CONSTRUCTING INTERPRETABLE FUZZY MODEL BASED ON REDUCTION METHODOLOGY

Xing Zongyi[a], Hu Weili[a], Jia Limin[b]

[a]*Automation Department, Nanjing University of Science and Technology, Jiangsu Nanjing 210094, China*
[b]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing, 100044, China*

Abstract: A systematic approach for constructing interpretable fuzzy model based on reduction methodology is proposed. Fuzzy clustering algorithm, combined with least square method, is used to identify initial fuzzy model with overestimated rule number. Orthogonal least square algorithm and similar fuzzy sets merging are then applied to remove redundancy of the fuzzy model. In order to obtain high accuracy, yet preserving interpretability, a constrained real coded genetic algorithm is utilized to optimize reduced fuzzy model. The proposed method was applied to automobile MPG prediction, and results show its validity.

Keywords: fuzzy modelling; fuzzy sets; input estimation; fuzzy model; genetic algorithms

1. INTRODUCTION

In recent years, data-driven fuzzy modelling techniques have become an active research area, and applied in many fields for classification, data mining, pattern recognition, simulation, analysis, prediction, control. Different from mathematical models and neural networks, fuzzy models provide with some distinctive features including knowledge expression using if-then rules and mechanism of human-like reasoning in linguistic manner, which are the most distinguishable features of fuzzy models.

Several approaches have been proposed to identify fuzzy models from measured data: fuzzy clustering (Gomez-Skarmeta, *et al*., 1999), neural-fuzzy systems (Lefteri, *et al*., 1997), genetic rule generation (Cordon, *et al*., 2001). All these methods utilize only the function approximation of fuzzy models, and more emphasis is put on the numerical performance. However, interpretability, which is the most prominent feature distinguished fuzzy systems from many other models, is often neglected. As a result, the fuzzy model generated is redundant, less generalized and uninterpretable.

In order to improve the interpretability of fuzzy models, some methods have been developed. Setnes, *et al.* proposed a set-theoretic similarity measure to quantify the similarity among fuzzy sets, and to reduce the number of fuzzy sets in the model. Yen, *et al*. introduce several orthogonal transformation techniques for selecting the most important fuzzy rules from a given rule base in order to construct a compact fuzzy model. Abonyi, *et al*. propose a combined method to create simple Takagi-Sugeno fuzzy model that can be effectively used to represent complex system. Roubos, *et al*. present an approach to identify compact and accurate fuzzy model based on iterative complexity reduction combined with fuzzy clustering and multi-objective genetic algorithm optimization.

This paper develops a systematic technique to construct interpretable, yet accurate fuzzy model based on reduction methodology. In section 2, the initial fuzzy model is identified based on fuzzy clustering, including fast input variable selection. Rule reduction and similar fuzzy sets merging are used to improve its interpretability in section 3. Then constrained genetic algorithm is utilized to reach higher precision performance in section 4. In section 5, the method is demonstrated on the automobile MPG prediction problem. Section 6 concludes the paper.

2. IDENTIFICATION OF FUZZY MODEL

2.1 Takagi-Sugeno Fuzzy model

The Takagi-Sugeno (TS) Fuzzy model is usually used as data-driven fuzzy model. A typical fuzzy rule of the model has the form:

$$\begin{aligned} R^i &: \text{IF } x_1 \text{ is } A_{i,1} \text{ and } \cdots \text{ and } x_p \text{ is } A_{i,p} \\ &\text{THEN } \hat{y}_i = f_i(x) \end{aligned} \quad (1)$$

where R^i is the ith rule, x_j are the input variables, A_{ij} are fuzzy sets defined on the universe of discourse of the input. $\hat{y}_i = f_i(x)$ is usually a linear polynomial function in the input variables.

In TS fuzzy model, each fuzzy rule describes a local linear model. All these local models combine to describe the global behaviours of a non-linear complex system. The output of the TS fuzzy model is computed using the normalized fuzzy mean formula:

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

where c it the number of rules, P_i is the normalized firing strength of the ith rule:

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

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

$$y = P\theta + e \quad (4)$$

where θ is consequents matrix of rules, and e is approximation error matrix.

