International Journal of Applied Science and Engineering
2019. 16, 3: 215-227

Fuzzy Rules Reduction Based on Sparse Coding

Huiqin Jiang[ab], Rung-Ching Chen[b*], Qiao-En Liu[b] and Su-Wen Huang[b]

[a]*Department of Information Engineering, Xiamen University of Technology*
Xiamen, China
[b]*Department of Information Management, Chaoyang University of Technology*
Taichung, Taiwan (R.O.C.)

Abstract: With high-dimensional data appearing, the number of fuzzy rules increases which degrade the interpretability and increases the computation complexity of the fuzzy rule-based system. In this paper, we proposed a rule-reduced algorithm. Through the sparse encoding of the fuzzy basis functions (FBFs), rules are reduced. Least angle regression algorithm is proposed here to select the important rules. Compared with other sparse encoding algorithm, Least angle regression algorithm has the advantage of lower computation complexity and better performance. The experimental results show that our proposed algorithm has excellent performance, especially for high-dimension data.

Keywords: Data-driven FISs; lasso; LARS; FCM; sparse enconding.

1. Introduction

In the past two-decade Fuzzy inference systems (FISs) have been successfully applied to different areas as medicine, engineer, management, behavioral science and so on. There are two main methods in FISs design: the expert-driven approach, where expert experience is used to construct fuzzy rules, and the data-driven approach, where knowledge is discovered and extracted from numerical data. Note that these two methods mentioned above are not mutually exclusive, in [1-3], hybrid approaches are explored. Indeed in many situations, the expert experience is lacking and difficult to acquire, and the data-driven approach can get knowledge directly from data which can save a lot of money and time. In addition, the data-driven FISs have strong learning ability [4]; The data-driven FISs have the ability to approximate any continuous function in arbitrary accuracy that means it is a universal approximator [5-7]; The data-driven FISs are necessary for the situation where the process is so complex that a deterministic deduction needs high development time [8]. All reasons mentioned above make the data-driven FISs prevalent system models. Many classical data-driven FISs have been proposed, which include SVD-QR-based approach [9], dynamic optimal training method [10], support vector learning method [1], hybrid learning algorithms based on different mechanisms [11]. Though the methods mentioned above were proved useful, they face the common problem: The rule-explosion problem.

With high-dimensional data appearing, the number of the rules increases exponentially with the dimensionality of the data [12]. The number of rules is of $O(n^k)$, where n is the maximum or an average number of fuzzy terms in each dimensionality, and k is the dimensionality of data. Explosion rules degrade the interpretability and increase the computation complexity of the FISs. Many algorithms have been suggested to overcome this problem.

Corresponding author; e-mail: crching@cyut.edu.tw
Received 30 July 2019
doi:10.6703/IJASE.201911_16(3).215
Revised 11 September 2019

Accepted 21 October 2019

In [12] hierarchical structures of TSK fuzzy classifier were recommended. The main design idea of hierarchical structures is to modify the rule base structure, which is not easy to interpret the output as the input of the next layer. These methods mainly include eigenvalue decomposition (ED) method [13], total least squares (TLS) method based on unbiased parameter estimation of local model [14], the orthogonal least squares (OLS) method [15], the singular value decomposition (SVD) of the fire strength matrix of the fuzzy reasoning system, tensor-form high-order singular value decomposition (HOSVD) method [16], and rule reduction method of fuzzy rule interpolation technology [17, 18] and so on. The common characteristics of these methods are to ignore the structural information of the FISs and to suppose the parameters of each fuzzy rule estimated as independent parameters.

Heuristic algorithms, especially genetic techniques are often applied in the FISs to rank the contribution of each rule and select the important one. The first hereditary rule selection method was introduced by Ishibuchi et al. in [19] for a fuzzy classification task. A genetic multi-selection process within the MOGUL GFS design framework was introduced in [20]. Niching procedure [21] was combined with a genetic selection algorithm in [22] to obtain a best fuzzy rule subset. The sparse representation of the fire strength matrix of rules can also reduce the rules effectively.

Initially, the algorithms were only focused on the accuracy of the system. The accuracy, computational complexity and the interpretability of the FISs have also become the focus of the algorithm. Selecting the best rule base (RB) of a fuzzy system can be seen as a multi-objective problem. Several evolutionary techniques have been proposed to concurrently optimize the accuracy and the complexity of a fuzzy rule base system (FRBS), including the nondominated sorting genetic algorithms [23], the Pareto archived evolution strategy [24], and the continuous ant colony optimization algorithm [25].

Given enough time, heuristic methods can get better system accuracy than noniterative ones. They need a large number of evaluations to converge the optimal RB. So in [26], the approach with the smallest possible number of the assessments based on a greedy search algorithm and followed by a rule removal procedure is suggested. It can only balance the accuracy and convergence speed to a certain extent and can't solve the problem.

