A Rough-Set-Based for fuzzy modeling with outlier

Chih-Ching Hsiao
Department of Electrical Engineering, Kao Yuan University, Taiwan, R.O.C
(Tel : ( 886)-(7)-6077286; E-mail: cchsiao@cc.kyu.edu.tw)

Abstract: For high nonlinearly or unknown systems, the interest is toward data-driven methods for obtaining the system model. Fuzzy-rule-based modeling is a suitable tool that combines good approximation properties with a certain degree of inspects ability. The rough set theory is successes to deal with imprecise, incomplete or uncertain for information system. 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-based fuzzy C-regression model (RFCRM), that define fuzzy subspaces in a fuzzy regression manner and also include Rough-set theory for TSK modeling with robust capability against outliers.

Keywords: fuzzy modeling, outlier, rough set.

1. INTRODUCTION

The TSK type of fuzzy models proposed in [1,2] has attracted a great attention of the fuzzy modeling community due to their good performance in various applications. Various fuzzy modeling approaches are proposed in the literature and can be found in textbooks, such as [3–7]. 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) [7] are suitable to define fuzzy subspaces for TSK fuzzy modeling. Although FCM is a very useful clustering method, 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 [8]. In real-data analysis, noise and outliers are unavoidable. Outliers are observations that deviate significantly from majority of observations. They may be generated by a different mechanism corresponding to normal data and may be due to sensor noise, process disturbance, instrument degradation or human-related errors. It is futile to do data based analysis when data are contaminated with outlier because outliers van lead to model misspecification, biased parameter estimation and incorrect analysis results. FCRM clustering algorithm [7] 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. However, the obtained fuzzy rules can be considered as a rough approximation to the desired TSK fuzzy model. The model is further adjusted by supervised learning algorithms to improve the modeling accuracy. This stage is referred to as the fine-tuning process. Besides, the FCRM clustering algorithms is based on the principle of least square error minimization and is easily affected by outliers, which should be degraded in the clustering process.

The rough set theory proposed by Pawlak [9-10] is a mathematical theory dealing with uncertainty in data. Rough sets rely on the notion of lower and upper approximations of a set and are applied to rough fuzzy control [11-13], modeling, system identification [14] and discovery of discussion rules from experimental data[15-18]. In [19], Lingras and West introduced a new clustering method called rough c-means (RCM), which describes a cluster by a prototype (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. 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 captures the distinct aspect of indiscernibility in knowledge, while the latter describes the inherent feature of vagueness in linguistics and decision-making. Recently, combining both rough and fuzzy sets[20-26], Mitra [23] proposed a new c-means algorithm, where each cluster is consist of a fuzzy lower approximation and a fuzzy boundary. Each object in lower approximation takes a distinct weight, which is its fuzzy membership value. 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 [24] proposed a generalized hybrid algorithm, which is termed as rough–fuzzy PCM (RFPCM), 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, vagueness, and incompleteness in class definition. The algorithm attempts to exploit the benefits of both probabilistic and possibilistic membership functions. It will avoid the problems of noise sensitivity of the FCM. 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. There is includes two stages . First, is data-clustering stage , it set initial values for clustering by applying fuzzy c-means to only input space. Next is regression-clustering stage by viewing each rule as a rough set.
The remaining part of the paper is outlined as follows. Section 2 describes the Fuzzy C mean and rough set. In Section 3, the Rough-based Fuzzy C-Regression modeling Algorithm is proposed to meaningfully define a TSK fuzzy model. Simulation results are presented in Section 4. Concluding remarks are presented in section 5.

2. FUZZY C MEAN AND ROUGH SET

The FCM algorithm is suitable to define fuzzy subspaces for TSK fuzzy modeling and its cost function is defined as

$$J_u = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 (d_y^2)$$

subject to $\sum_{i=1}^{C} u_{ij} = 1$, for $1 \leq j \leq N$, \hspace{1cm} (1)

where $C$ and $N$ are the numbers of fuzzy clusters and of the input training data, respectively. $u_{ij}$ is the membership of the $i$-th cluster for the $j$-th training pattern. The Euclidean distance measure is used, $d_y$ be the distance between the $j$-th input data and the center of the $i$-th cluster.

$$d_y^2 = \left[\sum_{i=1}^{n} (\tilde{x}(j) - \tilde{\theta}_i^j)\Sigma^{-1}(\tilde{x}(j) - \tilde{\theta}_i^j)^T \right]$$ \hspace{1cm} (2)

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

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 (d_y^2) - \sum_{j=1}^{N} \lambda_j \left( \sum_{i=1}^{C} u_{ij} - 1 \right) \hspace{1cm} (3)$$

The rough set theory is a mathematical theory dealing with uncertainty in data. Rough sets rely on the notion of lower and upper approximations of a set. A rough set $X$ is characterized by its lower and upper approximations $\underline{B}X$ and $\bar{B}X$, respectively, with the following properties.