In this paper, Gaussian membership functions are used to represent the fuzzy set A_{ij}

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

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

2.2 Input variable selection of TS fuzzy model

Input variable selection is the principal element in fuzzy modelling, which is a process to choose a small subset of input variables from a large set of input variable candidates, ideally necessary and sufficient to satisfy the needs of modelling.

Exhaust searching method of input variable subsets is impossible due to dimensional disaster in practice when there are many input variables, though it is feasible in theory. So a new input variable selection method is proposed in this paper.
Firstly, a zero-order TS fuzzy model is constructed using all input variable candidates. Then for each single input variable, the corresponding model output is calculated neglected other variables contribution. The model output change indicates the importance of the input variable. The larger the output change is, the more important the input variable is.

Given N input-output data, for a specific input variable x_i, there is output vector $Y^i = \left[y_1^i, y_2^i, \cdots, y_N^i\right]$. The output change indicating the importance of the ith input variable is defined:

$$\Delta O_i = \max(Y^i) - \min(Y^i) \quad (6)$$

The input variable selection is carried out according to the following step:
1) Define the normalized importance measure for the ith input variable by

$$M_i = \frac{\Delta O_i - \min\limits_{j=1,\cdots,q}(\Delta O_j)}{\max\limits_{j=1,\cdots,q}(\Delta O_j) - \min\limits_{j=1,\cdots,q}(\Delta O_j)} \quad (7)$$

Obviously, the larger value of M_i is, the more important of the ith input variable is. The variable with M_i =1 is the most important. A small value of M_i corresponds to a relatively unimportant input variable.

2) Given a threshold $\lambda \in (0,1)$. When M_i is less than the threshold, i.e. $M_i < \lambda$, the corresponding ith input variable is believed to be unimportant and should be removed.

3) Given a threshold $\tau \in (0,1)$. Calculate the correlation function between the selected input variables $\{x_i, x_j\}$ to recognize the closely related input variables by

$$\sigma(x_i, x_j) = \frac{\sum_{k=1}^{N}(x_i - \overline{x_i})(x_j - \overline{x_j})}{N\sqrt{\phi_{x_i}\phi_{x_j}}} \quad (8)$$

where $\overline{x_i}, \overline{x_j}, \phi_i, \phi_j$ are the means and variances of input variable x_i and x_j respectively. If $\sigma(x_i, x_j) > \tau$, then x_i is closely related to x_j, thus remove the one with smaller value of M_i.

4) As a result, the task of input variable selection is completed, and a subset of p significant input variables is selected for the fuzzy model.

2.3 Identification of TS Fuzzy Model

It is generally acknowledged that fuzzy clustering is a well-recognized paradigm to generate initial fuzzy model. The G-K algorithm (Gustafson and Kessel, 1979) is employed in this paper.

The objective function of G-K algorithm is described following:

$$J(Z;U,V) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m \| z_k - v_i \|_A^2 \quad (9)$$

where Z is the set of data, $U = [\mu_{ik}]$ is the fuzzy partition matrix, $V = [V_1, V_2, \cdots, V_c]^T$ is the set of centres of the clusters, c is the number of clusters, N is the number of data, m is the fuzzy coefficient fuzziness, μ_{ik} is the membership degree between the ith cluster and kth data, which satisfy conditions:

$$\mu_{ik} \in [0,1], \sum_{i=1}^{C} \mu_{ik} = 1 \quad (10)$$

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

$$D_{ik}^2 = \| z_k - v_i \|_{A_i}^2 = (z_k - v_i) A_i (z_k - v_i) \quad (11)$$

where

$$A_i = (\rho \det(F_i))^{1/n} F_i^{-1} \quad (12)$$

$$\det(A_i) = \rho \quad (13)$$

n is the dimension of data Z, F_i is the fuzzy covariance matrix of ith cluster.

$$F_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m (z_k - v_i)(z_k - v_i)^T}{\sum_{k=1}^{N} (\mu_{ik})^m} \quad (14)$$

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

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (D_{ik} / D_{jk})^{2/(m-1)}} \quad (15)$$

$$v_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m z_k}{\sum_{k=1}^{N} (\mu_{ik})^m} \quad (16)$$

The TS modelling methodology is essentially a multi-model approach in which simple local sub-models are coupled to describe the global behaviours of the system. Since one of the important motivations of using the TS model is to gain insights into the model, it is important to investigate the local interpretation issue of the TS model. Here weighted least square estimation is employed construct more interpretable local models.

