Update the premise parameters $\beta_{i_j,j}, p_{i_j,j}$ using equation (35), while the consequent parameters obtained in step 2 remain unchanged here.

Step 4. Calculate the combined performance measure (29), and go to step 2 until $J \le J_0$.

Step 5. Select the most important rules by applying the pivoted QR decomposition algorithm to the generated rule base.

6 Experimental Results

In this section we extensively evaluate the performance of the proposed learning scheme for constructing interpretable TS fuzzy models with satisfactory global model accuracy, in which statistical tests are conducted. Two existing interpretability-oriented fuzzy modeling methods, the PBM method [33][34] and the ComGLL scheme[9], are compared with the proposed approach. The first example is to recover the original signal from data highly contaminated by noise. For the sake of visualizing the interactions of TS local linear models with global model in a 2D plot, in the first example a TS fuzzy model with only one input variable and one output variable is considered due to the fact that it is impossible to visualize TS local linear models clearly in a 3D or higher dimensional plot. The statistical test method, *t-test*, is used to evaluate the performance of the proposed scheme in comparison with the two existing related methods. The second example involves a real-world problem, in which a TS fuzzy model with four input variables and one output variable is considered to

predict the steam heat exchange. In the second example, the generalization performances of the constructed TS fuzzy models will be evaluated by generating distinguishable input space partitions. The third example is to construct an interpretable TS model to identify the voltage time series produced by a nonlinear circuit. Global model accuracy is measured by root-mean-square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\big(d(k) - y(x(k))\big)^2} \tag{39}$$

where d is the desired output and y the real output of the constructed global model, whilst the extent of local models approaching to the global model locally is measured by the TS local model performance function (1) in the following form

$$RMSLIV = \sqrt{\sum_{i=1}^{L}\sum_{x(k)\in N^{(i)}} w_i(k)\big(y_i(k) - y(k)\big)^2} \tag{40}$$

where $y(k)$ is the global model output (8), $N^{(i)}$ is the neighborhood of the core center of the multidimensional MF, i.e., τ_i defined by (9). Eq. (40) is called TS root-mean-square local model index value (*RMSLIV*). According to (1) , a smaller *RMSLIV* value implies that the TS local models fit the global model better in the corresponding local regions, which indicates that better local model interpretability in the sense of Definition 1 can be preserved.

6.1 Noisy Signal Recovery

In the first example, the noisy signal is generated by

$$z = \begin{cases} \tilde{z} + n_1 & x \le 13 \\ \tilde{z} + n_2 & otherwise \end{cases} \tag{41}$$

$$\tilde{z} = 10^3(1-\cos(2\pi x/5))\sin(2\pi x)e^{-x/2} \tag{42}$$

where $\tilde{z}$ is the original signal, n_1 is a random Gaussian noise with zero mean and variance $\sigma_1^2 = 0.2$, and n_2 is another random Gaussian noise with zero mean and variance $\sigma_2^2 = 0.6$. The measured signal z is the sum of the original signal $\tilde{z}$ and the interference noise n_1 and n_2. However, we do not know the original signal $\tilde{z}$. The only signal available to us is the measured signal z. The task is to recover the original signal $\tilde{z}$ from the measured signal z by constructing an interpretable TS fuzzy model. Although the modelled system in this example seems simple with one input and one output, it is known in the signal processing community that it is rather challenging to recover original signal from data highly contaminated by noise withoug prior knowledge.

In order to extensively evaluate the performance of the proposed TS modeling method, 10 TS fuzzy models are constructed on 10 different datasets generated by running the data generation process (41) and (42) 10 times, each with 400 samples $\left\{\left(x^{(s)}(k), d^{(s)}(k)\right)\right\}_{k=1}^{N}$ (N=400, $s = 1,\cdots,10$), where $x^{(s)}(k) \in$ [11.5, 15.5] and $d^{(s)}(k)$ are obtained by (43). Furthermore, the proposed method is compared with the interpretability-oriented fuzzy modeling methods: the PBM method [33][34] and the ComGLL scheme [9]. The PBM method can be used to improve both Mamdani fuzzy model interpretability and TS fuzzy model interpretability whilst the ComGLL scheme specially aims at improving TS local model interpretability. The performance of the proposed scheme is then evaluated via *t-test* on the 10 TS models in terms of global model error-*RMSE* and local model interaction value *RMSLIV* respectively.

First, the input space should be initialised by an unsupervised clustering algorithm. In our experiments, the normalized kernel based FCM (NKFCM) clustering algorithm [42] is used. For each run, given the input-output data samples $\left\{\left(x^{(s)}(k), d^{(s)}(k)\right)\right\}_{k=1}^{N}$, the NKFCM algorithm generates fuzzy clusters according to a partition entropy measure [38]. These cluster centers on x domain are used as the core centers of initial fuzzy sets. The width of the linguistic modifiers, which determines the crossover points, is estimated by a nearest neighbor heuristic suggested by Moody and Daken [43]. In updating MFs and the local models, the learning steps are set as $\kappa = 0.1$ for $\beta_{i_j,j}$ and $\kappa = 1$ for $u_{i_j,j}$ in (40), and initial values of $p_{i_j,j}$ are all set as 1.1.

