

Fuzzy identification of nonlinear dynamic system based on selection of important input variables

LYU Jinfeng^{1,2}, LIU Fucui^{1,*}, and REN Yaxue¹

1. Engineering Research Center of the Ministry of Education for Intelligent Control System and Intelligent Equipment, Yanshan University, Qinhuangdao 066004, China; 2. School of Mathematics and Information Science and Technology, Hebei Normal University of Science and Technology, Qinhuangdao 066004, China

Abstract: Input variables selection (IVS) is proved to be pivotal in nonlinear dynamic system modeling. In order to optimize the model of the nonlinear dynamic system, a fuzzy modeling method for determining the premise structure by selecting important inputs of the system is studied. Firstly, a simplified two stage fuzzy curves method is proposed, which is employed to sort all possible inputs by their relevance with outputs, select the important input variables of the system and identify the structure. Secondly, in order to reduce the complexity of the model, the standard fuzzy c-means clustering algorithm and the recursive least squares algorithm are used to identify the premise parameters and conclusion parameters, respectively. Then, the effectiveness of IVS is verified by two well-known issues. Finally, the proposed identification method is applied to a realistic variable load pneumatic system. The simulation experiments indicate that the IVS method in this paper has a positive influence on the approximation performance of the Takagi-Sugeno (T-S) fuzzy modeling.

Keywords: Takagi-Sugeno (T-S) fuzzy modeling, input variable selection (IVS), fuzzy identification, fuzzy c-means clustering algorithm.

DOI: 10.23919/JSEE.2022.000027

1. Introduction

There are many forms of motion in nonlinear systems, most of which are complex and uncertain. In general, a large number of data points are used to build a model to describe the nonlinear system and solve the practical problems. Among the nonlinear models, the fuzzy model can deal with the relation of input and output, especially for a complex and imprecise system. Although any element that may affect the output can be considered as a possible input, some inputs may be noise. Therefore, in the fuzzy modeling of the nonlinear time-varying system, determin-

ing optimal input variables is a problem that must be solved. Input variable selection (IVS) is to use certain methods and criteria to find out significant variables from numerous candidate input variables [1]. Determining important input variables, eliminating redundant input variables and lessening the total number of input variables are of great significance for decreasing calculation consumption, reducing model dimensions and improving prediction accuracy. At present, IVS is mainly used in neural network prediction and feature extraction [2–5], which can effectively reduce the dimension of input variables and the complexity of the fuzzy model, and improve the generalization ability of the model. However, the selection of the input variables of the fuzzy model has been paid little attention.

The Takagi-Sugeno (T-S) fuzzy model by using the experimental data is regarded as an effective means for modeling the actual system [6–17]. The work of T-S fuzzy system identification consists of two parts: structure identification and parameter estimation. The structure identification consists of structure identification I (selecting and deciding system inputs) and structure identification II (setting the number of fuzzy rules and dividing the fuzzy space). Parameter identification contains antecedent and consequent parameter identification of fuzzy rules. According to Sugeno [7], each work of fuzzy identification has different influence on the identification result. Sugeno believed that structure recognition I is the most important, with a proportion of 100, followed by structure recognition II with a proportion of 10, and finally parameter identification with a proportion of only 1. Therefore, IVS plays a great role in improving the identification performance of the model. For a real complex system, there are many factors that affect the output of the system. If all the factors are considered as the input of the model, the fuzzy rules will grow exponentially with the increase of the system dimension, which is unrealis-

Manuscript received March 22, 2020.

*Corresponding author.

This work was supported by the Natural Science Foundation of Hebei Province (F2019203505).

tic in the practical system application. For the rule-based fuzzy identification problem, selecting important input variables is an important means to solve the “dimension disaster” problem. In the existing literature, the input variables of fuzzy models are mostly determined by experience. Therefore, developing a more robust IVS approach is an important means to solve the problem of system identification with the fuzzy system, which can accurately describe the interdependence between the candidate input variables and the output variables, and can quickly sort the input variables according to the correlation.

The commonly used IVS methods contain two categories, one is model-based, and the other is model-free [1]. Model-free IVS methods include correlation coefficient (CC), partial correlation coefficient (PCC), mutual information (MI), partial mutual information (PMI), covariance matrix (CM), random forest (RF), etc.; and model-based IVS methods include input omission (IO), combined neural path strength (CNPS), etc. In [1], four input variable selection methods (PC, PMI, IO and CNPS) were comprehensively compared. Moreover, the IVS based artificial neural network flow prediction model is verified on two water basins datasets. In [18], the method MI was utilized to select input variables, and the dynamic neural network model was established to obtain the feeding conditions of broilers. The input variables were selected by forward selection (FS), Gamma test (GT) and principal component analysis (PCA), and the artificial neural network model was established to predict the human response to odor perception in [19]. The prediction model of carbon content in fly ash and the classification model of crude oil samples were established by RF, respectively in [20] and [21]. The model has a higher accuracy and faster calculation speed. For the selection of input variables of the fuzzy model, there are few literature. In order to determine the input variables of the Takagi-Sugeno and Kang (TSK) fuzzy model, Banakar et al. adopted the loop algorithm and the genetic algorithm, and established the dynamic system model by combining the modified mountain clustering (MMC) algorithm and the structure tree (ST) algorithm [22]. Lin et al. used two-stage fuzzy curve and surface (TSFCS) to select important input variables for nonlinear system structure identification [23]. With these fuzzy curves and surfaces, the correlation between the candidate input variables and the output variables was determined. The correlation was used to sort candidate input variables, select important input variables with strong correlation, and eliminate the variables with poor correlation.

In existing researches on fuzzy modeling for complex nonlinear dynamic systems, the inputs of the model are mostly selected or determined by experience. Though the

determination of input variables can make a great contribution to improving the performance of fuzzy identification, the research results on this part are not perfect yet. In order to make up for this shortcoming, this paper will carefully study what kind of performance the fuzzy model will show under the effect of IVS. In order to verify the accuracy and effectiveness of the fuzzy identification method considering the IVS, this paper compares with previous research results in two aspects, one is the number of input variables and the other is the number of fuzzy rules. The research results of this paper have important theoretical and practical significance for the practical application of fuzzy systems in nonlinear dynamic systems. In order to improve fuzzy modeling accuracy of nonlinear time-varying system, a method of fuzzy identification based on IVS is raised. The TSFCS method in [23] is simplified to identify the important input variables of the T-S fuzzy model. On this basis, the traditional fuzzy c-means (FCM) clustering algorithm and the recursive least square (RLS) method are utilized to identify the premise parameters and conclusion parameters respectively. The algorithm preprocesses the possible input variables offline, which reduces complexity of structure identification and can obtain a high identification accuracy without complicated parameter optimization.