Given the input variable X, output y and fuzzy partition matrix U as following:

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_N^T \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, U_i = \begin{bmatrix} u_{i1} & 0 & \cdots & 0 \\ 0 & u_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{iN} \end{bmatrix} \quad (17)$$

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

$$X_e = \begin{bmatrix} X & 1 \end{bmatrix} \quad (18)$$

Then

$$\theta_i = [X_e^T U_i X_e]^{-1} X_e^T U_i y \quad (19)$$

is the solution of (4), i.e. the consequent parameters of fuzzy model.

The procedure of constructing initial fuzzy model is summarized as follows:
1) Choose the number of fuzzy rules and the weighting exponent, and the stop criterion $\varepsilon > 0$.
2) Generate the matrix U with the membership randomly; U must satisfy the condition (10).
3) Compute the centres of the clusters using (16) and fuzzy covariance matrix by (14).
4) Calculate norm of distance utilizing (11).
5) Update the partition matrix U using (15);
6) Stop if $\left\| U^{(l)} - U^{(l-1)} \right\| \le \varepsilon$, else go to 3).
7) Compute the consequence parameters of the fuzzy model using (19)

3. INTERPRETABLITY IMPROVEMENT

3.1 *Some issues about interpretability*

Apart from precision performance, interpretability is another important feature of fuzzy model. Though there is no formal definition about interpretability, some important elements can be concluded mainly as follows.

1) The fuzzy model should use least possible variables, and each rule should use as few variables as possible. 2) The number of rules should not exceed ten experientially. The rules must be consistent and not conflictive. The rule base should be complete, compact and has no redundancy. 3) The fuzzy membership functions should be both complete and distinguishable, so that a linguistic term could be associated with a fuzzy set.

3.2 *Rule reduction by orthogonal least square*

The initial fuzzy model is complete and consistent which are essential for interpretability. However, it contains redundant fuzzy rules hampering its interpretability. This problem is solved by rule reduction using orthogonal transformation method (Yen, *et al*., 1999).

Orthogonal transformation methods define importance measure for each rule to determine if it should be retained or eliminated. Singular value decomposition method and orthogonal least square algorithm are the most applied orthogonal transformation techniques. In Singular value decomposition method, the estimation of the effective rank of firing matrix influences important measures, and different estimation leads to different result. Orthogonal least square algorithm avoids this objective 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. With Gram–Schmidt orthogonalization procedure, firing strength matrix P is decomposed into

$$P = WA \quad (20)$$

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

Substituting (20) into (4) yields

$$y = WA\theta + e = Wg + e \tag{21}$$

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

$$y^T y = \sum_{i=1}^{M} g_i w_i^T w_i + e^T e \tag{22}$$

Dividing N on both side of (22), it can be seen that the part of the output variance $y^T y/N$ explained by the regressors is $\sum g_i w_i^T w_i / N$, and an error reduction ratio due to an individual rule is defined as

$$[err]^i = \frac{g_i^2 w_i^T w_i}{y^T y},\ 1 \le i \le M. \tag{23}$$

This 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 paying any attention to the premise structures, this method is possible to give a high importance to redundant fuzzy rules with high firing degrees due to their contributions to the output. This drawback can be solved by a simple modification to the method. 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 a high importance.

After rule reduction, the simplified fuzzy model is more interpretable for removing excessive fuzzy rules. Constrained genetic algorithm is then adopted to improve its precision.

3.3 *Similar fuzzy sets merging*

The simplified fuzzy model obtained above may contain redundant information in the form 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 making the fuzzy model as simple as possible are necessary.

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

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

For fuzzy sets A and B, a set-theoretic operation based similarity measure M (Setnes and Babuska, 1998) is defined as

$$S(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{24}$$

where $|\cdot|$ denotes the cardinality of a set, the $\cap$ and $\cup$ operators represent the intersection and union respectively. For $X = \{x_j \mid j = 1, 2, \cdots, m\}$, this can be rewritten as

$$S(A,B) = \frac{\sum_{j=1}^{m}[\mu_A(x_j) \wedge \mu_B(x_j)]}{\sum_{j=1}^{m}[\mu_A(x_j) \vee \mu_B(x_j)]} \tag{25}$$

where $\wedge$ 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 overlapping between fuzzy sets.

