Proceedings of the World Congress on Engineering and Computer Science 2013 Vol I
WCECS 2013, 23-25 October, 2013, San Francisco, USA

Takagi-Sugeno (T-S) Fuzzy Regression of Fuzzy Data

Peihong Wang, Yifan Wang, and Zhigang Su*

***Abstract*—The conventional Takagi-Sugeno (T-S) fuzzy model is an effective tool used to approximate the behaviors of uncertain nonlinear systems on the basis of precise observations. In many real-life situations, however, the observations can be imprecise due to limited precisions of devices. This paper presents a systematic method to design T-S fuzzy model when the observations are imprecise, represented as fuzzy data, and then proposes the so-called fuzzy T-S regression model (FTS). The consequents of FTS are identified by using a fuzzy EM algorithm, a fuzzy extension of EM algorithm. The antecedents of FTS are automatically constructed by using a data-driven strategy, considering both the accuracy and complexity of the produced FTS. The performance of FTS was illustrated by using some simulations.**

Index Terms—Takagi-Sugeno fuzzy model, fuzzy data, imprecision, regression, soft computing

I. Introduction

Takagi-Sugeno (T-S) fuzzy model recently has attracted most attention [14]. The T-S fuzzy model consists of IF-THEN rules with fuzzy antecedents and mathematical functions in the consequent part. The fuzzy antecedents partition the input space into a number of fuzzy regions, while the consequent functions describe the system's behavior in these regions [10]. The construction of a T-S fuzzy model is generally done in two steps. In the first step, the fuzzy sets in the rule antecedents are determined. In the second step, the parameters of the consequent functions are estimated.

There emerges a surge of versions of T-S fuzzy models since the birth of the original T-S fuzzy model. The popular two are the population stochastic algorithm based methods, for example, see in [1, 6, 8, 17], and the clustering algorithm based methods, e.g., see in [7, 11, 13, 15]. As far as we know, all the recent T-S fuzzy models assume a perfect knowledge of the value of the response x for the learning samples. That is to say, the observations are supposed to be precise (i.e., point-valued). However, in many real-life situations, we cannot obtain such standard observations. Quite often, information about the response is obtained through measuring devices, or sensors, with limited precision. Therefore, on the one hand, it is interesting to extend the T-S fuzzy model to deal with imprecise data. On the other hand, it is necessary to propose a new methodology in the imprecise setting. Up to date, there is few literatures on extending T-S fuzzy model to deal with imprecise data. In this paper, we suppose the imprecise data are represented by fuzzy data and investigate on the T-S fuzzy regression with crisp inputs and fuzzy output, represented by fuzzy data.

There exist two obstacles preventing the conventional T-S fuzzy model to deal with imprecise data. The first one is how to determine the fuzzy sets (i.e., to find fuzzy rules or structure of data) in the antecedents when the response x is fuzzy data. The second obstacle is how to identify the linear functions in the consequent when observations (of responses) are fuzzy data. To solve the first problem, we propose a data-driven automatic method. This method views the input data and output data separately, but it considers both the structure of input data and the performance of T-S fuzzy model so as to find optimal number of rules with an acceptable accuracy. More precisely, we firstly partition the input data space with an initial rule, and then the following other rules are iteratively produced when the performance of and number of existing rules commit some conditions. When a given maximal rule number or predetermined accuracy is committed, the iteration will be terminated. To identify the linear behaviors in these fuzzy regions (i.e., to solve the second problem), a novel algorithm used to estimating parameters in fuzzy setting is needed. Recently, a significant contribution is the extension of Expectation-Maximization (EM) algorithm [2] to fuzzy data, i.e., the so-called fuzzy EM (FEM) algorithm [3]. Use of such a contribution makes the parameter estimation in statistic models be implemented in the case where data are imprecise represented as fuzzy data. Using FEM algorithm makes the consequents in T-S fuzzy model can be identified when observations are fuzzy data. Therefore, the so-called fuzzy T-S regression model is derived and automatically driven by data. In addition, we observe that the investigations on T-S fuzzy regression of point-valued data can be covered when the FEM inversely degenerates to traditional EM algorithm and the output fuzzy data become to be point values. Based on our above observations, we therefore propose the so-called a classical data-driven T-S fuzzy regression model.

The rest of this paper is organized as follows. Section 2 recalls the preliminaries for the paper. Section 3 presents the proposed systematical methodology. Section 4 applies some numerical experiments to validate the performance of the proposed methods. The last section concludes the paper.

Manuscript received June 3, 2013, revised June 28, 2013. This work was supported in part by the National Natural Science Foundation of China (No. 51106025).

* Corresponding author.

Peihong Wang and Zhigang Su are both with the Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education, School of Energy and Environment, Southeast University, Nanjing, 210096, China. (e-mail: phwang@seu.edu.cn, zhigangsu@seu.edu.cn).

Yifan Wang is with VMware, Palo Alto, CA, 94304, US (e-mail: yifanwang@vmware.com).