1) An object $x_1$ can be part of at most one lower approximation. 
2) If $x_1 \in \underline{B}X$ of cluster $X$, then simultaneously $x_1 \in \bar{B}X$. 
3) If $x_1$ is not a part of any lower approximation, then it belongs to two or more upper approximations. This permits overlaps between clusters.

3. ROUGH-BASED FUZZY C-REGRESSION MODELING ALGORITHM

A novel approach, termed as Rough-based Fuzzy C-Regression modeling Algorithm, is proposed, a modified version of FCRM [7]. In the RFCRM algorithm, it is integrated two contemporary progresses in analyzing vague data. One is the fuzzy set and the other is rough set. There is includes two stages. First, is data-clustering stage and the next is regression-clustering stage.

In the data-clustering stage, it set initial values for clustering by applying fuzzy c-means to only input space. Next, the regression-clustering stage is applied. The theory of rough sets [9-10] has recently emerged as another major mathematical tool for managing uncertainty that arises from granularity in the domain of discourse—that is, from the indiscernibility between objects in a set. The intention is to approximate a rough (imprecise) concept in the domain of discourse by a pair of exact concepts, called the lower and upper approximations. These exact concepts are determined by an indiscernibility relation in the domain, which, in turn, may be induced by a given set of attributes ascribed to the objects of the domain. 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. In the RFCRM algorithm, the concept of FCRM is extended by viewing each rule as a rough set [13].

The cost function in Eq.(1) is replaced $d_y^2$ with $r_y^2$, and rewritten as $J_u = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 (r_y^2)$ subject to $\sum_{j=1}^{N} u_{ij} = 1$, for $1 \leq j \leq N$, all symbols are similar to previously definition. $r_y$ be the residual between the $j$-th desired output of the modeled system and the output of the $i$-th rule with $j$-th input data; i.e.,

$$r_y = y_j - f_i(\tilde{x}(j); \bar{a}^i)$$ \hspace{1cm} (4)

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

$$\bar{a}^i = \left[ X^T D_i X \right]^{-1} X^T D_i Y, \hspace{1cm} i = 1, 2, ..., C.$$

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 $u_{ij}$ as its $k$-th diagonal element. Beside the condition under which an object may belong to the lower or upper bound of a rule. Let $\tilde{x}(j)$ be an object at residual $r_y$ 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 residual $r_y - r_y, i \neq k$, can be used to determine whether $\tilde{x}(j)$
should belong to the lower or upper approximations of the rules. The algorithm is outlined as follows.

1) Assign each data object \( \tilde{x}(j) \) to the lower approximation \( \bar{B}_i^k \) or upper approximation \( \bar{B}_i^k \), \( \bar{B}_i^k \) of rule pairs \( R_i \) and \( R_k \) by computing the difference in its difference \( r_y - r_j \) from the rule pairs \( i \)-th rule and \( k \)-th rule.

2) Let \( r_y \) be minimum and \( r_j \) be the next to minimum.

\[ \text{If } r_y - r_j \text{ is less than some threshold, then } \tilde{x}(j) \in \bar{B}_i^k \text{ and } \tilde{x}(j) \in \bar{B}_i^k \text{ and } \tilde{x}(j) \text{ cannot be a member of any lower approximation,} \]

\[ \text{else } \tilde{x}(j) \in \bar{B}_i^k \text{ such that residual } r_y \text{ is minimum over the C rules.} \]

3) Compute new parameter vector for each rules using Equ. (5).

4) Repeat Steps 1)–3) until convergence.