This article is structured as below. Section 2 gives an outline of the T-S fuzzy model. In Section 3, a fuzzy identification method is proposed, which uses the FCM algorithm to identify antecedent parameters on the basis of determining significant input variables. In Section 4, two classical examples verify the performance of the established model. Furthermore, the discussed identification method considering IVS is applied to the identification of a real pneumatic loading system in Section 5. In the last section, the work of this paper is summarized.

2. Preliminaries

2.1 T-S fuzzy model

The T-S fuzzy model is a fuzzy rule based model, which approximates the nonlinear system by the local linear subsystem and realizes global nonlinearity by fuzzy reasoning. The premise of the rule is a fuzzy variable, and the conclusion is the linear function of input and output. The T-S fuzzy model with N groups of sample data $(x_{k1}, x_{k2}, \dots, x_{kr}, y_k)$ can be described as

$$R_i: \text{If } x_{k1} \text{ is } A_1^i, \dots, x_{kr} \text{ is } A_r^i, \\ \text{then } y_k^i = p_0^i + p_1^i x_{k1} + p_2^i x_{k2} + \dots + p_r^i x_{kr} \quad (1)$$

where $R_i (i = 1, 2, \dots, c)$ is the i th fuzzy rule, c stands for the fuzzy rule number, $\mathbf{p}^i = [p_0^i, p_1^i, p_2^i, \dots, p_r^i] \in \mathbf{R}^{r+1}$ represents the conclusion parameter vector of the i th fuzzy

rule, $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kr}]$ is the k th input vector, $A_1^i, A_2^i, \dots, A_r^i$ are premise fuzzy subsets and y_k^i is the local output variable.

ω_k^i is the matching degree, which is shown below:

$$\omega_k^i = \mu_{k1}^i(x_{k1}) \times \mu_{k2}^i(x_{k2}) \times \dots \times \mu_{kr}^i(x_{kr}) = \prod_{j=1}^r \mu_{kj}^i(x_{kj}) \quad (2)$$

where $\mu_{kj}^i(x_{kj})$ denotes the membership function of x_{kj} belonging to A_j^i . The Gaussian function is one of the most commonly used membership functions, and other types of membership functions include aggregate type membership function, trigonometric function and so on.

The global output of the fuzzy model is gained by weighted processing of local output and given by

$$\hat{y}_k = \sum_{i=1}^c \bar{\omega}_k^i y_k^i \quad (3)$$

where

$$\bar{\omega}_k^i = \frac{\omega_k^i}{\sum_{j=1}^c \omega_k^j}$$

is the validity function of the i th rule, $0 \leq \bar{\omega}_k^i \leq 1$ and $\sum_{i=1}^c \bar{\omega}_k^i = 1, \forall k = 1, 2, \dots, N; \forall i = 1, 2, \dots, c$.

From (1) and (3), the following equation can be obtained:

$$\begin{aligned} \hat{y}_k &= \sum_{i=1}^c [\bar{\omega}_k^i \cdot (p_0^i + p_1^i x_{k1} + p_2^i x_{k2} + \dots + p_r^i x_{kr})] = \\ &[\bar{\omega}_k^1 \bar{\omega}_k^1 x_{k1} \dots \bar{\omega}_k^1 x_{kr} \dots \bar{\omega}_k^c \bar{\omega}_k^c x_{k1} \dots \bar{\omega}_k^c x_{kr}] \cdot \\ &[p_0^1 p_1^1 \dots p_r^1 \dots p_0^c p_1^c \dots p_r^c]^T. \end{aligned} \quad (4)$$

The matrix equation (5) is obtained by substituting N pairs of sample data into (4).

$$\mathbf{Y} = \mathbf{X}\mathbf{P} \quad (5)$$

where $\mathbf{P} = [p_0^1 p_1^1 \dots p_r^1 \dots p_0^c p_1^c \dots p_r^c]^T$ represents the consequent parameter, which is an L -dimensional matrix, and $L = (r+1) \cdot c$; c is the number of fuzzy rules; r represents the number of input variables of the fuzzy model. $\mathbf{Y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N]$ is the output vector of $N \times 1$; \mathbf{X} is a matrix of $N \times L$, and

$$\mathbf{X}_k = [\bar{\omega}_k^1 \bar{\omega}_k^1 x_{k1} \dots \bar{\omega}_k^1 x_{kr} \dots \bar{\omega}_k^c \bar{\omega}_k^c x_{k1} \dots \bar{\omega}_k^c x_{kr}]$$

is the k th row of \mathbf{X} . $\mathbf{P}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ is the required parameter vector.

The purpose of consequent parameters identification is to determine the coefficient of linear regression model (1). In order to avoid inverse operation of the matrix, the RLS method is used to obtain the parameter estimation vector \mathbf{P} by iterative optimization, and the recursive algorithm is

$$\mathbf{P}_{l+1} = \mathbf{P}_l + \frac{\mathbf{S}_l \cdot \mathbf{X}_{l+1}^T \cdot (y_{l+1} - \mathbf{X}_{l+1} \cdot \mathbf{P}_l)}{1 + \mathbf{X}_{l+1} \cdot \mathbf{S}_l \cdot \mathbf{X}_{l+1}^T} \quad (6)$$

$$\mathbf{S}_{l+1} = \mathbf{S}_l - \frac{\mathbf{S}_l \cdot \mathbf{X}_{l+1}^T \cdot \mathbf{X}_{l+1} \cdot \mathbf{S}_l}{1 + \mathbf{X}_{l+1} \cdot \mathbf{S}_l \cdot \mathbf{X}_{l+1}^T} \quad (7)$$

where $l = 0, 1, 2, \dots, N-1, \mathbf{S}_l$ is a matrix for auxiliary calculation.

The initial condition is $\mathbf{P}_0 = 0, \mathbf{S}_0 = \alpha \mathbf{I}$. α is a constant greater than 10 000. \mathbf{I} is a unit matrix, which is $L \times L$ -dimensional. The optimal conclusion parameter matrix in the sense of mean square error (MSE) is obtained by using (6) and (7), and the conclusion parameter and MSE are output after recursion.

$$\text{MSE} = \sum_{k=1}^N (y_k - \hat{y}_k)^2 / N$$

where \hat{y}_k and y_k are the k th output of the model and actual system respective. In the next series of studies, $e_k = y_k - \hat{y}_k$ is used to evaluate the gap between the model output and the source output.

2.2 Fuzzy c-means clustering algorithm

There are many methods to divide fuzzy space and estimate premise parameters, such as the Gustafson-Kessel clustering algorithm [24], the Gath-Geva clustering algorithm [25], FCM and the fuzzy C-regression model (FCRM) [26–31]. Among them, the FCM algorithm is an effective method in practical application, which has the characteristics of simple implementation and a high identification accuracy.