II. Preliminaries: Fuzzy EM algorithm

Let $\boldsymbol{X}$, referred to as the *complete-data* vector, be a random vector, taking value in sample space Ω_X and describing the result of a random experiment. The probability density function (p.d.f.) of $\boldsymbol{X}$ is denoted by $g(\mathbf{x}, \boldsymbol{\psi})$, where $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_d)'$ is a column vector of unknown parameters with parameter space Ω_ψ, where superscript " ′ " indicates transposition. If $\mathbf{x}$, a realization of $\boldsymbol{X}$, is known exactly, we could compute the maximum likelihood estimate (MLE) of $\boldsymbol{\psi}$ as any value maximizing the complete-data likelihood function:

$$L(\boldsymbol{\psi};\mathbf{x}) = g(\mathbf{x};\boldsymbol{\psi}) \tag{1}$$

However, $\mathbf{x}$ is usually not observed precisely, e.g., only imprecise information about $\mathbf{x}$ is available in the form of a fuzzy subset $\tilde{\mathbf{x}}$ of Ω_X. Therefore, the complete-data likelihood function (1) should be extended. Given $\tilde{\mathbf{x}}$ and assume its membership function to be the Borel measurable, the probability of fuzzy set $\tilde{\mathbf{x}}$ can be computed according to Zadeh's definition of the probability of a fuzzy event [16]. Thus, the observed-data likelihood in the imprecise setting can then be defined as.

$$L(\boldsymbol{\psi};\tilde{\mathbf{x}}) = P(\tilde{\mathbf{x}};\boldsymbol{\psi}) = \int_{\chi} \mu_{\tilde{\mathbf{x}}}(\mathbf{x})\, g(\mathbf{x};\boldsymbol{\psi})\, d\mathbf{x} \tag{2}$$

In the special case where the complete data $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ is a realization of an independent identically distributed (i.i.d.) random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$, and assuming the joint membership function $\mu_{\tilde{\mathbf{x}}}$ to be decomposed in the product of $\mu_{\tilde{x}_i}$ $(i = 1, 2, \ldots, n)$, i.e.,

$$\mu_{\tilde{\mathbf{x}}}(\mathbf{x}) = \prod_{i=1}^{n} \mu_{\tilde{x}_i}(x_i) \tag{3}$$

the likelihood function (2) can be written as a product of n items:

$$L(\boldsymbol{\psi};\tilde{\mathbf{x}}) = \prod_{i=1}^{n} \int \mu_{\tilde{x}_i}(x) g(x;\boldsymbol{\psi}) dx \tag{4}$$

and the observed-data log likelihood is:

$$\log L(\boldsymbol{\psi};\tilde{\mathbf{x}}) = \sum_{i=1}^{n} \log \int \mu_{\tilde{x}_i}(x) g(x;\boldsymbol{\psi}) dx \tag{5}$$

The fuzzy EM algorithm approaches the problem of maximizing the observed-data log likelihood $\log L(\boldsymbol{\psi}, \tilde{\mathbf{x}})$ by proceeding iteratively with the complete-data likelihood $\log L(\boldsymbol{\psi}, \mathbf{x}) = \log g(\mathbf{x}, \boldsymbol{\psi})$. Each iteration of the fuzzy EM algorithm involves two steps called the expectation step (E-step) and the maximization step (M-step).

The E-step consists in the calculation of

$$Q\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right) = \frac{\int \mu_{\tilde{\mathbf{x}}} \log\left[L(\boldsymbol{\psi};\mathbf{x})\right] g\left(\mathbf{x},\boldsymbol{\psi}^{(q)}\right) d\mathbf{x}}{L\left(\boldsymbol{\psi}^{(q)};\tilde{\mathbf{x}}\right)} \tag{6}$$

where the expectation of $\log L(\boldsymbol{\psi}, \boldsymbol{X})$ is taken with respect to the conditional p.d.f. of $\mathbf{x}$ given $\tilde{\mathbf{x}}$, using vector $\boldsymbol{\psi}^{(q)}$:

$$g\left(\mathbf{x}\middle|\tilde{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right) = \mu_{\tilde{\mathbf{x}}}(\mathbf{x})\, g\left(\mathbf{x}\middle|\boldsymbol{\psi}^{(q)}\right) \Big/ \int \mu_{\tilde{\mathbf{x}}}(\mathbf{u})\, g\left(\mathbf{u}\middle|\boldsymbol{\psi}^{(q)}\right) d\mathbf{u}$$

The M-step requires the maximization of $Q(\boldsymbol{\psi}, \boldsymbol{\psi}^{(q)})$ with respect to $\boldsymbol{\psi}$ over the parameter space Ω_ψ. The fuzzy EM algorithm alternately repeats the E- and M-steps above until the increase of observed-data likelihood becomes smaller than some threshold.

