An Adaptive Fuzzy Clustering Approach for the Network Management

Amal Elmzabi, Mostafa Bellafkih, and Mohammed Ramdani

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. In addition, these rules are not explicit for the expert.

In this paper, we develop a method which generates Mamdani FIS, where the rules output is fuzzy. The method proceeds in two steps: first, it uses the subtractive clustering principle to estimate both the number of clusters and the initial locations of a cluster centers. Each obtained cluster corresponds to a Mamdani fuzzy rule. Then, it optimizes the fuzzy model parameters by applying a genetic algorithm. This method is illustrated on a traffic network management application. We suggest also a Mamdani fuzzy rules generation method, where the expert wants to classify the output variables in some fuzzy predefined classes.

Keywords—Fuzzy entropy, fuzzy inference systems, genetic algorithms, network management, subtractive clustering

I. 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 [2], [8], [16]. Chiu developed a fuzzy clustering approach [4], [5], called subtractive clustering, to extract 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 (Adaptive Neuro-Fuzzy Inference System) approach [11], [12], [13]. ANFIS uses backpropagation learning to determine premise parameters and least mean square estimation to determine the consequent parameters (to learn the parameters related to membership functions). In this paper, we recall initially the subtractive clustering, then we propose a new approach.

A. Elmzabi is with the Networks and Telecommunications laboratory, National Institute of posts and Telecommunications (INPT-Rabat), and the Department of Computer Science, Faculty of Sciences and technologies, Mohammedia, Morocco (mobile: 212-70-85-26-39; e-mail: elmzabi@yahoo.fr).

M. Bellafkih is with The National Institute of posts and Telecommunications (INPT), Networks and Telecommunications laboratory, 2 Allal Elfassi - Rabat-institutes, Rabat, Morocco (phone: 212-37-77-30-79 Fax: 212-37-77-30-44; e-mail: bellafkih@inpt.ac.ma).

M. Ramdani is with the Department of Computer Science, Faculty of Sciences and Technologies, B.P. 146, Mohammedia, Morocco, (phone: 212-23-31-53-52; Fax: 212-23-31-53-53; e-mail: ramdani@uh2m.ac.ma).

It is organised as follows: first, it uses the subtractive clustering and adapts it with a Mamdani Fuzzy Inference System (*FIS*) to generate fuzzy rules with a fuzzy output. Then, a genetic algorithm is proposed to outperform the FIS resulted. This approach is applied on 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 worked on the extracting Mamdani fuzzy rules problems, where the expert wants to classify the output variables in some fuzzy predefined classes. For this purpose, an extension of the previous method is proposed in section VI.

II. SUBTRACTIVE CLUSTERING

Among the fuzzy clustering algorithms existing in the literature we find for example, the fuzzy c-means, the mountain clustering, the subtractive clustering and the entropy-based fuzzy clustering [14], [17], [18]. In this paper we are interested in the subtractive clustering. The purpose of this algorithm is to estimate both the number and initial locations of cluster centers [4]. 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 $(D-p)$ outputs variable. The potential P_i of data point is defined by:

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

where $\alpha = 4/r^2$, r is the radius defining a W_i neighbourhood, and $\|.\|$ denotes the Euclidean distance.

Thus, the measure of the potential for a data point is a function of its distances to all other data points. A data point with many neighboring 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:

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

β is a positive constant defining the neighbourhood which will have measurable reductions in potential. Thus, we subtract an amount of potential from each data point as a function of its

distance from the first cluster center. The data points near the first cluster center will have greatly reduced potential, and therefore will unlikely be selected as the next cluster center.

When the potential of all data points has been revised according to Eq. (2), we select the data point with the highest remaining potential as the second cluster center.

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

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

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 β [4].

These clusters were exploited by Chiu in a Takagi-Sugeno Fuzzy Inference System (FIS) [3], where the rule output is a linear function of the inputs.

However in this system the output variable centers is 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.

III. GENERATING MAMDANI FUZZY RULES

In our method [6] 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.

IV. ADJUSTMENT OF THE FUZZY MODEL BY GENETIC ALGORITHMS

The genetic algorithm (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. GA creates a population of solutions and applies genetic 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 [10]. 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.

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. In this context, the application of the genetic algorithm is primarily used to modify the parameters of predicted rules, so as to make the error optimal.

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

$$\text{Error} = \sum_{k=1}^{N} f(y(k) - y_k) \qquad (4)$$

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 which computes the difference between two outputs values.

Representation of individuals

As underlined above, individuals representation usable by GAs should be defined and implemented.

