# A Conditional Fuzzy Clustering with Adaptive Method

**A. Elmzabi[1,2], M. Bellafkih[1], M. Ramdani[2], K. Zeitouni[3]**

*[1]Laboratoire de Réseaux &Télécoms*
*Institut National des Postes et Télécommunications*
*Av. Allal El Fassi – Madinat Al Irfane*
*Rabat – Maroc*

*[2]UFR Systèmes d'information réactifs & Ingénierie des Systèmes Intelligents*
*Dep Informatique - FSTM*
*B.P. 146*
*Mohammedia – Maroc*

*[3] Laboratoire PRISM-Université de Versailles-Saint-Quenti-en-Yvelines*
*45, av. Des Etats Unis*
*78035- Versailles Cedex*
*France*

## Abstract

The Chiu's method which generates a Takagi-Sugeno Fuzzy Inference System (FIS) is a method of fuzzy rules extraction. The rules output is a linear function of inputs. Those rules are not explicit for the expert.

This paper proposes a new method to generate Mamdani FIS, where the rules output is fuzzy. The method proceeds in two steps. The first step consists in using the subtractive clustering principle to estimate the number of clusters and the initial locations of cluster centers, each obtained cluster corresponds to a Mamdani fuzzy rule. The second step optimizes the fuzzy model parameters by using a genetic algorithm. This method has been implemented and tested in the framework of a traffic network management application. This method has been extended to generation of Mamdani fuzzy rules when fuzzy classes can be predefined by the expert.

**Keywords***:* Subtractive clustering, fuzzy inference systems, fuzzy entropy, genetic algorithms, network management.

## 1. Introduction

Clustering is a fundamental method in data mining and pattern recognition areas. Fuzzy clustering allows natural grouping of data in a large data set and provides a basis for constructing rule-based fuzzy model. Chiu developed a fuzzy clustering approach [5,6], called subtractive clustering, for extracting the Takagi–Sugeno fuzzy rules from data, where the rule output is not fuzzy but a linear function of inputs. To optimize these rules, Chiu used the ANFIS [11] approach (Adaptive Neuro-Fuzzy Inference System). ANFIS uses back propagation learning to determine premise parameters and least mean square estimation to determine the consequent parameters (i.e. to learn the parameters related to membership functions).

In this paper, we first recall the subtractive clustering, then, we propose a new approach. This is a two steps approach. First, it uses the subtractive clustering and adapts it with a Mamdani Fuzzy Inference System (FIS) in order to generate fuzzy rules with fuzzy outputs. Then, a genetic algorithm is proposed to adjust the obtained FIS in the second step. This approach is illustrated in the prediction of state and evolution of a local area traffic network, and results are compared with those obtained by the Chiu's approach.

Finally we study the problem of extracting Mamdani fuzzy rules allowing the expert to classify the data set in predefined fuzzy classes. An extension of the previous method is proposed for this purpose in section 6.

## 2. Subtractive clustering

The purpose of this method is to estimate the number and initial locations of cluster centers [5]. The choice of a cluster center is based on the density of the actual data points in its vicinity, estimated by a potential value introduced below.

Let a training set of $N$ data points $W_i$, $i=(1,..,N)$ in an $D$ dimensional space.

$W_i = (x_i, y_i)$, where $x_i$ represent the $p$ input variables and $y_i$ the $(D\text{-}p)$ output variables. The potential $P_i$ of data point is defined as:

$$P_i = \sum_{j=1}^{N} e^{-\alpha \|W_i - W_j\|^2} \qquad (1)$$

where $\alpha = 4/r^2$, $r$ is the radius defining a $W_i$ neighbourhood.

A data point with many neighbouring data points will have a high potential value. The data point with the highest potential value is chosen as the first cluster center. Let $W_1{*}$ be the location of the first cluster center and $P_1{*}$ be its potential value. To generate the cluster centers, the potential is revised of each data points $W_i$ by the formula:

$$P_i = P_i - P_1{*} \exp(-\beta \|W_i - W_1{*}\|^2) \qquad (2)$$

$\beta$ is a positive constant defining the neighbourhood which will have measurable reductions in potential. The second cluster center is placed at the data point with the highest reduced potential value.

More generally, once the $k$'th cluster center is identified, the potential of all data points is revised by the formula:

$$P_i = P_i - P_k{*} \exp(-\beta \|W_i - W_k{*}\|^2) \qquad (3)$$

Where $W_k{*} = (x_k{*}, y_k{*})$ is the location of the $k$'th cluster center and $P_k{*}$ is its potential value. The process of acquiring new cluster centers and revising potentials ends when the potential of all data points falls below some fraction of the potential of the first cluster center.

