Control and Cybernetics

vol. **25** (1996) No. 2

A best-fit based fuzzy forecasting algorithm: principle and formulation

by

Wai Man Tse* and Kwan Fai Cheung**

*HSBC Asset Management,
The Hong Kong University of Science and Technology
Hong Kong

** The Hong Kong University of Science and Technology,
Hong Kong

Abstract: This paper introduces a global optimization technique based on fuzzy c-means method to partition the output space with respect to a set of input/output quantitative data. The technique can identify the underlying data structure. And it can account for outliers. This coupled with the induced-input-partition preserving map result in a superb forecasting methodology.

1. Introduction

In this paper we introduce a forecasting technique utilizing the notion of global optimization meant for the definition of the input-output membership functions with respect to a set of input-output data. Sugeno (Sugeno and Yasukawa, 1993) studied extensively methods of identification of the structure of fuzzy model. His objective was to extract qualitative structure from quantitative data. By using "fuzzy c-means (FCM)" to partition output space, local clusters were formed. The input space partitions were directly inferred from the output space partition. The multi-dimensional input membership grade values were projected onto individual variable subspaces. The projected values in every subspace were then fitted by a collection of trapezoidal functions. Linguistic labels were then attached to each of this input/output membership function. Sugeno's methodology can arrive at a qualitative model characterized by rules being linguistic statements.

While Sugeno's method of inferring qualitative structure from quantitative data is appealing, the method cannot guarantee good fit between the model and the training data. Also, the constraint that sum of membership values be equal one is not addressed. Indeed, the output space partition by minimizing the data-cluster centroid distances may not result in an optimal model-data good fit.

This paper presents an extension of the work of our previous paper (Tse and Cheung, 1994) presented in IIZUKA'94. In this paper, we address these issues by imposing the requirement that the system's output matches that of the training data. The method is akin to FCM formulation. This paper is organized as follows: the next section will discuss the methodological formulation, section 3 presents the results and the concluding remarks appear in section 4.

2. Methodology

A useful forecasting structure is a result of the correct choice of estimation technology. This choice will govern the model-data parametric identification. Contingent upon the estimation technology but beyond its scope is structural identification which defines the candidates to be introduced into the model input-output relationship. Similar to Sugeno's, we infer input space partition from that of the output space. A parametric estimation technique using global optimization is introduced for the output space partition. The discussion of input space structural identification will follow. But we first discuss the optimization technology in the next section.

2.1. A parametric estimation technique

Consider a M-input-N-output fuzzy logic system (FLS). Given the values $\{\beta_i\}$ and $\{u_{it}\}$, the output value of the system y_t^* is calculated using center of gravity (COG) method where β_i is the centroid of the ith cluster and u_{it} is the membership grade of the tth data point in the ith cluster

$$y_t^* = \frac{\sum_{i=1}^{c} u_{it}^m \beta_i}{\sum_{i=1}^{c} u_{it}^m};\ 1 \leq t \leq n \tag{1}$$

Here, c and m are design parameters. In particular, c is used to determine the number of rules embedded in the system and m is used to determine the shapes of the cluster distributions. The distance from the tth data point to the ith cluster center is defined by

$$d_{it} = \|y_t - \overline{y}_i\|_2 \tag{2}$$

Our goal is to compute β_is such that the overall distance between y_t and y_t^* is minimized. Similarly to FCM formulation, the objective function is as follows:

$$\min \sum_{t=1}^{n} \|y_t - y_t^*\|^2;\ 1 \leq i \leq c, 1 \leq t \leq n \tag{3}$$

subject to

$$u_{1t} = \frac{1}{1 + \sum_{i=2}^{c} (\frac{d_{1t}}{d_{it}})^{\frac{2}{m-1}}}; \tag{4}$$

$$u_{it} = u_{1t}(\frac{d_{1t}}{d_{it}})^{\frac{2}{m-1}} \tag{5}$$

Here, (4) and (5) are constraints which lead to

$$\sum_{i=1}^{c} u_{it} = 1; \; 1 \leq t \leq n$$

Minimization of the sum of errors norm in (3) corresponds to solving of the matrix equation:

