ICIC Express Letters ICIC International ©2019 ISSN 1881-803X
Volume **13**, Number **1**, January **2019** pp. **51–58**

T-S FUZZY SYSTEM IDENTIFICATION BASED ON BAYESIAN INFERENCE

Limin Zhang[1], Hongjiu Yang[2] and Qin Mao[3]

[1]Department of Mathematics and Computer Science
Hengshui University
No. 1088, Heping West Road, Hengshui 053000, P. R. China
limin_zhang@yeah.net

[2]School of Electrical and Information Engineering
Tianjin University
No. 92, Weijin Road, Nankai District, Tianjin 300072, P. R. China
yanghongjiu@ysu.edu.cn

[3]China Petroleum Engineering & Construction Corporation
No. 28, Gulouwai Ave., Dongcheng Dist., Beijing 100120, P. R. China

Received June 2018; accepted September 2018

Abstract. *This paper introduces a new method for fuzzy modeling based on sparse Bayesian techniques. The proposed method is called sparse Bayesian fuzzy inference systems (B-sparseFIS). There are two main procedures in the paper. First, initial fuzzy rule antecedent part is extracted automatically by an AP clustering method; second, the system consequent parameters are identified and simplified with sparse Bayesian techniques such that more consequent parameters will approximate to zero. An example is provided to test the effectiveness of the proposed algorithm. Furthermore, the performances of the algorithm are validated through the results of statistical analyses including parameter estimate error and RMSE.*
Keywords: T-S fuzzy system, Sparse Bayesian inference, AP clustering

1. **Introduction.** Fuzzy model is one of the most important modeling approaches which belongs to fuzzy logic theory [1]. It is useful in nonlinear dynamic system modeling, prediction, and model-based control [2]. Among different fuzzy models, the Takagi-Sugeno (T-S) fuzzy model [3] can be easily applied in various kinds of nonlinear system modeling, which decomposes a nonlinear system into a collection of local linear models. If the T-S fuzzy system is selected as the equivalence model in our study, there are two problems to confront in the T-S fuzzy system identification: structure identification and parameter estimation [4]. Structure identification includes the selection of the number of rules. Parameter estimation respectively includes antecedent membership functions (AMFs) and the corresponding consequent parameters. We try to select the appropriate fuzzy rules and make the desired performance to be a meaningful problem in the T-S fuzzy system.

Fuzzy clustering methods are most common and widely used approaches in T-S fuzzy modeling. Fuzzy clustering provides a certain advantage over other techniques since the fuzzy partition of the input (or the product) space is obtained as a direct result [5]. In order to search structure rules, various clustering algorithms are widely used in [6, 7, 8, 9, 10, 11], such as k-means algorithm [6], fuzzy c-means algorithm [7, 8, 9], hierarchical fuzzy-clustering [10], and IT2-FCRM clustering approach [11].

In our paper, we focus on T-S fuzzy system consequent identification parameters using sparse Bayesian approach. Too many consequent parameters will increase its computation complexity. In order to reduce the consequent parameter number, sparse representation

DOI: 10.24507/icicel.13.01.51

methods were widely used in much literature. Here, block-structured sparse representation [12], as a successor of traditional sparse representation [13], was first introduced and investigated in the literature of the so-called group least-absolute-shrinkage-and-selection operator (LASSO). It provides a regression model, where many blocks of the regression coefficient with small contribution would shrink exactly to zero while keeping high prediction accuracy [14]. Motivated by sparse representation, we exploit a sparse Bayesian fuzzy inference systems, which is called B-sparseFIS. Based on the number of sparsed consequent parameters in each rule, number of the rules is fine turned.

The aim of this paper is to develop a systematic fuzzy modeling mechanism. The rest of this paper is organized as follows. Section 2 describes the Takagi-Sugeno fuzzy system. In Section 3, the selected fuzzy rules consequent parameters are identified and simplified with sparse Bayesian techniques such that more consequent parameters will approximate to zero. Section 4 provides an illustrative example for the performances of the proposed algorithm. Finally, we offer some concluding remarks in Section 5.

