Far East Journal of Experimental and Theoretical Artificial Intelligence
Volume 5, Numbers 1-2, 2010, Pages 45-55
Published Online: July 22, 2010
This paper is available online at http://pphmj.com/journals/fejetai.htm
© 2010 Pushpa Publishing House

ROBUST TSK MODELING FOR FUNCTION APPROXIMATION BASED ON ROUGH SETS

CHIH-CHING HSIAO

Electrical Engineering Department
Kao Yuan University
Taiwan
e-mail: cchsiao@cc.kyu.edu.tw

Abstract

Fuzzy set and the rough set theories turned out to be particularly adequate for the analysis of various types of data, especially, when dealing with inexact, uncertain or vague knowledge. In this paper, we propose a novel algorithm, which termed as rough-fuzzy *C*-regression model (RFCRM), that defines fuzzy subspaces in a fuzzy regression manner and also includes rough-set theory for function approximation with robust capability against outliers.

1. Introduction

In the last two decades, rough sets and fuzzy sets turned out to be two contemporary progresses in analyzing inexact, imprecise, uncertain, or vague knowledge. The former capture the distinct aspect of indiscernibility in knowledge, while the latter describe the inherent feature of vagueness in linguistics and decision-making. The rough-set theory proposed by Pawlak [1, 2] is a mathematical theory dealing with uncertainty in data. Rough sets rely on the notion of lower and upper

Keywords and phrases: TSK modeling, fuzzy set, rough set, function approximation.

This work is supported in part by National Science Council of Taiwan under Grant NSC 97-2221-E-244 -008.

Communicated by Shun-Feng Su

Received February 28, 2010

approximations of a set and are applied to control applications, system identification and discovery of rules from experimental data [3-6].

The TSK type of fuzzy models proposed in [7, 8] has attracted a great attention of the fuzzy modeling community due to their good performances in various applications. To construct a TSK fuzzy model, the fuzzy subspaces required for defining fuzzy partitions in premise parts and the parameters required for defining functions in consequent parts must be both obtained. The fuzzy c-mean (FCM) [9] is suitable to define fuzzy subspaces for TSK fuzzy modeling. However, the resulting membership values do not always correspond well to the degrees of belonging of the data, and it may be inaccurate in a noisy environment [10]. In [11], Lingras and West introduced a new clustering method called rough c-mean (RCM), which describes a cluster by a center and a pair of lower and upper approximations. The lower and upper approximations are different weighted parameters that are used to compute the new centers. Recently, combining both rough and fuzzy sets are proposed in the literature such as [12-16]. Mitra et al. [12] proposed a new c-mean algorithm where each cluster is consist of a fuzzy lower approximation and a fuzzy boundary. However, the objects in lower approximation of a cluster should have a similar influence on the corresponding centroid and cluster, and their weights should be independent of other centroids and clusters. Moreover, it is sensitive to noise and outliers. In [13], Maji and Pal proposed a generalized hybrid algorithm based on rough and fuzzy sets. While the membership function of the fuzzy sets enables efficient handling of overlapping partitions, the concept of lower and upper approximations of rough sets deals with uncertainty in class definition. It avoids the problems of noise sensitivity of the FCM.

In real-data analysis, noise and outliers are unavoidable. It is futile to do data based analysis when data are contaminated with outlier because outliers van lead to incorrect analysis results. FCRM clustering algorithm [17] finds a set of training data whose input-output relationship is somehow linear, and then, those training data can be clustered into one fuzzy subspace. Moreover, the model is further adjusted by supervised learning algorithms to improve the modeling accuracy and is easily affected by outliers. In [18, 19], the proposed clustering algorithms are modified from FCRM clustering algorithm by incorporating a robust mechanism and the obtained model will not be significantly affected by outliers.