As a result of subtractive clustering, we obtain $q$ cluster centers $W_i{*} = (x_i{*}, y_i{*})$ and $D$ corresponding spreads $s_i$, $i=(1,..,D)$, then we determine their membership functions. The spread is calculated according to $\beta$ [5].

These clusters were exploited by Chiu in a Takagi-Sugeno Fuzzy Inference System (FIS) [4], where the rule output is a linear function of the inputs.

However, in this system the output variable centers are not exploited.

In the following, we shall restrict ourselves to the case of a fuzzy model with $p$ input variables and a single output variable. The generalization of the result to multiple output variables raises no conceptual complications.

## 3. Generating Mamdani fuzzy rules

In our method [7] we transform each cluster center $(x_i{*}, y_i{*})$ and the corresponding spread $s_i$, obtained by the subtractive clustering, to a fuzzy rule of the form:

`If the input is` *`close to`* $x_i{*}$ `then the output is` *`close to`* $y_i{*}$.

The linguistic values {*close to* $x_i{*}$} and {*close to* $y_i{*}$} represent the fuzzy subsets whose membership functions premises and consequent are the Gaussian obtained from the position of the cluster center $(x_i{*}, y_i{*})$ and the corresponding spread $s_i$.

These rules are injected in a Mamdani fuzzy inference system.

## 4. Adjustment of the fuzzy model by genetic algorithms (GAs)

GA [10] is an iterative method of an optimization function, known as objective function. To use this algorithm, we must represent a solution to our problem as a chromosome. The genetic algorithm creates a population of solutions and applies operators such as mutation and crossover to evolve the solutions in order to find the best one(s).

The selection operator makes it possible to the individuals of a population to survive, to reproduce or to die. There are several methods for reproduction. The most known and used method is the wheel of lottery of Goldberg [9]. According to this method, each chromosome will be duplicated in a new population proportionally with its adaptation value. The crossover makes it possible to produce two new individuals (children) starting from two individuals (parents). The mutation operator consists in changing randomly the value of certain variables in an individual. In genetic algorithms, the mutation is considered as a secondary operator compared to the crossover.

The three most important aspects when using genetic algorithms are: definition of an objective function, definition and implementation of a genetic representation, and definition and implementation of genetic operators.

### 4.1. Adaptation of GAs to the fuzzy model

Our problem is to adjust the membership function of the fuzzy model parameters obtained by subtractive clustering, in order to minimize the error between the real output $y_k$ and the predicted output $y(k)$ obtained by the fuzzy model.

We chose as objective function, the total error between the two outputs for the training set, defined by:

$$\text{Error} = \sum_{k=1}^{N} f(y(k) - y_k)$$

Where $y(k)$, $y_k$ are respectively the predicted and the real outputs corresponding to the $k$'th data point, and $f$ is a function that computes the difference between two output values.

### 4.2. Representation of individuals

As emphasized above, the individuals should be mapped to a representation usable by GAs. This section describes this mapping and its implementation.

To optimize our fuzzy model, we look for the cluster centers and the corresponding spreads for which the objective function is minimal. Let

$$C = \begin{bmatrix} x_{11}^* & x_{12}^* \ldots\ldots\ldots & x_{1p}^* & y_1^* \\ x_{21}^* & x_{22}^* \ldots\ldots\ldots & x_{2p}^* & y_2^* \\ . & . \ldots\ldots\ldots & . & . \\ x_{q1}^* & x_{q2}^* \ldots\ldots\ldots & x_{qp}^* & y_q^* \\ s_1 & s_2 \ldots\ldots\ldots & s_p & s_{p+1} \end{bmatrix}$$

be the matrix obtained by the subtractive clustering, where the $q$ first lines correspond to the cluster centers and the last line represents spreads.

Then we transform this matrix in a vector of real values, of dimension $(p+1)*(q+1)$:

$T=(x_{11}*,\ldots,x_{1p}*,y_1*,\ldots,x_{q1}*,\ldots,x_{qp}*,y_q*, s_1,s_2,\ldots s_{p+1})$

Each vector $T$ represents an individual coded by a fixed size binary chain. We use the real coding.

### 4.3. Creation of the initial population

We apply $M$ times the subtractive clustering on the data set by varying the $r$ radius to obtain every time $q$ rules. Each obtained result represents an individual for the GA. These individuals form the initial population.

To form the new population, the wheel of lottery is chosen as selection operator. The choice of crossover rate $p_c$ and the mutation rate $p_m$ depends of problem's nature.