$$\mathbf{AY} = \mathbf{B} \tag{6}$$

where $\mathbf{A} = [a_{ij}]_{cxc}, \mathbf{Y} = [\beta_i]_{cx1}$ and $\mathbf{B} = [b_i]_{cx1}$. In particular,

$$a_{ij} = \sum_{t=1}^{n} \frac{u_{it}^m u_{jt}^m}{(\sum_{k=1}^{c} u_{kt}^m)^2}; 1 \leq i,j \leq c, 1 \leq t \leq n$$

$$b_j = \sum_{t=1}^{n} \frac{y_t u_{it}^m}{\sum_{k=1}^{c} u_{kt}^m}$$

Equation (6) can either be solved in one step, or recursively. The recursive algorithm is as follows:

- Step One: Initialize the variables $(\{\beta_i\}, \{u_{it}\})$.
- Step Two: Compute cluster centroids from equation 6.
- Step Three: Compute membership grades $\{u_{it}^P\}$ using constraint equations (4) and (5).
- Step Four: Calculate $\epsilon = \sum_i \sum_t |u_{it}^P - u_{it}|$.
- Step Five: If $\epsilon < TOL$ (set to 0.0001), then stop. Else, assign u_{it}^P to u_{it} and repeat steps two to five.

u_{it}^P denotes the updated u_{it}.

2.2. Structural identification

We impose the output space partition parameters: $(c, m, \{u_{it}\})$ on the input space inducing a partition by FCM. We replace the distance norm in (2) by an inner product induced norm metric of the form

$$d_{it} = (x_t - \overline{x}_i)^T M_i (x_t - \overline{x}_i); \; 1 \leq i \leq c, \; 1 \leq t \leq n$$

with M_i symmetric and positive-definite and $\{x_t\}_{k\times 1}, \{\overline{x}_i\}_{k\times 1}$. The centroids $\{\overline{x}_i\}_{1\leq i \leq c}$ and the membership grades are generated from equations (1), (4) and (5). We calculate the scaling matrix M_i from a fuzzy covariance matrix (Gustafson and Kessel, 1979)

$$P_{fi} = \frac{\sum_{t=1}^{n} u_{it}^m (x_t - \overline{x}_{fi})(x_t - \overline{x}_{fi})^T}{\sum_{t=1}^{n} u_{it}^m}; \ 1 \le i \le c$$

The scaling matrix can then be obtained from

$$M_i^{*-1} = (\frac{1}{\rho_i |P_{fi}|})^{\frac{1}{k}} P_{fi}$$

We set

$$\rho_1 = 1 \tag{7}$$

$$\rho_i = \frac{1}{n} \sum_{t=1}^{n} \{ (\frac{u_{1t}}{u_{it}})^{(m-1)/2} \frac{(x_t - \overline{x_1})^T |P_{f1}|^{1/k} P_{f1}^{-1} (x_t - \overline{x_1})}{(x_t - \overline{x_i})^T |P_{fi}|^{1/k} P_{fi}^{-1} (x_t - \overline{x_i})} \}^k \tag{8}$$

for $i = 2, \cdots, c$ to generate an induced-input partition preserving map, which minimizes iteration-led shocks. The output space membership grades will be perfectly preserved if all the individual summed terms in (8) are of identical values. To identify the input structure, we undertake the same evaluation criterion as Sugeno's. The regularity criterion is defined as follows:

$$RC = \frac{1}{2} [\sum_{t=1}^{n_A} \frac{(y_t^A - y_t^{AB})^2}{n_A} + \sum_{t=1}^{n_B} \frac{(y_t^B - y_t^{BA})^2}{n_B}] \tag{9}$$

where

n_A and n_B : the number of data of the groups A and B;
y_t^A and y_t^B : the model output data of the groups A and B;
y_t^{AB} : the model output for the group A input estimated by the model identified using the group B data;
y_t^{BA} : the model output for the group B input estimated by the model identified using the group A data.

The above construction helps to identify the input elements compatible with the output partition structure.