In this paper, the proposed approach is integrated two contemporary progresses in analyzing vague data. One is the fuzzy set and the other is rough set. The rough-fuzzy *C*-regression modeling algorithm for the TSK fuzzy modeling, it is employed to obtain the fuzzy subspaces and the parameters of functions in consequent parts for the TSK fuzzy models by viewing each rule as a rough set. The remaining part of the paper is outlined as follows. Section 2 describes the TSK fuzzy model and rough sets. In Section 3, the rough-fuzzy *C*-regression modeling algorithm (RFCRM) is proposed to meaningfully define a TSK fuzzy model. Simulation results are presented in Section 4. Concluding remarks are presented in Section 5.

2. TSK Fuzzy Modeling and Rough Sets

2.1. TSK fuzzy modeling

A TSK fuzzy model consists of IF-THEN rules that have the form

$$R^i : \text{If } x_1 \text{ is } A_1^i(\vec{\theta}_1^i) \text{ and } x_2 \text{ is } A_2^i(\vec{\theta}_2^i), \ldots, x_n \text{ is } A_n^i(\vec{\theta}_n^i)$$

$$\text{then } h^i = f_i(x_1, x_2, \ldots, x_n; \vec{a}^i) = a_0^i + a_1^i x_1 + \cdots + a_n^i x_n, \tag{1}$$

for $i = 1, 2, \ldots, C,$ where C is the number of rules, $A_l^i(\vec{\theta}_l^i)$ is the fuzzy set of the ith rule for x_l with the adjustable parameter set $\vec{\theta}_l^i$, and $\vec{a}^i = (a_0^i, \ldots, a_n^i)$ is the parameter set in the consequent part. The predicted output of the fuzzy model is inferred as $\hat{y} = \sum_{i=1}^{C} h^i w^i \Big/ \sum_{i=1}^{C} w^i,$ where h^i is the output of the ith rule, $w^i = \min_{l=1,\ldots,n} A_l^i(\vec{\theta}_l^i; x_l)$ is the ith rule's firing strength. In equation (1), both the parameters of the premise parts (i.e., $\vec{\theta}_l^i$) and of consequent parts (i.e., $\vec{a}^i$) of a TSK fuzzy model are required to be identified. Moreover, the number of rules must be specified. The FCM algorithm is proposed in [9] and its cost function is defined as

$$J_d = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 (d_{ij}^2), \quad \text{subject to } \sum_{i=1}^{C} u_{ij} = 1, \text{ for } 1 \le j \le N, \tag{2}$$

where C and N are the numbers of fuzzy clusters and of the input training data, respectively. u_{ij} is the membership of the ith cluster for the jth training pattern. The Euclidean distance measure is used, d_{ij} is the distance between the jth input data and the center of the ith cluster. The FCRM algorithm is a modified version of FCM.

The cost function in equation (2) is replaced d_{ij}^2 with r_{ij}^2, and rewritten as

$$J_r = \sum_{i=1}^{C} \sum_{j=1}^{N} \bar{u}_{ij}^2 (r_{ij}^2), \quad \text{subject to } \sum_{i=1}^{C} \bar{u}_{ij} = 1, \text{ for } 1 \le j \le N, \tag{3}$$

where r_{ij} is 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., $r_{ij} = y_j - f_i(\vec{x}(j); \vec{a}^i)$, and the parameter vector $\vec{a}^i$ for the consequent part of the *i*th rule is obtained as

$$\vec{a}^i = [X^T D_i X]^{-1} X^T D_i Y, \quad i = 1, 2, ..., C, \tag{4}$$

where $X \in R^{N \times (n+1)}$ is matrix with x_k as its $(k+1)$th row (entries in the first row of *X* are all 1), $Y \in R^N$ is a vector with y_k as its *k*th element and $D_i \in R^{N \times N}$ is a diagonal matrix with u_{ik}^2 as its *k*th diagonal element. The Lagrange multiplier method is applied to minimize J_r in equation (3).

2.2. Rough sets

The theory of rough sets [1, 2] has recently emerged as another major mathematical tool for managing uncertainty. The intention is to approximate a rough concept in the domain of discourse by a pair of exact concepts, called the lower and upper approximations. The lower approximation is the set of objects definitely belonging to the vague concept, whereas the upper approximation is the set of objects possibly belonging to the same. A rough set *X* is characterized by its lower and upper approximations $\underline{B}X$ and $\overline{B}X$, respectively, with the following properties:

(1) An object $\mathbf{x}_k$ can be part of at most one lower approximation.