## 5. Application to a management problem of a local area network

We apply this method to a problem of prediction of the state and the evolution of traffic in a local network [1,4]. We present the global system architecture for management network in figure1.

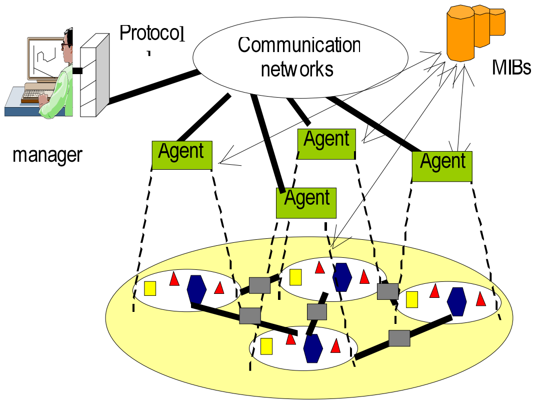


**Figure 1**: Network management architecture

The Management information (performance, traffic…) are collected in a MIB (Management Information dataBase) by using various protocols such as SNMP (Simple Network Information Protocol), and CMIP (Common Management Information Protocol).

The data abundance in this database requires tools for knowledge extraction to help the network manager. Indeed, for a large data set, the expert needs a decision support tool to manage his database. Such decision support tool is the goal of our application.

It is based on the above presented approach that first performs data clustering, then, builds a fuzzy model. This leads to cluster representation by a minimal number of fuzzy rules, which can be easily treated and interpreted by the network manager.

The training set is constituted by 100 examples; In this application, we have focused in two performance parameters: the *response time*, and the *collisions number* as input variables, and the *network charge* as output variable.

### 5.1. Presentation of the results with Mamdani FIS

We use the subtractive clustering with a radius $r$ = 0.75, four cluster centers and four corresponding spreads are found in the data set. We transform these four clusters into four fuzzy rules. Those rules are injected in a Mamdani FIS with the following characteristics: *min* operator for conjunction, *max* operator for union, *min* operator for implication, *max* operator for aggregation method, and *som* operator for deffuzzification. The obtained rules are shown in the figure 2.

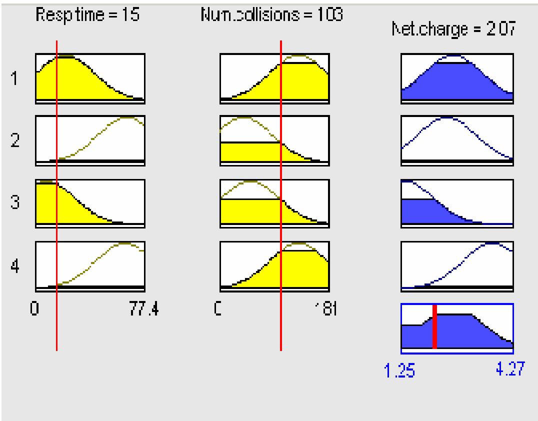


**Figure 2**: The initial fuzzy rules

We present a comparison between the real and the predicted outputs in figure 3.

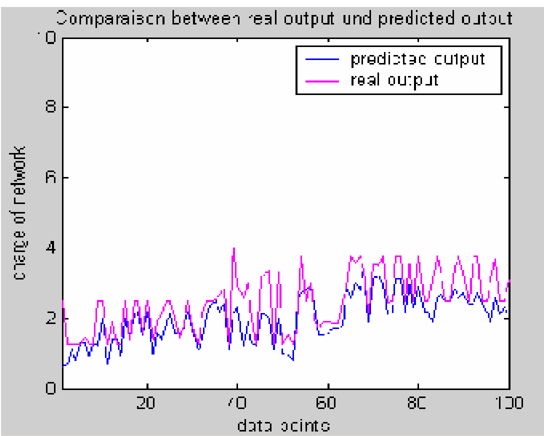


**Figure 3**: real and predicted outputs

This shows an important gap between the predicted output resulting from this fuzzy model and real output. The Root-Mean-Square Error (RMSE) is 0.0681.

In order to remedy to this problem, an adjustment of the membership functions parameters in the fuzzy rules is necessary.

At this end, we apply the second step of our method that uses the genetic algorithms.

### 5.2. Adjustment with GA

In our case, we intend to adjust the parameters of the four fuzzy rules obtained by the first step. Hence, an individual of the GA corresponds to a vector of 15 real variables, that are coded by a 150 bits binary chain.