Ten TS fuzzy models were constructed by the proposed method on the 10 datasets. As a comparison, the ComGLL scheme is also used to construct 10 TS fuzzy models based on the same 10 datasets. This scheme combines global learning and local learning by a global influence factor α and a local influence factor β with $\alpha + \beta = 1$. Using a smaller β the global model accuracy can be improved by the ComGLL scheme, but the local model interpretability will get worse, while a larger β leads to local models with better interpretability in the sense of Definition 1, but degrades global model accuracy. Different parameter values including $\beta = 0.1$ and 0.6 are separately used to evaluate the performance of the ComGLL scheme. Furthermore, the PBM method [33], another interpretability-oriented modeling scheme, is also used to improve the interpretability of the fuzzy models constructed by the initial fuzzy rules on the 10 same datasets.

Table 1 shows the experimental results by averaging the performances of the 10 TS models constructed by the proposed method, the ComGLL ($\beta = 0.6$), ComGLL ($\beta = 0.1$), and the PBM method respectively, in terms of global model (GM) accuracy index *RMSE* and local model (LM) interaction index *RMSLIV*. These averaging results indicate the good performance of the proposed method in producing interpretable TS fuzzy models whilst keeping

global model accuracy at a satisfactory level. In order to check whether the average differences in the performances of different modeling schemes are significant, a *t-test* between the proposed method and other known methods is conducted in the following.

TABLE 1. Average performances of the proposed method and the known ones in the first example

Algorithms	*GM Average RMSE*	*LM Average RMSLIV*
The proposed ($\lambda = 0.6$)	0.6678	0.1376
ComGLL ($\beta = 0.6$)	1.2089	0.3744
ComGLL ($\beta = 0.1$)	0.8458	0.5415
PBM	0.6283	3.1438

The first step of *t-test* is to specify a null hypothesis and an alternative hypothesis. In our experiments of testing differences between the average performances of the proposed method and other methods, the null hypothesis "the difference between means is zero" is used, i.e.,

$$H_0 : \mu_{new} - \mu_{other} = 0$$
$$H_1 : \mu_{new} - \mu_{other} \neq 0 \quad (43)$$

where μ_{new}, μ_{other} are the means of the results achieved by the proposed method and other method respectively, H_1 is a specified alternative hypothesis. The second step is to choose a significance level for the *t-test*. Usually, the significance level is chosen as α_0 =0.05. Given the samples, the *t*-value is calculated and a *p*-value can be determined according to the Student's *t*-distribution. If the *p*-value is less than α_0, then the null hypothesis H_0 is rejected, and the alternate hypothesis H_1 is accepted. Small *p*-value casts doubt on the validity of the null hypothesis. However, if the *p*-value was greater than the α_0 level, the hypothesis H_0 would be retained.

TABLE 2. *T-test* of the global model performances of the proposed method and the known ones in the first example (α_0 =0.05)

Algorithms	***t-value***	***p-value***	***Decision***
The proposed vs ComGLL ($\beta = 0.6$)	-19.5628	1.4103e-013	H_0 rejected
The proposed vs ComGLL ($\beta = 0.1$)	-5.3545	5.2557e-005	H_0 rejected
The proposed vs PBM	2.9805	0.0080	H_0 rejected

Table 2 illustrates the *t-test* results about the TS global model performances achieved by applying the proposed method and the known ones to the 10 datasets. For example, in the *t-test* of difference between the proposed method and ComGLL($\beta = 0.6$), the probability value (1.4103e-013) is less than the significance level (0.05), which implies that the difference between the two means is significant, so the null hypothesis H_0 is rejected. It is concluded that the average *RSME* (1.2089) by the ComGLL($\beta = 0.6$) is really higher than the one (0.6678) by the proposed method. Similar conclusions about the global model performances of the proposed method vs. other methods can also be drawn according to the *t-test* results in *Table* 2.

TABLE 3. *T-test* of the local model performances of the proposed method and the known ones in the first example (α_0 =0.05)

Algorithms	***t-value***	***p-value***	***Decision***
The proposed vs ComGLL ($\beta = 0.6$)	-25.2731	1.6393e-015	H_0 rejected
The proposed vs ComGLL ($\beta = 0.1$)	-33.7096	1.0181e-017	H_0 rejected
The proposed vs PBM	-4.9254	1.0914e-004	H_0 rejected

Furthermore, *t-test* is carried out to evaluate the differences of local model performances in terms of the *RMSLIV* values achieved by the proposed method and other known methods, *Table* 3 gives the corresponding experimental results. For example, in the *t-test* of difference between the proposed method and the PBM scheme, the probability value (1.0914e-004) is less than the significance level (0.05), which implies that the difference between the two *RMSLIV* means is significant, so the null hypothesis H_0 is rejected. Thus, it is concluded that the average *RMSLIV* (3.1438) achieved by the PBM scheme is really higher than the one (0.1376) by the proposed method. Similar conclusions about the local model performances of the proposed method vs. other methods can also be reached according to the *t-test* results in the *Table* 3. To summarize, the proposed method outperforms the ComGLL and PBM schemes in producing interpretable TS local models in terms of Definition 1. Particularly, both global model accuracy and local model interpretability obtained by the ComGLL scheme are worse than the proposed method. One possible reason is that the ComGLL learning scheme, specially aiming at improving TS local model interpretability, does not optimally adjust MFs to make the local models dominate the local behaviors of the system.

