A Case Study of Data-driven Interpretable Fuzzy Modeling1)

XING Zong-Yi[1] JIA Li-Min[2] ZHANG Yong[1] HU Wei-Li[1] QIN Yong[2]

[1](*Automation Department, Nanjing University of Science and Technology, Nanjing* 210094)

[2](*School of Traffic and Transportation, Beijing Jiaotong University, Beijing* 100044)

(E-mail: xingzongyi@tom.com)

Abstract An approach to identify interpretable fuzzy models from data is proposed. Interpretability, which is one of the most important features of fuzzy models, is analyzed first. The number of fuzzy rules is determined by fuzzy cluster validity indices. A modified fuzzy clustering algorithm, combined with the least square method, is used to identify the initial fuzzy model. An orthogonal least square algorithm and a method of merging similar fuzzy sets are then used to remove the redundancy of the fuzzy model and improve its interpretability. Next, in order to attain high accuracy, while preserving interpretability, a constrained Levenberg-Marquardt method is utilized to optimize the precision of the fuzzy model. Finally, the proposed approach is applied to a PH neutralization process, and the results show its validity.

Key words Fuzzy modeling, interpretability, fuzzy clustering

1 Introduction

During the past few years, fuzzy modeling techniques have attracted research interest due to their successful applications in areas such as classification, data mining, pattern recognition, simulation, prediction, and control. Compared to traditional mathematical models and pure neural networks, fuzzy models possess some distinct advantages, including the facility for explicit knowledge representation in the form of if-then rules, and the ability to approximate complicated nonlinear functions with simple models.

Several fuzzy modeling methods have been proposed, including fuzzy clustering-based algorithms[1], neuro-fuzzy systems[2,3] and genetic rules generation[4,5]. However all these methods only focus on fitting data with the highest possible accuracy, neglecting the interpretability of the obtained fuzzy models, which is a primary advantage of fuzzy systems and the most prominent feature that distinguishes fuzzy systems from many other models[6].

In order to improve interpretability of fuzzy models, some methods have been developed. Roubos *et al.*[7~11] stated several necessary conditions for interpretability. Angelov *et al.*[12~18] proposed detailed techniques to improve interpretability of fuzzy models, including the method of merging similar fuzzy sets, rule reduction by orthogonal least-square methods and genetic algorithms, global learning and local learning. Roubos *et al.*[7] obtained an initial redundant fuzzy model by fuzzy clustering, and adopted the method of merging similar fuzzy sets and genetic algorithms to reduce the initial fuzzy model iteratively, and finally used genetic algorithms to optimize all parameters of the fuzzy model. Xing *et al.*[19] proposed an approach to identify accurate and interpretable fuzzy models by input variables selection and model selection. Min *et al.*[20] identified an initial fuzzy model by fuzzy clustering algorithm and fuzzy partition validity indices, and simplified the fuzzy model by similarity measure. Delgado *et al.*[21] presented fuzzy modeling as a multi-objective decision-making problem, considering accuracy, interpretability and autonomy as goals. All these goals were handled via a single-objective ε–constrained decision making problem, which was solved by a hierarchical evolutionary algorithm.

This paper proposes a systematic technique to construct interpretable and accurate fuzzy models. First, interpretability issues in fuzzy models are analyzed. Secondly fuzzy cluster validity indices are adopted to determine the number of fuzzy rules. A modified fuzzy clustering algorithm, combined with a least square method, is used to identify the initial fuzzy model. Then an orthogonal least-square algorithm and the method of merging similar fuzzy sets are used to remove redundancy of the fuzzy model and improve its interpretability. Finally, a constrained Levenberg-Marquardt method is utilized to optimize the fuzzy model and improve its precision. The proposed approach is applied to a PH neutralization process, and the results show its validity.

1) Supported by National Natural Science Foundation of P. R. China (60332020) and Scientific Research Foundation of Nanjing University of Science and Technology (2005)
Received July 27, 2004; in revised form July 18, 2005

2 Preliminaries

2.1 Takagi-Sugeno fuzzy model

The typical fuzzy rule of the Takagi-Sugeno (TS) fuzzy model[22] has the form:

$$\begin{aligned} R^i: &\text{IF } x_1 \text{ is } A_{i,1} \text{ and } \cdots \text{ and } x_p \text{ is } A_{i,p} \\ &\text{THEN } y_i = a_{i0} + a_{i1}x_1 + \cdots + a_{ip}x_p \end{aligned} \tag{1}$$

where x_j is the j-th input variable, A_{ij} is the fuzzy set of the j-th input variable in the i-th rule, y_i are output of the i-th fuzzy rule.

