Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

A SYSTEMATIC APPROACH TO LINGUISTIC FUZZY MODELING BASED ON INPUT-OUTPUT DATA

Hossein Salehfar

Department of Electrical Engineering
Campus Box 7165
University of North Dakota
Grand Forks, ND 58202, U.S.A.

Nagy Bengiamin

Department of Electrical and Computer Engineering
Campus Box EE94
California State University - Fresno
Fresno, CA 93740-8030, U.S.A.

Jun Huang

Department of Electrical Engineering
Campus Box 7165
University of North Dakota
Grand Forks, ND 58202, U.S.A.

ABSTRACT

A new systematic algorithm to build adaptive linguistic fuzzy models directly from input-output data is presented in this paper. Based on clustering and projection in the input and output spaces, significant inputs are selected, the number of clusters is determined, rules are generated automatically, and a linguistic fuzzy model is constructed. Then, using a simplified fuzzy reasoning mechanism, the Back-Propagation (BP) and Least Mean Squared (LMS) algorithms are implemented to tune the parameters of the membership functions. Compared to other algorithms, the new algorithm is both computationally and conceptually simple. The new algorithm is called the Linguistic Fuzzy Inference (LFI) model.

1 INTRODUCTION

Fuzzy logic modeling techniques can be classified into three categories, namely the linguistic (Mamdani-type), the relational equation, and the Takagi, Sugeno and Kang (TSK). In linguistic models, both the antecedent and the consequence are fuzzy sets while in the TSK model the antecedent consists of fuzzy sets but the consequence is made up of linear equations. Fuzzy relational equation models aim at building the fuzzy relation matrices according to the input-output process data.

Based on the TSK model, an Adaptive Network based Fuzzy Inference System (ANFIS) has been introduced by Jang (Jang 1993). This model is mostly suited to the modeling of nonlinear systems. It combines the recursive least-square estimation and the steepest descent algorithms for calibrating both premise and consequent parameters iteratively. This algorithm is limited in incorporating human knowledge. In contrast, Linguistic fuzzy models are effective in embedding the human knowledge and have simpler forms. Systematic approaches to building linguistic fuzzy models are proposed in (Emami 1998, Sugeno. 1993). These approaches, however, involve nonlinear programming and are computationally cumbersome. This paper addresses this problem and proposes a new systematic and simple algorithm to build and tune models directly from the input-output data. Like ANFIS (Jang 1993) the new algorithm takes advantage of Neural Networks training techniques and it uses projection methods (Emami 1998, Sugeno 1993) to build the fuzzy rules. The new algorithm consists of two procedures. The first one is for fuzzy structure identification, in which the inputs, membership functions and fuzzy rules are determined. The second one is for fuzzy parameter identification, in which training algorithms are used to tune the parameters of the membership functions.

2 GENERAL LINGUISTIC FUZZY MODEL

The general Linguistic Fuzzy Model of a Multi-Input Single-Output (MISO) system is interpreted by rules with multi-antecedent and single-consequent variables such as the following:

Rule l: IF U_1 is B_{l1} AND U_2 is B_{l2} AND U_r is B_{lr}
THEN V is D_l , $l = 1,2,\ldots,n$ (1)

Where U_1 , U_2 ,……, U_r are input variables and V is the output, B_{ij} (i=1, …..,n, j=1,…..,r) and Di (i=1, …..,n) are fuzzy sets of the universes of discourse $X_1, X_2,\dots, X_r$, and Y of U_1 , U_2 ,……, U_r and V respectively. The above rule can be interpreted as a fuzzy implication relation

$B_l = B_{l1} \text{ x } B_{l2} \text{ x}\dots\text{x } B_{lr} \rightarrow D_l$ in $(X = X_1 \text{ x } X_2 \text{ x}\dots\text{x } X_r)\text{x}Y$:

$$R_l(x,y) = T(B_l(x), D_l(y)),\; B_l(x) = T'(B_{l1}(x), B_{l2}(x), \cdots B_{lr}(x)) \tag{2}$$

Where T and T' are the t-norm operators and may be different from each other.