III. The proposed fuzzy T-S fuzzy regression model

In this section, we consider the problem of *approximating* a continuous multiple input single output (MISO) function to clarify the basic ideas of the presented method, since the extension of the method to a multiple output function is straightforward. Let $\mathbf{u} = (u_1, u_2, \ldots, u_p)'$ be a p-dimensional input variable vector, and x be the associated output variable. The T-S fuzzy model comprises of a set of IF-THEN fuzzy rules having the following form:

R_k: IF u_1 is $\tilde{A}_{k1}$ and … and u_p is $\tilde{A}_{kp}$, THEN

$$x^k = b_{k0} + b_{k1}u_1 + b_{k2}u_2 + \ldots + b_{kp}u_p \tag{7}$$

where $k = 1$ to M, x^k is the output of the kth rule, $\mathbf{b}_k = (b_{k0}, b_{k1}, \ldots, b_{kp})'$ is the consequent coefficients in the kth rule, and $\tilde{A}_{kj}$ $(j = 1, 2, \ldots, p)$ are fuzzy sets in antecedent defined:

$$\mu_{\tilde{A}_{kj}}(u_j) = \exp\left(-\left(u_j - v_{kj}\right)^2 \big/ 2s_{kj}^2\right) \tag{8}$$

where v_{kj} and s_{kj} ($k = 1$ to M, $j = 1$ to p) are *respectively* the center and width of the jth memberships in the kth rule.

Given arbitrary input data $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{ip})'$, each rule provides a *predicted* output. The overall output of the T-S fuzzy model is computed as follow:

$$x_i = \sum_{k=1}^{M} \tau_i^k x_i^k \Big/ \sum_{k=1}^{M} \tau_i^k \tag{9}$$

where $\tau_i k$ is *the* firing strength of R_k for the ith input, which is defined as:

$$\tau_i^k = \tilde{A}_{k1}(u_{i1}) \times \tilde{A}_{k2}(u_{i2}) \times \cdots \times \tilde{A}_{kp}(u_{ip}) \tag{10}$$

A. *Identification of antecedents in FTS*

This section presents the strategy used to identify antecedent of fuzzy T-S fuzzy model when the observations are in the following form:

$$T = \left\{ e_i \middle| e_i = (\mathbf{u}_i, \tilde{x}_i), i = 1, 2, \cdots, n \right\} \tag{11}$$

where $\tilde{x}_i$ is the imprecisely observed values for response x, represented as fuzzy data.

The basic idea of the proposed strategy is similar to the method in the context of RBF network [4], which is firstly extended to construct classical T-S fuzzy model by Rezaee and Fazel Zarandi [9]. Here, we further extend it to deal with fuzzy data. To present the strategy, a new performance measure is first needed, which, called *mean square fuzzy expectation error,* is defined as follow:

$$MSE_{fuzzy} = \frac{1}{n}\sum_{i=1}^{n} MSE_{fuzzy}\_i = \frac{1}{n}\sum_{i=1}^{n} \left\| \mathrm{E}[\tilde{x}_i] - x_i \right\|^2 \tag{12}$$

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

where $\mathsf{E}[\tilde{x}_i]=\int_{\Omega_X}\omega\mu_{\tilde{x}_i}(\omega)d\omega$ is the fuzzy expectation associated to $\tilde{x}_i$. When $\tilde{x}_i$ degenerate to be point values, criteria (12) will degenerate to be *mean square error* (MSE).

The proposed strategy is an iterative procedure satisfying the following two termination conditions: (1) the approximate accuracy is higher than a given acceptable performance (ε), i.e., $MSE_{fuzzy} > \varepsilon$; and (2) the size of rule base is not bigger than a given maximum rule number (R_{max}), i.e., $M \le R_{max}$. These two conditions let the designer(s) adjust the desirable tradeoff between accuracy and the size of rule base according to his or her intuition or expertise. For instance, if the designer knows the accuracy as a prior, ε can be set to be the user-given threshold and R_{max} can be set to be a large number. In this case, the rule number increases until the performance of the model meets the use-given accuracy. This case is useful when the designer has nothing priori knowledge on the rule number. On the contrary, if the designer prefers to have a model with certain rule number, he or she can preset a small value for ε. The four-step proposed strategy is interpreted as follows.

Firstly, data preprocess is an import step for model construction. Preprocessing involves both identifying and eliminating the outliers in the data set and selecting the significant input variables among the candidates. However, we do not focus on the solution to this step and left it as a further study.

Secondly, an initial fuzzy rule is generated in the rule base. The initial fuzzy rule is extracted by a simple method. The fuzzy sets $\tilde{A}_{kj}$ in antecedent of the initial fuzzy rule are determined by:

$$v_{1j}=\frac{1}{n}\sum_{i=1}^{n}u_{ij} \tag{13}$$

$$s_{1j}=\sqrt{\frac{1}{n-1}\sum\nolimits_{i=1}^{n}\left(u_{ij}-v_{1j}\right)^2} \tag{14}$$

In this original initial rule, the behavior in consequent is identified using fuzzy EM algorithm (FEM), which will be detailed in the consequent section. Here, we assume we have obtained the linear behavior in this rule. This assumption is held in this section.