The output of the TS fuzzy model is computed using the normalized fuzzy mean formula:

$$y(k) = \sum_{i=1}^{c} p_i(x)\hat{y}_i \tag{2}$$

where c is the number of rules, and P_i is the normalized firing strength of the i-th rule:

$$P_i(x) = \frac{\prod_{j=1}^{p} A_{ij}(x_j)}{\sum_{i=1}^{c}\prod_{j=1}^{p} A_{ij}(x_j)} \tag{3}$$

Given N input-output data pairs $\{x_k, y_k\}$, the model in (2) can be written as a linear regression problem

$$\boldsymbol{y} = \boldsymbol{P\theta} + \boldsymbol{e} \tag{4}$$

where $\boldsymbol{\theta}$ is the consequent matrix of rules, and $\boldsymbol{e}$ is the approximation error matrix.

Gaussian membership function is used to represent the fuzzy set A_{ij}

$$A_{ij}(x_j) = \exp\left(-\frac{1}{2}\frac{(x_j - v_{ij})^2}{\sigma_{ij}^2}\right) \tag{5}$$

where v_{ij} and σ_{ij} represent the center and the variance of the Gaussian function respectively.

2.2 Interpretability issues in fuzzy models

Interpretability refers to the ability of the fuzzy model to express the behavior of a system in a human understandable way. This is a subjective property that depends on several factors. Although there is no formal definition for interpretability, several characteristics are believed to be essential. These are described as follows[7~11]:

1) The number of variables: a high-dimensional fuzzy model is difficult to interpret. The model should use as few variables as possible.

2) The number of rules: a fuzzy model with a large rule base is less interpretable than a fuzzy system containing few rules. Experientially, the number of fuzzy rules of an interpretable model is no more than ten, which is determined by the limit of human intelligence.

3) Completeness, consistency and compactness of fuzzy rules: for each effective input variable combination, there must be at least one fuzzy rule being fired, *i.e.*, fuzzy rules cover the whole input space. The fuzzy rules in the rule base should be consistent. If there are rules that are contradictory to one another, it is hard to understand the fuzzy rules. There must be no rule whose antecedent is a subset of another rule, and no rule may appear more than once in the rule base.

4) Characteristics of membership functions: convexity and normality are the two principle characteristics which are satisfied naturally for most widely-used membership functions, *e.g.*, the Gaussian function, and the triangle function. The fuzzy partition of all input variables should be complete to prevent unpredictable system outputs. Fuzzy sets should be distinguishable in order to assign linguistic terms to the fuzzy sets. Usually, a minimum/maximum degree of overlapping between fuzzy sets must be enforced.

3 Design of data-driven interpretable and accurate fuzzy model

3.1 Fuzzy cluster validation indices

It is essential to determine the number of rules, *i.e.*, the number of fuzzy clusters. This can be accomplished by validation analysis using fuzzy cluster validity indices.

There are two categories of fuzzy cluster validity indices. The first category uses only the membership values of the fuzzy partition of data. The second one involves both the partition matrix and the data itself.

The partition coefficient (PC) and the partition entropy coefficient (PE)[23~25] are the typical cluster validity indices of the first category.

$$PC(c) = \frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^2 \tag{6}$$

$$PE(c) = -\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}\log_a \mu_{ik} \tag{7}$$

where $PC(c) \in [1/c, 1]$, $PE(c) \in [0, \log_a c]$. With an increase of c, the values of PC and PE are decreased/increased, respectively. The number corresponding to a significant knee is selected as the optimal number of rules.

The above mentioned cluster validity indices are sensitive to fuzzy coefficient m. When $m \to 1$, the indices give the same values for all c. When $m \to \infty$, both PC and PE exhibit significant knee at $c = 2$.

The compactness and separation validity index proposed by Xie and Beni (XB)[26] is a representation of the second category. The smallest value of XB indicates the optimal clusters.

$$XB(c) = \frac{\displaystyle\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^m\|x_k - v_i\|^2}{n \cdot \min_{i,k}\|v_i - v_k\|^2} \tag{8}$$

There are several other cluster validity indices, including fuzzy hyper volume (FHV), average partition density (PA), partition density index (PD), *etc*[27].

3.2 Modified Gath-Geva fuzzy clustering algorithm

Fuzzy clustering is a well-recognized technique to identify fuzzy models. The fuzzy C-means algorithm[23] and the Gustafson-Kessel algorithm[28] are the widely-used methods in fuzzy modeling. However, there are two main drawbacks for these algorithms. First, only clusters with approximately equal volumes can be properly identified, which is frequently difficult to satisfy in real systems. Second, clusters obtained are generally axes-oblique rather than axis-parallel; consequently, a decomposition error is made in their projection onto the input variables. To circumvent these problems, a modified Gath-Geva algorithm[29] is applied in this paper.