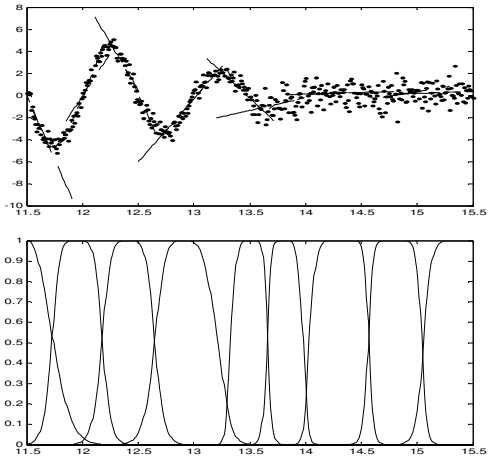

Fig. 3 TS model with local models (top) and corresponding MFs (bottom) obtained by the proposed method: (top)-dotted lines represent the global models (recovered signal), dashed lines represent the local models

Promisingly, the distinguishability of the input space partition generated by the proposed method is improved simultaneously due to the use of linguistic modifiers. *Fig. 3* shows one TS model with local models and the corresponding MFs produced by the proposed method. As a comparison, *Fig. 4* illustrates the TS model with local models and the corresponding MFs generated by the PBM method applying to the same dataset as the one used in *Fig. 3*, which indicates that the input space partitions obtained by the proposed method possess better distinguishability than the ones achieved by the PBM method, and the TS submodes achieved by the proposed method exhibit better interpretabtion of the interactions with the global model than the ones obtained by the the PBM method. Because the ComGLL does not consider the improvement of the distinguishability of input space partitions, the input space partition is not illustrated here.

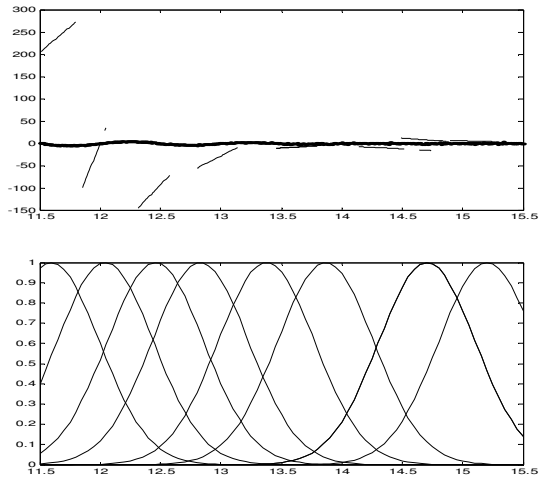

Fig. 4 TS model with local models (top) and corresponding MFs (bottom) obtained by the PBM method: (top)-dotted lines represent the global models (recovered signal), and dashed lines represent the local models

Next, the performance of the pivoted QR decomposition algorithm for selecting the most important fuzzy rules is evaluated in comparison with the SVD-QR with column pivoted method [36][37]. The two methods are separately applied to the firing strength matrices of the 10 TS fuzzy models produced by the proposed method. *Table* 4 summarizes the average *RMSEs* and *RMSLIVs* of the TS models constructed by 8 most important fuzzy rules selected in terms of the R-values and singular values of fuzzy rules, in which the R-values of fuzzy rules are obtained by the pivoted QR decomposition algorithm, whilst the singular values of fuzzy rules are calculated by the SVD-QR with column pivoting method (*r=4*). In order to test whether the difference between the two rule ranking methods over the available data sets is non-random, the *t-test* is performed, in which the null hypothesis "the difference between means is zero" is used, i.e.,

$$\begin{aligned} H_0 &: \mu_{qr} - \mu_{svd-qr} = 0 \\ H_1 &: \mu_{qr} - \mu_{svd-qr} \neq 0 \end{aligned} \tag{44}$$

where μ_{qr}, μ_{svd-qr} are the means of the results achieved by the QR method and the SVD-QR method respectively, H_1 is the specified alternative hypothesis, and the significance level is chosen as α_0 =0.05.