2. **Takagi-Sugeno Fuzzy System.**

The T-S fuzzy system can be described by IF-THEN fuzzy rules. Each rule consists of fuzzy rule antecedent and consequent.

$$\text{Rule } i: \text{ If } x_{i1}^t \text{ is } A_{i1}, x_{i2}^t \text{ is } A_{i2}, \ldots, x_{in}^t \text{ is } A_{in}, \text{ then}$$
$$y_i^t = \theta_{i0} + \theta_{i1}x_{i1}^t + \cdots + \theta_{in}x_{in}^t$$

where $i = 1, 2, \ldots, r$, $t = 1, 2, \ldots, N$. r is the number of fuzzy rules. $\mathbf{x}^t = \left[x_{i1}^t, x_{i2}^t, \ldots, x_{in}^t\right]^T$ is the system input and n is the input dimension, t is the number of inputs. y_i^t is output of the ith rule. $\theta_i = [\theta_{i0}, \theta_{i1}, \ldots, \theta_{in}]^T$, θ_{in} is the consequent parameter of the ith rule.

The final output of the T-S fuzzy model can be expressed by a weighted mean defuzzification as follows:

$$\hat{y}_i^t = \sum_{i=1}^{r} \phi_i(\mathbf{x}^t) y_i^t(\mathbf{x}^t) \tag{1}$$

$$\phi_i(\mathbf{x}^t) = \frac{\mu_i(\mathbf{x}^t)}{\sum_{k=1}^{r} \mu_k(\mathbf{x}^t)}, \quad \mu_i(\mathbf{x}^t) = \prod_{j=1}^{n} \mu_{A_{ij}}(x_{ij}^t) \tag{2}$$

$$\mu_{A_{ij}}(x_{ij}^t) = \exp\left[\left(\frac{x_{ij}^t - c_{ij}}{\sigma_{ij}}\right)^2\right] \tag{3}$$

where c_{ij} and σ_{ij} denote the mean and the variance of the corresponding bell-shaped membership function, respectively.

As the definitions in [15], $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_r]$,

$$\Phi_i = diag\left(\phi_i(\mathbf{x}^1), \phi_i(\mathbf{x}^2), \ldots, \phi_i(\mathbf{x}^N)\right)\left[\mathbf{1}, \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N\right]^T, \ \mathbf{1} = [1, 1, \ldots, 1] \in R^N$$

Let $\hat{\mathbf{y}}^t = \hat{y}_1^t + \hat{y}_2^t, \ldots, \hat{y}_r^t$, $\hat{\mathbf{y}} = \left[\hat{\mathbf{y}}^1, \hat{\mathbf{y}}^2, \ldots, \hat{\mathbf{y}}^N\right]^T$, $\theta = \left[\theta_1^T, \theta_2^T, \ldots, \theta_r^T\right]^T \in R^{r(N+1)}$.

Φ is called to be dictionary of the T-S fuzzy system and Φ_i is the subdictionary of the ith fuzzy rule. The fuzzy model output $\hat{\mathbf{y}}$ can be expressed as follows

$$\hat{\mathbf{y}} = \Phi\theta = \sum_{i=1}^{r} \Phi_i\theta_i \tag{4}$$

3. **Identification of Takagi-Sugeno Fuzzy System Using Sparse Bayesian Technology.** In this section, we use the affinity propagation (AP) clustering algorithm to produce an input space fuzzy partition [16]. The consequent parameters are identified by using sparse Bayesian technology. The number of fuzzy rules is fine turned again based on consequent parameters. In the subsections, the above introduction will be analyzed in detail.

3.1. **Fuzzy rule antecedent rough selection.** Fuzzy rules can be obtained with some prior knowledge; nevertheless, unsupervised clustering methods are extensively exploited to roughly partition the input space and determine the fuzzy rule antecedent. We use the clustering algorithm based on AP, which is an unsupervised learning algorithm for exemplar-based clustering. AP clustering algorithm outputs a set of data points that best represent the data (exemplars), and assignments of each non-exemplar point to its most appropriate exemplar based on input similarities between data points. Then the input data is partitioned into clusters.