Let the fuzzy set A in the universe of discourse X be the input to the fuzzy system of (1). Then, each fuzzy IF-THEN rule determines a fuzzy set F_l in Y:

$$F_l\ (y) = T(R_l(x,y), A(x)) \tag{3}$$

For a crisp input $x^* = (x_1^*, x_2^*, \cdots, x_r^*)$,

$$A_i(x) = \begin{cases} 1, & \text{if } x_i = x_i^* \\ 0, & \text{if } x_i \neq x_i^* \end{cases} \tag{4}$$

Then

$$\begin{aligned} F_l\ (y) &= T(R_l(x,y), A(x)) \\ &= T(B_l(x), A(x), D_l(y)) \\ &= T(B_l(x^*), D_l(y)) \end{aligned} \tag{5}$$

where $B_l(x^*)$ is called the Degree Of Firing (DOF) of rule l:

$$B_l(x^*) = T'(B_{l1}(x_1^*), B_{l2}(x_2^*), \cdots, B_{lr}(x_r^*)) \tag{6}$$

The output fuzzy set F of the fuzzy system is the t-conorm of the n fuzzy sets $F_l\ (l = 1,2,\cdots n)$:

$$F(y) = S[F_1\ (y), F_2\ (y), \cdots F_n\ (y)] \tag{7}$$

Where, S denotes the t-conorm operator. To obtain a crisp value of the output, the commonly used Center of Area (COA) method, may be used.

$$y^* = \frac{\int_{y_0}^{y_1} yF(y)dy}{\int_{y_0}^{y_1} F(y)dy} \tag{8}$$

Where, the real interval $Y = [y_0, y_1]$ is the universe of discourse for the output.

The fuzzy system is usually not analytical, but analytical formulation is essential for the use of training algorithms like BP and LMS. We, therefore, use the following simplified fuzzy inference engine: First, T-norm and T-conorm operators are chosen to be the multiplication and addition operators, respectively. Then equation (7) becomes,

$$F(y) = \sum_{l=1}^{n} F_l(y) = \sum_{l=1}^{n} B_l(x^*) \cdot D_l(y) \tag{9}$$

Obviously, the summation brings the output fuzzy set F(y) out of the unit interval. However, it doesn't have an effect on the defuzzified value. By substituting for F(y) in (8) we get the COA defuzzified value:

$$y^* = \frac{\int_{y_0}^{y_1} y \sum_{l=1}^{n} B_l(x^*) D_l(y) dy}{\int_{y_0}^{y_1} \sum_{l=1}^{n} B_l(x^*) D_l(y) dy} = \frac{\sum_{l=1}^{n} B_l(x^*) \left\{ \int_{y_0}^{y_1} y D_l(y) dy \middle/ \int_{y_0}^{y_1} D_l(y) dy \right\}}{\sum_{l=1}^{n} B_l(x^*)} = \frac{\sum_{l=1}^{n} B_l(x^*) y_l^*}{\sum_{l=1}^{n} B_l(x^*)} \tag{10}$$

Where the y_l^*'s are the centroids of the fuzzy sets D_l .

The defuzzified value y^* is determined by the weighted average of the centroids of the individual consequent fuzzy sets [2]. Using a symmetric triangular membership function, the fuzzy system becomes,

$$y^* = f(x^*) = \frac{\sum_{l=1}^{n} y_l^* \left(\prod_{i=1}^{r} 1 - \frac{|x_i - c_{li}|}{b_{li}}\right)}{\sum_{l=1}^{n} \left(\prod_{i=1}^{r} 1 - \frac{|x_i - c_{li}|}{b_{li}}\right)},\; c_{li} - b_{li} \le x_i \le c_{li} + b_{li} \tag{11}$$

Where c_{li} and b_{li} are the center and the half-width of the triangular membership function respectively.

3 CLUSTERING AND FUZZY STRUCTURE IDENTIFICATION

The essence of the fuzzy structure identification method is in the clustering and the projection (Emami 1998). First, the output space is partitioned using a fuzzy clustering algorithm like Fuzzy C-Mean clustering. Second, the partitions (clusters) are projected onto the space of the input variables. The output partition and its corresponding input partitions are the consequents and antecedents, respectively.