The purpose of FCM is to minimize the objective function

$$\begin{aligned} J_m(U, V) &= \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m (d_{ik})^2, \\ \text{s.t. } &\sum_{i=1}^c u_{ik} = 1. \end{aligned} \quad (8)$$

In the above equation, N is the sample quantity, c is the number of fuzzy rules, m represents the weight of the membership function, usually an integer greater than 1. The value of m will affect the accuracy of recognition, and $m = 2$ is taken usually. u_{ik} characterizes the membership of the k th sample point $\mathbf{x}_k = (x_{k1}, x_{k2}, \dots, x_{kr})$ in the i th clustering and $u_{ik} \geq 0$; \mathbf{v}_i is the i th cluster center. $\forall i = 1, \dots, c, u_{ik}$ and \mathbf{v}_i are calculated by (9)–(11):

$$\mathbf{v}_i = \sum_{k=1}^N (u_{ik})^m \mathbf{x}_k \Big/ \sum_{k=1}^N (u_{ik})^m \quad (9)$$

$$u_{ik} = 1 \Big/ \sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)} \quad (10)$$

$$d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\| \geq 0, \quad \forall i \text{ and } k,$$

$$\text{If } d_{ik} = 0, \text{ then } u_{ik} = 1, u_{jk} = 0, j \neq i. \quad (11)$$

$U = (u_{ik})$ is a fuzzy partition matrix and $V = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c)^T$ is the matrix composed of cluster centers.

According to the given data point (\mathbf{x}_k, y_k) , the following steps give a detailed offline calculation method:

Step 1 Initialize the number of iterations $l = 0$, set the value of c , weight index m , stop criterion ε , and initialize the matrix U randomly.

Step 2 Update the matrix V by (9).

Step 3 Use (11) to update the distance d_{ik} of each cluster.

Step 4 Update the matrix U by (10).

Step 5 If $\|U^{(l)} - U^{(l-1)}\| < \varepsilon$, stop, otherwise skip back to Step 2.

3. Fuzzy identification method based on IVS

This part shows in detail the innovations of this article. Firstly, the algorithm of IVS is discussed to determine the important input variables for structure identification. Secondly, the FCM algorithm is used for parameter identification. Then, the fuzzy model is constructed by combining IVS with FCM. Compared with the previous fuzzy models, the advantage of this model can avoid the complex iterative optimization process of fuzzy model parameters and reduce the calculation amount on the premise of high fuzzy modeling accuracy. It is suitable for online identification of practical system and has important theoretical and practical guiding significance for the practical application of fuzzy system in nonlinear dynamic system.

3.1 Improved IVS

The IVS method based on TSFCS is more suitable for the nonlinear system with strong interdependence between input variables. The two-stage fuzzy curve gives the correlation between a single input variable and the output, and the two-stage fuzzy surface describes the interdependence between two different input variables relative to the output. In order to simplify the model, the correlation between input variables is not considered temporarily, and the difference calculation in the original algorithm is abandoned. The simplified TSFCS method is called the two stage fuzzy curve (TSFC) method, through which all possible input and output correlation indexes are obtained and sorted. The detailed process is discussed as below. See Fig. 1 for the specific algorithm flow chart.

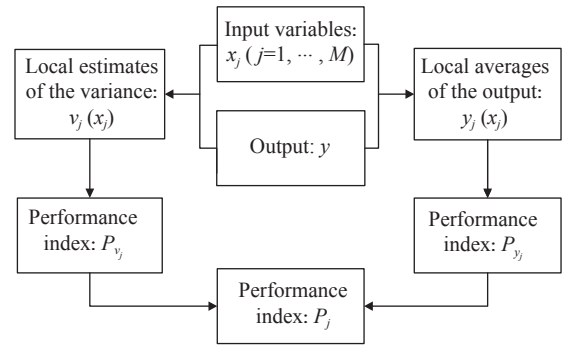


Fig. 1 Schematic diagram of TSFC

3.1.1 The first stage fuzzy curves

Based on the design concept that the more important the inputs are, the more relevant the outputs are, the first stage fuzzy curves are designed. Suppose x_1, x_2, \dots, x_M are the possible input variables of the fuzzy system model, y is the output, and N is the sample quantity.

First, in each $x_i - y$ space, a Gaussian membership function $\mu_{ki}(x_i)$ is constructed:

$$\mu_{ki}(x_i) = \exp\left(-\left(\frac{x_{ki} - x_i}{\beta_i}\right)^2\right) \quad (12)$$

where $x_{k1}, x_{k2}, \dots, x_{kM}, y_k$ is a sample point in $x_i - y$ space, $k = 1, 2, \dots, N$; β_i is the width of the membership function, in general, $\beta_i = 0.2(\max_{1 \leq k \leq N} x_{ki} - \min_{1 \leq k \leq N} x_{ki})$.

Then, for each x_i , a fuzzy curve $\tilde{y}_i(x_i)$ is defined by

$$\tilde{y}_i(x_i) = \frac{\sum_{k=1}^N [y_k \cdot \mu_{ki}(x_i)]}{\sum_{k=1}^N \mu_{ki}(x_i)}. \quad (13)$$

The fuzzy curves \tilde{y}_i in the above formula represents the local weighted mean of y_k for each x_i , and β_i decides the range of the local neighborhood. If an input variable x_i is more relevant to the output y_k than x_j , then the difference between $\tilde{y}_i(x_{ki})$ and output y_k is less than that between $\tilde{y}_j(x_{kj})$ and y_k , which indicates the input x_i is more significant than x_j . Therefore, a performance index $P_{\tilde{y}_i}$ is introduced to describe the importance of variables x_i :

$$P_{\tilde{y}_i} = \frac{1}{N v_y} \sum_{k=1}^N (\tilde{y}_i(x_{ki}) - y_k)^2 \quad (14)$$

where \bar{y} and $v_y = \sum_{k=1}^N (y_k - \bar{y})^2 / N$ is mean and variance of all output $\{y_1, y_2, \dots, y_N\}$ respectively. A smaller $P_{\tilde{y}_i}$ means that x_i is more important. Therefore, sorting $P_{\tilde{y}_i}$ in the descending order, a list of all candidate input variables also in the descending order of importance can be obtained.