By the iterative AP clustering algorithm, each cluster is associated with one fuzzy rule. Each dimension of input corresponds to a bell-shaped membership function A_{ij} of fuzzy rule antecedent. The cluster center and variance in each dimension are written as c_{ij} and σ_{ij} for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, n$. In order to control the bandwidth of bell-shaped membership function, σ_{ij} should be in a reasonable range. We set $\sigma_{ij} = \max\{\sigma_{ij}, \varepsilon\}$, and ε is a small positive integer. After the rules determined, the mathematic model can be represented by the T-S fuzzy system, the corresponding consequent parameters of which are identified with regression model.

3.2. **Sparsity representation for the consequent parameter.** After the premise fuzzy set parameters have been obtained, we start to compute the consequent parameters θ. The flowchart of the sparse Bayesian T-S fuzzy system identification is in Figure 1. Here, we want to go a step further and make the consequent parameters also sparse, again without losing model accuracy. This is achieved by using sparse Bayesian approach and evokes the useful weights selection, as some dimension of inputs with low weights in the rules may be ignored [17]. This reduces computational complexity of the consequent parameters, which are less significant in these regions. Based on Algorithm 1, we exploit an effective method for the T-S fuzzy system identification, which is called B-sparseFIS.

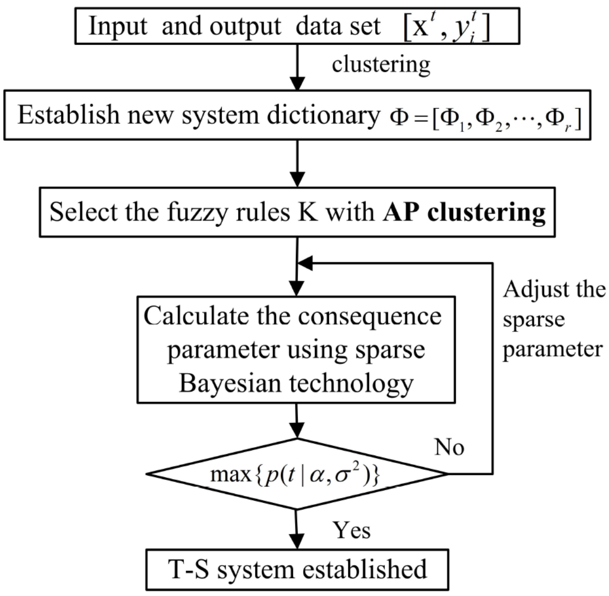


Figure 1. Flowchart of the B-sparseFIS algorithm

Algorithm 1 Sparse Bayesian T-S fuzzy system identification algorithm

Input:

system dictionary $\Phi = [\Phi_1, \Phi_2, \dots, \Phi_p]$ from input data.

1: **for** $p = 1, 2, \dots$ **do**

2: Fuzzy rules selection: select K important fuzzy rule together with the AP clustering algorithm.

3: Consequent parameters sparsity representation: Estimate consequent parameters with sparse Bayesian technology.

4: **end for**

Output:

The optimal sparse consequent parameters are determined with the minimal rules.

There are many documents that solution for θ is given using least-squares, maximum likelihood estimation methods. These methods easily result in overfitting and large computational complexity. So instead of the earlier regularization weight penalty, we now define a prior distribution of θ as follows

$$p\left(\theta|\alpha_1, \dots, \alpha_M\right) = \prod_{m=1}^{M} \left[(2\pi)^{-\frac{1}{2}} \alpha_m^{\frac{1}{2}} \exp\left(-\frac{1}{2}\alpha_m \theta_m^2\right)\right] \tag{5}$$

We now have M hyperparameters $\alpha = (\alpha_1, \dots, \alpha_M)$, one α_m independently controlling the variance of each weight θ_m.

