Robust TSK Fuzzy Modeling with Proper Clustering Structure

Chih-Ching Hsiao¹, Shun-Feng Su²

¹ Department of Electrical Engineering, Kao Yuan Institute Of Technology
No.1821, Chung Shan Rd., Lu Chu Hsiang, Kaohsiung County 821, Taiwan R.O.C.
E-mail: s95116@ms10.hinet.net

² Department of Electrical Engineering, National Taiwan University of Science and Technology
E-mail: su@orion.ee.ntust.edu.tw

Abstract: Traditional approaches for modeling TSK fuzzy rules are trying to adjust the parameters in models, and not considering the training data distribution. Hence it will result in an improper clustering structure, especially, when outliers exist. In this paper, a clustering algorithm termed as Robust Proper Structure Fuzzy Regression Algorithm (RPSFR) is proposed to define fuzzy subspaces in a fuzzy regression manner and also data clustering with robust capability against outliers.

Keywords: TSK fuzzy rules, outlier, clustering, proper structure.

1. Introduction

The Takagi–Sugeno–Kang (TSK) type of fuzzy models has attracted a great attention of the fuzzy modeling community due to its good performance in various applications. Fuzzy models can be constructed from given nonlinear system equations by the idea of local approximation [1,2]. Such a way is similar to local linearization technique, so each fuzzy subspace present the individual distribution state, that is such kind behavior very properly describe the characteristic of data clustering. This structure of fuzzy model is called the proper clustering structure. Another way of constructing fuzzy models is to directly use input-output training data. Traditional learning algorithms [3-12] for this kind of approaches are to tune system parameters according to modeling error, and without considering the training data distribution. Thus, the structure of the fuzzy model, after learning, especially when the training data with outliers, will make its specific fuzzy subspaces form the disorderly distribution state. Such kind structure of fuzzy model usually will be improper clustering structure. As training data insufficient or not distribute evenly, the fuzzy model of this kind of training result is unable to describe the prototype system precisely. Thus, in this paper, we propose a novel algorithm for robust fuzzy modeling, called Robust Proper Structure Fuzzy Regression Algorithm (RPSFR), that define fuzzy subspaces in a fuzzy regression manner and also data clustering with robust capability against outliers.

This paper integrates the modeling error and data cluster in a novel fuzzy modeling approach. The
original RCA clustering algorithm [13,14] adopts the idea of robust statistics to reduce the effects of outliers and the concept of competitive agglomeration to determine the proper number of clusters. The RFRA clustering algorithm [15] is modified from RCA. Instead of using clustering concept in partitioning the fuzzy subspaces, it attempted to find fuzzy regressions for fuzzy rules. It is not considered the data distribution and usually, the resultant fuzzy partition may be improper clustering structure. The proposed approach is integrated these two competitive agglomeration processes. Firstly, it called data clustering process, that is, the prototype parameters are the centers and variances of clusters. Next, called regression clustering process, the prototype parameters are the parameter vector in linear regression. Thus, the Robust Proper Structure Fuzzy Regression Algorithm (RPSFR) can not only obtaining the TSK fuzzy model that against the outlier, and also with proper clustering structure. Moreover, the accuracy of the TSK modeling is also to improve.

If more precision is required, it also adopts a so-called Annealing Robust Backpropagation (ARBP) [16] learning algorithm to fine-tuning.

The remaining part of the paper is outlined as follows. Section 2 describes the various clustering algorithms for TSK fuzzy modeling. In Section 3, the RPSFR algorithm is proposed to meaningfully define a proper structure TSK fuzzy model. Simulation results are presented in Section 4. Concluding remarks are presented in section 5.

2. Traditional Clustering Algorithms

The considered problem is to obtain a model $\hat{f}$ from a set of observations, $\{(\tilde{x}(1),y_1),(\tilde{x}(2),y_2),...,\tilde{x}(N),y_N)\}$ with $\tilde{x}(i) = [x_1(i),x_2(i),\cdots,x_n(i)]$ is the $i$-th input vector, and $y_i$ is the desired output for the input $\tilde{x}(i)$. Suppose that those observations are obtained from an unknown function $y = f(x_1,x_2,\cdots,x_n)$. Ideally, we want to construct an $\hat{f}$ that can accurately represent $f$ in term of input-output relationships. A TSK fuzzy model consists of IF-THEN rules that have the form