From above algorithm, we can adjust the lower approximation (or/and upper approximation) to weaken the outlier effect. Assume that Gaussian membership functions are used in the premise parts, (i.e., \( \mathcal{A}_i(\theta_j, \theta_r) = \exp \left( -\frac{(x_i - \theta_j)^2}{2\theta_r^2} \right) \), where \( \theta_j \) and \( \theta_r \) are two adjustable parameters of the \( l \)-th membership function of the \( i \)-th fuzzy rules. Then, we have two update equations as follows

\[ \theta_j = \frac{\sum x_i(k) u_i}{\sum u_i^2}, \quad \text{(6)} \]

\[ \theta_r = \frac{\sum u_i \left( x_i(k) - \theta_j \right)^2}{\sum u_i^2}. \quad \text{(7)} \]

The proposed RFCRM Algorithm is described in the following.

[Step 1] : Assign initial means for the \( C \) clusters by FCM.

[Step 2] : Assign each data object \( \tilde{x}(j) \) to the lower approximation \( \bar{B}_i^k \) or upper approximation \( \bar{B}_i^k \), \( \bar{B}_i^k \) of rule pairs \( R_i \) and \( R_k \) by computing the difference in its difference \( r_y - r_j \) from the rule pairs \( i \)-th rule and \( k \)-th rule by using Equ. (4).

[Step 3] : Let \( r_y \) be minimum and \( r_j \) be the next to minimum.

If \( r_y - r_j \) is less than some threshold, then \( \tilde{x}(j) \in \bar{B}_i^k \) and \( \tilde{x}(j) \in \bar{B}_i^k \) and \( \tilde{x}(j) \) cannot be a member of any lower approximation.

else \( \tilde{x}(j) \in \bar{B}_i^k \) such that residual \( r_y \) is minimum over the \( C \) rules..


[Step 5] : Update the center \( \theta_j \) and variance \( \theta_r \) by Equ. (6),(7).

[Step 6] : Repeat Steps 2)–5) until convergence.

The TSK fuzzy model obtained by the RFCRM algorithm has been with a level of accuracy. Furthermore, to improve the modeling accuracy, it should to adjust the parameters in fuzzy rules by the fine-tuning process. Hence a robust learning algorithm called ARBP [27] is employed to adjust these parameters of TSK fuzzy rules. In this algorithm, it is simply embedding an annealing process into the learning process, and a deterministic annealing process is used to replace the scale estimator. Based on this idea, the robust cost function of ARBP is defined as:

\[ E_{ARB} = \sum_{i=1}^{N} \sigma(e_i, \beta(\tau)), \quad \text{(8)} \]

where \( \sigma(\cdot) \) is the loss function. Thus the parameters of premise parts are updated as

\[ \Delta \theta_j = \eta \varphi(e_i, \beta) \left( y_i - \hat{y}_i \right) \frac{1}{w} \sum \frac{\partial \sigma}{\partial e_i}, \quad \text{(9)} \]

where \( e_i = y_i - \hat{y}_i \), \( \varphi(\cdot) \) is the derivative of \( \sigma(\cdot) \) (i.e. \( \varphi = \frac{d}{de_i} \)). The parameters of consequent parts are updated as

\[ \Delta \theta_j = \zeta \varphi(e_i, \beta) \frac{w x_j}{\sum_{i=1}^{N} \frac{\partial \sigma}{\partial e_i}}, \quad \text{(10)} \]

where \( \zeta \) is the learning constant.

4. SIMULATION EXAMPLE

The \( \text{sinc} \) function is considered. \( \text{sinc}(x) \) is defined as

\[ y = \frac{\sin(x)}{x}, \quad -10 \leq x \leq 10. \]

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 \( H \sim N(0,1) \). Based on the above procedure, there are some results for approximating the sinc function. The FCRM and RFCRM have similar results as Fig. 1. When add the gross error model that is used for modeling outliers, the approximated results of each approach are presented in Fig. 2. It can show that the RFCRM is better than the FCRM. Since RFCRM can weaken the outlier effect. After Fine-tuning, the RFCRM also has the better simulated results that are shown in Fig. 3 and Fig. 4.

5. CONCLUSION

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-based fuzzy C-regression model (RFCRM), that define fuzzy subspaces in a fuzzy regression manner and also include Rough-set theory for TSK modeling with robust capability against outliers. In the simulation results, it can show that the RFCRM is better than the FCRM. Since RFCRM can weaken the outlier effect. After Fine-tuning, the RFCRM also has the better
REFERENCES