3.1 Clustering

Fuzzy C-mean (FCM) clustering method clusters the data by minimizing the total "distance" of each data point to the cluster centers. For detailed, the reader is referred to

(Emami 1998). A critical problem for the FCM algorithm is how to determine the optimal number of clusters. Xie-Beni index (or S-function) (Xie 1991), computed as the ratio of compactness and separation of clusters, can be used as a validity measure. If the smallest S is found, the optimal number of clusters is then determined. However, it is intuitive that more clusters and more rules usually provide a more accurate model. Therefore, the optimal number of clusters determined by the smallest S only gives the minimum number of clusters in order to achieve an acceptable model.

The weighting exponent "m" controls the extent of membership sharing between fuzzy clusters in the data set. The larger the m, the greater the extent of membership sharing between fuzzy clusters. A general rule as introduced in (Emami 1998) suggests that m should be far enough from both of its limits, one and infinity.

Another clustering method "Subtractive Clustering", which can automatically determine the number of clusters, is also used in the proposed algorithm of this paper. For details on subtractive clustering, the reader is referred to (Chiu 1994).

3.2 Input Selection

Since the fuzzy rules are constructed by clustering and projection procedures, a simple and effective method to determine the significant inputs can evolve (Emami 1998). The so-called non-significance index π_j, defined for each input variable x_j, is used.

$$\pi_j = \prod_{i=1}^{n} \frac{\Gamma_{ij}}{\Gamma_j}, \qquad j = 1,2,\cdots,r_0 \tag{12}$$

Where Γ_{ij} is the range of x_j in which its membership function $B_{ij}(x_j)$ is one in the i[th] partition (or cluster), Γ_j is the entire range of the variable x_j, n is the number of rules, and r_0 is the number of input variables. The smaller the non-significance index, the more significant the corresponding input variable. However, the non-significance index is usually too small to be a good index. The reason is that the range in which the membership is one is typically very small. This can be improved by calculating Γ_{ij} as the range of clusters as is the case in the new proposed algorithm. If an input variable is not significant, the clusters are intended to evenly cover the whole range of this input variable, which makes its non-significance index close to one. Therefore, by removing the input variables whose non-significance indices is close to one or greater than that of the others, one can determine the significant input variables.

3.3 Membership Function Assignment and Rule Generation

First, partition the output space using the FCM clustering method. The number of clusters is determined by calculating the Xie-Beni index. Then by projecting each cluster onto each input variable we get temporary clusters in the input space. However, these temporary clusters are probably not well formed and should be clustered again into several sub-clusters. This is implemented by using the subtractive clustering method that automatically determines the number of sub-clusters in each temporary cluster.

After obtaining all the sub-clusters in all input variables and all clusters in the output space, the next step is to assign a membership function to each cluster. First the data points whose membership grades are among the highest are chosen. The mid-point of these data points is assigned grade of one, which is the vertex of the membership function. Then a membership grade C (0<C<1) is assigned to the points at the edge of the cluster. The membership function is shown in the Figure 1, where c_{li} and b_{li} are the center and the half-width of the membership function respectively. And x is the average distance of the vertex to the left and the right edges. Thus, we have:

$$\frac{x}{b_{li}} = \frac{1-C}{1} \Rightarrow b_{li} = \frac{x}{1-C} \tag{13}$$

C is a parameter to be assigned. This C is usually determined by experience, although some optimization techniques may be used. Typical values of CM vary from 0.5 to 0.8.

After partitioning the input and output spaces and assigning the membership functions, the next step is to construct the rules.

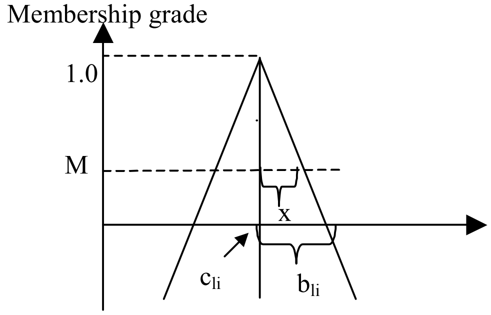

Figure 1: The Triangular Membership Function

4 PARAMETER IDENTIFICATION

This procedure optimizes the parameters of the membership functions of the model to minimize the performance index:

$$PI = RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (14)$$

Where, $\hat{y}_i$ is the model's output, y_i is the real output, and N is the number of data points.