(2) If $\mathbf{x}_k \in \underline{B}X$ of cluster *X*, then simultaneously $\mathbf{x}_k \in \overline{B}X$.

(3) If $\mathbf{x}_k$ is not a part of any lower approximation, then it belongs to two or more upper approximations.

That is, the lower approximation $\underline{B}X$ is the union of all the elementary sets which are subsets of *X*, and the upper approximation $\overline{B}X$ is the union of all the elementary sets which have a nonempty intersection with *X*. The $\overline{B}X - \underline{B}X$ is the representation of an ordinary set *X* in the approximation space called the rough set of *X*.

3. Rough-fuzzy *C*-regression Modeling Algorithm

A novel approach, termed as rough-fuzzy *C*-regression modeling algorithm (RFCRM), is proposed, a modified version of FCRM. In the RFCRM algorithm, it is integrated fuzzy set and rough set. This includes two stages, first is data-clustering stage and the next is regression-clustering stage.

In the data-clustering stage, it sets initial values for clustering by applying fuzzy c-means. Next, the regression-clustering stage is applied. In the RFCRM algorithm, the concept of FCRM is extended by viewing each rule as a rough set. Beside the condition under which an object may belong to the lower or upper bound of a rule. Let $\vec{x}(j)$ be an object at membership value u_{ij} between the *j*th desired output of the modeled system and the output of the *i*th rule with the *j*th input data. The difference in membership value $u_{ij} - u_{kj}$, $i \neq k$, can be used to determine whether $\vec{x}(j)$ should belong to the lower or upper approximations of the rules. Assume that Gaussian membership functions are used in the premise parts, $A_l^i(\theta_{cl}^i, \theta_{vl}^i) = \exp\{-(x_l, \theta_{cl}^i)^2/2\theta_{vl}^i\}$, where θ_{cl}^i and θ_{vl}^i are two adjustable parameters of the *l*th membership function of the *i*th fuzzy rules. Then, we have two update equations as equation (5) and equation (6). Here, $\underline{B}R_i$ and $\overline{B}R_i$ indicate the lower and upper approximations of rule R_i, respectively. $\overline{B}R_i - \underline{B}R_i$ is the rough boundary between the two approximations. The parameters $\underline{w}$ and $\overline{w}$ correspond to the relative importance of the lower and upper approximations, respectively. When a rule contains data objects in both its lower and upper approximations, these are weighted by $\underline{w}$, $\overline{w}$ and $\underline{w} + \overline{w} = 1$ depending on their importance during modeling. The proposed **RFCRM Algorithm** is described in the following:

[**Step 1**]. Assign initial means for the *C* clusters by FCM.

[**Step 2**]. Compute u_{ij} for the *C* rules and data objects.

[**Step 3**]. Assign each data object $\vec{x}(j)$ to the lower approximation $\underline{B}R_i$ or upper approximation $\overline{B}R_i$, $\overline{B}R_k$ of rule pairs R_i and R_k by computing the difference in its difference $u_{ij} - u_{kj}$ from the rule pairs *i*th rule and *k*th rule.

[**Step 4**]. Let u_{ij} be maximum, u_{kj} be the next to maximum and ρ be a pre-define threshold.

If $u_{ij} - u_{kj} < \rho$,

then $\vec{x}(j) \in \overline{B}R_i$ and $\vec{x}(j) \in \overline{B}R_k$ and $\vec{x}(j)$ cannot be a member of any lower approximation,

else $\vec{x}(j) \in \underline{BR}_i$ such that residual u_{ij} is maximum over the C rules.

[**Step 5**]. Compute new parameter vector for each rule using equation (4).

[**Step 6**]. Update the center θ_{cl}^{i} and variance θ_{vl}^{i} by equations (5), (6).