We chose an initial population formed by 32 individuals, and apply the wheel of lottery in order to create a new population. The crossover is defined with a rate $p_c$= 0.8.

The quadratic average error is chosen as objective function:

$$\text{Error} = \frac{\sum_{k=1}^{N} (y(k) - y_k)^2}{N}$$

To stop the process, two criteria are used: the stagnation of fitness and the maximum number of generations. The latest is set to 50. The stop of research of optimum intervenes since one condition is satisfied.

#### 5.2.1. Results after optimization

We apply again the fuzzy inference system with the resulting parameters of the genetic algorithm. Those parameters correspond to the best cluster centers and the corresponding spreads. The RMSE between real and predicted outputs after adjustment with GA is 0.0322. This value shows a clear improvement of the results by comparison to the previous RMSE. The **figure 4** presents the obtained results.

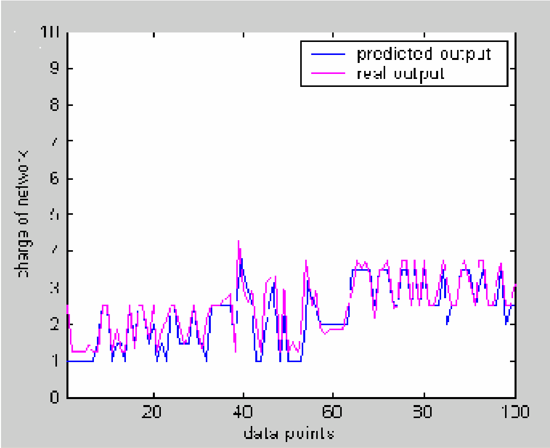


**Figure 4**: real and predicted outputs after adjustment with GA

In order to validate our approach, we also compare our results with those obtained by the Chiu method. The RMSE for this method is 0.0228.

Although our results are lightly less good than those obtained by the Chiu method, they are more meaningful for the expert, because he could interpret them. Indeed, they are based on the logical Mamdani fuzzy rules, whereas the Chiu's method produces functional Takagi-Sugeno fuzzy rules, where the output is a linear function of the input.

## 6. Extracting fuzzy rules with a priori fuzzy output values

In this part, we study the problem of generation of Mamdani fuzzy rules, where the expert wants to classify the output variable in $K$ fuzzy predefined classes. Theses classes are defined by some Gaussian membership functions

Let be a training set with $N$ data points, with $p$ input variables and one output variable.

Let $C= \{C_1, C_2,\ldots,C_K\}$ be the fuzzy classes set. These fuzzy classes are predefined by the expert, and they constitute a fuzzy partition on the value space $Y$ of the output variable.

In order to have $K$ classes in output, we vary the $r$ radius, in the first part of our approach, in the way to obtain every time $K$ clusters, and thereafter $K$ fuzzy rules. The output variable of these rules has $K$ fuzzy values defined with Gaussian membership functions.

Let $C'=\{C'_1, C'_2,\ldots, C'_K\}$ be the set of $K$ fuzzy values of the predicted output classes.

For every value of $r$ used in our approach, we compute the fuzzy entropy $E^*(C'/C)$ defined [16] by:

$$E^*(C'/C)=\sum_{i=1}^{K} p^*(C_i)E^*(C'/C_i)$$

with

$$E^*(C'/C_i)=-\sum_{j=1}^{K} p^*(C'_j/C_i)\log p^*((C'_j/C_i))$$

where $p^*(C_i)$ is the fuzzy probability of $Ci$ defined [21] by:

$$p^*(C_i) = \sum_{y\in Y} p^*(y) = \sum_{y\in Y} f_{C_i}(y)p(y)$$

$f_{Ci}$ is the membership function of $C_i$, and $p(y)$ is esteemed by the frequency of $y$ in the training set, the conditional fuzzy probability is defined by:

$$p^*(C_j'/C_i)=\frac{p^*(C_j'\cap C_i)}{p^*(C_i)}$$

where:

$$p^*(C_i\cap C_j')=\sum_{y\in Y}\min(f_{C_i}(y), f_{C_j'}(y))p(y)$$

$p^*(C_i)$ could be valued in the training set by:

$$p^*(C_i)=\frac{|C_i|}{N}\text{ , where } |C_i|=\sum_{y\in Y} f_{C_i}(y)$$

We use the genetic algorithm in order to minimize the fuzzy entropy of predicted outputs with regard to the outputs suggested by the user. The minimization is computed according to the radiuses. This approach gives us good $C'_i$, in the sense to be enough satisfactory for the user and respecting the real data at the same time.