Thirdly, a new rule is constructed and is added to the original rule base. The vector that has the worst MSE_{betp}_i, denoted by $\mathbf{u}_{i'}$, is considered as the candidate center for this new rule:

$$\mathbf{u}_{i'}=\left\{\mathbf{u}_j \,\middle|\, MSE_{fuzzy}\_j=\max_{i=1,2,\ldots,n}\left\{MSE_{fuzzy}\_i\right\}\right\} \tag{15}$$

Because the candidate rule is only based on performance error, it is possible for an outlier to be considered as a new rule's center. Although the preprocessing of data maybe detects and eliminates the outliers, it is still need to reduce the effects of the noisy data and exclude the chance of an outlier to become a rule center. In addition, we do not want the new candidate center is too close to the existing centers. In this regard, the following conditions should be satisfied:

$$\begin{cases}(1)\sum\nolimits_{i=1}^{n}\mu_{i,i'}^2\ge\Delta_1\\(2)\min\limits_{k=1,2,\ldots,M}\left\|\mathbf{u}_{i'}-\mathbf{v}_k\right\|^2\ge\Delta_2\end{cases} \tag{16}$$

where Δ_1 and Δ_2 are constants, and $\mu_{i,i'}$ is the membership degree of the ith data belonging to the i'th cluster, determined in the following way [5]:

$$\mu_{i,i'}=\left(\sum\nolimits_{k=1}^{M}\frac{d_{i,i'}^2-\min d_{i,*}^2}{d_{i,k}^2-\min d_{i,*}^2}\right)^{-1} \tag{17}$$

where $d_{i,k}$ is distance between the ith sample and the kth center, defined as $d_{ik}=\left\|\mathbf{u}_i-\mathbf{v}_k\right\|$, and the item $\min d_{i,*}^2$ is defined:

$$\min d_{i,*}^2=\min_i\{d_{i,k}^2 \mid k=1,2,\ldots M\}-\gamma,\gamma>0 \tag{18}$$

As can be seen from conditions (1) and (2), a low $\sum_{i=1}^{n}\mu_{i,i'}^2$ indicates the i'th data is far away from other data and it cannot be considered as a new rule's center, and a low $\min_{k=1,2,\ldots,M}\left\|\mathbf{u}_{i'}-\mathbf{v}_k\right\|^2$ indicates the i'th data locates too close to one existing rule center. Therefore, the role of condition (1) is to prevent an outlier to be a new rule's center, and the condition (2) ensures that the new rule's center is not located very close to the other existing rule centers. In these viewpoints, the constant Δ_1 and Δ_2 can be defined respectively as follows:

$$\Delta_1=\eta\sum\nolimits_{k=1}^{M}\sum\nolimits_{i=1}^{n}\mu_{i,k}^2\Big/M \tag{19}$$

$$\Delta_2=\sum\nolimits_{i=1}^{n}\mu_{i,i'}^2\left\|\mathbf{u}_i-\mathbf{u}_{i'}\right\|^2\Big/\sum\nolimits_{i=1}^{n}\mu_{i,i'}^2 \tag{20}$$

where $0\le\eta\le1$ is a *soft factor* used to control the effects of the average membership degrees of all data over all rules. Through our experiments, we found that sometimes there are no points satisfying $\sum\nolimits_{i=1}^{n}\mu_{i,i'}^2\ge\sum\nolimits_{k=1}^{M}\sum\nolimits_{i=1}^{n}\mu_{i,k}^2\Big/M$. In other words, we usually can only obtain one rule, i.e., the initial rule. In this case, we can soft the constraint by using small η.

If the selected vector $\mathbf{u}_{i'}$ satisfies (16), then it is declared as the center of a new rule. Otherwise, it is marked as an outlier and the process of selecting the vector that has the worse performance is repeated without considering the outliers. When none of the existing vectors satisfy (16), the procedure is terminated to avoid over-fitting. The fuzzy antecedents $\tilde{A}_{new,j}$ of the generated new rule is then characterized by ($v_{new,j}$, $s_{new,j}$), which are defined by

$$v_{new,j}=u_{i',j},\ j=1,2,\cdots,p\ \ or\ \ \mathbf{v}_{new}=\mathbf{u}_{i'} \tag{21}$$

$$s_{new,j}=\sqrt{\sum\nolimits_{i=1}^{n}\mu_{i,i'}^2\left\|u_{i,j}-v_{new,j}\right\|^2\Big/\sum\nolimits_{i=1}^{n}\mu_{i,i'}^2} \tag{22}$$

Finally, once the new fuzzy rule is added to the rule base, the rule number increases one, i.e., $M = M + 1$, and we have $v_{M,j} = v_{new,j}$, $s_{M,j} = s_{new,j}$. Due to the added rule, the rule base should be updated. The centers $\mathbf{v}_k$ of the previous M-1 rules

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

existing in the rule base can be maintained whereas their widths s_{kj} (k = 1, 2, …, M-1) should be updated and can be updated according to (37) only by replacing index i' with the index k for k = 1, 2, …, M-1.

B. Identification of consequents in FTS

For the sequent discussion, we firstly transform overall output (9) to the following vector or matrix form:

$$x_i = \mathbf{h}_i'\mathbf{b}, i = 1,2,\cdots,n \text{, or } \mathbf{x} = \mathbf{H}'\mathbf{b} \tag{23}$$

where **h**, **H** and **b** are defined respectively as follows

$$\begin{aligned} &\mathbf{h}_i = \left[w_i^1, w_i^1 u_{i1}, \cdots, w_i^1 u_{ip}, \cdots, w_i^M, w_i^M u_{i1}, \cdots, w_i^M u_{ip} \right]' \\ &\mathbf{H} = \left[\mathbf{h}_1, \mathbf{h}_i, \cdots, \mathbf{h}_n \right], \quad \mathbf{b} = \left[\mathbf{b}_1', \mathbf{b}_2', \cdots, \mathbf{b}_M' \right]' \end{aligned} \tag{24}$$

where w_ik is the weight of the ith data in the kth rule, defined as $w_i^k = \tau_i^k \Big/ \sum_{j=1}^{M} \tau_i^j$.