We assume independent Gaussian noise: $t_n \sim N\left(y(x_n; \theta), \sigma^2\right)$, then $t_n = y(x_n; \theta) + \varepsilon_n$, where ε_n is an independent stochastic variable satisfying Gaussian distribution with zero mean value and variance σ_ε^2, i.e., $p(\varepsilon) \sim N\big(0, \sigma_\varepsilon^2\big)$. The corresponding likelihood can be written as:

$$p\big(t|\theta, \sigma^2\big) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2}||t - \Phi\theta||^2\right) \tag{6}$$

Using the Bayesian techniques, we obtain the posterior distribution over all unknowns:

$$p\left(\theta, \alpha, \sigma^2|t\right) = \frac{p\left(t|\theta, \alpha, \sigma^2\right) p\left(\theta, \alpha, \sigma^2\right)}{p(t)} \tag{7}$$

$$p\left(t|\theta, \alpha, \sigma^2\right) = \frac{p\left(t|\theta, \sigma^2\right) p(\theta|\alpha)}{p\left(t|\alpha, \sigma^2\right)} = (2\pi)^{-\frac{N+1}{2}} |\Sigma^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta - \mu)^T(\theta - \mu)\right) \tag{8}$$

$$\begin{aligned} \Sigma &= \left(\sigma^{-2}\Phi^T\Phi + A\right)^{-1} \\ \mu &= \sigma^{-2}\Sigma\Phi^T t \\ A &= diag(\alpha_1, \dots, \alpha_M) \end{aligned} \tag{9}$$

Again we will adopt the type-II maximum likelihood approximation where we maximize $p(t|\alpha, \sigma^2)$

$$\begin{aligned} p(t|\alpha, \sigma^2) &= \int p(t|\theta, \alpha)p(\theta|\alpha)d\theta \\ &= (2\pi)^{-\frac{N}{2}} \left|\sigma^2 I + \Phi A^{-1}\Phi^T\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}t^T \left(\sigma^2 I + \Phi A^{-1}\Phi^T\right) t\right) \end{aligned} \tag{10}$$

Considering Equation (10) and making $\sigma^{-2} = \beta$, the log of the evidence is then given as follows

$$\ln p\left(t|\alpha, \sigma^2\right) = -\frac{1}{2}\Big[\ln\left|\beta^{-1}I + \Phi A^{-1}\Phi^T\right| + t^T\left(\beta^{-1}I + \Phi A^{-1}\Phi^T\right)^{-1} t\Big] - \frac{N}{2}\ln(2\pi) \tag{11}$$

$$|A|\left|\beta^{-1}I + \Phi A^{-1}\Phi^T\right| = \left|\beta^{-1}I\right|\left|A + \beta\Phi\Phi^T\right|$$

$$\ln\left|\beta^{-1}I + \Phi A^{-1}\Phi^T\right| = -\ln(\det A) + \frac{L}{2}\ln(|\Sigma|) - N\ln(\beta) \tag{12}$$

Using the Woodbury inversion identity:

$$\left(\beta^{-1}I + \Phi A^{-1}\Phi^T\right)^{-1} = \beta I - \beta\Phi\left(A + \beta\Phi\Phi^T\right)^{-1}\Phi^T\beta \tag{13}$$

$$t^T\left(\beta^{-1}I + \Phi A^{-1}\Phi^T\right)^{-1}t = \beta t^T t - \beta t^T\Phi\Sigma\Phi^T\beta t = \beta t^T(t - \Phi\mu)$$

$$\begin{aligned}\beta t^T(t-\Phi\mu) &= \beta||(t-\Phi\mu)||^2 + \beta t^T\Phi\mu - \beta\mu^T\Phi^T\Phi\mu \\ &= \beta||(t-\Phi\mu)||^2 + \mu^T\Sigma^{-1}\mu - \beta\mu^T\Phi^T\Phi\mu \\ &= \beta||(t-\Phi\mu)||^2 + \mu^T A\mu\end{aligned}$$

differentiate Equation (11) with respect to α, β respectively and setting to zero. Recalculate α, β using their value in the last computation.

The iterative re-estimation formula can be represented as

$$\alpha_i^{re} = \frac{\gamma_i}{\mu_i^2}, \quad \left(\sigma^2\right)^{re} = \frac{||(t-\Phi\mu)||^2}{N - \sum_{i=1}^{M}\gamma_i} \tag{14}$$

where

$$\gamma_i = 1 - \alpha_i^{re}\Sigma_{ii} \tag{15}$$

The optimal hyper-parameters values α_{MP}, β_{MP} are obtained until convergence in the re-estimation.