TABLE 4. Average performances of the QR method and the SVD-QR with column pivoted method in the first example

Algorithms	*GMl Average RMSE*	*LM Average RMSLIV*
QR	0.8352	0.1474
SVD-QR	1.5200	0.1493

TABLE 5. *T-test* of the global model and local model performances of QR method and SVD-QR method in the first example (α_0 =0.05)

Global/local models	*t-value*	*p-value*	*Decision*
GM performance	-6.1107	8.9913e-006	H_0 rejected
LM performance	-0.1873	0.8535	H_0 retained

Table 5 illustrates the *t-test* results on global model performances and local model performances achieved by the two rule selection methods. In the *t-test* of difference between the QR method and SVD-QR method on global model performance, the probability value (8.9913e-006) is less than the significance level (0.05), which implies that the difference between the two *RMSE* means is significant, so the null hypothesis H_0 is rejected. The average *RMSE* (0.8352) achieved by the pivoted QR method is significantly smaller than the one (1.5200) achieved by the SVD-QR with column pivoting method. However, in the *t-test* of difference between the two methods on local model performance, the probability value (0.8535) is greater than the significance level (0.05), which implies that the difference between the two *RMSLIV* means is not significant, so the null hypothesis H_0 is retained. That is to say, given the available TS models constructed by the proposed method, the pivoted QR method is comparable with the SVD-QR with column pivoting method in further improving the TS local model interpretability. To summarize, when applied to the TS models constructed by the proposed method, the pivoted QR method can achieve significantly better global model accuracy than the SVD-QR with column pivoting method, but the two methods achieve the same level of performance in further improving the TS local model interpretability.

Now let us see whether the practically generated TS models can verify the above claim. *Fig. 5* shows two TS models constructed by the 8 most important rules in terms of *R*-values and singular values respectively, which clearly indicate that the SVD-QR with column pivoted method greatly degrades the global model accuracy, but there is not much difference of local model interpretations between the two TS models. These results also justify that the *RMSLIV* really possesses the capability of characterizing the status of TS local model interaction with global model.

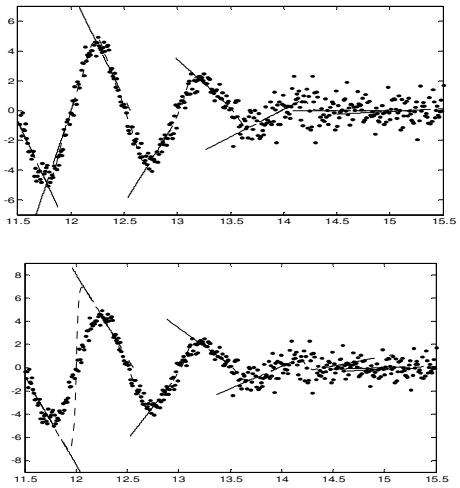

Fig. 5 The TS models and local models obtained by 8 important rules selected by the pivoted QR method (top) and SVD-QR with column pivoting method (bottom): dotted line represents the global model (recovered signal), and dashed lines represent the local models

6.2 Steam Heat Exchanger

The second example considers a liquid-saturated steam heat exchanger [44], where water is heated by pressurized saturated steam through a copper tube. The process plant is shown in *Fig. 6*, in which the output is the outlet liquid temperature, and the inputs are the liquid flow rate, the steam temperature, and the inlet liquid temperature. In this experiment the steam temperature and the inlet liquid temperature are kept constant to their nominal values, so only the liquid flow rate is considered as plant input variable. The main motivation for the choice of the heat exchange process is that this plant is a significant benchmark for nonlinear control design purposes, because it is characterized by a non-minimum phase behavior which makes the design of suitable controllers particularly challenging even in a linear design context [44]. Hence, it is highly expected to construct TS fuzzy model with good local linear model interpretation to predict the system behaviors.

In our experiment, 1000 heat exchanging samples are used to build up a TS fuzzy system model with 4 inputs, *i.e.*, $v_t = f(v_{t-1}, v_{t-2}, v_{t-3}, u_t)$, where v_t is the outlet liquid temperature at time *t*, and u_t is the liquid flow rate at time *t*. And 10-fold cross validation is used to evaluate the performance of the TS model, which works as follows:

- Divide the 1000 instances into 10 disjoint data subsets, each containing 100 heat exchanging samples;
- Form a testing sample set with each data subset;
- Form a training sample set for every testing set with the remaining 900 instances;
- Train and test the TS fuzzy model by the proposed method using each of the pairs of training and testing sets;
- Record and average the results for the testing sets to determine the model generalization performance, *i.e.*, the RMSEs of predicting the outlet liquid temperatures by the trained TS model on the test samples.

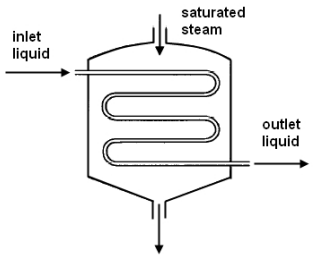

Fig. 6 The steam heat exchanger plant

Firstly, an initial input space partition is obtained by using the NKFCM clustering algorithm [42] on the available input-output samples, which generates 2 clusters as the core centers of initial fuzzy sets. For the 4 input variables $v_{t-1}, v_{t-2}, v_{t-3}$ and u_t, 16 fuzzy rules are generated in the initial TS model. In the experiment, the learning rates for updating the antecedent parameters $\beta_{i_j,j}$ and $u_{i_j,j}$ are adapted dynamically with κ=0.1 and 1 separately, the initial value of $p_{i_j,j}$ is set as 1.1, and λ=0.6 is used to evaluate the performance of the proposed method. The experimental results are summarized in *Table* 6 and *Table* 7, which indicate that the TS models constructed by the proposed method achieve good generalization performance, whilst their local models possess good interpretability in terms of Definition 1. Moreover, the generated fuzzy sets by the linguistic modifiers exhibit good distinguishability in the input space partition as illustrated in *Fig. 7*.