2.3. Forecasting

Having identified the parameters and the structure of the system, the formulation of the forecasting strategy is a natural and easy outcome. Applying FCM on $(c, m, \{\overline{x}_i\}_{1 \le i \le c})$ using the new input data x_N for $N > n$, a unique set of membership grades $\{u_{iN}\}_{1 \le i \le c}$ is produced. The predicted output value can then be constructed as:

$$y_N^* = \frac{\sum_{i=1}^{c} u_{iN}^m \beta_{bi}}{\sum_{i=1}^{c} u_{iN}^m} \tag{10}$$

where β_{bi} is the centroid of the ith cluster calculated from data points over $1 \le t \le n$.

	c_{ji}			d_i
i/j	1	2	3	
1	0.7	1.6	2.3	-0.8
2	0.5	1.3	1.7	0.9
3	0.2	1.4	2.6	1.8

Table 1.

3. Results

3.1. Data generation

Our testing data were generated by a three input-variable and three-rule FLS. The defuzzified outputs are calculated by equation 1. The tth input data point of the jth variable $x_{jt} = (x_{jit})_{1 \le i \le 3}$ was generated by

$$x_{jit} = (c_{ji} - a_i) + 2a_i rand(n_{it}); 1 \le i, j \le 3 \tag{11}$$

where $\sum_i n_{it} = n$ and the random numbers are generated uniformly in $[-1, 1]$. The ith input membership functions of the variable x_{jt} is

$$f(x_{jit}) = c_{ji} + a_i(x_{jit} - c_{ji});\ 1 \le i, j \le 3$$

The truth value of the antecedent of every ith rule is obtained by multiplications:

$$u_{it} = \prod_{j=1}^{3} f(x_{jit});\ 1 \le i \le 3,\ 1 \le t \le n$$

And the output value y_t is calculated using COG formula in (1) where d_i is the centroid of the ith cluster in the output space generated by the testing data. Table 1 lists the parameter values used to produce our data.

The generated data series is plotted in Figure 1. Distinct outliers occur at $y_6 = 4.7664$ and $y_{41} = 6.7727$. Less distinct outliers occur at $y_{16} = -2.4620, y_{24} = -2.5718$ and $y_{77} = -1.9923$.

3.2. Fuzzy logic optimization

First we want to explore the ability of the global optimization technique to identify the number of rules inherent in the data set. Then we will examine the goodness of fit of the technique compared to the FCM. We choose the fine-tuning parameter m to be 3. In Table 2 we tabulate c versus R^2 (given $m = 3$).[1] It shows that R^2 increases to a maximum at $c = 7$. When $c > 8$, **A**

[1] $R^2 = \frac{\sum_t (y_t^* - \overline{y})^2}{\sum_t (y_t - \overline{y})^2}$ where $\overline{y} = \frac{\sum_t y_t}{\sum_t 1}$.

c	R^2
1	0.0000
2	0.5936
3	0.8670
4	0.9482
5	0.9782
6	0.9799
7	0.9883
8	0.9879

(a) ($m = 3$)

m	R^2
1.2	0.9706
1.5	0.9701
1.8	0.9721
2.1	0.9903
2.3	0.9904
2.4	0.9905
2.5	0.9872
3.0	0.9879

(b) ($c = 8$)

Table 2.

is getting ill-posed when ϵ becomes small. This indicates that some cluster centroids overlap. The eight cluster centers are $(\beta_i)_{1 \leq i \leq 8}$ = (**-2.3504**, -0.1848, 0.2045, 0.5029, 0.9437, 2.1672, **4.7674**, **6.7728**). Clearly the three highlighted centers account for outliers. Removing these three centers yields the centers: $(-0.1848, 0.2045, 0.5029, 0.9437, 2.1672)$. Hence, the three additional rules assigned to outliers account for a much improved model-data goodness of fit. By setting $c = 8$, the technique is tested on various m's. The relationship between R^2 and m is tabulated in Table 3. It shows that R^2 increases to a maximum at $m = 2.4$. When m increases, the distributions around centroids spread out. There exists a critical value m beyond which the clusters overlap so much that the prespecified c centroids cannot be all identified. In our case, the matrix $\mathbf{A}$ becomes ill-posed when m exceeds 3. A plot of the data and predicted values series is illustrated in Figure 2 with $m = 2.4$ and $c = 8$. Clearly, the fitting is excellent.

A corresponding plot is shown in Figure 3. It plots the data and predicted series using the FCM. The R^2 is 0.9595. The comparison of the figures delineates the superiority of the global optimization technique.