[**Step 7**]. **Repeat** Steps 2-6 **until** convergence.

$$\theta_{cl}^{i} = \begin{cases} \underline{w}\dfrac{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2 x_l(k)}{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2} + \overline{w}\dfrac{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2 x_l(k)}{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2}, & \text{if } \underline{BR}_i \neq \varnothing \text{ and } \overline{B}R_i - \underline{BR}_i \neq \varnothing, \\[2ex] \dfrac{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2 x_l(k)}{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2}, & \text{if } \underline{BR}_i \neq \varnothing \text{ and } \overline{B}R_i - \underline{BR}_i \neq \varnothing, \\[2ex] \dfrac{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2 x_l(k)}{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2}, & \text{otherwise,} \end{cases} \tag{5}$$

$$\theta_{vl}^{i} = \begin{cases} \underline{w}\dfrac{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2 (x_l(k)-\theta_{cl}^{i})^2}{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2} + \overline{w}\dfrac{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2 (x_l(k)-\theta_{cl}^{i})^2}{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2}, & \text{if } \underline{BR}_i \neq \varnothing \text{ and } \overline{B}R_i - \underline{BR}_i \neq \varnothing, \\[2ex] \dfrac{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2 (x_l(k)-\theta_{cl}^{i})^2}{\sum_{\vec{x}(j)\in(\overline{B}R_i-\underline{BR}_i)}(u_{ik})^2}, & \text{if } \underline{BR}_i = \varnothing \text{ and } \overline{B}R_i - \underline{BR}_i \neq \varnothing, \\[2ex] \dfrac{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2 (x_l(k)-\theta_{cl}^{i})^2}{\sum_{\vec{x}(j)\in\underline{BR}_i}(u_{ik})^2}, & \text{otherwise.} \end{cases} \tag{6}$$

From above algorithm, the larger value of threshold ρ, the more likely is $\vec{x}(j)$ to lie within the rough boundary of a rule. The small value of ρ implies that more patterns are allowed to belong to any of the lower approximations. In other words, ρ represents the size of granules of rough-fuzzy clustering, thus, if we set $\rho = 1$, then $\underline{B}R_i = \varnothing$, which is equivalent to FCRM. Thereby, we can adjust the weighting parameter $\underline{w}$ and/or lower approximation (or/and upper approximation) to weaken the outlier effect. It is reasonable to select $0.5 < \underline{w} < 1$.

The TSK fuzzy model obtained by the RFCRM algorithm has been with a level of accuracy. Furthermore, to improve the modeling accuracy, a robust learning algorithm called ARBP [20] is employed to adjust these parameters of TSK fuzzy rules. In this algorithm, it is simply embedding an annealing process into the learning process, the robust cost function is defined as $E_{ARBP} = \sum_{i=1}^{N} \sigma(e_k, \beta(\tau))$, where $\sigma(\cdot)$ is the loss function. Thus the parameters of premise parts are updated as

$$\Delta\theta_{jl}^{i} = \eta\varphi(e_k, \beta)(y_k^i - \hat{y}_k)\frac{1}{\sum_{i=1}^{C} w^i}\frac{\partial w^i}{\partial \theta_{jl}^{i}}, \tag{7}$$

where $e_k = y_k - \hat{y}_k$, $\varphi(\cdot)$ is the derivative of $\sigma(\cdot)$. The parameters of consequent parts are updated as

$$\Delta a_j^i = \zeta\varphi(e_k, \beta)\frac{w^i x_j}{\sum_{i=1}^{C} w^i}, \tag{8}$$

where ζ is the learning constant.

4. Simulation Examples

Two examples of using the proposed RFCRM approach are illustrated. Both of them are added gross error models 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 ε, 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 Example 1, a function is considered as $y = x^{2/3}$, $-2 \leq x \leq 2$, 201 input-output data that added the gross error are used. In this example, the number of clusters is set as 3. For comparison, the FCRM clustering algorithm is considered in this study. The rough approximated results of each approach are presented in Figure 1. Furthermore, the final results of the TSK fuzzy models after fine-tuning process are also shown in Figure 2. It can show that the RFCRM with robust capability against outliers.