TABLE 6. Global model performance comparisons of the proposed method with the known ones in the second example

Algorithms	*Training RMSE*	*Variance*	*Testing RMSE*	*Variance*
The proposed ($\lambda = 0.6$)	0.2433	0.0052	0.2531	0.0332
ComGLL ($\beta = 0.6$)	16.5659	0.2453	17.1984	3.2568
ComGLL ($\beta = 0.05$)	0.8854	0.0220	0.9296	0.1615
PBM	0.2478	0.0031	0.2617	0.0327

As a comparison, the ComGLL method is also used to construct TS fuzzy model via similar 10-fold cross validation on the same data subsets. The parameter values of β=0.05 and 0.6 are separately used to evaluate the performance of the ComGLL method. The experimental results are summarized in *Table* 6 and *Table* 7 as well. Because the ComGLL scheme does not consider improving the distinguishability of the input space partition, the fuzzy sets used remain unchanged as in the initial partition. Furthermore, the PBM method is used to enhance the distinguishability of the fuzzy model constructed by the 16 initial fuzzy rules. The 10-fold cross validation is also used to evaluate the performance of the PBM method

applying to the same data subsets, and the experimental results are shown in *Table* 6 and *Table* 7 as well. *Fig. 8* illustrates one improved input space partition by the PBM method. It can be seen that the proposed method achieves better performance in improving the TS local model interpretability with satisfactory generalization performances on the steam heat exchanger problem than other known interpretability-oriented fuzzy modeling methods. What is more, the distinguishability of input space partition is simultaneously improved by the proposed method.

TABLE 7. Local model performance comparisons of the proposed method with the known ones in the second example

Algorithms	*Average RMSLIV*
The proposed ($\lambda = 0.6$)	0.1939
ComGLL ($\beta = 0.6$)	0.5217
ComGLL ($\beta = 0.05$)	1.9686
PBM	2.3242

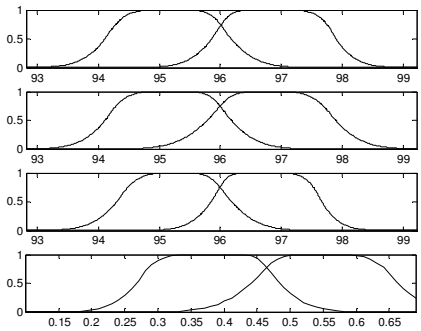

Fig. 7 The MFs generated by the proposed method in the second example

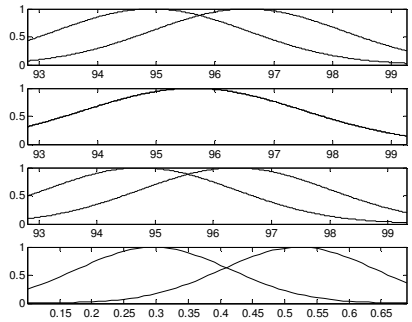

Fig. 8 The MFs generated by the PBM method in the second example

Next, the performance of the pivoted QR decomposition algorithm is evaluated on the steam heat exchanger problem via 10-fold cross validation in comparison with the SVD-QR with column pivoted method. By applying the pivoted QR decomposition algorithm to the firing strength matrix of the 16 fuzzy rules produced by the proposed method, the *R*-values are obtained to rank the fuzzy rules as illustrated in *Table* 8. The rule ranking results obtained by the SVD-QR with column pivoted method are also depicted in *Table* 8, which indicate that the SVD-QR with column pivoted method heavily depends on the selection of the efficient rank parameter *r*. By setting fs_0 =2.0, 9 most important fuzzy rules are selected in terms of *R*-values to construct a TS model. This TS model achieves good model performance: training *RMSE*=0.2496 with variance 0.0038, testing *RMSE*=0.2578 with variance 0.0381, and average *RMSLIV*=0.2445. However, the TS model constructed in terms of the 9 most important fuzzy rules selected by the SVD-QR with column pivoted method with *r*=6 achieves training *RMSE* =0.8906 with variance 0.5204, testing *RMSE* =1.0659 with variance 0.7235, and average *RMSLIV*=0.2804. From the above results, it can be seen that the pivoted QR decomposition method is more efficient in identifying influential fuzzy rules than the SVD pivoted QR decomposition in constructing parsimonious TS models.