3.1.2 The second stage fuzzy curves

If $i \neq j$, $P_{\tilde{y}_i}$ is equal to $P_{\tilde{y}_j}$, which makes it impossible to identify which input variable is more important. The fuzzy curve $\tilde{v}_i(x_i)$ in the second stage aims to solve the problem. The second stage fuzzy curve is given through fuzzy curves in (13) as follows:

$$\tilde{v}_i(x_i) = \frac{\sum_{k=1}^N [(\tilde{y}_i(x_i) - y_k)^2 \cdot \mu_{ki}(x_i)]}{\sum_{k=1}^N \mu_{ki}(x_i)}. \quad (15)$$

The fuzzy curve $\tilde{v}_i(x_i)$ is a weighted estimation of the output variance along the x_i -axis. If there is a poor correlation between an input x_i and the output y , then there should be $\tilde{v}_i(x_{ki}) \approx v_y$ for all x_{ki} ; on the contrary, if there is a strong correlation between the input and the output, then the value $\tilde{v}_i(x_i)$ at any point is significantly different from the global variance v_y . The second stage fuzzy curve can be used to calculate the performance index function equation (16), the function of which reflects the contribu-

tion of all possible inputs to the output.

$$P_{\tilde{v}_i} = \frac{1}{N \cdot (v_y)^2} \sum_{k=1}^N [\tilde{v}_i(x_{ki}) - v_y]^2 \quad (16)$$

Different from $P_{\tilde{y}_i}$ in (14), a larger $P_{\tilde{v}_i}$ means that x_i is more significant. For the convenience of application, $P_{\tilde{y}_i}$ and $P_{\tilde{v}_i}$ are unified into a performance index P_i ($0 < P_i < 1$):

$$P_i = \frac{P_{\tilde{y}_i}}{1 + P_{\tilde{v}_i}}. \quad (17)$$

The variable list sorted by importance from large to small can be obtained by calculating the performance index function P_i and arranging it in the descending order. If $P_i = 0$, it means a good performance; if $P_i = 1$, it means a poor performance.

3.2 The proposed T-S fuzzy identification approach

The T-S fuzzy identification approach discussed in this article is illustrated in Fig. 2, which combines the TSFC method and the FCM algorithm.

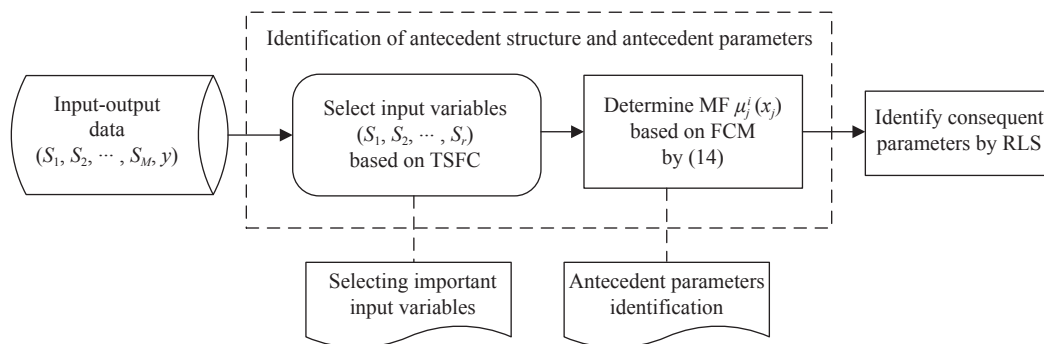


Fig. 2 T-S fuzzy system identification approach based on TSFC

The detailed steps of the fuzzy modeling are described as below:

Step 1 Use the TSFC method to sort input variables;

Step 2 Set termination threshold ε and the number of input variable r , conduct fuzzy subsets partition (determine c);

Step 3 Calculate the antecedent parameters $\mu_{kj}^i(x_{kj})$ according to (10);

Step 4 Get X from (5);

Step 5 Solve P through (6) and (7);

Step 6 Calculate the model performance evaluation index MSE. If MSE meets the identification accuracy, then the identification algorithm will terminate; otherwise, add c and go to Step 3.

4. Simulation research

The main goal of this section is to verify the effectiveness of IVS based on the TSFC method through several typical nonlinear system models. To determine the reliable

dynamic characteristics, two methods, FCM and TSFC+FCM algorithms, are used to carry out a comparative study.

4.1 Example 1: Mackey-Glass chaotic system

4.1.1 IVS based on TSFC

The Mackey-Glass chaotic system [32] is considered as a standard case, which is widely applied to research of fuzzy model performance. It is obtained by

$$\frac{dx(t)}{dt} = \frac{0.2x(t-17)}{1+x^{10}(t-17)} - 0.1x(t). \quad (18)$$

These 1000 datasets are obtained by (18). Applying the TSFC method mentioned in Section 3, letting $s_i = x(t-i)$ ($i = 1, 2, \dots, 18$), $y = x(t+1)$, calculating the index function P_i of each variable s_i and sorting all variables according to the value of P_i , the first six important input variables s_1, s_2, s_3, s_4, s_5 and s_{18} are screened out, and the corresponding values P_i are 0.0611, 0.1164, 0.1967,

0.3003, 0.4203 and 0.4960 respectively. The first six input variables represented by them are $x(t-1)$, $x(t-2)$, $x(t-3)$, $x(t-4)$, $x(t-5)$ and $x(t-18)$.

4.1.2 Experimental results

One thousand sets of data obtained by (18) are equally divided into two groups to verify the prediction performance of the model. The first group of data is used for model training, and the latter group of data is used for model testing.

In general, the six variables $x(t-1)$, $x(t-2)$, $x(t-3)$, $x(t-4)$, $x(t-5)$ and $x(t-6)$ are selected as model inputs. In this article, the above selected variables $x(t-1)$, $x(t-2)$, $x(t-3)$, $x(t-4)$, $x(t-5)$ and $x(t-18)$ are used for fuzzy modeling. Set the initial values of $x(0)$ to $x(17)$ all as 1.2 and the rule number $c = 2$. The performance of our fuzzy model is shown in Fig. 3, where Fig. 3 (a) exhibits the comparison of the original output and the predicted output of the model, and Fig. 3 (b) depicts respective errors.