The objective function based on the minimization of the sum of weighted squared distances between the data points and cluster centers is described in the following:

$$J(\boldsymbol{Z}; \boldsymbol{U}, \boldsymbol{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m D_{ik}^2 \tag{9}$$

where $\boldsymbol{Z}$ is the set of data, $\boldsymbol{U} = [\mu_{ik}]$ is the fuzzy partition matrix, $\boldsymbol{V} = [\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_c]^{\mathrm{T}}$ is the set of centers of the clusters, c is the number of clusters, N is the number of data, m is the fuzzy coefficient, μ_{ik} is the membership degree between the ith cluster and kth data, which satisfies conditions:

$$\mu_{ik} \in [0, 1];\ \sum_{i=1}^{C}\mu_{ik} = 1 \tag{10}$$

The Lagrange multiplier is used to optimize the objective function (9). The minimum of (U, V) is calculated as follows:

$$\mu_{ik} = \frac{1}{\displaystyle\sum_{j=1}^{c}(D_{ik}/D_{jk})^{2/(m-1)}} \tag{11}$$

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m \boldsymbol{z}_k}{\sum_{k=1}^{N} (\mu_{ik})^m} \tag{12}$$

The variance of the Gaussian function is:

$$\sigma_{ij}^2 = \frac{\sum_{k=1}^{N} \mu_{ik}(x_{jk} - v_{jk})^2}{\sum_{k=1}^{N} \mu_{ik}} \tag{13}$$

The norm of distance between the ith cluster and kth data is

$$\frac{1}{D_{ik}^2} = \prod_{j=1}^{n} \frac{\sum_{k=1}^{N} \mu_{ik}}{N\sqrt{2\pi\sigma_{ij}^2}} \exp(-\frac{1}{2}\frac{(x_{jk} - v_{ij})^2}{\sigma_{ij}^2} \cdot \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(-\frac{(y_k - \hat{y}_k)^{\mathrm{T}}(y_k - \hat{y}_k)}{2\sigma_i^2}) \tag{14}$$

Given the input variable X, output y and fuzzy partition matrix U as follows.

$$X = \begin{bmatrix} X_1^{\mathrm{T}} \\ X_2^{\mathrm{T}} \\ \vdots \\ X_N^{\mathrm{T}} \end{bmatrix}, \ \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \ U_i = \begin{bmatrix} \mu_{i1} & 0 & \cdots & 0 \\ 0 & \mu_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{iN} \end{bmatrix} \tag{15}$$

appending a unitary column to X gives the extended matrix X_e:

$$X_e = [X \quad 1] \tag{16}$$

Then

$$\boldsymbol{\theta}_i = [X_e^{\mathrm{T}} U_i X_e] X_e^{\mathrm{T}} U_i \boldsymbol{y} \tag{17}$$

is the consequent parameter of the fuzzy model.

The procedure of constructing a fuzzy model based on the modified Gath-Geva fuzzy clustering algorithm is summarized as follows:

1) Choose the number of fuzzy rules, the weighting exponent, and the stop criterion $\varepsilon > 0$.
2) Generate the matrix U randomly. U must satisfy condition (10).
3) Compute the parameters of the model using (12), (13) and (17).
4) Calculate the norm of distance utilizing (14).
5) Update the partition matrix U using (11).
6) Stop if $\|\boldsymbol{U}^{(l)} - \boldsymbol{U}^{(l-1)}\| \leqslant \varepsilon$, else go to 3).

3.3 Rule reduction based on the orthogonal least-square method

In the process of cluster validity analysis, it is possible that noise or abnormal data are clustered to create redundant or incorrect fuzzy rules. This problem can be solved by rule reduction using orthogonal transformation methods[30].

Orthogonal transformation methods assign importance to each rule to determine if it should be retained or eliminated. The singular value decomposition (SVD) method and the orthogonal least square (OLS) method are the most-applied orthogonal transformation techniques. In SVD, the estimation of the effective rank of the firing matrix influences the importance, and different estimations lead to different results. The OLS algorithm avoids this subjective error and is adopted in this paper.

The orthogonal least square method transforms the columns of the firing matrix into a set of orthogonal basis vectors. Using the Gram-Schmidt orthogonalization procedure, the firing strength matrix P is decomposed into

$$P = WA \tag{18}$$

where W is a matrix with orthogonal columns w_i, and A is an upper triangular matrix with unity diagonal elements.