To solve the above regression with fuzzy data as output, it can be assumed that each component x_i of the complete-data vector **x** is a realization of a normal random variable X_i with mean $\mathbf{h}_i'\mathbf{b}$ and standard deviation σ, and observer encodes his/her imprecise knowledge of x_i in the form of pieces of fuzzy data $\tilde{x}_i$. In other words, the observed data **x** is multivariate Gaussian with mean $\mathbf{H}'\mathbf{b}$ and variance $\sigma^2 I_{p+1}$ associated with $\tilde{x}_i$, where I_{p+1} denotes the (p+1)-dimensional identity matrix. With such assumption, the complete parameter vector is thus ψ = (**b'**, σ)' that should be identified in the case where only $\tilde{x}_i$ can be observed.

According to above interpretations, the complete-data p.d.f. can therefore be defined as

$$g(\mathbf{x};\psi) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\left(x_i - \mathbf{h}_i'\mathbf{b} \right)^2 \Big/ 2\sigma^2 \right) \tag{25}$$

By using the complete-data p.d.f., the complete-data log likelihood is thus:

$$\begin{aligned} \log L(\psi;\mathbf{x}) &= \sum_{i=1}^{n} \log g(x_i;\psi) \\ &= -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\left(\mathbf{x}'\mathbf{x} - 2\mathbf{b}'\mathbf{H}\mathbf{x} + \mathbf{b}'\mathbf{H}\mathbf{H}'\mathbf{b} \right) \end{aligned} \tag{26}$$

Taking the expectation of $\log L(\psi, \mathbf{x})$ conditionally on the observed $\tilde{x}_i$ and using the fit $\psi^{(q)}$ of ψ to perform the E-step, it can get

$$\begin{aligned} Q\left(\psi,\psi^{(q)}\right) &= \mathsf{E}_{\psi^{(q)}}\left(\log\left[L(\psi;\mathbf{x}) \right] \middle| \tilde{\mathbf{x}} \right) = -\frac{n}{2}\log(2\pi) \\ &\quad - n\log\sigma - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} \alpha_i^{(q)} - 2\mathbf{b}'\mathbf{H}\boldsymbol{\beta}^{(q)} + \mathbf{b}'\mathbf{H}\mathbf{H}'\mathbf{b} \right) \end{aligned} \tag{27}$$

where $\alpha_i^{(q)} = \mathsf{E}_{\psi^{(q)}}\left(X_i^2 \middle| \tilde{x}_i \right)$ and $\beta_i^{(q)} = \mathsf{E}_{\psi^{(q)}}\left(X_i \middle| \tilde{x}_i \right)$, and $\boldsymbol{\beta}^{(q)}$ = ($\beta_1^{(q)}$, …, $\beta_n^{(q)}$)′ are expectations conditioning on fuzzy data. The detail computations of these expectations can refer to our previous work [12] for completeness.

The M-step requires maximizing $Q_{FE^2M}(\psi, \psi^{(q)})$ with respect to ψ. This can be achieved by differentiating $Q_{FE^2M}(\psi, \psi^{(q)})$ with respect to **b** and σ, which results in:

$$\frac{\partial Q\left(\psi,\psi^{(q)}\right)}{\partial \mathbf{b}} = -\frac{1}{\sigma^2}\left(-\mathbf{H}\boldsymbol{\beta}^{(q)} + \mathbf{H}\mathbf{H}'\mathbf{b} \right)$$

$$\frac{\partial Q\left(\psi,\psi^{(q)}\right)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left(\sum_{i=1}^{n} \alpha_i^{(q)} - 2\mathbf{b}'\mathbf{H}\boldsymbol{\beta}^{(q)} + \mathbf{b}'\mathbf{H}\mathbf{H}'\mathbf{b} \right)$$

Equating these derivatives to zero and solving for **b** and σ, we get the following unique solution:

$$\mathbf{b}^{(q+1)} = \left(\mathbf{H}\mathbf{H}' \right)^{-1} \mathbf{H}\boldsymbol{\beta}^{(q)} \tag{28}$$

$$\sigma^{(q+1)} = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n} \alpha_i^q - 2\mathbf{b}^{(q+1)'}\mathbf{H}\boldsymbol{\beta}^{(q)} + \mathbf{b}^{(q+1)'}\mathbf{H}\mathbf{H}'\mathbf{b}^{(q+1)} \right)} \tag{29}$$

When the iteration terminates, we can obtain the regression coefficients **b** and thus obtain the multiple regression model with crisp inputs and fuzzy output.

IV. Experiments

In the experiment, we suppose that the domain of input Ω_u = [0, 10]. The true output x_is were generated using the following specific nonlinear model:

$$x = u\sin u, \qquad u \in [0,10] \tag{30}$$

To model the situation where response x can only be imprecisely observed, triangular fuzzy data (see equation () in Appendix) is adopted. The core and support of such kind of fuzzy observations were generated according to the following two-step strategy:

Step 1: Generate the cores x_i of fuzzy observations $x_i = f(u_i) + \varepsilon_i$, where $\varepsilon_i \sim N(0, \delta_{\max})$.