Edwin Lughofer [27] proposed the SparseFIS model, which first applied the sparse coding in FIS to reduce the fuzzy rules. Minnan Luo improved SparseFIS model and suggested H-sparseFIS algorithm in [28] for rules reduction. Orthogonal matching pursuit (OMP) algorithm is applied to seek the sparsest solution in [28]. Rather than minimizing an objective function, OMP constructs a dispersed solution to a given problem by iteratively building up an approximation. OMP is a heuristic method, in some cases [29, 30] it works marvellously. However, in a specific situation, l_1 norm minimization does better than OMP [31, 32]. And there exists a wide range of cases where l_1 norm minimization can find the sparsest solution while OMP fails [29, 31, 32]. General-purpose optimizers are much too slow for l_1 norm minimization in a much large-scale application. So quasi-l_1 minimization running as fast as OMP was suggested to seek the sparse solution [33]. Efron proposed the least angle regression (LARS) algorithm [34] in 2004. Experimental results showed the solution obtained by the basic LARS procedure is often identical to the l_1 norm minimization solution. So, LARS was suggested to solve l_1 norm minimization problem.

In this paper, we cast the fuzzy system identification to the sparse coding problem. Through l_1 norm constrain the fuzzy rules are reduced. LARS is applied to seek the most dispersed solution. The main contribution of the LARS-based sparse coding algorithm is as follows. (1) Through l_1 norm regularization, the proposed algorithm can select the essential rules and remove the redundant rules. (2) Rule reduction improves the interpretability of the system while taking into account the accuracy of the system. (3) LARS algorithm is suggested for the solution method for lasso which can complete the solving process in several steps equal to the complexity of the Least squares. The proposed algorithm is especially suited for the high-dimensionality data.

This paper is organized as the following: Section 2 is the introduction of the related work. The suggested fuzzy model is introduced in section 3. Section 4 is the experiment results. The last section is the conclusion.

2. Related Work

2.1 The sparse coding

Sparse coding was proposed by Olshausen and Field, which is across the field of neurobiology, cognitive science, psychology, and computational science. Sparse coding is widely used in image denoising, super-resolution reconstruction, image classification, foreground-background segmentation, and feature selection.

According to the theory of sparse representation, an n-dimensionality signal $Y=(y_1,y_2...y_N)$ can be approximately represented by a set of fixed bases $\{d_i \in R^N\}_{i=1}^{K}$ as:

$$Y \approx \sum_{i=1}^{K} x_i d_i \tag{1}$$

Where $X=(x_1,x_2...x_N)$ is a coefficient vector. Fixed bases are also called dictionary atoms. They are also called over-complete bases, where m □ n. For feature coding, dictionaries are pre-determined. For image and audio signals, their dictionaries can be fast Fourier transform basis, discrete wavelet transform basis, discrete cosine transform basis, curvature basis, and Haar dictionary, or directly select all or part of the input signal set to form a dictionary. If it is sparse, then some dictionary atoms can be used to represent signals. Therefore the process of solving their sparse coefficient vectors is usually called sparse coding.

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} \approx \begin{pmatrix} d_{11} \\ d_{12} \\ \cdots \\ d_{1n} \end{pmatrix} \begin{pmatrix} d_{21} \\ d_{22} \\ \cdots \\ d_{2n} \end{pmatrix} \cdots \begin{pmatrix} d_{K1} \\ d_{K2} \\ \cdots \\ d_{Kn} \end{pmatrix} \begin{pmatrix} 0 \\ : \\ \beta_i \\ : \\ \beta_j \\ : \\ \beta_l \\ : \\ 0 \end{pmatrix} \mathrm{S}$$

Figure 1. Sparse coding.

In Figure 1, S is the number of the non-zero coefficient β_m. And, $S << K$ which means the coefficients are sparse. The sparse representation is to obtain the optimal sparse representation coefficient of the signal in a given dictionary base. The mathematical model of sparse coding can be defined as Eq. (2).

$$\min_X \|DX - Y\| + \lambda \ell(x) \tag{2}$$

The first term in Eq. (2) is the sparse reconstruction term; the second term is the sparse constraint term. Eq. (2) is an under-determined linear equation. The solution of coefficient vector is not unique. Only under regularization constraint, an optimal sparse solution can be found. It can be a norm, where the definition of the norm is as Eq. (3).

$$\|X\|_p = (\sum_i x_i{}^p)^{\frac{1}{p}} \tag{3}$$

When p=1, Eq. (3) is changed as formular (4).

$$\min_X \|DX - Y\| + \lambda \|X\|_1 \tag{4}$$

Eq. (4) is known as Lasso, the classical optimization theory [34] can prove that λ, t, ε satisfy some conditions. The Eq. (4) is equivalent to formular 5 and 6.

$$\min_X \|DX - Y\|, \ s.t. \ \|X\|_1 \le t \tag{5}$$

$$\min_X \|X\|_1, \ s.t. \ \|DX - Y\| \le \varepsilon \tag{6}$$

2.2 Least Angle Regression

LARS [34] was proposed by Efron in 2004 to solve lasso problem, LARS makes a compromise to the forward stagewise algorithm and forwards selection algorithm, preserving a certain degree of accuracy of the forward stagewise algorithm and reducing the computational burden of the forward stagewise algorithm.

Figure 2 shows the LARS algorithm in the case of 2 covariates $X = (x_1, x_2)$. We find the covariate x_1 most correlated with the response y. We take the most significant step $\hat{\gamma}_1$ in the possible direction of the covariate x_1 until covariate x_2 has as much correlation with the current residual. Instead of continuing along the direction of x_1, LARS proceeds in the direction equiangular between the two covariates x_1, x_2.