Substituting (18) into (4) yields

$$\boldsymbol{y} = WA\boldsymbol{\theta} + \boldsymbol{e} = W\boldsymbol{g} + \boldsymbol{e} \tag{19}$$

where $\boldsymbol{g} = A\boldsymbol{\theta}$. Since the columns $\boldsymbol{w}_i$ of W are orthogonal, the sum of squares of $y(k)$ can be written as

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} = \sum_{i=1}^{M} g_i^2 \boldsymbol{w}_i^{\mathrm{T}} \boldsymbol{w}_i + \boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} \tag{20}$$

Both sides of (20) are divided by N, and it can be seen that the part of the output variance $y^{\mathrm{T}}y/N$ explained by the regressors is $\sum g_i w_i^{\mathrm{T}} w_i/N$, and an error reduction ratio due to an individual rule is defined as

$$[err]^i = \frac{g_i^2 \boldsymbol{w}_i^{\mathrm{T}} \boldsymbol{w}_i}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}, \quad 1 \leqslant i \leqslant c \tag{21}$$

The ratio offers a simple means for seeking a subset of important rules in a forward-regression manner. If it is decided that r rules are used to construct a fuzzy model, then the first r rules with the largest error reduction ratios will be selected.

Without regarding to the premise structures, the OLS method may assign high importance to the redundant fuzzy rules with high firing degrees. This is a drawback that can be solved by a simple modification to the OLS. Each time a new rule is selected, its corresponding column vector is analyzed. If the vector is a linear combination of the firing vectors corresponding to the previously selected rules, then it should not be assigned high importance.

If the importance of the fuzzy rule is far less than other′s, and deletion of this rule dose not deteriorates precision, then this fuzzy rule may be picked out to improve interpretability of the fuzzy model.

3.4 Simplification of the fuzzy model by merging similar fuzzy sets

The simplified fuzzy model obtained above may contain redundant information in terms of similarity between fuzzy sets. The similarity of fuzzy sets makes the fuzzy model uninterpretable, for it is difficult to assign qualitatively meaningful labels to similar fuzzy sets. In order to acquire an effective and interpretable fuzzy model, elimination of redundancy and simplification of the fuzzy model are necessary.

There are three types of redundant or similar fuzzy sets in the fuzzy model: 1) a fuzzy set is similar to the universal set, 2) a fuzzy set is similar to the singleton set, and 3) the fuzzy set A is similar to the fuzzy set B.

If a fuzzy set is similar to the universal set or the singleton set, it should be removed from the corresponding fuzzy rule antecedent. As for two similar fuzzy sets, a similarity measure is utilized to determine if the fuzzy sets should be combined.

For fuzzy sets A and B, a set-theoretic operation based similarity measure[31] is defined as

$$S(A,B) = \frac{\sum_{j=1}^{m} [\mu_A(x_j) \wedge \mu_B(x_j)]}{\sum_{j=1}^{m} [\mu_A(x_j) \vee \mu_B(x_j)]} \tag{22}$$

where $X = \{x_j | j = 1, 2, \cdots, m\}$ is the discrete universe, $\wedge$ and $\vee$ are the minimum and maximum operators respectively. S is a similarity measure in [0,1]. $S = 1$ means the compared fuzzy sets are equal, while $S = 0$ indicates that there is no overlap between the fuzzy sets.

If similarity measure $S(A, B) > \tau$, *i.e.*, the fuzzy sets are very similar, then the two fuzzy sets A and B should be merged to create a new fuzzy set C, where τ is a predefined threshold. It should be pointed out that threshold τ influences the model performance significantly. A small threshold leads to a fuzzy model with low accuracy and highly interpretability. In a general way, $\tau = [0.4 - 0.7]$ is a good choice.

For the Gaussian type of fuzzy sets used in this paper, the parameters of newly merged fuzzy set C from A and B are defined as

$$\begin{cases} v_c = (v_A + v_B)/2 \\ \sigma_c = \sqrt{\sigma_A^2 + \sigma_B^2}/2 \end{cases} \tag{23}$$

The method of merging similar fuzzy sets is carried out iteratively. For each iteration, the similarity measures between all pairs of fuzzy sets for each variable are calculated. The pair of highly similar fuzzy sets with $S > \tau$ is merged to create a new fuzzy set. The rule base of the fuzzy model is updated by substituting the new fuzzy set for the two highly similar fuzzy sets. This process continues until there are no fuzzy sets for which $S > \tau$. Then the fuzzy sets that have similarity to the universal set or the singleton set are removed.