TABLE 8. Rule ranking results by pivoted QR decomposition and SVD-QR with column pivoting in the second example

Methods	*Rule Ranking Results*
QR	15, 16, 1, 2, 7, 8, 3, 4, 13, 14, 9, 10, 11, 12, 5, 6
SVD-QR (r=4)	4, 1, 11, 10, 5, 6, 7, 8, 9, 2, 3, 12, 13, 14, 15, 16
SVD-QR (r=5)	8, 4, 1, 10, 12, 6, 7, 3, 9, 2, 11, 5, 13, 14, 15, 16
SVD-QR (r=6)	8, 9, 4, 1, 12, 10, 7, 3, 2, 6, 11, 5, 13, 14, 15 ,16

6.3 Nonlinear Circuit System

The third example is a benchmark nonlinear circuit creating a time series of voltage. The theoretical model of this circuit is described in [45]. The voltage recordrings from the nonlinear circuit have been collected. The aim is to construct a fuzzy model with good interpretability which is capable of reproducing the voltage time series.

In our experiment, 1000 voltage samples are used to build up a TS fuzzy system model with 4 inputs, *i.e.*, $v_t = f(v_{t-1}, v_{t-2}, v_{t-3}, v_{t-4})$, where v_t is the voltage value at time *t*. And 5-fold cross validation is used to evaluate the performance of the TS model. Similar scheme for initializing the input space was used as before. 2 clusters are generated for the 4 input variables, and thus 16 fuzzy rules are used in the initial TS model. The experimental results are summarized in *Table* 9 and *Table* 10. It can be seen that the TS models constructed in this example by the proposed method achieve good generalization performance, whilst their local models possess good interpretability in terms of Definition 1.

TABLE 9. Global model performance comparisons of the proposed method with the known ones in the third example

Algorithms	*Average Training RMSE*	*Average Testing RMSE*
The proposed ($\lambda = 0.6$)	0.0011	0.0013
ComGLL ($\beta = 0.6$)	0.0446	0.0457
PBM	0.0011	0.0014

TABLE 10. Local model performance comparisons of the proposed method with the known ones in the third example

Algorithms	*Average RMSLIV*
The proposed ($\lambda = 0.6$)	0.2953

ComGLL ($\beta = 0.6$)	0.5611
PBM	1.1307

6.4 Further Discussions

Two issues might arise about the proposed method. One is that facing the same problem, different initial input space partitions may lead to TS models with different local model interpretability. The other is the proposed method involves hyperparameters κ, J_0, and fs_0. In (38), parameter $\kappa(>0)$ is used to determine the speed of the learning algorithm, which is similar to the parameter determining learning rates commonly used in gradient-based machine learning algorithms. Parameter J_0 is used as a threshold or target value of the objective function *J*, which corresponds to the stopping criterion used in machine learning algorithms. Hence, parameters κ and J_0 are the common ways of controlling a machine learning algorithm. Comparatively, only parameter fs_0 is specially designed in the proposed method as a parameter of threshold for most influential fuzzy rules, which makes the constructed TS models parsimonious. The choice of these parameter values is data dependent. With specific data, trial-and-error procedures are appropriate in determining the values for hyperparameters with the objective of achieving satisfactory results. For readers interested in trying the proposed method, the source codes are available from the authors[2].

7 Discussions and Conclusions

There are several aspects of TS fuzzy model interpretability that are worth being addressed [8][27][29][30]. This paper just focuses on one of them, i.e., good interactions of TS local linear models with global model that make the local models dominate the system behaviors locally and separately as indicated in Definition 1 [8][26][27]. Among the existing schemes that are capable of improving TS model interpretability in the sense of Definition 1, most methods focus on either improving the distinguishability of input space partitions without optimizing local linear models to fit global model in local regions, or optimizing local linear models to fit global model locally without considering the improvement of the distinguishability of input space partitions. Interestingly, one advantage of the proposed method with linguistic modifiers in this paper lies in its ability to fulfil the two objectives together in one model structure, i.e., local linear model interpretability in the sense of Definiton 1 can be improved and distinguishable input space partition can be simultaneously produced. The experimental results have shown that by using the proposed method, the produced input space partitioning has good distinguishablity and the local models match the global model well in the corresponding local regions. As a result, good model interpretability has been achieved while the global model accuracy remains at a satisfactory level. Another contribution of this paper is to have applied the pivoted QR decomposition algorithm to fuzzy rule ranking to produce more transparent and parsimonious TS fuzzy models.

Due to the promising capability of the proposed method in construcing TS fuzzy models with comprehensible linear submodels in the sense of Definition 1, the proposed method would have potential applications to fuzzy modeling for nonlinear state estimation and control problems. Generally speaking, for highly nonlinear systems, many rules would be required to characterize them. Some interesting issues include formal stability analysis of TS models with interpretable submodels, possibility of applying reinforcement techonology to interpretability improvement of TS fuzzy models in case of no input-output training samples available. These topics merit further reseach in the future.

[2] http://dces.essex.ac.uk/staff/jqgan/gan.htm; http://www.cci.dmu.ac.uk /index.php?i=5&id=5

References

[1] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. on Sys., Man and Cyber.*, Vol. 15, No. 1, pp. 116-132, 1985.