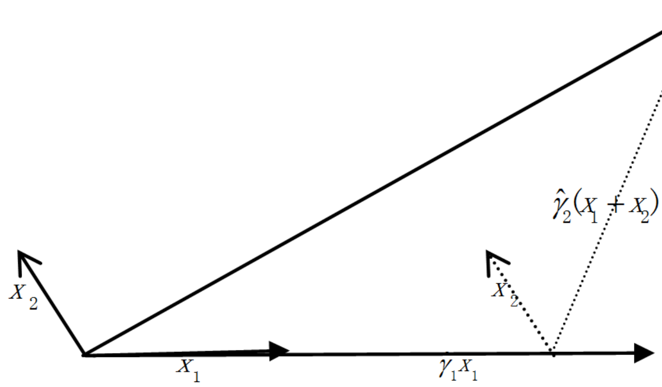

Figure 2. The geometric principle of LARS.

Extend the case of 2-dimensional data to the case m-dimensional data. LARS builds up the estimate $\hat{\mu} = X\hat{\beta}$, where X is m-dimensional data and $\hat{\beta}$ is the coefficient parameter of the regression model, in successive steps, each step adding one covariate to the model.

Algorithm 1 is the steps of the LARS algorithm. The active set A is a subset of indices {1,2,...m}.The initial A is empty. At each step, one of the indexes is added to A. $\hat{\mu}_k$ is the estimate of y at the kth steps. After m steps, the selection process is finished.

Algorithm **1.** The steps of the LARS algorithm.

(1) k=0: begin with the active set A empty which is a subset of indices $\{1,2,...m\}$

(2) k=k+1: the correlation coefficient is $\hat{c}_j = x_j'(y-\mu_{k-1})$ to choose the most significant correlate coefficient $\hat{C} = \max\{|\hat{c}_j|\}$, and add j to $A = \{j : |\hat{c}_j = \hat{C}|\}$

(3) Computing $X_A = (...,s_j x_j,...)_{j\in A}$, $A_A = (1_A' G_A^{-1} 1_A)^{-\frac{1}{2}}$ and $u_A = X_A A_A G_A^{-1} 1_A$, where $s_j = sign\{\hat{c}_j\}$ and $G_A = X_A' X_A$. 1_A is a vector of 1's of length equal to $|A|$, and $a \equiv X' u_A$.

(4) Updating $\hat{\mu}_k = \hat{\mu}_{k-1} + \hat{\gamma}_k u_A$, $\hat{\gamma}_k = \min_{j\in A^c}{}^{+} \left\{ \frac{\hat{C}-\hat{c}_j}{A_A - a_j}, \frac{\hat{C}+\hat{c}_j}{A_A + a_j} \right\}$, $\min^{+}$ indicates that the minimum is taken over only positive components.

(5) Repeating from 2 to 4, until the end.

LARS can be viewed as moderately greedy forward stepwise procedure. LARS moves along the most compromise direction, the equiangular vector. Efron [34] proposed a modified LARS algorithm to produce Lasso solution. Under the Lasso modification, the LARS yields all Lasso solutions [34]. The advantage of LARS [34] answer for Lasso is:

(1) LARS is particularly suitable for cases where the characteristic dimensionality n is much higher than the sample number m.
(2) The algorithm's worst computational complexity is similar to that of the least square method, but its computational speed is almost the same as that of the forward selection algorithm.
(3) The complete path that can produce piecewise linear results is beneficial in cross-validation of the model.

3. Material and method

3.1 The Architecture of Our Suggested Model

The architecture of our suggested model is shown in Figure 3: FCM clustering algorithm is used to partition the input space and extract the antecedent of the fuzzy rules, through weighted defuzzifier, we get the representation function of the output variant with consequent parameters. By solving the optimization problem of minimizing model errors, the following parameters of fuzzy rules are estimated. Usually, the least square algorithm is used to reduce the model error. In our model, the least square algorithm with l_1 constraint, also called Lasso, is adopted to solve the optimization problem. LARS, which is one of the solution methods for lasso problem, are low computational complexity and highly efficient solution for the following parameters.

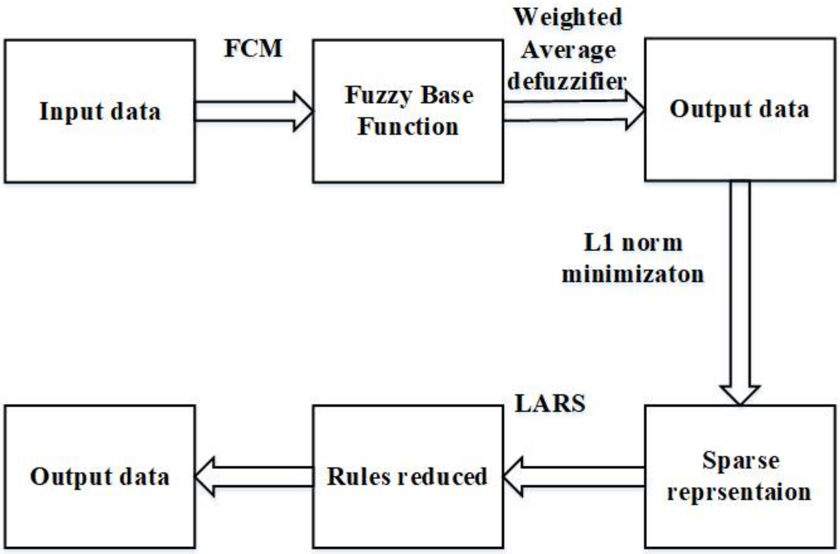

Figure 3. The architecture of the proposed model.