Example 2 is the *sinc* function $y = \sin(x)/x$, $-10 \leq x \leq 10$, 201 input-output data that added the gross error are also used. The rough approximated results of each approach are presented in Figure 3. After Fine-tuning, the simulated results are shown in Figure 4. It also shows that the RFCRM with better approximated result.

5. Conclusions

Fuzzy set and the rough set theories turned out to be particularly adequate for the analysis of various types of data with inexact, uncertain or vague knowledge. In this paper, we propose a novel algorithm, which termed as rough-fuzzy *C*-regression model (RFCRM), that defines fuzzy subspaces in a fuzzy regression manner and also includes rough-set theory for function approximation with robust capability against outliers. In the simulated examples, it can show that the RFCRM with good approximated results.

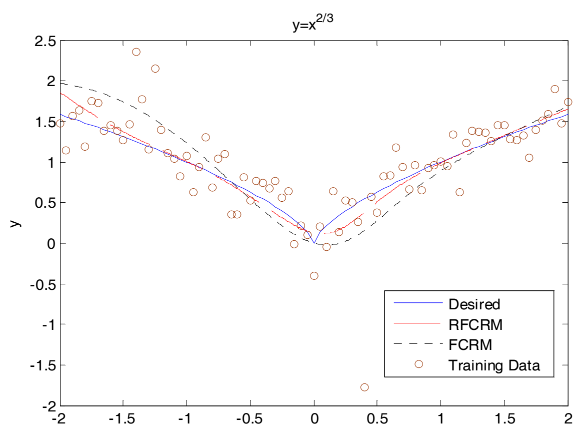

Figure 1. The approximated results with outlier for Example 1.

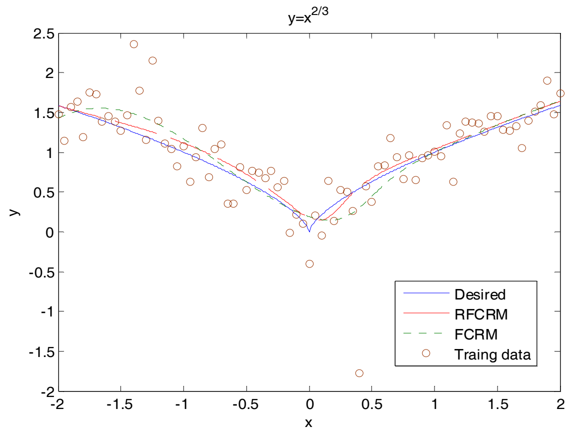

Figure 2. The approximated results with outlier after fine-tuning for Example 1.

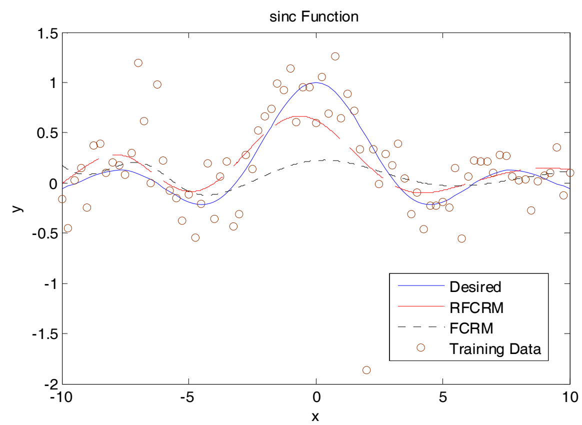

Figure 3. The approximated results with outlier for Example 2.

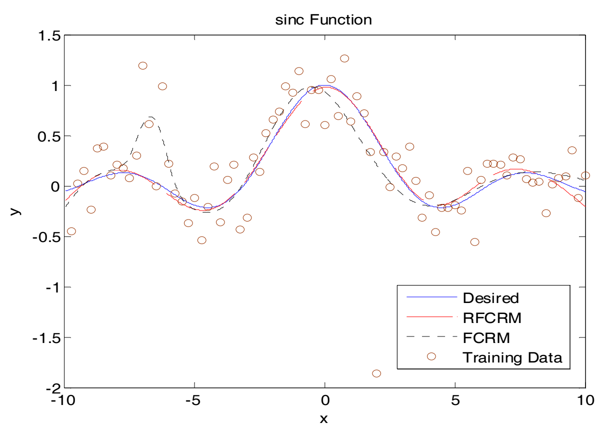

Figure 4. The approximated results with outlier after fine-tuning for Example 2.