4.1 Back-Propagation (BP)

The fuzzy system can be represented by a three-layer feed-forward network as shown in Figure 2, where the membership functions $B_{li}, l = 1,\cdots n; i = 1,\cdots r$ are triangular functions,

$$B_{li}(x_i) = \begin{cases} 1 - \dfrac{|x_i - c_{li}|}{b_{li}}, c_{li} - b_{li} \le x_i \le c_{li} + b_{li} \\ 0 \qquad , otherwise \end{cases} , l = 1,\cdots n; i = 1,\cdots r$$

(15)

Other variables are:

$$z_l = \prod_{i=1}^{r} B_{li}(x_i);\ a = \sum_{l=1}^{n}(\bar{y}_l z_l);\ b = \sum_{l=1}^{n} z_l;\ f = a/b$$

The back-propagation algorithm may be used to train this system. Suppose that we are given an input-output pair $(x^k, d^k), x^k = x_1^k, x_2^k, \cdots x_r^k$, our aim is then to determine a fuzzy logic system such that:

$$e^k = \frac{1}{2}\left[f(x^k) - d^k\right]^2 \quad (16)$$

is minimized.

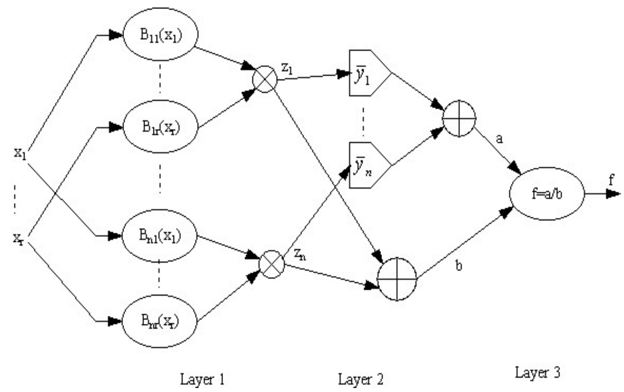

Figure 2: Three-Layer Feed-Forward Network Representation of Fuzzy System

Three parameters, $\bar{y}_l, c_{li}, b_{li}$, need to be adjusted. The training procedure for $\bar{y}_l$ is:

$$\bar{y}_l(k+1) = \bar{y}_l(k) - \alpha \frac{\partial e}{\partial \bar{y}_l}\Big|k \quad (17)$$

$$\frac{\partial e}{\partial \bar{y}_l} = (f-d)\frac{\partial f}{\partial a}\frac{\partial a}{\partial \bar{y}_l} = (f-d)\frac{1}{b}z_l \quad (18)$$

Where α is a step size.

To train c_{li}, the following adaptive rules are used:

$$c_{li}(k+1) = c_{li}(k) - \alpha \frac{\partial e}{\partial c_{li}}\Big|k \quad (19)$$

$$\frac{\partial e}{\partial c_{li}} = (f-d)\frac{\partial f}{\partial z_l}\frac{\partial z_l}{\partial c_{li}} = \begin{cases} (f-d)\dfrac{\bar{y}_l - f}{b}\dfrac{1}{b_{li}}\dfrac{z_l}{B_{li}(x_i)}, & c_{li} + b_{li} \ge x_i > c_{li} \\ (f-d)\dfrac{\bar{y}_l - f}{b}\dfrac{-1}{b_{li}}\dfrac{z_l}{B_{li}(x_i)}, & c_{li} - b_{li} \le x_i < c_i \\ 0 & , otherwise \end{cases}$$

(20)