3.2 The Antecedent Extraction Of Fuzzy Rule

For a T-S Fuzzy system, the ith fuzzy rules is presented as:

$$R^i \text{ if } \mathrm{x}_1 \text{ is } \mathrm{A}_{i1},\ \mathrm{x}_2 \text{ is } \mathrm{A}_{i2} \ldots x_{in} \text{ is } \mathrm{A}_{in} \text{ then } \mathrm{y}=\beta_{i1}x_1+\ldots+\beta_{in}x_n \quad i=0,1,2\ldots r \tag{7}$$

Where $X=(x_1,x_2\ldots x_n)^T \in R^n$. For a zero-order T-S system, Eq. (7) is changed as Eq. (8).

$$R^i \text{ if } \mathrm{x}_1 \text{ is } \mathrm{A}_{i1},\ \mathrm{x}_2 \text{ is } \mathrm{A}_{i2} \ldots x_{in} \text{ is } \mathrm{A}_{in} \text{ then } \mathrm{y}=\beta_i \tag{8}$$

Assume fuzzy set A_{ij} has the Gauss membership function.

$$\mu_{A_{ij}}(x_j)=\exp[-(\frac{x_j-c_{ij}}{\sigma_{ij}})^2] \tag{9}$$

c_{ij} and σ_{ij} respectively denote the mean and the standard deviation of the Gauss member function.

Clustering approaches are extensively proposed to partition the input space and determine the membership functions (MFs) of the input variable. Such examples include the subtractive clustering algorithm, fuzzy c-means (FCM) algorithm, mountain clustering algorithm, Gustafson–Kessel (GK) algorithm, and Gath–Geva (GG) algorithm. In this paper, the FCM clustering algorithm is used to partition the input space and extract the antecedent parameters of the fuzzy rules.

Through FCM of input variants, we obtain the membership function matrix $U=(u_{ik})\in R^{r\times N}$, which contains the grade of membership of each data point in each cluster. The membership function matrix should satisfy the following condition:

$$\sum_{i=1}^{r} u_{ik}=1$$

$$u_{ik}\in\{0,1\} \tag{10}$$

The fuzzy fire strength of the ith fuzzy rule is:

$$A_i(x) = \prod_{j=1}^{n} A_{ij}(x_j) = \exp\left(-\frac{1}{2} \times \sum_{j=1}^{n} \left(\frac{x_j - v_{ij}}{\sigma_{ij}}\right)^2\right) \tag{11}$$

The mean c_{ij} and the standard deviation σ_{ij} is computed by Eq. (12, 13)

$$c_{ij} = \frac{\sum_{k=1}^{N} \mu_{ik} x_{kj}}{\sum_{k=1}^{N} \mu_{ik}} \tag{12}$$

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

3.3 The Spare Code of Fuzzy Consequent Parameters

With the weighted average defuzzifier, the ith fuzzy basis functions (FBFs) is denoted by p_i $R^n \to R$ r is the number of the fuzzy rule:

$$p_i = \frac{A_i(x)}{\sum_{i=1}^{r} A_i(x)} \tag{14}$$

The output of singleton zero-order T-S fuzzy inference system $\hat{y}(x)$ is the linear combination of FBFs:

$$\hat{y}(x) = \sum_{i=1}^{r} p_i(x)\beta_i \tag{15}$$

Assume there is a sample dataset with N pairs of input variants and output variants :

$$Z = \{(x_k, y_k) | x_k = (x_{k1}, x_{k2} ... x_{kn})^T \in R^n, y_k \in R, k = 1, 2...N\} \tag{16}$$

x_k and y_k respectively denote the kth n-dimension input variant and the kth output variant. N output variants of singleton FISs are indicated as Eq. (17).

$$\hat{y} = (\hat{y}_1, \hat{y}_2 ... \hat{y}_N)^T \in R^N \tag{17}$$

$\hat{y}$ can be computed by Eq. (18).

$$\hat{y} = P\beta \tag{18}$$

Here P is $N \times r$ dimensionality, representing the FBFs of the N input variants, and $\beta = (\beta_1, \beta_2 ... \beta_r)$ is the vector of the consequent parameters.

Traditionally the least square (LS) is often used to estimate the consequent parameters W by optimizing system：

$$\beta = (\beta_1, \beta_2 ... \beta_r)^T = \arg\min \|y - P\beta\| \tag{19}$$

As we mentioned above, LS with l_1 norm regularization, often called lasso, can select the essential features and avoid the overfitting. The selection function of lasso is utilized in this paper to choose the important rules and remove the redundant rule.

Modification LARS algorithm is proposed in this paper to solve the lasso problem, which can complete the rules selection in m steps, m equal to the number of fuzzy rules. Contrast to the methods proposed before, the algorithm has less complexity and higher computational efficiency.