Step 2: The supports of the fuzzy observations $\tilde{x}_i$ are defined as $[x_i - \delta_i, x_i - \delta_i]$, where $\delta_i \sim rand[\delta_{\min}, \delta_{\max}]$.

By taking the nonlinear function (30) as the example in this section, we analyze the performance of the proposed fuzzy T-S fuzzy model when the imprecision (δ_i) takes different ranges of values: [0.2, 2.2], [0.1, 1.1], [0.01, 0.1], [0, 0]. Note that, when $\delta_i \in$ [0, 0], the study on classical T-S fuzzy model is covered. In these four cases, we suppose γ = 1, accuracy threshold $\varepsilon = 10^{-5}$ and maximum rule number $R_{\max}$ = 4, 5, 6, and 7.

In each case study, one hundred data sets $T^{(l)}$, l = 1, 2, …, 100 were generated using above strategy. The sample size of each data $T^{(l)}$ is n = 21. The FTS are identified on each set $T^{(l)}$. To measure the prediction accuracy of the identified model on each training set, we regularly generate number n_t of testing samples from the input domain according to:

$$\begin{aligned} &u_i^t = 0.1(i-1), i = 1,2,\ldots,101 \\ &x_i^t = u_i^t \sin u_i^t \end{aligned} \tag{31}$$

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE 2
APPROXIMATION AND PREDICTION ACCURACIES IN DIFFERENT RANGES OF IMPRECISION *1

Imprecision	Error type (mean ± std)	Maximum rule number R_{max} 4	5	6	7	Remark
$\delta_i \in [0.2, 2.2]$	Approximation	**1.1331 ± 0.6761**	1.0284 ± 0.6192	0.9329 ± 0.2984	0.7256 ± 0.3761	M = 4
	Prediction	**0.6768 ± 0.2704**	0.8090 ± 0.3957	1.0402 ± 0.3620	1.1059 ± 0.5102	Over-fitting
$\delta_i \in [0.1, 1.1]$	Approximation	0.3008 ± 0.0877	0.3489 ± 0.1656	0.2061 ± 0.0897	**0.1589 ± 0.0643**	M = 7
	Prediction	0.2828 ± 0.1213	0.2936 ± 0.1091	0.3239 ± 0.2346	**0.2692 ± 0.0810**	
$\delta_i \in [0.01, 0.1]$	Approximation	0.1076 ± 0.0388	0.0296 ± 0.0173	0.0123 ± 0.0028	**0.0049 ± 0.0025**	M = 7
	Prediction	0.1057 ± 0.0396	0.0310 ± 0.0191	0.0128 ± 0.0017	**0.0052 ± 0.0027**	
$\delta_i \in [0, 0]$	Approximation	0.1405 ± 0	0.0422 ± 0	0.0088 ± 0	**0.0017 ± 0**	M = 7
	Prediction	0.1383 ± 0	0.0438 ± 0	0.0091 ± 0	**0.0018 ± 0**	

*1 The bold number indicates best performance without considering over-fitting, and the shadow cells indicate over-fitting cases.