$$
R^i: \text{If } x_1 \text{ is } A^i_1(\tilde{\theta}^i_1) \text{ and } x_2 \text{ is } A^i_2(\tilde{\theta}^i_2) \cdots x_n \text{ is } A^i_n(\tilde{\theta}^i_n) \text{ then } h^i = f_1(x_1,x_2,\cdots,x_n;\tilde{a}^i) = a^i_0 + a^i_1x_1 + \cdots + a^i_nx_n
$$

for $i=1,2,\ldots,C$, where $C$ is the number of rules, $A^i_l(\tilde{\theta}^i_l)$ is the fuzzy set of the $i$-th rule for $x_l$ with the adjustable parameter set $\tilde{\theta}^i_l$, and $\tilde{a}^i = (a^i_0,\ldots,a^i_n)$ is the parameter set in the consequent part. The predicted output of the fuzzy model is inferred as $\hat{y} = \frac{\sum_{i=1}^{C} h^iw^i}{\sum_{i=1}^{C} w^i}$, where $h^i$ is the output of the $i$-th rule, $w^i = \min_{i=1,\ldots,n} A^i_l(\tilde{\theta}^i_l;x_l)$ is the $i$-th rule’s firing strength, which is obtained as the minimum of the fuzzy membership degrees of all fuzzy variables. In Eq. (1), both the parameters of the premise parts (i.e. $\tilde{\theta}^i_l$) and of consequent parts (i.e. $\tilde{a}^i$) of a TSK fuzzy model are required to be identified. Moreover, the number of rules must be specified.

There are three kinds of TSK fuzzy modeling from training data frequently used in the literature. (1).Fixed clusters type: In this kind approach, the cluster number must be defined in advance. Fuzzy C-Regression Model (FCRM) clustering algorithm [5] is to find a set of training data whose input-output relationship is somehow linear, and then, those
training data can be clustered into one fuzzy subspace.

(2). Growing Clusters type: This kind clustering is generated a new rule or merged the rules by some degree measure for each incoming pattern. SONFIN [6] is basically a fuzzy modeling approach being equipped with structure learning capability. In the approach, fuzzy subspaces are defined by clustering the input portion of the training data. The initial consequent is a singleton and added the input variables formed as linear regression by degree measure.

(3) Competitive Agglomeration Type: This kind algorithm starts by partitioning the data set into a large number of small clusters which reduces its sensitivity to initialization. As the algorithm progresses, adjacent clusters compete for points, and clusters that lose the competition gradually vanish. Thus, it can determine the number of clusters via a process of competitive agglomeration by itself [14]. The RFRA clustering algorithm [16] is modified from RCA, Instead of using clustering concept in partitioning the fuzzy subspaces, it attempted to find fuzzy regressions for fuzzy rules.

For any real-world applications, the obtained training data are always subject to noise or maybe outliers. When noise becomes large or outliers exist, the firstly two kinds modeling approaches may try to fit those improper data in the training process and thus, the learned systems are corrupted. In other words, if there exist outliers and the training process lasts long enough, the obtained systems may have overfitting phenomena. The competitive agglomeration algorithm includes the robust learning algorithms to overcome this problem. However, all of these are only considering the modeling error, and then the resulting TSK fuzzy model in an improper structure.

3. RPSFR Algorithm for TSK Modeling

A novel approach, termed as Robust Proper Structure Fuzzy Regression (RPSFR) Algorithm, is proposed. In the RPSFR algorithm, it is integrated two competitive agglomeration processes. One is the data-clustering process and the other is regression-clustering process. For the data-clustering process, the RCA algorithm is applied and its cost function is defined as

$$J_d = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 \rho_i \left( d_{ij}^2 \right) - \alpha \sum_{i=1}^{C} \left[ \sum_{j=1}^{N} w_{ij} u_{ij} \right]^2$$, \hspace{1cm} (1)

subject to \( \sum_{j=1}^{C} u_{ij} = 1 \), for \( 1 \leq j \leq N \), where \( C \) and \( N \) are the numbers of fuzzy rules and of the training data, respectively. \( u_{ij} \) is the firing strength of the \( i \)-th rule for the \( j \)-th training pattern, \( \rho_i \) is a robust loss function associated with cluster \( i \), \( w_{ij} \) is the weight function and obtained as \( w_{ij} = \partial \rho \left( d_{ij}^2 \right) / \partial d_{ij}^2 \), and \( \alpha \) is the parameter usually called the agglomeration parameter. The Euclidean distance measure is used, \( d_{ij} \) be the distance between the \( j \)-th input data and the center of the \( i \)-th cluster (rule).