4. Experimental results

4.1 Nonlinear function modeling

Let us consider the nonlinear function defined as the equation:

$$F(x,y)=\exp\left(-\frac{x^2}{4}\right)+\exp\left(-\frac{y^2}{4}\right)$$

$$\forall x,y\in[-5,5];F(x,y)\in[0,2] \tag{20}$$

We use the suggested algorithm to model this function. We get 400 data points by uniformly sampling the 2-dimension input space. We use 200 data points for training and 200 data points for testing.

Performance indices root mean square error (RMSE) and mean square error (MAE) as shown Eq. (21) are utilized compare to the genfis2 model, genfis3 model, ANFIS and the proposed model.

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2, \qquad \text{RMSE}=\sqrt{MSE}$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{21}$$

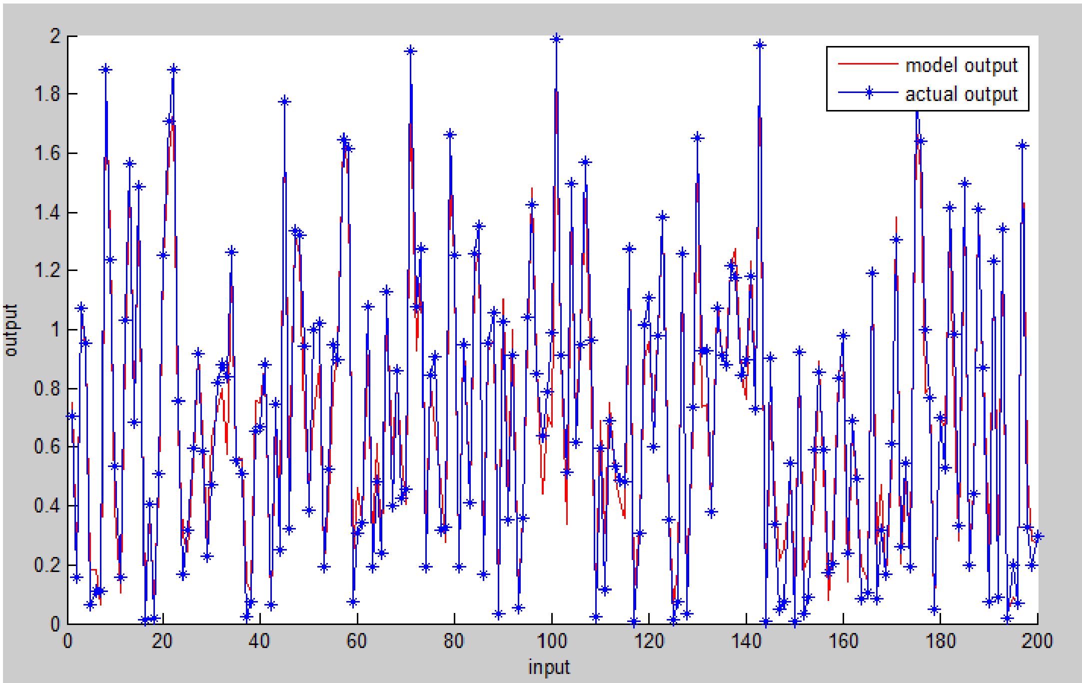

Figure 4. The comparison of model output and actual output.

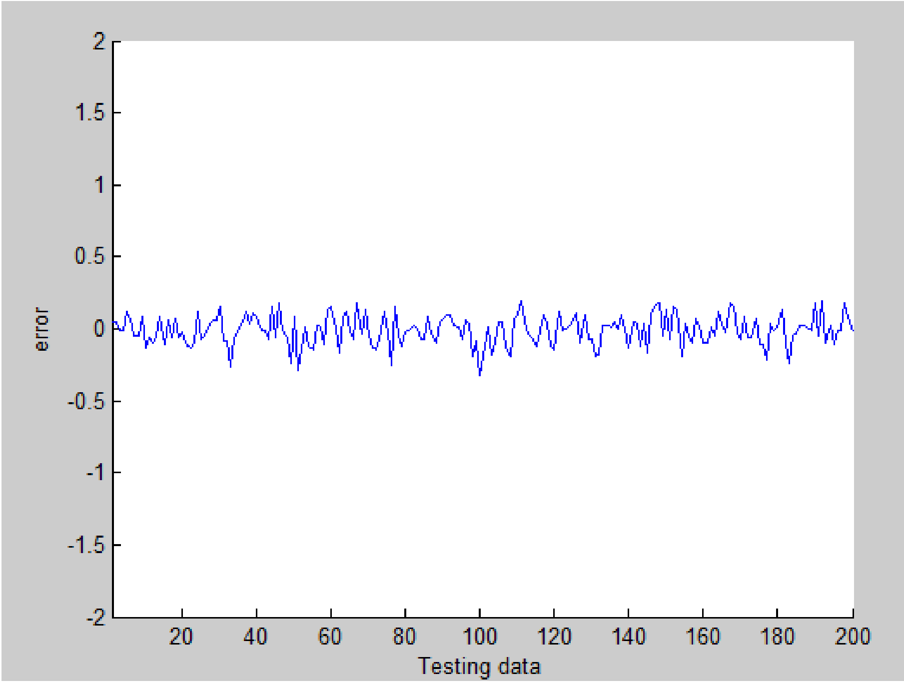

Figure 5. The MAE of model output and actual output.

Figure 4 is the comparison of model output, and actual production and Figure 5 is the MAE of model output and actual output.