In the re-estimation above, we generally find that many of α_i tend to infinity or become numerically indistinguishable from infinity. From (7), this implies that $p(\theta_i|t,\alpha,\sigma^2)$ becomes highly peaked at zero. The corresponding consequent parameters thus can be pruned, and sparsity is realized.

The optimal sparse conseqence parameters are written as

$$\theta^{MP} = \beta_{MP}\Sigma\Phi^T t \tag{16}$$

Remark 3.1. *Sparse Bayesian technology is used in computing the consequent parameters* θ*. Some dimension of inputs with low weights in the rules may be ignored. Consequent parameters are simplified in the model identification. We can obtain the relative simple parameter but not lose the prediction accuracy too much. The advantage of this kind of methods is that the consequent parameters sparsity improves scalability of algorithm for a large-scale training dataset in Bayesian inference framework.*

4. **Numerical Example.** In this section, the chaotic Mackey-Glass time series example is studied to evaluate the advantages and the effectiveness of the proposed identification method. Root mean square error (RMSE) criterion is considered, which is measure of deviation between the true values and estimated values.

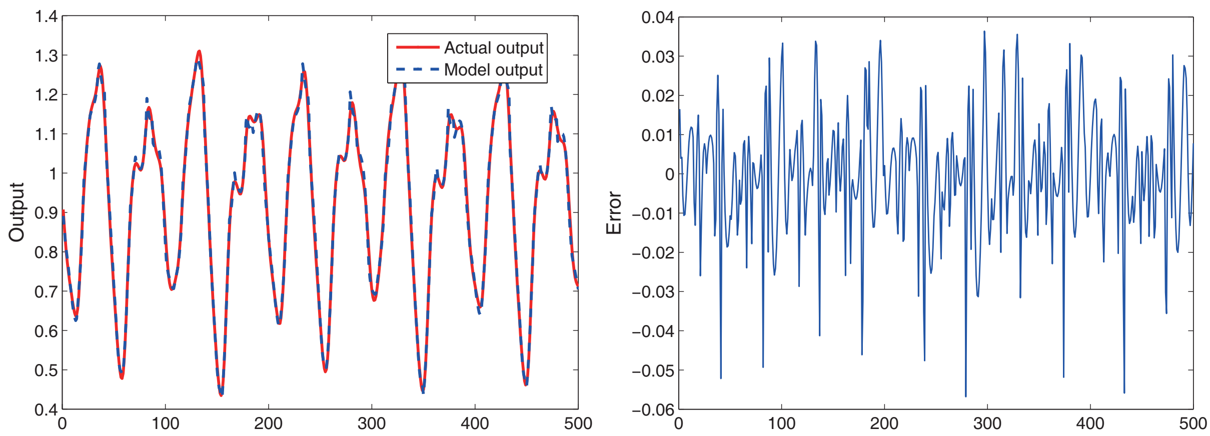
$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y(t) - \hat{y}(t))^2}$$

$y(t)$ is the true output value without noise. $\hat{y}(t)$ is the estimate value.

Chaotic Mackey-Glass time series is generated by the chaotic Mackey-Glass differential delay equation [18] of the form:

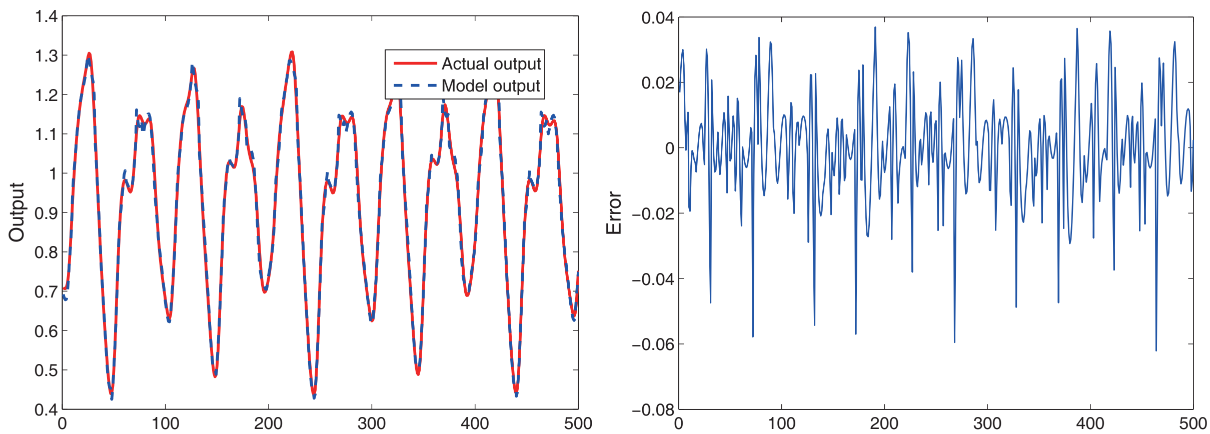
$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1 + x^{10}(t-\tau)} - 0.1x(t)$$

τ is parameter of time delay, which is set $\tau = 17$ in the simulations. A set of $N = 1000$ input-output samples are generated from the Mackey-Glass time series $x(t)$, where 118 $\leq t \leq 1117$. The performance of the proposed model is contrasted with some other models existing in the literature. The first 500 pairs were used as the training samples while the remaining 500 pairs were the testing samples for assessing the predictive performance. We consider the RMSE as the performance index. Using the proposed algorithm we developed a fuzzy model which estimates the future point $x(k+6)$, given the four inputs $x(k-18)$, $x(k-12)$, $x(k-6)$, $x(k)$. Initialized by setting number of cluster $k(1) = 30$, which means that the algorithm is started from a random cluster number. The maximal selected rule number $\max k = 8$. In order to compare the results with [6, 7, 8, 9, 10, 11], we also try to control the maxmum number of rules $\max k$ equal to (or less than) their number of rules. The fuzzy models are trained for 100 epochs. The average values of the premise parameter and corresponding identified consequent parameters are listed in Table 1. The zero values of θ_i in Table 1 are the dimensions ignored. Figures 2(a) and 3(a) show the comparison of the actual output and the model output produced by the system, while Figures 2(b) and 3(b) show the respective error. We can see that the errors are small in the acceptable range. In our method, we focus on the consequent parameters sparsity and its use in



(a) (b)

FIGURE 2. Comparison of our model and the real system for the Mackey-Glass system using training data



(a) (b)

FIGURE 3. Comparison of our model and the real system for the Mackey-Glass system using testing data

Table 1. Parameter values for the Mackey-Glass model

Fuzzy rule		Parameter				
R^1	c_1	1.093	1.187	0.994	0.852	
	σ_1	0.100	1.100	1.100	1.100	
	θ_1	2.858	−0.800	0	−1.596	0.520
R^2	c_2	0.896	1.138	1.184	0.852	0.651
	σ_2	0.110	0.100	0.100	0.124	
	θ_2	1.152	0	0	−0.849	0.592
R^3	c_3	1.181	0.966	0.726	0.742	
	σ_3	0.100	0.100	0.100	0.112	
	θ_3	1.694	−1.028	0.778	−0.974	0.660
R^4	c_4	0.625	0.963	1.074	1.144	
	σ_4	0.119	0.100	0.100	0.100	
	θ_4	1.766	0	0	−1.408	0.641
R^5	c_5	0.821	0.641	0.914	1.136	
	σ_5	0.105	0.100	0.100	0.106	
	θ_5	0.980	0	0	−0.441	0.547
R^6	c_6	1.119	0.662	0.602	1.027	
	σ_6	0.100	0.116	0.100	0.100	
	θ_6	0	0	0.232	0.573	0.583

Table 2. Comparison results for the Mackey-Glass model

Model	Number of inputs	Number of rules	Number of parameters	RMSE (test)
Wang and Mendel [6]	9	121	–	0.01
ANFIS [7]	4	16	104	0.0016
Chen et al. [8]	4	9	81	0.0264
Kukolj [9]	4	9	117	0.0061
H-TS [10]	4	6	78	0.0041
Zou et al. [11]	–	10	–	–
Our model	4	6	68	0.067

adjusting the cluster number (rule number). Going through the statistical analysis of the prediction results in Table 2, it becomes apparent that the number of parameters equals 6 in our model, which is much less than the other methods. The RMSE value equals 0.067, which can meet demand in the practical application.