$$d_{ij} = \left| \sum_{r=1}^{M} \left( \tilde{x}(r) - \tilde{\theta}^i \right) \left( \Sigma^i \right)^{-1} \left( \tilde{x}(r) - \tilde{\theta}^i \right)^T \right|$$ \hspace{1cm} (2)

\( i=1, 2, \ldots, C \) and \( j=1, 2, \ldots, N \). \( \tilde{\theta}^i = [\theta^i_{c1}, \ldots, \theta^i_{cn}] \) is the center vector of cluster \( \Theta^i \) and \( \Sigma^i \) is its diagonal covariance matrix with diagonal elements \( [\theta^i_{c1}, \ldots, \theta^i_{cn}] \). To minimize \( J_d \) in Eq. (1), the Lagrange multiplier method is applied. The Lagrange function is defined as

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 \rho_i \left( d_{ij}^2 \right) - \alpha \sum_{i=1}^{C} \left[ \sum_{j=1}^{N} w_{ij} u_{ij} \right]^2 - \sum_{j=1}^{N} \lambda_j \left( \sum_{i=1}^{C} u_{ij} - 1 \right)$$ \hspace{1cm} (3)
Next, the RFRA algorithm is adjacent to do the regression-clustering process. The cost function in Eq.(1) is replaced $d^2_{ij}$ with $r^2_{ij}$, and rewritten as

$$J_r = \frac{C}{2} \sum_{i=1}^{C} \sum_{j=1}^{N_i} (\tilde{y}_j - \tilde{f}_i(\tilde{u}; \tilde{a}^i))^2 - \alpha \sum_{i=1}^{C} \left[ \sum_{j=1}^{N_i} \tilde{w}_j \tilde{u}_j \right]^2$$

subject to

$$\sum_{i=1}^{C} \tilde{u}_j = 1, \text{ for } 1 \leq j \leq N, \text{ all symbols are similar to previously definition.}$$

$r_{ij}$ be the residual between the $j$-th desired output of the modeled system and the output of the $i$-th rule with the $j$-th input data; i.e.,

$$r_{ij} = y_j - f_i(\tilde{u}(j); \tilde{a}^i),$$

and the parameter vector $\tilde{a}^i$ for the consequent part of the $i$-th rule is obtained as

$$\tilde{a}^i = [X^T D_i X]^{-1} X^T D_i y, \quad i = 1, 2, ..., C,$$  \hspace{1cm} (5)

where $X \in \mathbb{R}^{N \times (k+1)}$ is matrix with $x_k$ as its $(k+1)$-th row (entries in the first row of $X$ are all 1), $y \in \mathbb{R}^N$ is a vector with $y_k$ as its $k$-th element and $D_i \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $u^2_{ik}w_{ik}$ as its $k$-th diagonal element. $N_i = \sum_{k=1}^{N} \pi_{ik} \pi_{ik}$ is called the robust cardinality of cluster $i$. The robust cardinality is a measure about whether the considered cluster can be merged into its adjacent cluster; i.e. the agglomeration process. When the robust cardinality is less than a pre-specified constant $N_i$, cluster $i$ is discarded. Assume that Gaussian membership functions are used in the premise parts, (i.e.,

$$\mu_i(\theta_{ci}, \theta_{vi}) = \exp \left\{ - \frac{(x_i - \theta_{ci})^2}{2 \theta_{vi}^2} \right\},$$

where $\theta_{ci}$ and $\theta_{vi}$ are two adjustable parameters of the $l$-th membership function of the $i$-th fuzzy rules. To combine those two processes, a good choice of these two parameters is a weighted average of them, that is, the new parameters are

$$\tilde{u}_{ik} = \lambda u_{ik} + (1 - \lambda) \bar{u}_{ik} \quad \text{ (6)}$$

$$\tilde{w}_{ik} = \lambda w_{ik} + (1 - \lambda) \bar{w}_{ik} \quad \text{ (7)}$$

where $\lambda \in [0, 1]$ is a weighted factor that is selected by designer. To get much accuracy, a reasonable choice is $\lambda > 0.5$. Then, we have two update equations as follows

$$\theta'_{ci} = \frac{\sum_{k=1}^{N} (\tilde{u}_{ik})^2 \tilde{w}_{ik} x_i(k)}{\sum_{k=1}^{N} (\tilde{u}_{ik})^2 \tilde{w}_{ik}} \quad \text{ (8)}$$

$$\theta'_{vi} = \frac{\sum_{k=1}^{N} (\tilde{u}_{ik})^2 \tilde{w}_{ik} (x_i(k) - \theta_{vi})^2}{\sum_{k=1}^{N} (\tilde{u}_{ik})^2 \tilde{w}_{ik}} \quad \text{ (9)}$$

The proposed RPSFR Algorithm is described in the following.