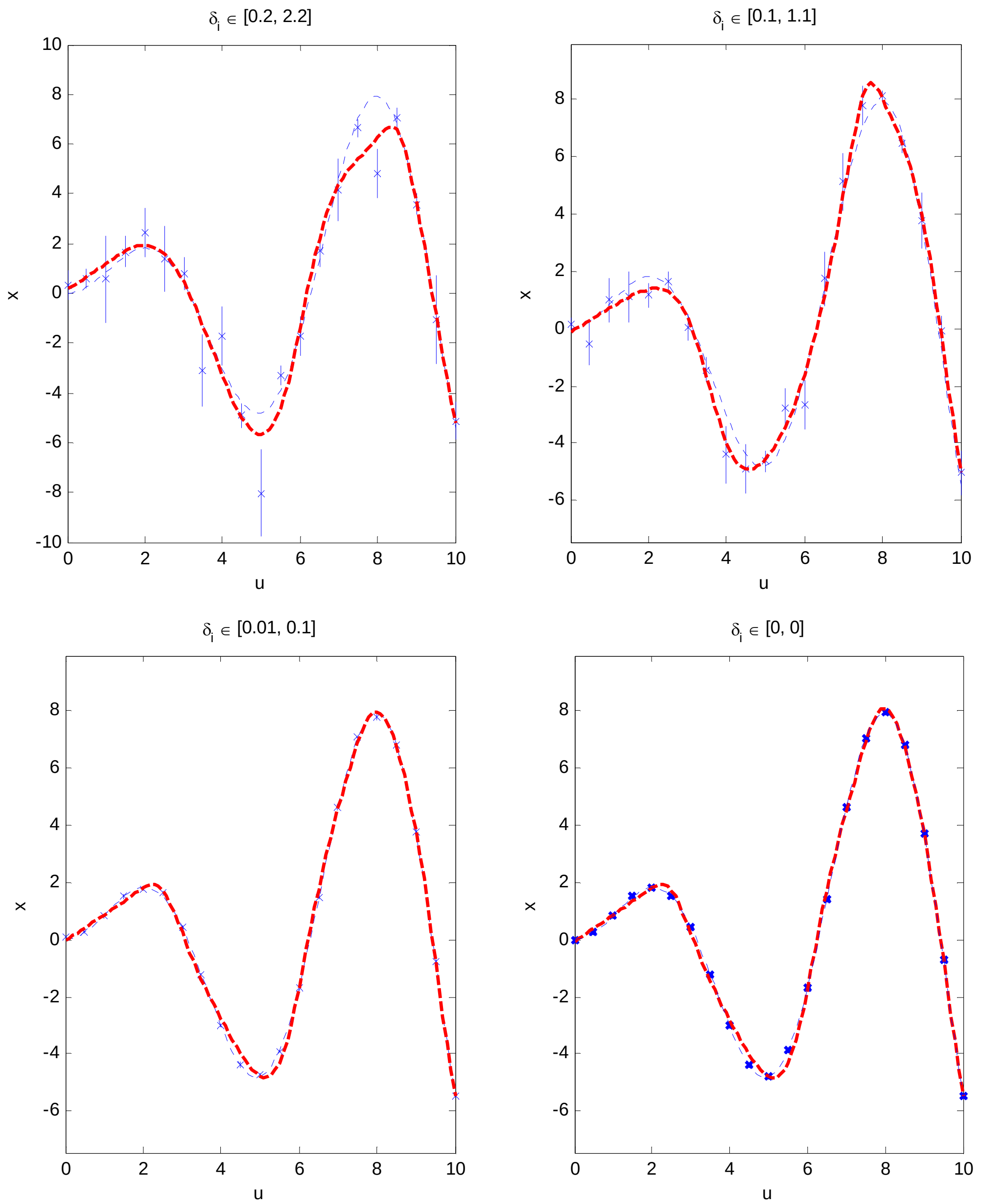

Fig. 1 Four random results over 100 trials for the specific nonlinear system (30). The dot curve indicates the true output values x_i as a function of inputs u_i. The cross denotes the core b_i of fuzzy number. Each vertical segment represents the support $[b_i - \delta_i, b_i + \delta_i]$. The dash line indicates the prediction results.

The error was computed as the mean squared differences between the true output x_i and model prediction $\hat{x}_i^{(l)}$:

$$MSE^{(l)} = \frac{1}{n_t}\sum_{i=1}^{n_t}\left(x_i - \hat{x}_i^{(l)}\right)^2 \tag{32}$$

The numerical results are presented in Table 1, and four graphical results randomly selected from the 100 trials in each case are shown in Fig. 1.

As an illustration, the FTS model structure for one training data present in upper-left subplot of Fig. 1 is presented as following if we preset $R_{max} = 5$ and $\varepsilon = 0.001$, we can derive the following fuzzy T-S regression model:

R_1: IF u_1 is $\tilde{A}_{11}$, THEN $x^1 = -2.9282 -0.7753u_1$;

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

R_2: IF u_1 is $\tilde{A}_{21}$, THEN x^2 = 79.8857-8.5144u_1;
R_3: IF u_1 is $\tilde{A}_{31}$, THEN x^3=-4.0260 + 0.2162u_1;
R_4: IF u_1 is $\tilde{A}_{41}$, THEN x^4 = 0.1216 +0.8476u_1;
R_5: IF u_1 is $\tilde{A}_{51}$, THEN x^5= 8.7062-3.1560u_1;

where the fuzzy sets $\tilde{A}_{k1}$ (k = 1, 2, …, 5) in antecedents are shown in Fig. 2.

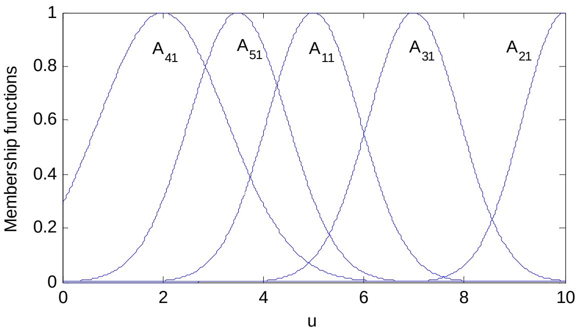

Fig. 2 The membership function of antecedent fuzzy sets in each rule corresponding to Fig. 2. The centers of $\tilde{A}_{k1}$ (k = 1, 2, …,5) is 5, 10, 7, 2, and 3.5, and the bandwidths of $\tilde{A}_{k1}$ (k = 1, 2, …,5) is 0.9183, 0.8462, 0.9092, 1.2668 and 0.9719.

Fig. 1 intuitively illustrates the prediction results of fuzzy T-S model in different cases. It can be seen that the predicted curves can approach to the true behavior. The difference between the predicted curves and true behavior becomes smaller with decreasing of imprecision. Especially, such difference approaches to zero in the precise and certain case (see in the below-left subplot in Fig. 1).