If similarity measure $S(A,B) > \lambda$, i.e. fuzzy sets are very similar, then the two fuzzy sets A and B should be merged to create a new fuzzy set C, where λ is a predefined threshold. It should be pointed out that threshold λ influences the model performance significantly. Small threshold generates low accurate and high interpretable fuzzy model. In a general way, $\lambda = [0.4 - 0.7]$ is a good choice.

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

$$\begin{cases} v_c = (v_A + v_B)/2 \\ \sigma_c = \sqrt{\sigma_A^2 + \sigma_B^2}\,/2 \end{cases} \tag{26}$$

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

4. GENETIC ALGORITHM OPTIMIZATION

After rule reduction and similar fuzzy sets merging, the obtained fuzzy model is rude and inaccurate. In order to improve the precision of the fuzzy model, while preserve its interpretability, a constrained genetic algorithm (GA) based method is applied to optimize parameters of antecedent and consequent simultaneously.

This paper adopts a real coded genetic algorithm due to its high searching efficiency and intuitionistic representation of chromosome. The main aspects of a

GA include: chromosome representation of fuzzy model, genetic operators and constraints handling.

With a predefined population size, the parameters of fuzzy model in a chromosome that describes antecedent and consequent parameters are encoded sequentially. The initial population is generated randomly within the chromosome constraints. Roulette wheel selection, simple arithmetical crossover and uniform mutation are chosen as genetic operators.

The optimization is subject to search space constraints in the interest of reserving interpretability. The premise parameters are limited to change in a range of $\pm\alpha\%$ around their initial value in order to preserve the distinguishability of the fuzzy sets. For the sake of maintaining the local interpretability of fuzzy model, the consequent parameters are restricted to vary $\pm\beta\%$ of the corresponding consequent parameters.

5. EXAMPLES

In order to illustrate the performance of the proposed technique of constructing interpretable fuzzy model, the well-known automobile MPG (Miles Per Gallon) prediction benchmark is demonstrated in this section. Automobile MPG prediction is a typical nonlinear regression problem. The goal is to predict the fuel consumption of an automobile based on five features including displacement, horsepower, weight, acceleration and model year. The data consisting of 392 measurements is divided into two parts equally: the training set and the checking set.

A zero-order TS fuzzy model is constructed with all five features to select input variables. The normalized importance measures of five input variables are [0.1204, 0.0177, 1.0000, 0, 0.1507]. So the importance sequence of input variables is: weight, year, displacement, horsepower, and acceleration. Obviously, weight and year can be selected as the significant input variable candidates. Weight has little relation with year because the correlation coefficient between them is 0.2992. Clearly, the features of weight and year are used as input variables for the following fuzzy model.

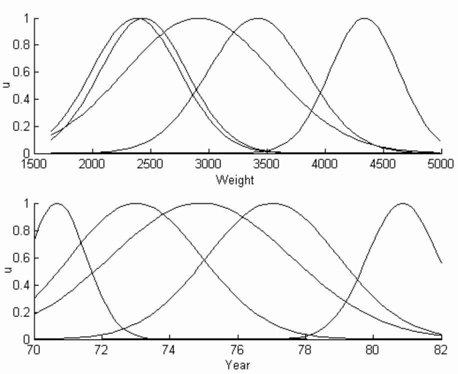

Fig.1. Membership functions of the model with 5 rules

The initial fuzz model is identified with five rules by clustering algorithm and weighted least square method. Fig.1 shows heavily overlapped membership functions of rules, and it indicates that the initial fuzzy model contains excessive rules.

The orthogonal least square method is used to select a set of important fuzzy rules. As result, error reduction ratio of rules are [0.1025, 0.3702, 0.4153, 0.0466, 0.0415] respectively. Obviously, the second and the third rule are the most important rules and are selected to compose a more compact and interpretable fuzzy model. This fuzzy model is then optimized using constrained GA algorithm. The root mean-square error (RMSE) of the model is 2.86 and 2.97 for training data and checking data respectively.

Fig.2 presents membership functions of the fuzzy model with two obtained fuzzy rules. Evidently, membership functions of “year” are much overlapped. The similarity measure between them is 0.8297, so the two fuzzy sets of “year” are merged into a new fuzzy set that is similar to universal set, and is removed from the antecedent of rules.