[Step 1]: Set initial conditions, step-size and the stop criterion.

[Step 2]: Compute Euclidean distance measure $d^2_{ij}$ by using Eq. (2).

[Step 3]: Update the weights $w_{ij}$, $\alpha(t)$ and $u_{ij}$.

[Step 4]: Compute the consequent parameter sets $\tilde{a}^i$ and $r_{ij}$.

[Step 5]: Update the weights $\bar{w}_{ij}$, $\bar{a}(t)$ and $\bar{u}_{ij}$.

[Step 6]: Update the center $\theta'_{ci}$ and variance $\theta'_{vi}$.

[Step 7]: Compute the robust cardinality $N_i = \sum_{k=1}^{N} \pi_{ik} \pi_{ik}$. Deciding whether cluster $i$ is discarded or not.

[Step 8]: Update the tuning parameter.

[Step 9]: If the stop criterion is satisfied, then stop; otherwise go to [Step 2].

4. Simulation Example

The $sinc$ function is considered. $sinc(x)$ is defined as

$$y = \frac{\sin(x)}{x}, \quad -10 \leq x \leq 10.$$  \hspace{1cm} (10)

201 input-output data are used. The gross error model
is used for modeling outliers. The gross error model is defined as

$$F = (1 - \varepsilon)G + \varepsilon H$$

where $F$ is the added noise distribution and $G$ and $H$ are probability distributions that occur with probability $1 - \varepsilon$ and $\varepsilon$, respectively. The values used in the gross error model are $\varepsilon = 0.05$, $G \sim N(0, 0.05)$ and $H \sim N(0, 1)$. For comparison, three algorithms of constructing TSK fuzzy models are also implemented in our study. One is the FCRM clustering algorithm with BP learning algorithm, Self-Constructing Neural Fuzzy Inference Network (SONFIN) and the other is the RFRA clustering algorithm with robust BP learning algorithm. These three algorithms are selected for comparison because they are all TSK fuzzy modeling approaches and possess the structure learning capability, which means their rules are dynamically generated in the training process. After training with fine-tuning, the RPSFR is the best one in RMSE that can refer to Table. I and Fig. 1, and the structure is a proper clustering structure, the others are not, as shown in Fig. 2.

### Table I.

<table>
<thead>
<tr>
<th>sinc function</th>
<th>Algorithm</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RPSFR (6 rules)</td>
<td>0.0706</td>
</tr>
<tr>
<td></td>
<td>RPSFR (6 rules) with ARBP</td>
<td>0.0264</td>
</tr>
<tr>
<td></td>
<td>RFRA (6 rules)</td>
<td>0.2706</td>
</tr>
<tr>
<td></td>
<td>RFRA (6 rules) with ARBP</td>
<td>0.0351</td>
</tr>
<tr>
<td></td>
<td>FCRM(6 rules)</td>
<td>0.3033</td>
</tr>
<tr>
<td></td>
<td>FCRM(6 rules) with BP</td>
<td>0.0924</td>
</tr>
<tr>
<td></td>
<td>SONFIN(6 rules)</td>
<td>0.3072</td>
</tr>
<tr>
<td></td>
<td>SONFIN(6 rules) with BP</td>
<td>0.0918</td>
</tr>
</tbody>
</table>

### 5. Conclusions

Various TSK modeling approaches have been proposed in the literature. However, all of them only consider the modeling error and result an improper clustering structure. In this paper, a robust TSK fuzzy modeling approach termed as Robust Proper Structure Fuzzy Regression (RPSFR) Algorithm is proposed. It is proposed to simultaneously define fuzzy subspaces and find the parameters in the consequent parts of TSK rules. This clustering algorithm not only finds regression instead of clustering for rules, but also has robust capability against outliers and with a proper
clustering structure that is very suitable to describe a dynamic system. The proposed robust TSK fuzzy modeling approach is tested for example and indeed showed superior performance in our simulation.

References