3.3. Structural identification

The identified parameters $(c, m, \{u_{it}\}_{c \times n})$ are carried forward from output space to input space. We first partition the data set $(\{x\}_{k \times n}, \{u\}_{c \times n})$ randomly into two data series. And then we apply the regularity criterion. To examine the identification power of the regularity test and our fuzzy logic procedure, we add in two new variables $\{x_{4t}\}_{1 \times n}$ and $\{x_{5t}\}_{1 \times n}$ also generated by (11) with $c_{ji} = a_i, 1 \leq i, j \leq 3$. The results of forward and backward tree searches are contained in Table 4.

The forward and backward tree searches confirm that (x_1, x_2, x_3) is compatible to the model structure inferred from output space using our global opti-

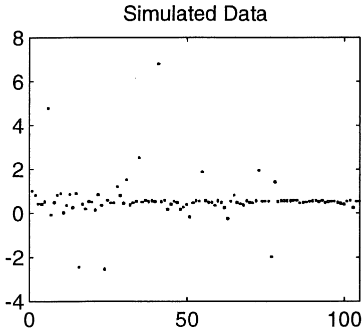

Figure 1

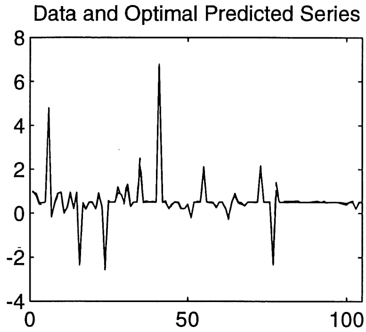

Figure 2

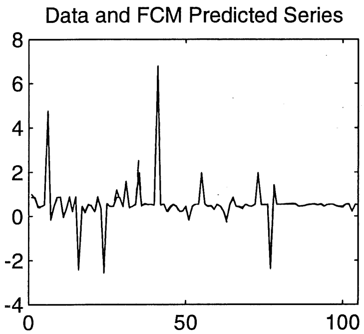

Figure 3

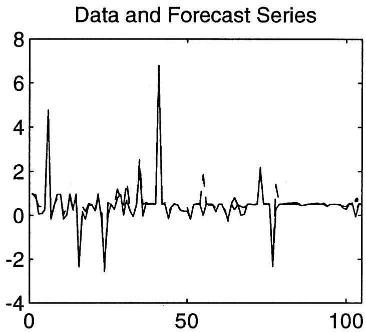

Figure 4

mization technique.

3.4. Forecasting

To understand the predictive power of our technology, we re-insert our input data $\{x\}_{k\times n}$ into the forecasting structure discussed previously. We tabulate the results in Figure 4. $R^2 = 1.0178$. It exceeds one because the mean of $\{y_t^*\}_{t=n+1}^{2n}$ (=0.4475)is different from the sample mean, i.e. the mean of $\{y_t\}_{t=1}^{n}$ (=0.5596), and $\sum_t y_t^*(y_t^* - y_t)$ $(= -1.2840)$ is not zero.

An ordinary least squares (OLS) regression is applied to the data series. By construction, the mean of the OLS predicted values $\{\hat{y}_t\}_{t=1}^{n}$ coincides with the sample mean and $\sum_t \hat{y}_t(\hat{y}_t - y_t) = 0$. The equation of the predicted series is

$$\hat{y} = 0.6734 - 0.1465x_1 - 0.0233x_2 + 0.1032x_3.$$

$R^2 = 0.0222$. The actual and the OLS predicted series appear in Figure 5. And the absolute residual errors, respectively, from our optimization method and the OLS method appear in Figure 6. Obviously, our forecasting series has a significantly superior predictive power.