To train b_{li}, the following adaptive rule are used:

$$b_{li}(k+1) = b_{li}(k) - \alpha \frac{\partial e}{\partial b_{li}}\Big|k \quad (21)$$

$$\frac{\partial e}{\partial b_{li}} = (f-d)\frac{\partial f}{\partial z_l}\frac{\partial z_l}{\partial b_{li}} = (f-d)\frac{\bar{y}_l - f}{b}\frac{|x_i - c_{li}|}{b_{li}^{\ 2}}\frac{z_l}{B_{li}(x_i)}$$

(22)

4.2 Least Mean Squared (LMS)

In fuzzy systems, if the antecedent fuzzy sets B_{li} (l=1,2,…,n; I=1,2,…r) are known (the type of membership functions and their parameters are determined), the normalized DOF v_l are also known,

$$v_l = \frac{\prod_{i=1}^{r} B_{li}(x_i)}{\sum_{l=1}^{n} \prod_{i=1}^{r} B_{li}(x_i)} \qquad (23)$$

then

$$y^* = f(x^*) = \sum_{l=1}^{n} y_l^* v_l \qquad (24)$$

To minimize the error (16), the Least Mean Squared (LMS) algorithm that recursively updates the values of the parameters y_l^* in the direction of greatest decrease of the error e is used:

$$y_l^*(k+1) = y_l^*(k) - \alpha \frac{\partial e^k}{\partial y_l^*} = y_l^*(k) - \alpha\, v_l e^k \qquad (25)$$

where α is the step size.

5 TEST RESULTS

To illustrate the validity of the proposed algorithm, three functions are tested. Due to its highly variable characteristics, the *Sinc* function is a typical benchmark for identification (Chen 1999). The second test function is a two-dimensional nonlinear static map, which has been studied in (Emami 1998, Sugeno 1993). The third one is the Mackey-Glass chaotic time-series generated by an underlying nonlinear dynamic system, which has been studied in (Jang 1993, Wang 1992). Due to lack of space, we present the results of the *Sinc* function only.

For convenience, we give the following definitions: The ANFIS method combined with the grid partition method is called ANFIS-GRID. The ANFIS method combined with the subtractive clustering method is called ANFIS-SUB. The method in (Chen 1999) is called ANFIS-EFCM.

The 1-D *sinc* function is defined as:

$$y = Sinc(x) = \frac{\sin(x)}{x}, x \in [-10, 10] \qquad (26)$$

As in (Chen 1999), 121 data points are uniformly sampled between [-10,10]. By calculating the Xie-Beni index, the optimal number of clusters is found to be 2. CM is set to 0.62 and four rules are built. The final rule base of the system is:

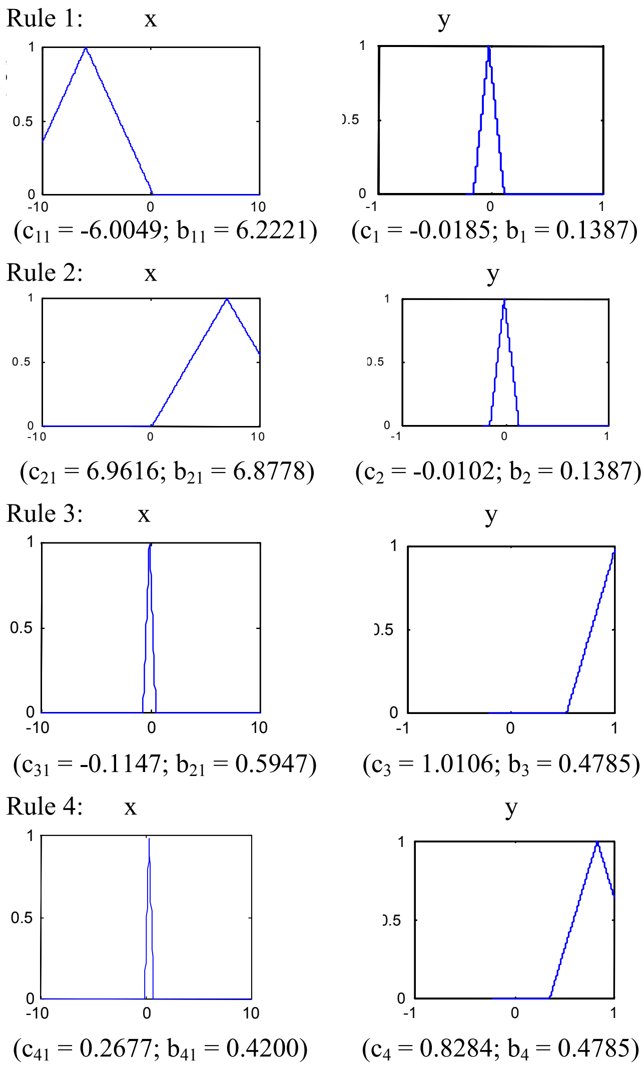

Figure 3: The Final Rule Base of the *Sinc* Function

The proposed algorithms (LFI) produced a performace index of 0.0735. Figure 4 depicts the LFI model's output compared to the raw output data. The LFI is thus superior to ANFIS-EFCM (RMSE = 0.09416), ANFIS-GRID (RMSE = 0.21132) and ANFIS-SUB (RMSE = 0.21221).

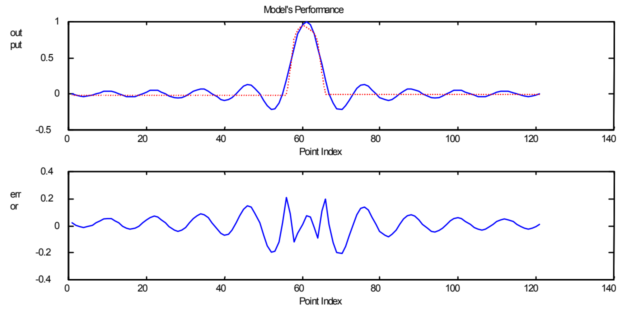

Figure 4 The Performance of LFI with 4 Rules (Upper Part: Dashed Line is Model's Output, Solid Line is Real Output; Lower Part: Residuals)

When the number of rules is increased to 8, the LFI model's performance is improved to 0.04438 while the performance index of ANFIS-GRID with 8 rules is 0.1288. Their performances are shown in Figures 5 and 6, respectively.

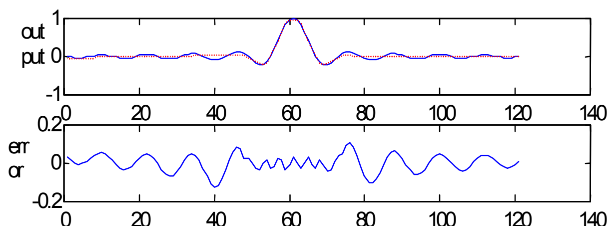

Figure 5 The Performance of the Proposed LFI with 8 Rules (Upper Part: Dashed Line is Model's Output, Solid Line is Real Output; Lower Part: Residuals)

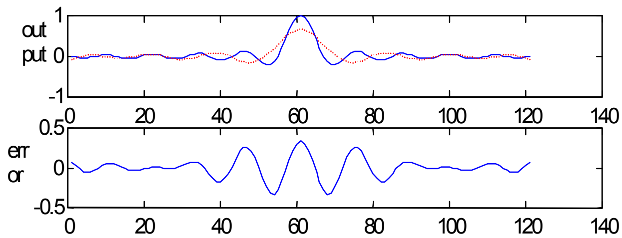

Figure 6 The Performance of ANFIS-GRID with 8 Rules (Upper Part: Dashed Line is Model's Output, Solid Line is Real Output; Lower Part: Residuals)

6 CONCLUSION

A new algorithm to build linguistic fuzzy models directly from input-output data is introduced. The proposed method is simple because of its pure linguistic nature. It uses symmetric triangular membership functions and a simplified fuzzy reasoning method. This algorithm can achieve either the same or better level of accuracy comapred to ANFIS and the methods of (Emami 1998, Sugeno 1993). For the test Sinc function, the proposed LFI model proved superior to the three different ANFIS algorithms. Although the TSK model is generally more discriptive than the pure linguistic model, sometimes it seems that it indulges into the insignificant details of the system while the LFI model allways retrieves the most important characteristics of the systems. Compared with the methods in of (Emami 1998, Sugeno 1993), which also aim at building pure linguistic models, the LFI model of this paper is much simpler both in computation and in form.