3.5 Optimization

After rule reduction and merge of similar fuzzy sets, the interpretability of the initial fuzzy model is improved, while its precision is reduced. It is essential to optimize the reduced fuzzy model to improve its precision, while preserving its interpretability.

The precision and parameters of the fuzzy model are strongly nonlinear, so a robust optimization technique should be applied in order to assure a good convergence. The constraint Levenberg-Marquardt (LM)[3] method is adopted in this paper. The premise parameters are limited to change in a range of $\pm\alpha\%$ around their initial values in order to preserve the distinguishability of the fuzzy sets. For the sake of maintaining the local interpretability of the fuzzy model, the consequent parameters are restricted to vary $\pm\beta\%$ of the corresponding consequent parameters.

4 Example

PH neutralization[32] is a typical nonlinear system. We build interpretable and accurate fuzzy models for PH neutralization using techniques developed in Section 3.

In order to determine the number of fuzzy rules, cluster validity indices including PC, PE, XB and PA are adopted. Fig. 1 displays the results of the cluster validity indices. All the cluster validity indices indicate that the optimal number of rules is three.

The modified Gath-Geva fuzzy clustering algorithm is used to construct a fuzzy model with three fuzzy rules. The training error and validation error of the obtained fuzzy model are 0.3791 and 0.3160, respectively, which are shown in Table 1. The fuzzy sets of membership functions are illustrated in Fig. 2(b). The fuzzy sets are distinguishable, and it is easy to assign understandable linguistic term to each fuzzy set. We can see that the obtained fuzzy model with three rules is interpretable.

Table 1 Comparison of different fuzzy models

Number of fuzzy rules	Number of fuzzy sets	Training error	Validation error	Validation error descend
2	2	1.6290	1.7547	–
2	4	0.4020	0.3433	411.13%
3	6	0.3791	0.3160	8.64%
4	8	0.3716	0.3119	1.31%
5	10	0.3739	0.3099	0.65%

The OLS method is adopted to pick out unnecessary fuzzy rules. The importance of the fuzzy rules are [0.2249 0.0250 0.7438] respectively, where the importance of the second fuzzy rule is far less than the others. Without considering precision, the second fuzzy rule can be deleted. Fig. 2(a) displays the fuzzy sets of membership with two rules. Table 1 gives the training and validation errors of the fuzzy model with two rules, which are 0.4020 and 0.3433, respectively.

The method of merging similar fuzzy sets is carried out for the reduced fuzzy model with two rules sequentially. The similarity measure between fuzzy sets of flow rate is 0.3420, and the measure between fuzzy sets of PH is 0.0943. Without considering precision and reality, the fuzzy sets of flow rate can be merged to a new fuzzy set. After similar fuzzy sets are merged, the training error and validation errors of the model are 1.6290 and 1.7547, respectively. The fuzzy sets of flow rate cannot be merged because precision deteriorates, so the model with two fuzzy rules and without merge of similar fuzzy sets is adopted in this paper.