[2] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. on Sys., Man and Cyber.*, Vol. 23, No. 3, pp. 665-685, 1993.

[3] L-X. Wang, *Adaptive Fuzzy Systems and Control*, Englewood Cliffs, NJ: Prentice-Hall, 1994.

[4] C. J. Harris, X. Hong, and J. Q. Gan, *Adaptive Modeling, Estimation and Fusion from Data - A Neurofuzzy Approach*, Springer, 2002.

[5] A. Gegov, *Complexity Management in Fuzzy Systems-A Rule Base Compression Approach*, Springer, 2007.

[6] S. Guillaume, "Designing fuzzy inference systems from data: an interpretability-oriented review," *IEEE Trans. on Fuzzy Systems*, Vol. 9, No. 3, pp. 426-443, 2001.

[7] R. Alcalá, J. Alcalá-Fdez, M.J. Gacto and F. Herrera, "On the use of multiobjective genetic algorithms to improve the accuracy-interpretability trade-off of fuzzy rule-based systems," In: A. Ghosh, S. Dehuri, S. Ghosh(Eds.): *Multi-objective Evolutionary Algorithms for Knowledge Discovery from Data Base*, Vol. 98, Springer Verlag, Heidelberg, Germany, 2008.

[8] J. R. Cano, F. Herrera, M. Lozano, "Evolutionary stratified training set selection for extracting classification rules with trade-off precision-interpretability," *Data and Knowledge Engineering*, Vol. 60, pp. 90-108, 2007.

[9] J. Yen, L. Wang, and C. W. Gillespie, "Improving the interpretability of TSK fuzzy models by combining global learning and local learning," *IEEE Trans. on Fuzzy Systems*, Vol. 6, No. 4, pp. 530-537, 1998.

[10] S. –M. Zhou and J. Q. Gan, "Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modeling," *Fuzzy Sets and Systems*, 2008 (Accepted).

[11] R. Jensen and Q. Shen, "Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 16, No. 12, pp. 1457-1471, 2004.

[12] Y. Jin, "Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement," *IEEE Trans. on Fuzzy Systems*, Vol. 8, No. 2, pp. 212-221, 2000.

[13] A. Hamilton-Wright and D. W. Stashuk, "Transparent decision support using statistical reasoning and fuzzy inference," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 18, No. 8, pp. 1125-1137, 2006.

[14] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Trans. on Fuzzy Systems*, Vol. 9, No. 4, pp. 516-524, 2001.

[15] W. Pedrycz, J. C. Bezdek, R. J. Hathaway, and G. W. Rogers, "Two nonparametric models for fusing heterogeneous fuzzy data," *IEEE Trans. on Fuzzy Systems*, Vol. 6, No. 3, pp. 411-425, 1998.

[16] J. V. de Oliveira, "Semantic constraints for membership function optimization," *IEEE Trans. on Sys., Man and Cyber.-Part A*, Vol. 29, No. 1, pp. 128-138, 1999.

[17] J. Espinosa and J. Vandewalle, "Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm," *IEEE Trans. on Fuzzy Systems*, Vol. 8, No. 5, pp. 591-600, 2000.

[18] J. Victor and A. Dourado, "On-line interpretability by fuzzyrule-base simplification and reduction," *Proc. EUNITE Symposium*, Aachen, 2004.

[19] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets and Systems*, Vol. 141, No. 1, pp. 59-88, 2004.

[20] F. Hoppner and F. Klawonn, "Obtaining interpretable fuzzy models from fuzzy clustering and fuzzy regression," *Proc. of the 4th Int. Conf. on Knowledge-based Intelligent Eng. Syst. and Allied Tech (KES)*, Brighton, UK, 2000, pp. 162-165.

[21] C. A. Penna-Reyes and M. Sipper, "Fuzzy CoCo: balancing accuracy and interpretability of fuzzy models by means of coevolution," in J. Casillas, O. Cordon, F. Herrera, and L Magdalena (Eds.), *Accuracy Improvements in Linguistic Fuzzy Modeling, Vol. 129 of Studies in Fuzziness and Soft Computing*, Springer, 2003, pp. 119-146.

[22] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum, 1981.

[23] R.L. de Mantaras and L. Valverde, "New results in fuzzy clustering based on the concept of indistinguishability relation," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, Vol. 10, No. 5, pp. 754-757, 1988.

[24] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, Vol. 11, No. 7, pp. 773-781, 1989.

[25] F. Hoppner and F. Klawonn, "A new approach to fuzzy partitioning," *Proc. of the Joint 9th IFSA Congress and 20th NAFIPS Int. Conf.*, Vancouver, Canada, 2001, pp. 1419-1424.

[26] J. Abonyi and R Babuska, "Local and global identification and interpretation of parameters in Takagi-Sugeno fuzzy models," *Proc. of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, San Antonio, USA, 2000, pp. 835-840.

[27] T. A. Johansen and R. Babuska, "Multi-objective identification of Takagi-Sugeno fuzzy models," *IEEE Trans. on Fuzzy Systems*, Vol. 11, No. 6, pp. 847-860, 2003.