REFERENCES

Emami, M.R. Turksen, Burhan. Goldenberg A.A., 1998. Development of A Systematic Methodology of Fuzzy Logic Modeling, *IEEE transaction on Fuzzy Systems*, 6(3).

Yager, Ronald R. Filev Dimitar P. 1994. *Essentials of Fuzzy Modeling And Control*, John Wiley & Sons, New York.

Takagi, T., and M. Sugeno. 1985. Fuzzy Identification of systems and its applications to modeling and control, *IEEE transactions on Systems, man, and Cybernetics*, 15(1): 116-132.

Chiu, S. 1994. Fuzzy Model Identification based on Cluster Estimation, *Journal of Intelligent &Fuzzy System*, 2(3).

Jang, Jyh-Shing Roger. 1993. ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE transactions on Systems, man, and Cybernetics*, 23(3): 665-680.

Wang, L., and Jerry M. Mendel, 1992. Generating Fuzzy Rules by learning from Examples, *IEEE transactions on Systems, man, and Cybernetics*, 22(6): 1414-1427.

Sugeno, M., and Takahiro Yasukawa, 1993. A Fuzzy-Logic-Based Approach to Qualitative Modeling, *IEEE transaction on Fuzzy Systems*, 1(1): 7-31.

Chen, M., and Shinn-Wen Wang, 1999. Fuzzy Clustering Analysis for Optimizing fuzzy Membership Functions, Fuzzy Sets and Systems, 103: 239-254.

Xie, X.L., and G.Beni, 1991. A validity measure for fuzzy clustering, *IEEE trans. Pattern Anal. Mach. Intell.* 3(8): 841-846.

AUTHOR BIOGRAPHIES

HOSSEIN SALEHFAR received his B.S. in Electrical Engineering from University of Texas at Austin, and his M.S. and Ph.D. in Electrical Engineering from Texas A&M University. He was a research assistant with the Electric Power Institute at Texas A&M University during 1985-1990. He was an Assistant Professor of Electrical Engineering at Clarkson University, New York, during 1990-1995. In 1995 he joined the Department of Electrical Engineering at University of North Dakota where he is an Associate Professor. Dr. Salehfar was a consultant to the

New York Power Pool and he has worked on several projects for electric utilities in the state of North Dakota. His research interests currently include power electronics, power systems, controls, fuzzy logic and neural net applications to power electronics and systems, and reliability. Dr. Salehfar is a member of several IEEE societies including Power Electronics, Power Engineering, Industrial Electronics, Industry Applications, Mechatronics, Reliability, and Engineering Education Societies. He is also a member of the American Society for Engineering Education (ASEE). His email and web addresses are <h_salehfar@und.nodak.edu> and <http://power.ee.und.nodak.edu>.

NAGY BENGIAMIN is a Professor and Chairman of the Department of Electrical and Computer Engineering at California State University - Fresno. He received his B.S. in electrical engineering and B.S. in applied mathematics from Alexandria University, Egypt in 1971, M.S. in systems engineering from Carleton University, Canada in 1974, and the Ph.D. in electrical engineering from the University of Calgary, Canada in 1979. His research interests include control systems, robotics, and electric power systems operation and control. He has consulted for the electric power and manufacturing industries and published numerous technical articles on nonlinear control with applications to robot manipulators, fuzzy logic control, electric power distribution and energy management. He is a member of IEEE, ASEE, and Tau Beta Pi. His email and web addresses are <bengiami@csufresno.edu> and <www.engr.csufresno.edu/~bengiami/>.

JUN HUANG received the B.S. and M.S. degrees in Measurement and Testing Engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1995 and 1998, respectively. He is now completing the M.S. degree in Electrical Engineering at University of North Dakota, Grand Forks, ND, While working as a research assistant. His research interest is in the applications of fuzzy logic, artificial intelligence and other intelligent techniques.