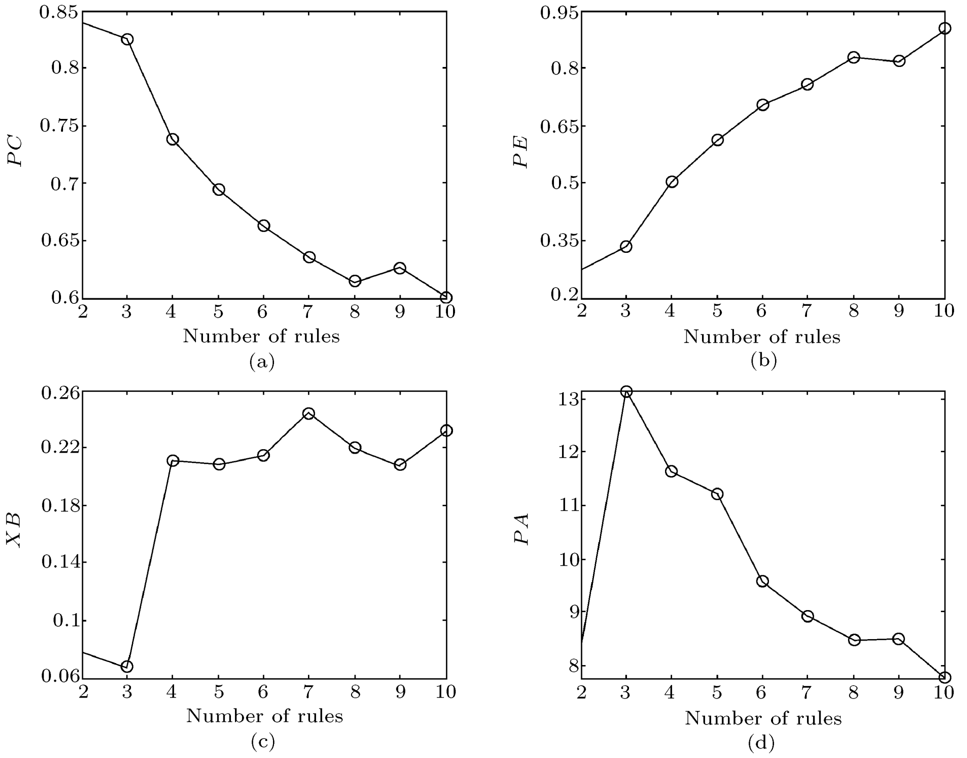

Fig. 1 Cluster validity indices

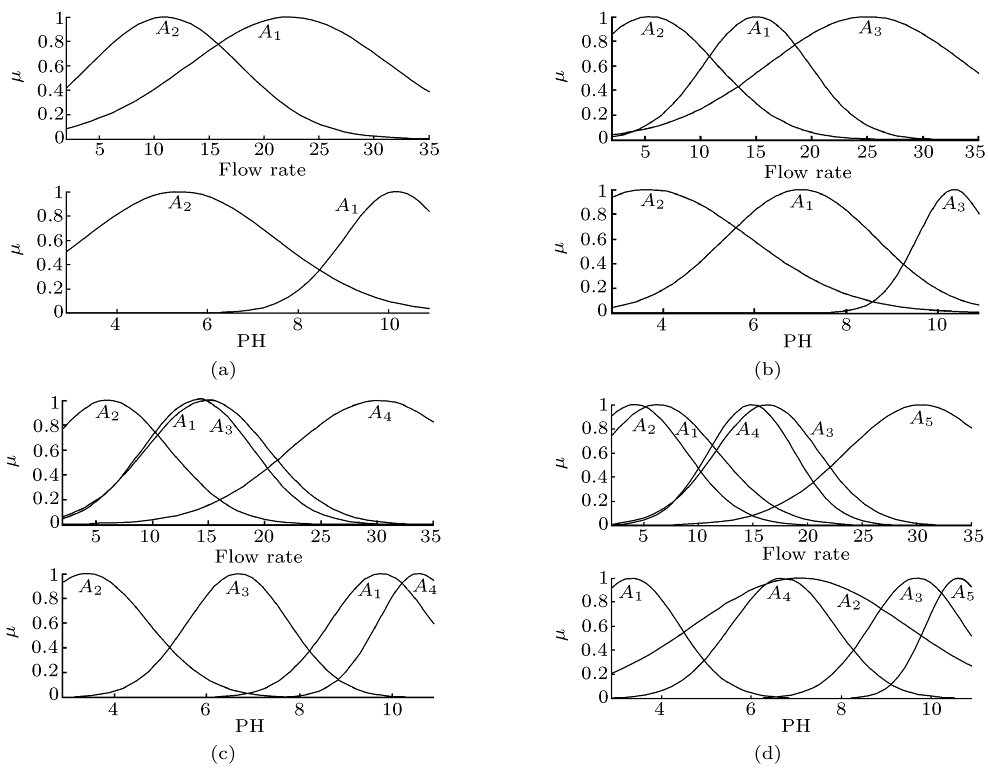

Fig. 2 Fuzzy sets of different fuzzy models

In order to illustrate interpretability and precision of different fuzzy models, Fig. 2 displays fuzzy sets of models with 2, 3, 4 and 5 fuzzy rules, and Table 1 shows the corresponding errors and number of rules/fuzzy sets.

With the increase of rules, the precision is improved, while the interpretability is reduced. When the number of rules is 4 and 5, the fuzzy sets are heavy overlapped, while the precision is only increased a little.

The fuzzy model is optimized by the constrained Levenberg-Marquardt method. The constraint of antecedent parameters is 5%, and the constraint of consequent parameters is 15%. The optimized model is

$$\begin{aligned} R^1 &: \text{If } FR(k) \text{ is High and } PH(k) \text{ is High} \\ &\quad \text{Then } PH(k+1) = 0.0321FR(k) + 0.8614PH(k) + 0.6659 \\ R^2 &: \text{If } FR(k) \text{ is Low and } PH(k) \text{ is Low} \\ &\quad \text{Then } PH(k+1) = 0.1103FR(k) + 0.7521PH(k) + 0.2614 \end{aligned} \tag{24}$$

where $FR(k)$ is the value of flow rate, and $PH(k)$ is the value of PH. Fig. 3(a) diagrams the fuzzy sets of the model.