Data and OLS Series

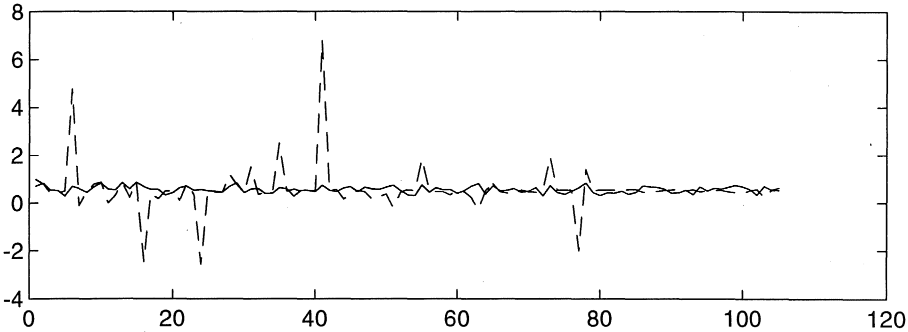

Figure 5

Absolute Residual Errors

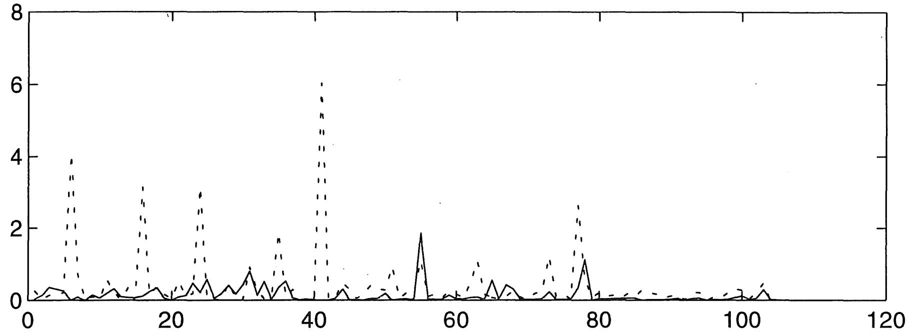

Figure 6

Forward Search		Backward Search	
Input Variables	RC	Input variables	RC
x_1	2.6131	$x_1 - x_2 - x_3 - x_4 - x_5$	**0.0591**
x_2	**1.5296**		
x_3	1.8502	$x_1 - x_2 - x_3 - x_4$	0.6600
x_4	2.2984	$x_1 - x_2 - x_3 - x_5$	**0.2183**
x_5	2.5432	$x_1 - x_3 - x_4 - x_5$	0.2960
		$x_2 - x_3 - x_4 - x_5$	0.2559
$x_1 - x_2$	0.6794	$x_1 - x_2 - x_4 - x_5$	0.3336
$x_2 - x_3$	**0.6584**		
$x_2 - x_4$	0.7711	$x_1 - x_2 - x_3$	**0.0561***
$x_2 - x_5$	0.7248	$x_1 - x_2 - x_5$	0.5154
		$x_1 - x_3 - x_5$	0.1350
$x_1 - x_2 - x_3$	**0.0561***	$x_2 - x_3 - x_5$	0.4428
$x_2 - x_3 - x_4$	0.4018		
$x_2 - x_3 - x_5$	0.4428	$x_1 - x_2$	0.6794
		$x_1 - x_3$	1.1443
$x_1 - x_2 - x_3 - x_4$	0.6600	$x_2 - x_3$	**0.6584**
$x_1 - x_2 - x_3 - x_5$	**0.2183**		
		x_2	**1.5296**
$x_1 - x_2 - x_3 - x_4 - x_5$	**0.0591**	x_3	1.8502

Table 3.

4. Conclusions

Sugeno (1993) provided a useful framework for identification of qualitative and quantitative structures of a set of quantitative data. However, the output space was not partitioned to make the model-data goodness of fit optimal. In this paper, we introduce a technique to ensure optimal fit of output data. Simulation results show that the technique can identify well the qualitative structure inherent in the data. Moreover, the technique can take special care of explanation of the outlier behavior.

References

GUSTAFSON, D.E., and KESSEL, W.C. (1979) Fuzzy Clustering with a Fuzzy Covariance Matrix. *Proceedings IEEE CDC*, San Diego, CA, 761-766.

SUGENO, M., and YASUKAWA, T. (1993) A Fuzzy-Logic-Based Approach to Qualitative Modeling. *IEEE Trans. Fuzzy Systems*, **1**, 1, 7-31.

TSE, W.M., and CHEUNG, K.F. (1994) A Parametric Fuzzy Logic Estimation Technique. *Proceedings IIZUKA'94*, 617-618.