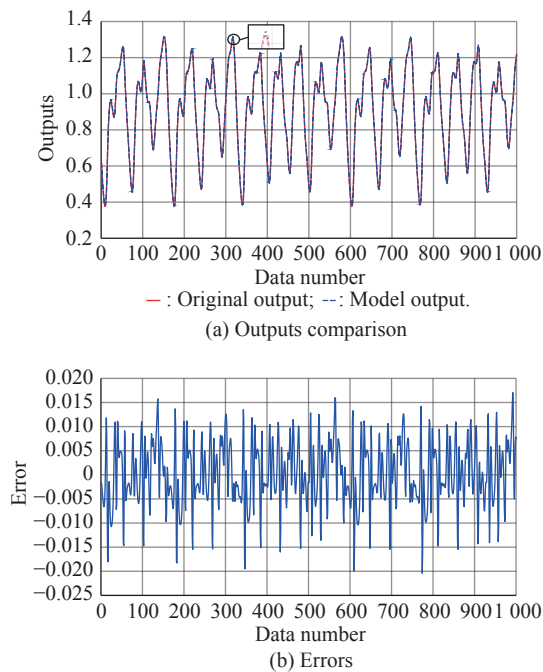


Fig. 3 Contrast of established model with actual model for Example 1

The traditional FCM algorithm and the proposed algorithm TSFC+FCM algorithm are respectively used for modeling comparison of Example 1. Under the condition that the fuzzy rules are both 2, the modeling accuracy (MSE1) of the FCM algorithm and the TSFC algorithm are 4.5655×10^{-5} and 4.1627×10^{-5} , and the prediction accuracy (MSE2) are 4.6050×10^{-5} and 4.2957×10^{-5} , respectively. From the experimental results, it can be concluded that the prediction performance of fuzzy identifi-

cation based on the TSFC method is better than that without IVS.

In addition, we compare our model performance with other results in literature in Table 1, and our model has a higher identification accuracy. The training MSE is 4.1627×10^{-5} and the obtained prediction MSE is 4.2957×10^{-5} for the test data.

Table 1 Performance contrast of different fuzzy models for Example 1

Model	Number of rules	Training MSE ($\times 10^{-4}$)	Testing MSE ($\times 10^{-4}$)
Guo et al. [11]	–	9.61	10.24
Masoumi et al. [33]	–	4.84	7.29
Zou et al. [13]	10	5.0216	6.7449
Liu [34]	2	0.6377	0.6518
Our model	2	0.41627	0.42957

4.2 Example 2: Box-Jenkins system

4.2.1 IVS based on TSFC

The Box-Jenkins data set [35] includes 296 sets of input and output measured values of the gas furnace process. At time k , the gas flow of the system input is expressed with $u(k)$ and CO_2 concentration of the system output is $y(k)$. The purpose of this experiment is to select the optimal inputs and predict $y(k+1)$. For this purpose, given

$$s_i = \begin{cases} u(k-i+1), & i = 1, 2, \dots, 5 \\ y(k-i+5), & i = 6, 7, 8 \end{cases}$$

The importance index P_i of each input to be selected s_i is calculated according to (17) based on the TSFC method and arranged in the descending order. Here, the first six input variables: s_6 , s_5 , s_7 , s_4 , s_8 and s_3 are screened out, and the values P_i of them are 0.0705, 0.1494, 0.1605, 0.2604, 0.3061 and 0.4295 respectively, and the corresponding actual inputs are $y(k-1)$, $u(k-4)$, $y(k-2)$, $u(k-3)$, $y(k-3)$, $u(k-2)$.

4.2.2 Experimental results

In this section, the model generalization is tested through two situations. In the first case, a full set of data is devoted to establish the model to verify its superiority; in the second case, the data set is equally divided into two parts, the former is used for model training and the latter is for examining the prediction ability of the model.

In most available literature, the input variables of the Box-Jenkins system are $u(k)$, $u(k-1)$, $u(k-2)$, $y(k-1)$, $y(k-2)$ and $y(k-3)$. The variables selected above are used as the input of the model. Fig. 4 and Fig. 5 show the model approximation performance in two cases respectively.

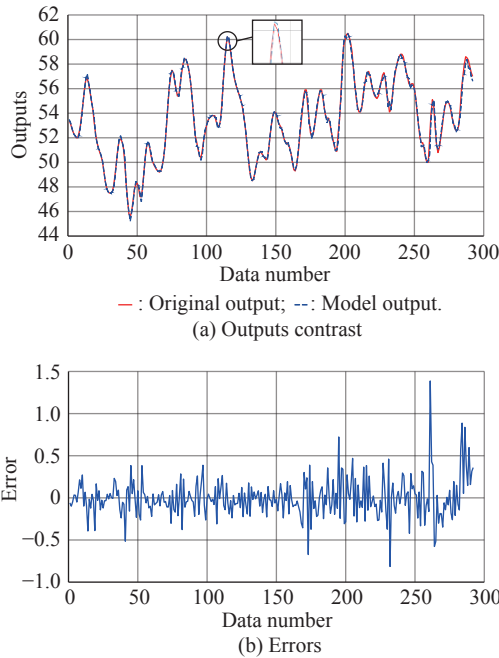


Fig. 4 Contrast of established model with actual model for Example 2 (Case 1)

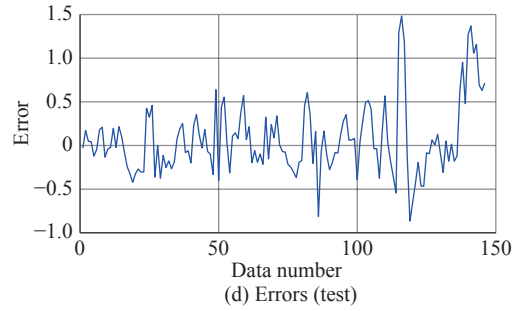
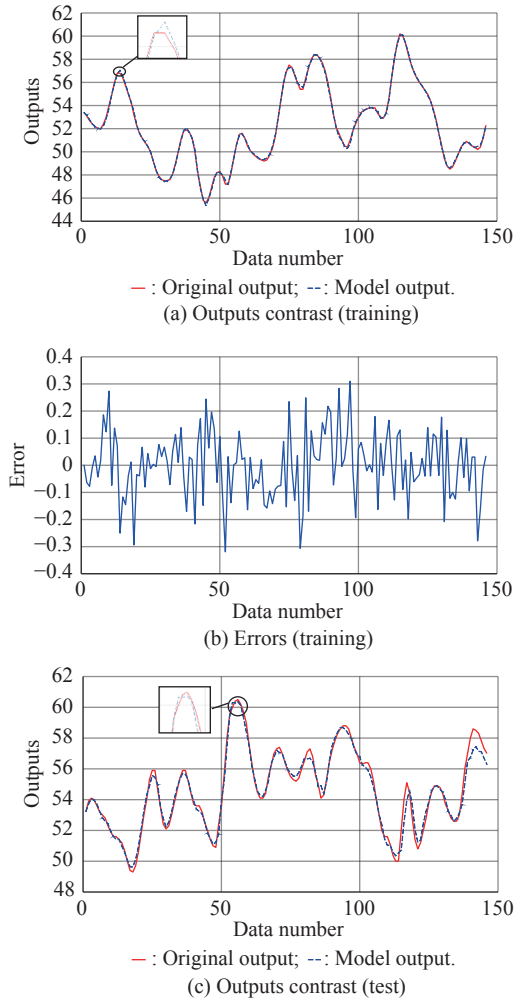


Fig. 5 Contrast of established model with actual model for Example 2 (Case 2)

Aiming at the first case, it is available from Fig. 4 (b) that the modeling error between 200–290 data groups is relatively large compared with that between 1–200 data groups. In [29], a classic fuzzy identification literature, and [8], the error of Box-Jenkins gas furnace modeling data also had an increasing trend in the same range. For the second case, in the test model stage of [36], the test error was also increasing in the 160–290 range. It can be concluded that since the data of gas furnace is the measurement data obtained through experiments, there will inevitably be noise interference, which will affect the modeling results.