There are only two input variables and one output variables in this example. We compare the genfis2 algorithm, genfis3 algorithm, ANFIS algorithm, and our proposed algorithm, we can see from Table 1, for low dimensionality data. Our proposed model has no distinct advantage. ANFIS has better performance at the cost of more rules. But with the dimensionality increasing as shown in the next example about MPG. We will find when the input variants are increased to 6, the Matlab will give the warning: genfis1 has created a large rulebase in the FIS. MATLAB may run out of memory if this FIS is tuned using ANFIS.

Table 1. The comparison of the different system.

model	**No. Of fuzzy rules**	MSE_{train}	MSE_{test}
Genfis2	5	0.0266	0.0371
Genfis3	7	0.2434	0.2740
ANFIS	25	2.3805e-04	6.2860e-04
Our model	8	0.0280	0.0313

4.2 MPG

In this section, a numerical experiment is carried out to illustrate the effectiveness of the suggested method. The data set of automobile fuel consumption in miles per gallon (Auto-Mpg) is obtained from the University of California, Irvine(UCI) Machine learning Repository (http://archive.ics,uci.rdu/ml/datasets.html). The five input variables x_1, x_2, x_3, x_4 and x_5 stand for cylinders, displacement, horsepower, weight, and acceleration respectively as the input variable $X=(x_1,x_2,x_3,x_4,x_5)$ of the system. The fuel consumption in miles per gallon, y, is the output of the system.

FCM is applied to extract the antecedents. We set 10 fuzzy rules, and initial FBFs is then gained through equation (10). We use LARS to solve equation (14). The number 1, 2...10 marked in Figure 6 represent rule 1, rule 2 ...rule 10 respectively. The horizontal coordinate s is defined as the l_1 size of beta at each step in the range [0, 1]. The vertical coordinate beta is the regression coefficient parameters.

Figure 7 shows the selection process. There are only ten steps which completed the regression process. Only one rule was added to the active set A at each step. At the first step, rule 10 was added to the active set A, then rule 7, rule 8, rule 1, rule 6, rule 4, rule 9, rule 3, rule 2, rule 5 was added to the active A in turn. Through Akaike's Information Criterion(AIC) and Bayesian Information Criterion (BIC) evaluation, the rule 5, rule 2, rule 3 are the redundant rules and are removed. Only seven rules are retained. Figure 8 shows the comparison of model output and actual output.

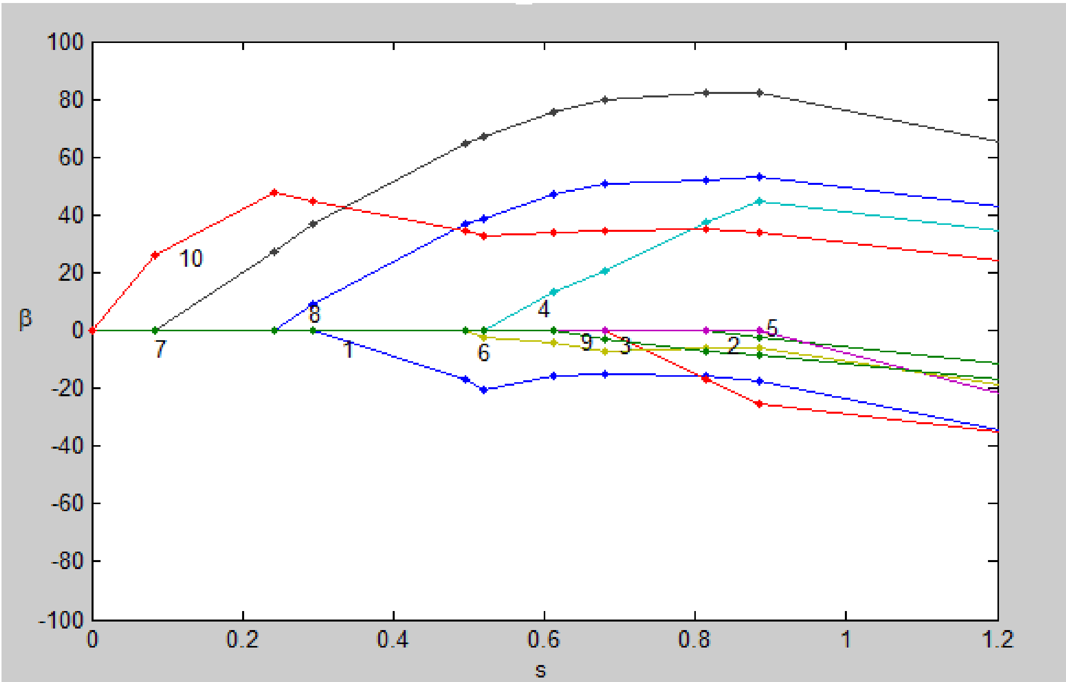

Figure 6. The selection process of fuzzy rules through LARS.

Step	Added	Active set size
1	10	1
2	7	2
3	8	3
4	1	4
5	6	5
6	4	6
7	9	7
8	3	8
9	2	9
10	5	10

Figure 7. The ten steps of rules added.