After optimization, the training error and validation error of the model are 0.3711 and 0.3088 respectively. Without decreasing interpretability, the precision of the model is improved. Fig. 3(b) illustrates the comparison of model outputs and measured outputs.

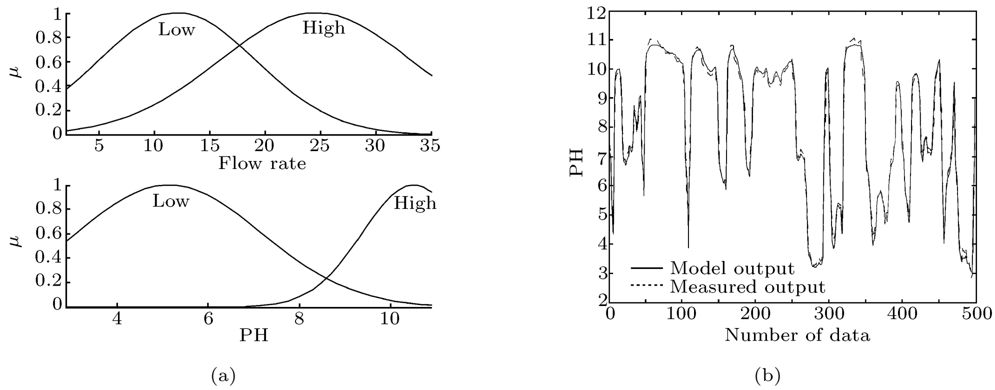

Fig. 3 Fuzzy sets of the model and comparison of model output and measured output

Table 2 compares the fuzzy model, neural network[33] and linear regression model[34]. The linear regression model owns the highest interpretability and the worst precision. The neural network has no interpretability, but has the best precision. The fuzzy model has acceptable interpretability and precision.

Table 2 Comparison of different types of models

Type	Training error	Validation error	Interpretability
Linear regression model	0.4883	0.3701	High
Fuzzy model	0.3711	0.3088	Medium
Neural network	0.3043	0.2796	None

5 Conclusion

This paper studies a case of interpretable fuzzy modeling. Preliminaries, including the TS fuzzy model and the interpretability issues, are stated first. Then a novel fuzzy modeling technique is proposed: a modified fuzzy clustering algorithm is used to build an initial fuzzy model, and an orthogonal least-square method and the method of merging similar fuzzy sets are carried out to reduce the fuzzy model and improve its interpretability. Finally, a constrained Levenberg-Marquardt method is adopted to optimize the fuzzy model and improve its precision. The simulation results on PH neutralization illustrate validity of the method.

References

1 Gomez-Skarmeta A F, Delgado M, Vila M A. About the use of fuzzy clustering techniques for fuzzy model identification. *Fuzzy Sets and Systems*, 1999, **106**(2): 179∼188

2 Lefteri H T, Robert E U. Fuzzy and Neural Approaches in Engineering. New York: Wiley, 1997

3 Jang J S R, Sun C T, Mizutani E. Neuro-Fuzzy and Soft Computing. New Jersey: Prentice Hall, 1996

4 Cordon O, Herrera F, Hoffmann F, Magdalena L. Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Rule Bases. Singapore: World Scientific, 2000

5 Cordon O, Gomide F, Herrera F, Hoffmann F, Magdalena L. Ten years of genetic fuzzy systems: Current framework and new trends. *Fuzzy Sets and Systems*, 2004, **141**(1): 5∼31

6 Babuska R, Bersini H, Linkens D A, Nauck D, Tselentis G, Wolkenhauer O. Future Prospects for Fuzzy Systems and Technology. ERUDIT Newsletter, Aachen, Germany, 6(1), 2000. Available: http://www.erudit.de/erudit/newsletters/news61/page5.htm

7 Roubos H, Setnes M. Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *IEEE Transactions on Fuzzy Systems*, 2001, **9**(4): 516∼524

8 Nauck D D. Fuzzy data analysis with NEFCLASS. *Approximate Reasoning*, 2003, **32**: 103∼130

9 Jin Y. Fuzzy modeling of high-dimensional systems complexity reduction and interpretability improvement. *Fuzzy Sets and Systems*, 2000, **8**(2): 212∼221

10 Casillas J, Cordón O, Herrera F, Magdalena L. Interpretability improvements to find the balance interpretability-accuracy in fuzzy modeling: an overview. Chapter of Interpretability Issues in Fuzzy Modeling. Springer, 2003, 3∼22