In both Case 1 and Case 2, the simulation results obtained by the TSFC + FCM algorithm and the FCM algorithm are displayed in Table 2 and Table 3. From Table 2 and Table 3, the identification performance of the model based on TSFC is obviously better than those without IVS under the condition of the same number of input variables. In both cases, detailed comparisons of our model performance with others are given in Table 4 and Table 5, which show that our model has a higher accuracy than other models in literature. It should be noted that those results in Table 4 and Table 5 are mostly obtained by using the complex parameter optimization process. In [10], the T-S fuzzy system with the least fuzzy rules was established by using the iterative vector quantization clustering method, and in [37], based on the establishment of fuzzy structure by FCRM, the number of fuzzy rules was optimized.

Table 2 Performance comparison of IVS based on TSFC for Example 2 (Case 1)

Model	Input variables	Number of rules	MSE
FCM	$y(k-1), u(k)$	2	0.4243
TSFC+FCM	$y(k-1), u(k-4)$	2	0.1655
FCM	$y(k-1), y(k-2), u(k), u(k-1)$	2	0.0757
TSFC+FCM	$y(k-1), y(k-2), u(k-3), u(k-4)$	2	0.0609
FCM	$y(k-1), y(k-2), y(k-3), u(k), u(k-1), u(k-2)$	2	0.0570
TSFC+FCM	$y(k-1), y(k-2), y(k-3), u(k-2), u(k-3), u(k-4)$	2	0.0545

Table 3 Performance contrast of different fuzzy models for Example 2 (Case 1)

Model	Input	c	MSE1 (Train)	MSE2 (Test)
FCM	$y(k-1), u(k)$	2	0.3252	0.9611
TSFC+FCM	$y(k-1), u(k-4)$	2	0.0801	0.3803
FCM	$y(k-1), y(k-2), u(k), u(k-1)$	2	0.0305	0.5387
TSFC+FCM	$y(k-1), y(k-2), u(k-3), u(k-4)$	2	0.0164	0.1679
FCM	$y(k-1), y(k-2), y(k-3), u(k), u(k-1), u(k-2)$	2	0.0171	0.1438
TSFC+FCM	$y(k-1), y(k-2), y(k-3), u(k-2), u(k-3), u(k-4)$	2	0.0152	0.1707

Table 4 Performance comparison of IVS based on TSFC for Example 2 (Case 2)

Model	Number of inputs	c	Training MSE
Tsekouras [28]	6	8	0.075
Liu [34]	6	2	0.0561
Li et al. [8]	6	4	0.0498
Li et al. [37]	6	3	0.0560
Li et al. [36]	6	3	0.0534
Our model	6	2	0.0545

Table 5 Performance comparison of different fuzzy models for Example 2 (Case 2)

Model	Number of rules	MSE1(Train)	MSE2(Test)
Tsekouras [28]	7	0.022	0.236
Li et al. [37]	3	0.0159	0.1255
Li et al. [36]	3	0.0150	0.1470
Luo et al. [10]	2	0.0254	0.1243
Yan et al. [15]	2	0.0168	0.1402
Li et al. [14]	3	0.0149	0.1324
Our model	2	0.0152	0.1707
Our model	3	0.0150	0.1567

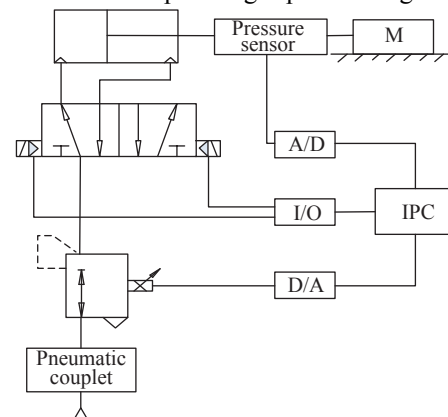
The gravitational search algorithm (GSA) was used to optimize the antecedent parameters of the fuzzy model in [14,36]. Yan et al. presented an improved hybrid backtracking search algorithm (IHBSA) to search the optimal number of fuzzy clusters and the corresponding cluster centers at the same time in [15]. However, the algorithm proposed in this paper avoids the complex iterative optimization process, which is the advantage of this model.

5. Application to the variable load pneumatic loading system

Due to the characteristic of simple structure, small volume, and without pollution and wide applicability, the pneumatic loading system has attracted much attention in the latest research. It has been widely applied to industrial automation, aerospace, health care and other fields [38].

However, the complexity of the gas flow, the charge and exhaust characteristics of the two chambers of the cylinder, the nonlinear flow of the electric proportional valve and other factors, lead to the pneumatic loading system having strong nonlinearity and strong coupling, which makes modeling and controlling the variable load pneumatic loading system become a difficult problem in industrial application. There are two approaches to build the system model in practical application. One is to establish the mechanism model according to the dynamic law of system operation; the other is to establish the data model by collecting experimental data of system operation. In our research, the fuzzy model of the pneumatic loading system is built based on data drive.

The pneumatic variable load test machine is suitable for the precise loading of small load and the continuous variable loading of the numerical value. It can simulate the movement behavior of the moving interface of the ground and space mechanism, and realize the friction and wear simulation test between various moving interfaces under the gravity and microgravity environment. The structure block diagram of pneumatic loading system controlled by electric proportional valve is shown in Fig. 6. The hardware system is mainly composed of air source, pneumatic couplet, electric proportional valve, industrial control computer, pressure sensor, data acquisition card, cylinder, etc. Among them, the pilot electric proportional valve of SMC ITV2050 is used to convert the electric signal into the pressure signal; the MCL-L pull and press type sensor is used as the detection element of the system; the D/A and A/D are composed of Advantech PCI1710 and Advantech PCI1710. The system controller is on IPC-610h industrial computer, which is used to collect, process, calculate and output the gas pressure signal.