5. **Conclusion.** In this study, a Bayesian identification framework for T-S fuzzy system is proposed. The main advantage of this approach is that the sparse Bayesian techniques are applied in consequent parameters identification and consequent parameters number sparsity. Many consequent parameters approximate to zero, which can be ignored. We present the numerical example to demonstrate the method. The results in the simulation have verified the validity of the proposed algorithms for T-S fuzzy system. In the future, noninformative prior Bayesian identification framework for T-S fuzzy system is an interesting topic.

Acknowledgment. This work is supported by National Natural Science Foundation (NNSF) of China under Grant (61703149), the Top Talent of the Youth of Project in Hebei Education Department (BJ2017106), Open Projects in Hebei (HBSJQ0705), Hengshui University Project (jg2018012).

REFERENCES

[1] L. A. Zadeh, Fuzzy sets, *Information and Control*, vol.8, no.3, pp.338-353, 1965.

[2] E. Lughofer, *Evolving Fuzzy Models: Incremental Learning, Interpretability, and Stability Issues, Applications*, VDM Publishing, 2008.

[3] T. Takagi and M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans. Systems, Man, and Cybernetics*, no.1, pp.116-132, 1985.

[4] M. Sugeno and T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, *IEEE Trans. Fuzzy Systems*, vol.1, no.1, pp.7-31, 1993.

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

[6] L. Wang and J. M. Mendel, Generating fuzzy rules by learning from examples, *IEEE Trans. Systems, Man, and Cybernetics*, vol.22, no.6, pp.1414-1427, 1992.

[7] A. Kroll, Identification of functional fuzzy models using multidimensional reference fuzzy sets, *Fuzzy Sets and Systems*, vol.80, no.2, pp.149-158, 1996.

[8] J. Chen, Y. Xi and Z. Zhang, A clustering algorithm for fuzzy model identification, *Fuzzy Sets and Systems*, vol.98, no.3, pp.319-329, 1998.

[9] D. Kukolj, Design of adaptive Takagi-Sugeno-Kang fuzzy models, *Applied Soft Computing*, vol.2, no.2, pp.89-103, 2002.

[10] G. Tsekouras, H. Sarimveis, E. Kavakli et al., A hierarchical fuzzy-clustering approach to fuzzy modeling, *Fuzzy Sets and Systems*, vol.150, no.2, pp.245-266, 2005.

[11] W. Zou, C. Li and N. Zhang, A T-S fuzzy model identification approach based on a modified inter type-2 FRCM algorithm, *IEEE Trans. Fuzzy Systems*, 2017.

[12] A. C. Lozano, G. Swirszcz and N. Abe, Grouped orthogonal matching pursuit for variable selection and prediction, *Advances in Neural Information Processing Systems*, pp.1150-1158, 2009.

[13] T. Bai, Y. Li and Y. Tang, Structured sparse representation appearance model for robust visual tracking, *IEEE International Conference on Robotics and Automation (ICRA)*, pp.4399-4404, 2011.

[14] T. Zhang, On the consistency of feature selection using greedy least squares regression, *Journal of Machine Learning Research*, vol.10, pp.555-568, 2009.

[15] M. Luo, F. Sun and H. Liu, Hierarchical structured sparse representation for T-S fuzzy systems identification, *IEEE Trans. Fuzzy Systems*, vol.21, no.6, pp.1032-1043, 2013.

[16] B. Frey and D. Dueck, Clustering by passing messages between data points, *Science*, vol.315, no.5814, pp.972-976, 2007.

[17] M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*, vol.1, pp.211-244, 2001.

[18] M. C. Mackey and L. Glass, Oscillation and chaos in physiological control systems, *Science*, vol.197, no.4300, pp.287-289, 1977.