11 Jin Y. Advanced Fuzzy Systems Design and Applications. New York: Physical-Verl, 2003

12 Angelov P P. An evolutionary approach to fuzzy rule-based model synthesis using indices for rules. *Fuzzy Sets and Systems*, 2003, **137**(3): 325∼338

13 Kaynak O, Jezernik K, Szeghegyi A. Complexity reduction of rule based models: A survey. In: Proceedings of IEEE International Conference on Fuzzy Systems, Honolulu, Hawai: IEEE Press, 2002. 1216∼1222

14 Yen J, Wang L, Gillespie C W. Improving the interpretability of TSK fuzzy models by combining global learning and local learning. *IEEE Transactions on Fuzzy Systems*, 1998, **6**(4): 530∼537

15 Guillaume S. Designing fuzzy inference systems from data: an interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 2001, **9**(3): 426∼443

16 Sudkamp T, Knapp A, Knapp J. Model generation by domain refinement and rule reduction. *IEEE Transactions on Systems, Man, and Cybernetics*, 2003, **33**(1): 45∼55

17 Setnes M, Hellendoorn H. Orthogonal Transforms for Ordering and Reduction of Fuzzy Rules. In: Proceedings of IEEE International Conference on Fuzzy Systems, San Antonio, USA: IEEE Press, 2000. 700∼705

18 Yen J, Wang L. Simplifying fuzzy modeling by both gray relational analysis and data transformation methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 1999, **29**(1): 13∼24

19 Xing Z-Y, Hu W-L, Jia L-M. A study on the tradeoff between precision and interpretability in fuzzy modeling. *Journal of Southeast University*, 2004, **20**(4): 472∼476

20 Min Y C, Linkens D A. Rule-base self-generation and simplification for data-driven fuzzy models. *Fuzzy Sets and Systems*, 2004, **142**(2): 243∼265

21 Delgado M R, Zuben F V, Gomide F. Multi-Objective Decision Making: Towards Improvement of Accuracy, Interpretability and Design Autonomy in Hierarchical Genetic Fuzzy Systems. In: Proceedings of IEEE International Conference on Fuzzy Systems, Honolulu, Hawai: IEEE Press, 2002. 1222∼1227

22 Takagi T, Sugeno M. Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 1985, **15**(1): 116∼132

23 Bezdek J C. Pattern Recognition with Fuzzy Objective Algorithm. New York: Plenum Press, 1981

24 Bezdeck J C. Cluster validity with fuzzy sets. *Cybernetics*, 1974, **3**(3): 58∼73

25 Bezdeck J C, Ehrlich R, Full W. FCM: Fuzzy C-means algorithm. *Computers and Geoscience*, 1984, **10**(2-3): 191∼203

26 Xie X L, Beni G. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, **13**(8): 841∼847

27 Gath I, Geva A B. Fuzzy clustering for the estimation of the parameters of the components of mixtures of normal distributions. *Pattern Recognition Letters*, 1989, **9**: 77∼86

28 Gustafson D, Kessel W. Fuzzy clustering with a fuzzy covariance matrix. In: Proceedings of IEEE Conference on Decision and Control. San Diego, USA: IEEE Press, 1979. 761∼766

29 Abonyi J, Babuska B, Szeifert F. Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Systems, Man, and Cybernetics*, 2002, **32**(5): 612∼621

30 Wang L X, Mendel J M. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, 1992, **3**(5): 807∼814

31 Setnes M, Babuska R, Kaymak U, Lemke H R N. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1998, **28**(3): 376∼386

32 Babuska R. Fuzzy Modeling for Control. Boston: Kluwer Academic Publishers, 1998

33 Neural network toolbox. Matlab Release13, The Mathworks Inc., 2003

34 System Identification Toolbox. Matlab Release13, The Mathworks Inc., 2003

XING Zong-Yi Postdoctor at Nanjing University of Science and Technology. His research interests include fuzzy modeling and intelligent control of industry process.

JIA Li-Min Professor at Beijing Jiaotong University. His research interests include fuzzy set and system, intelligent control and application, and railway intelligent transportation system.

ZHANG Yong Ph. D. candidate at Nanjing University of Science and Technology. His research interests include intelligent control and system, nonlinear control theory and application.

HU Wei-Li Professor at Nanjing University of Science and Technology. His research interests include intelligent control and intelligent system, nonlinear control theory and application, and digital AC servo system.

QIN Yong Associate professor at Beijing Jiaotong University. His research interests include fuzzy control and application, and intelligent transportation system.

(Continued on Page 959)