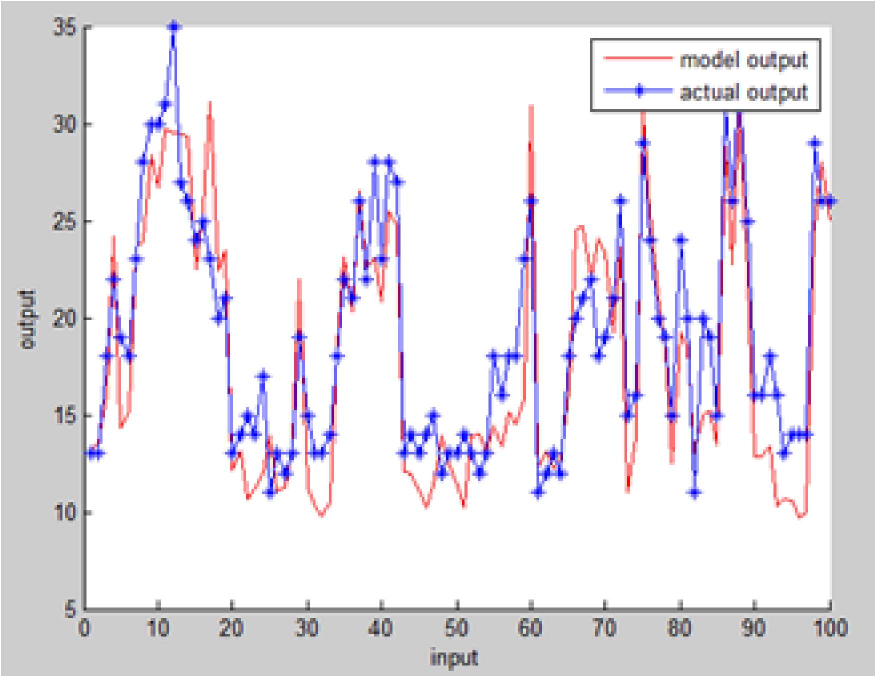

Figure 8. The comparison of model output and actual output.

Algorithm Genfis2 and Genfis3 are chosen as the comparison of our proposed method. These two algorithms also use clustering algorithm for the extract of fuzzy antecedent parameters and the linear regression for the estimate of the consequent parameter. Table 2 is the comparison of different system for errors.

Table 2. The comparison of different system.

model	No. Of fuzzy rules	RMSE	MAE
Genfis2	5	4.5987	3.7573
Genfis3	20	4.5823	3.7253
Our model	7	3.0279	2.5674

5. Conclusion

In this paper, we proposed a LARS-based rule reduced algorithm to solve rule explosion problem. Through the sparse encoding of the fuzzy basis functions (FBFs), important rules are picked and the redundant rules are reduced. LARS algorithm is proposed here to select the rules. LARS algorithm is a convex relaxation approach, also referred to the l_1 norm minimization problem. Compared with other sparse encoding algorithm, LARS algorithm has the advantage of lower computation complexity, equal to the complexity of least square algorithm, and better performance. We applied our algorithm to model nolinear function and the example of Auto-MPG. From the experimental result, we can see for low dimensionality data, the proposed algorithm has no distinct advantage. ANFIS has better performance at the cost of more rules. But with the dimensionality increasing as shown in MPG. When the dimensionality of input variant are increased to 6, the Matlab will give the warning: ANFIS has created a large rule base in the FIS. MATLAB may run out of memory if this FIS is tuned using ANFIS. The suggested algorithm is especially suit for high dimensional data. Sparse encoding has also attracted attention of the scholars in the field of signal processing and image processing. In the future, we will investigate further this issue and extend our algorithm to real signal processing and image processing applications.

Acknowledgment

This work supported by the Ministry of Science and Technology, Taiwan, under Grant MOST-107-2221-E-324-018-MY2. This research is also partially sponsored by Chaoyang University of technology (CYUT) and Higher Education Sprout Project, Ministry of Education, Taiwan, under the project name: "The R&D and the cultivation of talent for Health-Enhancement Products."

.

References

[1] Ji, R., Yang, Y. P. and Zhang, W. D. Zhang, 2014. TS-fuzzy modeling based on ε-insensitive smooth support vector regression. *J. Intell. Fuzzy Syst*, 24, 4:805–817.

[2] Su, M. T., Chen, C. H. and Jian, C. 2011. A rule-based symbiotic modified differential evolution for self-organizing neuro-fuzzy systems. *Appl. Soft Computing*, 11, 8:4847–4858.

[3] Gerardo, M. L. M. et al. 2014. Orthogonal-least-squares and backpropagation hybrid learning algorithm for interval A2-C1 singleton type-2 Takagi-Sugeno-Kang fuzzy logic systems. *Int. J. Hybrid Intell. Syst*, 11, 2:125–135.

[4] Hu, X. S., Li, S. E. and Yang, Y. 2017. Advanced machine learning approach for lithiumion battery state estimation in an electric vehicle. *IEEE Trans. on Transportation Electrification*, 2, 2:140-149.

[5] Wang, L. X. 1992. Fuzzy systems are universal approximators, in Proc. *IEEE Int. Conf. On Fuzzy Systems, San Diego*, 1163-1169.

[6] Buckley, J. J. 1993. Sugeno type controllers are universal controllers. *Fuzzy Sets and Systems,* 53:299-304.

[7] Castro, J. L. 1995. Fuzzy logic controllers are universal approximators. *IEEE Trans. On SMC*, 25:629-635.