## 7. Conclusion and discussion

We suggested a method in order to generate the Mamdani fuzzy inference systems. This method uses the results of the subtractive clustering in order to generate the Mamdani fuzzy rules and the genetic algorithms for the parameters optimization of these rules.

The obtained results are satisfactory and more explicit for the expert.

We also studied a problem of generation of Mamdani fuzzy rules, where the expert wants to classify the output variables in *K* fuzzy predefined classes. To do this, we minimize the fuzzy entropy between the predefined output classes with regard to the predicted output classes.

The expert will have the choice of using the first approach without taking into account the suggested classes or the second approach that allows for the predefined classes.

In the perspective of data mining application with a large database, we shall study the scalability of the proposed method with the data volume growth.

**REFERENCES**


[1]. M. Bellafkih, B. Idrissi Kaitouni, F. Elhebil, M. Ramdani. Un système à base de connaissances neuro-flou pour la gestion du trafic d'un réseau local. MCSEAI'2002, Annaba- Algeria, 4-6 May 2002.

[2]. B. Bouchon-Meunier, L. Foulloy, M. Ramdani, Exercice de la logique floue et applications, Ed. Cepadius, 1998.

[3]. B. Bouchon-Meunier, M. Ramdani, C. Marsala. Inductive learning and fuzziness, Scientia Iranica, International Journal of Science and Technology, volume 2,Number 4, pp. 289-298, 1996.

[4]. E. Carvalho, A. Dias Belchior, J Neuman de Souza. A fuzzy logic application applied to local area network management. Networking and Information Systems Journal, volume 2, Number 3, Hermes 1999.

[5]. S. L Chiu. Extracting fuzzy rules from data for function approximation and pattern classification. Chapter 9 in Fuzzy Information Engineering: a Guided Tour of applications, ed. D. Dubois, H. Prade, and R. Yager, John Wiley & Sons, 1997.

[6]. S. L Chiu. Fuzzy model identification based on cluster estimation. J. of Intelligent and Fuzzy Systems 2, 267-278, 1994.

[7]. A. Elmzabi, M. Bellafkih, M. Ramdani, F. Elhebil. Méthode adaptative pour générer des règles floues: Application à la gestion du réseau. CoPSTIC'03, Rabat, pp 107-110, Décembre 2003.

[8]. P. Y. Glorennec. Algorithmes d'apprentissage pour systèmes d'inférence floue. HERMES Science Publications, Paris, 1999.

[9]. D. E. Goldberg. Genetic Algorithms in search, optimization, and machine learning. Addison-Wesley, 1989.

[10] J. H. Holland. Adaptation in Natural and Artificial Systems, The University of Michigan Press, 1975.

[11]. J.-S. R Jang. ANFIS : adaptive-network-based fuzzy inference system. IEEE Trans. on Systems, Man and Cybernetics, vol. 23, no. 3, pp. 665-685, May 1993.

[12]. C. Marsala, M. Ramdani. Connaissances expertes floues et système d'apprentissage descendant. Rencontres francophones sur la logique floue et ses applications, pp 163-169, Paris 27 novembre 95.

[13]. R. Pytelkova. Modelling and control with neuro-fuzzy system. Helsinki Summer School 2001.

[14]. R.P. Quach, A. Dandrieux, G. Dray, D.W. Peason, G.Dusserre. Modelling gas dispersion by subtractive clustering and fuzzy c-means. IPMU, vol 3, pp 1328-1334, Madrid July 3-7, 2000.

[15]. J. R. Quinlan. Simplifying Decision Trees. Machine Learning, vol 1, pp 81-106, 1986.

[16]. M. Ramdani. Système d'induction formelle à base de connaissances imprécises. Thèse de doctorat, Paris VI. 28 février 1994.

[17]. R. A. Resende, S. M. Rossi, A. Yamakami, L. H Bonani, E. Moschim. Traffic engineering with MPLS using fuzzy logic for application in IP networks. Fuzzy IEEE, St Louis-USA, 25-28 May 2003.

[18]. T. Takagi, M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. IEEE Trans. On Systems, Man, and Cybernitics, 15:116-132, 1985.

[19]. R. Yager and D. Filev. Generation of fuzzy rules by mountain clustering. J. Of Intelligent and Fuzzy Systems 2, 209-219, 1994.

[20]. J Yao, M. Dash, S.T. Tan, H. Liu. Entropy-based fuzzy clustering and fuzzy modeling. Fuzzy Sets and Systems 113:381-388, 2000.

[21]. L. A. Zadeh. Probability measure of fuzzy events. J Math. Anal. App, 23:421-356, 1992.