**Fig. 6 Structural diagram of the pneumatic system**

Within the allowable scope of the dynamic system, this system uses square wave load as the excitation signal, as shown in (19):

$$u(t) = \begin{cases} 30, & 10k \leq t \leq 10k + 5 \\ 150, & 10k + 5 \leq t \leq 10k + 10 \end{cases} \quad (19)$$

It is continuously applied to the system in the opening loop state, and the input and output data of the system are collected for offline modeling. The sampling cycle and sampling duration are 0.1s and 100s respectively, and 1 000 sample data $[u(t), y(t)]$ are acquired, where $u(t)$ is the current input of the pneumatic proportional valve and $y(t)$ is the actual pressure output of the pneumatic loading system.

First, the appropriate input variables are selected by using 1 000 sets of training data denoted as

$$s_i = \begin{cases} u(t-11+i), & i = 1, \dots, 11 \\ y(t+11-i), & i = 12, \dots, 21 \end{cases}$$

According to the method of TSFC, the first six important input variables are screened out in sequence: s_{12} , s_9 , s_8 , s_{10} , s_{13} and s_7 , the values P_i of which are 0.0212, 0.0262, 0.0375, 0.0577, 0.0598 and 0.0702 respectively. The corresponding actual variables are $y(t-1)$, $u(t-4)$, $y(t-2)$, $u(t-3)$, $y(t-3)$, $u(t-2)$. The first four and six variables are separately chosen to establish the T-S fuzzy model, and the rule number is taken as 3. Then, FCM and RLS are used for identification of the premise parameters and conclusion parameters.

The approximation of the established model output to the actual system output is displayed in Fig. 7 (a), and the error between the two is shown in Fig. 7 (b). The performance comparison of the input variable selection model of the pneumatic loading system is shown in Table 6. Obviously, compared with the model with traditional input variables, this model has a higher accuracy. When the first six or the first four variables are adopted as model inputs, the MSE of the fuzzy model is 2.0168 or 3.0505, respectively. However, when six or four conventional variables are taken as model inputs, the MSE is 14.8809 or 24.4626. The approach error evaluation index of the proposed fuzzy modeling method is obviously reduced. The model based on IVS can effectively overcome the impact on system modeling created by time delay, which has great importance to the actual dynamic system modeling.

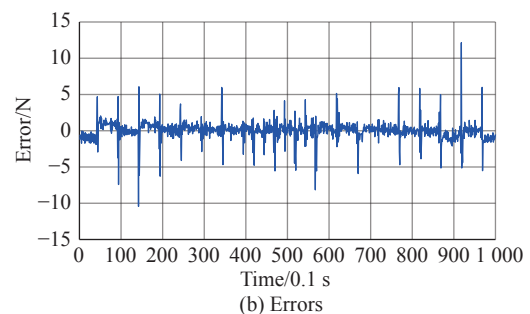
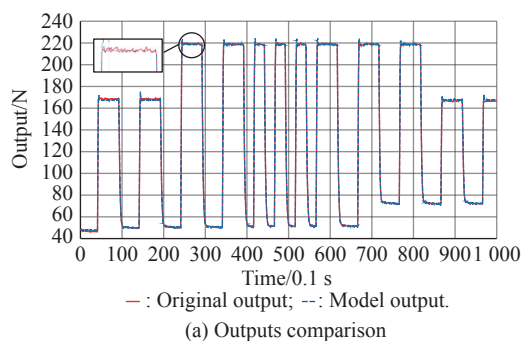


Fig. 7 Comparison of established model and actual system for pneumatic system

Table 6 Performance comparison of IVS based on TSFC for pneumatic system

Model	Input	Number of rules	MSE
FCM	$u(t-1), u(t), y(t-1), y(t-2)$	3	24.4626
TSFC-FCM	$u(t-1), u(t-2), u(t-3), y(t-1)$	3	3.0505
FCM	$u(t-2), u(t-1), u(t), y(t-1), y(t-2)$	3	14.8809
TSFC-FCM	$u(t-1), u(t-2), u(t-3), u(t-4), y(t-1)$	3	2.0168

6. Conclusions

In this study, a fuzzy model with important input variables is established to improve the identification accuracy of fuzzy system and reduce the complexity of the model. First of all, the improved IVS algorithm is adopted, that is, TSFC is used to screen the necessary input variables for model identification. Then, FCM is used for identification of premise parameters and the RLS method is used for identification of conclusion parameters. Finally, we apply the algorithm to the fuzzy identification benchmark problems and an actual system. The research results reveal that the identification performance of the T-S fuzzy model based on TSFC can be effectively improved. Compared with the previous fuzzy model, the combination of TSFC and FCM algorithms preprocess the important input variables off-line, which can simplify the complexity of fuzzy recognition and achieve a higher recognition accuracy without optimizing parameters.

References

- [1] SNIEDER E, SHAKIR R, KHAN U T. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. *Journal of Hydrology*, 2019. DOI: 10.1016/j.jhydrol.2019.124299.
- [2] CHEN S F, GU H, TU M Y, et al. Robust variable selection based on bagging classification tree for support vector machine in metabonomic data analysis. *Journal of Chemometrics*, 2018, 32(11): e2921.
- [3] KANG R, ZHANG X K, LIU H J, et al. Selection of optimal combinations of inputs in a partial least squares model for