References

[1] Z. Pawlak, Why rough set?, Proceedings of IEEE International Conference on Fuzzy Systems, New Orleans, Louisiana, USA, 1996, pp. 738-743.

[2] Z. Pawlak, Rough sets, Internat. J. Comput. Inform. Sci. 11(5) (1982), 341-356.

[3] E. Czogała, A. Mrózek and Z. Pawlak, The idea of a rough fuzzy controller and its application to the stabilization of a pendulum-car system, Fuzzy Sets and Systems 72(1) (1995), 61-73.

[4] Y. Cho, K. Lee, J. Yoo and M. Park, Autogeneration of fuzzy rules and membership functions for fuzzy modeling using rough set theory, IEE Proc. Control Theory Appl. 145(5) (1998), 437-442.

[5] S. K. Pal, S. Mitra and P. Mitra, Rough-fuzzy MLP: modular evolution, rule generation, and evaluation, IEEE Trans. Knowl. Data Eng. 15(1) (2003), 14-25.

[6] X. Wang, Eric C. C. Tsang, S. Zhao, D. Chen and Daniel S. Yeung, Learning fuzzy rules from fuzzy samples based on rough set technique. Inform. Sci. 177(20) (2007), 4493-4514.

[7] T. Takagi and M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Trans. Syst. Man Cybern. 15 (1985), 116-132.

[8] M. Sugeno and T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, IEEE Trans. Fuzzy Syst. 1 (1993), 7-31.

[9] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms, With a foreword by L. A. Zadeh, Advanced Applications in Pattern Recognition, Plenum Press, New York, London, 1981.

[10] R. Krishnapuram and J. M. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst. 1(2) (1993), 98-110.

[11] P. Lingras and C. West, Interval set clustering of web users with rough *K*-means, J. Intell. Inf. Syst. 23(1) (2004), 5-16.

[12] Sushmita Mitra, Haider Banka and Witold Pedrycz, Rough-fuzzy collaborative clustering, IEEE Trans. Syst. Man Cybern. Part B 36(4) (2006), 795-805.

[13] Pradipta Maji and Sankar K. Pal, Rough set based generalized fuzzy *C*-means algorithm and quantitative indices, IEEE Trans. Syst. Man Cybern. Part B 37(6) (2007), 1529-1540.

[14] D. Chen, Q. He and X. Wang, FRSVMs: Fuzzy rough set based support vector machines, Fuzzy Sets and Systems 161 (2010), 596-607.

[15] Robert Nowicki, On combining neuro-fuzzy architectures with the rough set theory to solve classification problems with incomplete data, IEEE Trans. Knowl. Data Eng. 20 (2008), 1239-1253.

[16] Y. Zhao and J. Sun, Rough v-support vector regression, Expert Systems with Applications 36 (2009), 9793-9798.

[17] E. Kim, M. Park, S. Ji and M. Park, A new approach to fuzzy modeling, IEEE Trans. Fuzzy Syst. 5(3) (1997), 328-337.

[18] C. C. Chuang, S. F. Su and S. S. Chen, Robust TSK fuzzy modeling for function approximation with outliers, IEEE Trans. Fuzzy Syst. 9 (2001), 810-821.

[19] C. C. Chuang, J. T. Jeng and C. W. Tao, Hybrid robust approach for TSK fuzzy modeling with outliers, Expert Systems with Application 36(5) (2009), 8925-8931.

[20] C. C. Chuang, S. F. Su and C. C. Hsiao, The annealing robust backpropagation (ARBP) learning algorithm, IEEE Trans. Neural Networks 11(5) (2000), 1067-1077.