[28] J. Q. Gan and C. J. Harris, "Fuzzy local linearization and local basis function expansion in nonlinear system modeling," *IEEE Trans. on Sys., Man and Cyber. - Part B*, Vol. 29, No. 4, pp. 559-565, 1999.

[29] A. Fiordaliso, "A constrained Takagi-Sugeno fuzzy system that allows for better interpretation and analysis," *Fuzzy Sets and Systems*, Vol. 118, pp. 307-318, 2001.

[30] T. G. Amaral, V. F. Pires and M. M. Crisóstomo, "An approach to improve the interpretability of neuro-fuzzy systems," *Proceeding of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Vancouver, BC, Canada, 2006, pp. 8502-8509.

[31] M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Trans. on Systems, Man and Cybernetics-Part B*, Vol. 28, No. 3, pp. 376-386, 1998.

[32] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. Man, "Agent-based evolutionary approach to interpretable rule-based knowledge extraction," *IEEE Trans. on Systems, Man, and Cybernetics-Part C* , Vol. 29, No. 2, pp. 143-155, 2005.

[33] C. Mencar, G. Castellano, and A. M. Fanelli, "Distinguishability quantification of fuzzy sets," *Information Sciences*, Vol. 177, pp. 130-149, 2007.

[34] H. A. Hefny, "Comments on 'distinguishability quantification of fuzzy sets,' " *Information Sciences*, 2007 (accepted).

[35] A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*, Academic Press: New York, 1975.

[36] G. C. Mouzouris and J. M. Mendel, "Designing fuzzy logic systems for uncertain environments using a singular-value-QR decomposition method," *Proc. 5th IEEE Int. Conf. Fuzzy Syst.*, New Orleans, LA, 1996, pp. 295–301.

[37] M. Setnes and R. Babuska, "Rule base reduction: some comments on the use of orthogonal transforms," *IEEE Trans. on Sys., Man and Cyber. - Part C*, Vol. 31, No. 2, pp. 199-206, 2001.

[38] S.-M. Zhou and J. Q. Gan, "Constructing accurate and parsimonious fuzzy models with distinguishable fuzzy sets based on an entropy measure," *Fuzzy Sets and Systems*, Vol. 157, No. 8, pp. 1057-1074, 2006.

[39] T. Furuhashi and T. Suzuki, "On interpretability of fuzzy models based on conciseness measure," *Proc. of the 10th IEEE Int. Conf. on Fuzzy Sets* (FUZZ-IEEE'01), Melbourne Australia, 2001.

[40] A. De Luca and S. Termini, "A definition of non-probabilistic entropy in the setting of fuzzy set theory," *Information Control*, Vol. 20, pp. 301–312, 1972.

[41] J. Yen and L. Wang, "Application of statistical information criteria for optimal fuzzy model construction," *IEEE Trans. on Fuzzy Systems*, Vol. 6, No. 3, pp. 362-372, 1998.

[42] S.-M. Zhou and J. Q. Gan, "An unsupervised kernel based fuzzy c-means clustering algorithm with kernel normalization," *Int. Journal of Computational Intelligence and Applications*, Vol. 4, No. 4, pp. 355-373, 2004.

[43] M. J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, Vol. 1, No. 2, pp. 281–294, 1989.

[44] S. Bittanti and L. Piroddi, "Nonlinear identification and control of a heat exchanger: a neural network approach," *Journal of the Franklin Institute*, Vol. 334, No. 1, pp. 135-153, 1997.

[45] N. F. Rulkov, L. S. Tsimring and H. D. I. Abarbanel, "Tracking unstable orbits in chaos using dissipative feedback control," Phy. Rev. E, Vol.50, No.1, pp. 314-324, 1994.

Shang-Ming Zhou received the BSc degree in mathematics from Liaocheng University, Shandong Province, China, the MSc degree in applied mathematics from Beijing Normal University, Beijing, China, and the PhD degree in computer science from the University of Essex, UK respectively. Currently, he is with the Centre for Comutational Intelligence, Department of Informatics, De Montfort University in the UK. His research interests include fuzzy logic systems (type-1 and type-2) and applications; decision support systems under uncertainty; model interpretability and transparency in data-driven neurofuzzy systems; machine learning (neural networks, kernel models and particle swarm optimization), intelligent signal and image processing. He has published extensively on these topics.

John Q. Gan received the BSc degree in electronic engineering from Northwestern Polytecnic University, China, in 1982, the M.Eng degree in automatic control and the PhD degree in biomedical electronics from Southesast University, China, in 1985 and 1991, respectively. He is a reader in Computer Science at the University of Essex, UK. He has co-authored a book and published over 150 research papers. He is Associate Editor for IEEE Transactions on Systems, Man and Cybernectics-Part B and in editorial board of other journals. His research interests are neurofuzzy computation and machine intelligence, brain-computer interface, robotics and intelligent systems, pattern recognition, signal processing, and data fusion.