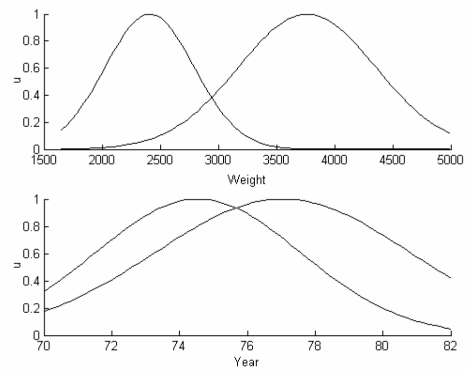

Fig.2. Membership functions of the model with 2 rules

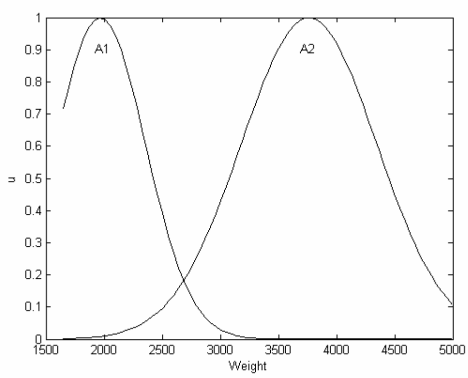

Fig.3. Membership functions of the final model

After deleting input variable of “year” from the antecedent, the RMSEs of the fuzzy model deteriorate to 5.64 and 5.36 for training data and checking data. Constrained genetic algorithm is then unutilized to improve accuracy of fuzzy model, while preserving its interpretability. The final fuzzy model, with improved precision of 3.28 and 3.17, is described as follows:

R^1 : If *weight* is A_1,

Then MPG=-0.0053$weight + 2.0071 year - 46.3304$

R^2 : If *weight* is A_2,

Then MPG=-0.0071$weight + 0.6220 year - 2.8766$

where A_1 and A_2 are showed in Fig.3.

Table 1 illustrates the comparison results of neural networks and linear regression model with obtained fuzzy model. It shows that the linear regression model is the most interpretable and worst precise model, while neural network is absolutely uninterpretable and highest accurate. The fuzzy model constructed is interpretable, and owns an acceptable precision.

Table 1 Comparison of different methods

method	training error	checking error
Linear model	3.54	3.44
neural networks	2.69	2.87
our method	3.28	3.17

6. CONCLUSIONS

In this paper, a systematic technique to construct interpretable, yet accurate fuzzy model based on reduction methodology is presented.

A new input variables selection algorithm is proposed to select the most influenced variables. Fuzzy clustering, combined with least square method, is used to identify initial fuzzy model. Orthogonal least square method and similar fuzzy sets merging are used to remove redundancy of fuzzy model and improve its interpretability. In order to obtain high precision model, yet containing interpretability, a constrained real coded genetic algorithm is utilized to optimize fuzzy model. The approach is applied to Automobile MPG prediction benchmark, and the result shows its validity.

ACKNOWLEDGEMENT

This paper is supported by National Natural Science Foundation of China and Jiangsu Planned Projects for Postdoctoral Research Funds.

REFERENCES

Abonyi, J., J.A. Roubos, M. Oosterom and F. Szeifert (2001). Compact TS-fuzzy models through clustering and OLS plus FIS model reduction. *Proc. of IEEE int. conf. on fuzzy systems*. Sydney, 1420-1423.

Cordon, O., F. Herrera, F. Hoffmann and L. Magdalena (2001). *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Rule Bases*. World Scientific, New York.

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

Gustafson, D. and W. Kessel (1979). Fuzzy clustering with a fuzzy covariance matrix. *Proc. of IEEE conf. on decision and control*. San Diego, 761-766.

Lefteri, H.T. and E.U. Robert (1997). *Fuzzy and Neural Approaches in Engineering*. Wiley, New York.

Roubos, H. and M. SETNES (2001). Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *IEEE Trans. on Fuzzy Systems*, **9**, 516-524

Setnes, M., R. Babuska, U. Kaymak and H.R.N. Lemke (1998). Similarity Measures in Fuzzy Rule Base Simplification. *IEEE Trans. on Systems Man and Cybernetics*, **28**, 376-386.

Yen, J. and L. Wang (1999). Simplifying fuzzy modelling by both gray relational analysis and data transformation methods. *IEEE Trans. on Systems Man and Cybernetics*, **29**, 13-24.