[8] Lughofer, E. D. 2008, FLEXFIS: A Robust Incremental Learning Approach for Evolving Takagi–Sugeno Fuzzy Models [J]. *IEEE Transactions on Fuzzy Systems*, 16, 6:1393-1410.

[9] Chen, C. C. and Wong, C. C. 2005. Significant fuzzy rules extraction by an SVD-QR-based approach. *Cybern. Syst.*, 36, 6: 597–622.

[10] Efe, M. O. and Kaynak, O. 2001. A novel optimization procedure for the training of fuzzy inference systems by combining variable structure systems technique and Levenberg–Marquardt algorithm. *Fuzzy Sets Syst,* 122, 1:153–165.

[11] Gerardoet, M. L. M. al., 2014, Orthogonal-least-squares and backpropagation hybrid learning algorithm for interval A2-C1 singleton type-2 Takagi-Sugeno-Kang fuzzy logic systems. *Int. J. Hybrid Intell. Syst,* 11, 2:125–135.

[12] Wang, L. X. 1999, Analysis and design of hierarchical fuzzy systems. *IEEE Trans. Fuzzy Syst*, 7, 5:617–624.

[13] Yen, J. and Wang, L. 1999. Simplifying fuzzy rule-based models using orthogonal transformation methods. *IEEE Transactions on Systems, Man and Cybernetics*, 29, 1:13-24.

[14] Jakubek, S., Hametner. C. and Keuth N. 2008. Total least squares in fuzzy system identification: an application to an industrial engine. *Engineering Applications of Artificial Intelligence*, 21, 8:1277-1288.

[15] Destercke, S., Guillaume, S. and Charnomordic, B. 2007. Building an interpretable fuzzy rule base from data using orthogonal least squares: application to a depollution problem. *Fuzzy Sets and Systems*, 158, 18:2078-2094.

[16] Szeidl, L. and Varlaki, P. 2009. HOSVD based canonical form for polytopic models of dynamic systems. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 13, 1:52- 60.

[17] Tikk, D. and Baranyi, P. 2000. Comprehensive analysis of a new fuzzy rule interpolation method, *IEEE Transactions on Fuzzy Systems*, 8, 3:281-296.

[18] Baranyi, P., Koczy, L. T. and Gedeon, T. D. 2004. A generalized concept for fuzzy rule interpolation. *IEEE Transactions on Fuzzy Systems*, 12, 6:820-832.

[19] Ishibuchi, H., Nozaki, K., Yamamoto, N. and Tanaka, H. 1995. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3, 3:260–270.

[20] Cordón, O., Jesús, M. J., Herrera, F. and Lozano, M. 1999. MOGUL: a methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach. *International Journal of Intelligent Systems*. 14, 11:1123–1153.

[21] Eiben, A. E., Smith, J. E. 2003. Introduction to Evolutionary Computing. *Springer-Verlag*

[22] Cordón, O. and Herrera, F. 1997. A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples. *International Journal of Approximate Reasoning*, 17, 4:369–407.

[23] Gorzałczany, M. B. and Rudzinski, F. 2016. A multi-objective genetic optimization for fast, fuzzy rule-based credit clasfication with balanced accuracy and interpretability. *Appl. Soft Comput*, 40:206–220.

[24] Antonelli, M., Ducange, P. and Marcelloni, F. 2014. A fast and efficient multiobjective evolutionary learning scheme for fuzzy rule-based classifiers, *Inf. Sci*, 283:36–54.

[25] Juang,C. F., Jeng, T. L. and Chang, Y. C. 2016. An interpretable fuzzy system learned through online rule generation and multiobjective ACO with a mobile robot control application. *IEEE Trans. Cybern*, 46, 12 :2706–2718.

[26] Dutu, L. C., Mauris, G. and Bolon, P. 2018. A Fast and Accurate Rule-Base Generation Method. *IEEE Tran, On Fuzzy Systems*, 26, 2:715-731.

[27] Lughofer, K., and Kindermann, S. 2010. SparseFIS: Data-driven learning of fuzzy systems with sparse constraints. *IEEE Trans. Fuzzy Syst*, 18, 2:396–411.

[28] Luo, M. N., Sun, F. and Liu, H. 2014. Hierarchical Structured Sparse Representation for T-S Fuzzy Systems Identification. *IEEE Trans. On Fuzzy Systems*, 21, 6:1032-1043.

[29] Tropp, J. A. 2004, Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50, 10:2231–2242.

[30] Tropp, J. A. and Gilbert, A. C. 2005. Signal recovery from partial information by orthogonal matching pursuit. *unpublished manuscript*.

[31] Donoho, D. L. 2006. For most large underdetermined systems of linear equations, the minimal l_1 solution is also the sparsest solution. *Commun.Pure Appl. Math*, 59, 7:907–934.

[32] Chen, S. S., Donoho, D. L. and Saunders, M. A. 1999. Atomic decomposition by basis pursuit. *SIAM J. Sci Comp*, 20, 1:33–61.

[33] Donoho, D. L. and Tsaig, Y. 2008. Fast solution of $\ell 1$-norm minimization problems when the solution may be sparse. *IEEE Trans. Inf. Theory*, 54, 11:4789–4812.

[34] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. 2004. Least Angle Regression. *The Annals of Statistics*, 32, 2:407-45.