Tables 1 presents the approximate and prediction accuracies when maximum rule number $R_{\max}$ takes different values in different ranges of imprecision and uncertainty. They numerically show the performance of FTS model. For a given range of imprecision, the $R_{\max}$ corresponding to the highest approximate accuracy is determined as the rule number without considering over-fitting. For instance in the first case $\delta \in$ [0.2, 2.2] in Table 1, the highest approximate accuracy appears when $R_{\max}$ = 4, therefore, number of rule in rule base is 4, i.e., M = 4. We call a model over-fitting if its approximate accuracy becomes small whereas its associated prediction accuracy approaches to high, see the shadow cells in Table 1. The over-fitting always occurs in the cases when low reliability and/or high imprecision exist. In this regard, it suggests constructing FTS with small size of rule base in the high imprecision and uncertainty cases.

In a word, the FTS model can deal with imprecise data, and its performance is determined by the ranges of imprecision: the lower the imprecision IS, the higher approximate and predication accuracies are.

V. Conclusions

This paper proposes a T-S fuzzy regression method used to deal with problems when response cannot be precisely observed and can only be represented by fuzzy data. In this approach, both the performance accuracy and size of rule number in rule base (i.e., the complexity of produced model) are considered simultaneously. The proposed approach starts with an initial rule, and then a new rule is added to the rule base. The antecedents of the new rule are constructed by using an automatic data-driven strategy, and the consequents are identified by using fuzzy EM algorithm. This procedure is terminated when the accuracy and rule number of the proposed approach meet the preset stopping criterions.

The proposed approach can not only deal with regression of fuzzy data, but also can be used in crisp setting. With some simplifications, we therefore derive the so-called a classical data-driven automatic T-S fuzzy model. The proposed models are validated by using some numerical experiments. The experimental results suggest that the proposed models have high prediction accuracy and can be used to interpret nonlinear system when the observations of response are fuzzy data.

References

[1] C.C. Chen, "A PSO-based method for extracting fuzzy rules directly from numerical data," Cybern. Syst., Vol. 37, no. 7, pp. 707-723, Dec. 2006.
[2] A.P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society B, vol. 39, no. 1, pp. 1–38, 1977.
[3] T. Denoeux, "Maximum likelihood estimation from fuzzy data using the EM algorithm," Fuzzy sets and systems, vol. 183, no. 1, pp. 72-91, Nov. 2011.
[4] B. Fritzke, "Fast learning with incremental RBF networks," Neural Process. Lett., vol. 2, no. 1, pp. 2-5, Mar. 1994.
[5] F. Hoppner, and F. Klawonn, "Improved fuzzy partitions for fuzzy regression model," International Journal of Approximate Reasoning, vol. 32, no. 2-3, pp. 85-102, Feb. 2003.
[6] S.J. Kang, C.H. Woo, H.S. Hwang, and K.B. Woo, "Evolutionary design of fuzzy rule base for nonlinear system modeling and control," IEEE Transaction on Fuzzy Systems, vol. 8, no. 1, pp. 37-45, Feb. 2000.
[7] N. Li, S. Y. Li and Y.G. Xi, "Multi-model modeling method based on satisfactory clustering," Control theory & application, vol. 20, no. 5, pp. 783–787, May 2003.
[8] W. Pedrycz, and M. Reformat, "Evolutionary fuzzy modeling," IEEE Transaction on Fuzzy Systems, vol. 11, no. 5, pp. 652-665, Oct. 2003.
[9] B. Rezaee, and M.H. Fazel Zarandi, "Data-driven fuzzy modeling for Takagi-Sugeno-Kang fuzzy system," Information Sciences, vol. 180, no. 2, pp. 241-255, Jan. 2010.
[10] M. Sugeno, and G.T. Kang, "Fuzzy modeling and control of multilayer incinerator," Fuzzy sets and systems, vol. 18, no. 3, pp. 329-346, April 1986.
[11] Z.G. Su, P.H. Wang, J. Shen, Y.F. Zhang, and L. Chen. "Convenient T-S fuzzy model with enhanced performance using a novel swarm intelligent fuzzy clustering technique," Journal of Process Control, vol. 22, no. 1, pp. 108-124, Jan. 2012.
[12] Z.G. Su, P.H. Wang, and Z.L. Song. "Kernel based nonlinear fuzzy regression model," Engineering Applications of Artificial Intelligence, vol. 26, no. 2, pp.724-738, Feb. 2013.
[13] M. Sugeno, and T. Yasukawa, "A fuzzy logic based approach to qualitative modeling," IEEE Transaction on Fuzzy Systems, vol. 1, no. 1, pp. 7–31, Feb. 1993.
[14] T. Takagi, and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," EEE Transactions on Systems, Man and Cybermetics, vol. 15, no. 1, pp. 116–132, Jan. 1985.
[15] C.W. Xu, and Z. Yong, "Fuzzy model identification and self-learning for dynamic systems," IEEE Transactions on Systems, Man and Cybermetics, vol. 17, no. 4, pp. 683–689, Jul. 1987.
[16] L.A. Zadeh, "Probability measures of fuzzy events," Journal of Mathematical Analysis and Applications, vol. 23, no. 2, pp. 421-427, 1968.
[17] L. Zhao, F. Qian, Y.P. Yang, et al., "Automatically extracting T-S fuzzy models using cooperative random learning particle swarm optimization," Applied soft computing, vol. 10, no. 3, pp. 938-944, Jun. 2010.

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)