- prediction of soil organic matter. *Spectroscopy Letters*, 2018, 51(7): 373–381.
- [4] WANG Y X, JIA Z H, YANG J. A variable selection method of the significance multivariate correlation competitive population analysis for near-infrared spectroscopy in chemical modeling. *IEEE Access*, 2019, 7: 167195–167209.
- [5] KIM G H, KIM S H. Variable selection for artificial neural networks with applications for stock price prediction. *Applied Artificial Intelligence*, 2019, 33(1): 54–67.
- [6] TAKAGI T, SUGENO M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on Systems, Man, and Cybernetics*, 1985, 15(1): 116–132.
- [7] SUGENO M, YASUKAWA T. A fuzzy-logic-based approach to qualitative modeling. *IEEE Trans. on Fuzzy Systems*, 1993, 122(1): 7–31.
- [8] LI C S, ZHOU J Z, XIANG X Q, et al. T-S fuzzy model identification based on a novel fuzzy c-regression model clustering algorithm. *Engineering Applications of Artificial Intelligence*, 2009, 22(4/5): 646–653.
- [9] CHANG C W, TAO C W. A novel approach to implement Takagi–Sugeno fuzzy models representation. *IEEE Trans. on Cybernetics*, 2017, 47(9): 2353–2361.
- [10] LUO M N, SUN F C, LIU H P. Hierarchical sparse representation for T-S fuzzy systems identification. *IEEE Trans. on Fuzzy Systems*, 2013, 21(6): 1032–1043.
- [11] GUO F, LIN L, XIE X, et al. Novel hybrid rule network based on TS fuzzy rules. *Neural Network World*, 2015, 25(1): 93–116.
- [12] ZHAO K, LI S R, KANG Z J. Takagi-Sugeno fuzzy modeling and control of nonlinear system with adaptive clustering algorithms. *Proc. of the 10th International Conference on Modelling, Identification and Control*, 2018: 1–6.
- [13] ZOU W, LI C S, ZHANG N. A T-S fuzzy model identification approach based on a modified inter Type-2 FRCM algorithm. *IEEE Trans. on Fuzzy Systems*, 2018, 26(3): 1104–1113.
- [14] LI C S, ZOU W, ZHANG N, et al. An evolving T-S fuzzy model identification approach based on a special membership function and its application on pump-turbine governing system. *Engineering Application of Artificial Intelligence*, 2018, 69: 93–103.
- [15] YAN S Q, ZHOU J Z, ZHENG Y, et al. An improved hybrid backtracking search algorithm based T-S fuzzy model and its implementation to hydroelectric generating units. *Neurocomputing*, 2018, 275: 2066–2079.
- [16] CHROUTA J, FARHANI J, ZAAFOURI A, et al. A methodology for modelling of Takagi-Sugeno fuzzy model based on multi-particle swarm optimization: application to gas furnace system. *Proc. of the 6th International Conference on Control, Decision and Information Technologies*, 2019: 23–26.
- [17] LI R C, GUO Y Q, NGUANG S K, et al. Takagi-Sugeno fuzzy model identification for turbofan aero-engines with guaranteed stability. *Chinese Journal of Aeronautics*, 2018, 31(6): 1206–1214.
- [18] JOHANSEN S V, BENDTSEN J D, JENSEN M R, et al. Broiler weight forecasting using dynamic neural network models with input variable selection. *Computers and Electronics in Agriculture*, 2019, 159: 97–109.
- [19] CHANG F, HEINEMANN P H. Optimizing prediction of human assessments of dairy odors using input variable selection. *Computers and Electronics in Agriculture*, 2018, 150: 402–410.
- [20] WANG Q, FAN C H, BAI J Y, et al. SO₂ emission characteristic modeling based on variable selection and SVM. *Thermal Power Generation*, 2018, 47(3): 68–75. (in Chinese)
- [21] LOVATTI B P O, NASCIMENTO M H C, NETO A C, et al. Use of random forest in the identification of important variables. *Microchemical Journal*, 2019, 145: 1129–1134.
- [22] BANAKAR A, AZEEM M F. Input selection for TSK fuzzy model based on modified mountain clustering. *Proc. of the 3rd International IEEE Conference on Intelligent Systems*, 2006: 295–299.
- [23] LIN Y H, CUNNINGHAM G A, COGGESHALL S V, et al. Nonlinear system input structure identification: two stage fuzzy curves and surfaces. *IEEE Trans. on System, Man, and Cybernetics—Part A: Systems and Humans*, 1998, 28(5): 678–684.
- [24] GUSTAFSON D E, KESSEL W C. Fuzzy clustering with a fuzzy covariance matrix. *Proc. of the 17th IEEE Conference on Decision Control*, 1979: 761–766.
- [25] GATH I, GEVA A B. Unsupervised optimal fuzzy clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 773–781.
- [26] TSAI S H, CHEN Y W. A novel identification method for Takagi-Sugeno fuzzy model. *Fuzzy Sets and Systems*, 2018, 338: 117–135.
- [27] ZHU L F, WANG J S, WANG H Y. A novel clustering validity function of FCM clustering algorithm. *IEEE Access*, 2019, 7: 152289–152315.
- [28] TSEKOURAS G E. On the use of the weighted fuzzy C-means in fuzzy modeling. *Advance Engineering Software*, 2005, 36(5): 287–300.
- [29] KIM E, PARK M, JI S, et al. A new approach to fuzzy modeling. *IEEE Trans. on Fuzzy Systems*, 1997, 5(3): 328–337.
- [30] JIANG W, YANG T, SHOU Y H, et al. Improved evidential fuzzy c-means method. *Journal of Systems Engineering and Electronics*, 2018, 29(1): 187–195.
- [31] LIU Z, CHEN X H, LIU J, et al. Source number estimation method based on fuzzy C-means clustering. *Systems Engineering and Electronics*, 2019, 41(2): 244–247. (in Chinese)
- [32] LO J C, YANG C H. A heuristic error-feedback learning algorithm for fuzzy modeling. *IEEE Trans. on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 1999, 29(6): 686–691.
- [33] MASOUMI M S, KHANJANI M, QADERI K. Uplift capacity prediction of suction caisson in clay using a hybrid intelligence method (GMDH-HS). *Applied Ocean Research*, 2016, 59: 408–416.
- [34] LIU F C. Fuzzy model identification of nonlinear system and its application. Beijing: National Defense Industry Press, 2006. (in Chinese)
- [35] BOX G E, JENKINS G M. Time series analysis: forecasting and control. San Francisco: Holden-Day Inc., 1976.
- [36] LI C S, ZHOU J Z, FU B, et al. T-S fuzzy model identification with a gravitational search-based hyper-plane clustering algorithm. *IEEE Trans. on Fuzzy Systems*, 2012, 20(2): 305–317.
- [37] LI C S, ZHOU J Z, LI Q Q, et al. A new T-S fuzzy-modeling approach to identify a boiler-turbine system. *Expert Systems with Applications*, 2010, 37(3): 2214–2221.
- [38] LIU F C, WANG L X, JIA X J, et al. The application of linear/nonlinear active disturbance rejection switching control in variable load pneumatic loading system. *Journal of Mechanical Engineering*, 2018, 54(12): 225–232. (in Chinese)

Biographies



LYU Jinfeng was born in 1977. She is currently a Ph.D. candidate in Automation Department of Yanshan University and an associate professor in Hebei Normal University of Science and Technology. Her main research interests are identification and control of nonlinear systems.
E-mail: xfy0308@163.com



REN Yaxue was born in 1995. She received her bachelor degree in measurement and control technology and instrument from Hebei University, and master degree in control science and engineering from Yanshan University. She is currently a Ph.D. student in instrument science and technology at Yanshan University. Her main research interest is plane array capacitance detection technology.
E-mail: 2592863354@qq.com



LIU Fucui was born in 1966. He received his master degree from the Automatic Control Department of Northeast Institute of Heavy Machinery in 1994, and his Ph.D. degree from the Control Science and Engineering Department of Harbin Institute of Technology in 2003. He is now the director of Automation Department of Yanshan University. His main research interests are nonlinear system fuzzy identification and predictive control, and space robot control technology.
E-mail: lfc@ysu.edu.cn