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

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

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 $(p+1)*(q+1)$:

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

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

Creation of the initial population

We apply *M* times the subtractive clustering on 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 the problem's nature.

V. 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], [3], [7]. We present general system architecture for network management
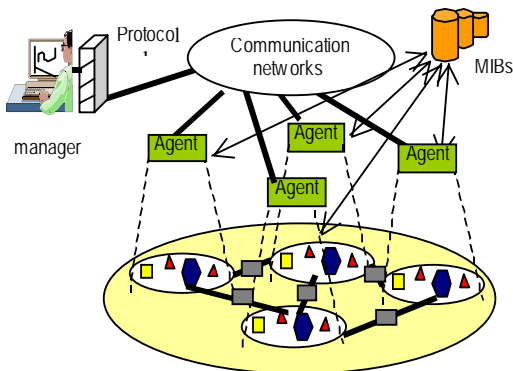

Fig. 1 Network management architecture

in figure 1.

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

The data abundance in this database requires tools for knowledge extraction to help the network manager. Then for a large data set, the expert needs an intelligent tool to manage his base [15].

Our application attempts to develop a decision support tool, based on grouping data and providing a basis for constructing a fuzzy model. In other words, we represent this grouping by a minimal number of fuzzy rules which can be treating and interpreting by the network manager. For that we retained two performance parameters (response time and the collisions number).

The training set is constituted by 100 examples, with two inputs variables: the *collisions number* and the *response time*, the *network charge* is the output variable.

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.
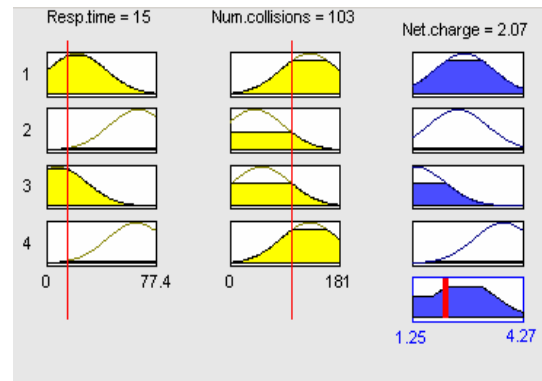

Fig. 2 The initial fuzzy rules

The first column represents the Gaussian membership functions (MFs) corresponding to the first input variable (response time), the second column shows the Gaussian MFs of the second input variable (collisions number), whereas the MFs corresponding to the output variable (network charge) are represented in the third column. This last column contains also the predicted output value of the fuzzy system after deffuzzification.

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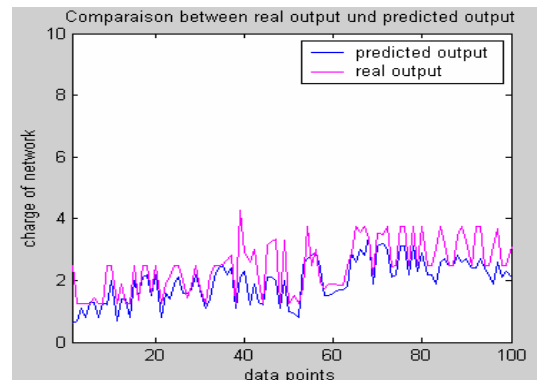

Fig. 3 Real and predicted outputs

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.

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, which 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:

$$\text{Error} = \frac{\sum_{k=1}^{N}(y(k)-y_k)^2}{N} \qquad (5)$$

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.

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 clearly the obtained results.
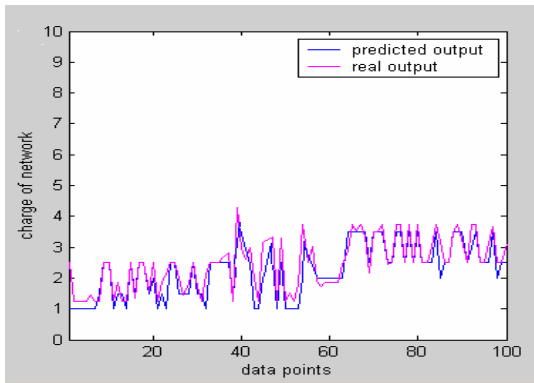

Fig. 4 Real and predicted outputs after adjustment with GA

The experimental results prove that the fuzzy model is satisfactory in the forecast of the output variable values. And to validate our approach, we compared 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 method produces functional Takagi-Sugeno fuzzy rules, where the output is a linear function of the input. For example, we obtain by our method the fuzzy rules of the form: if "response time is average" and "collisions number is average" then "network charge is saturated", whereas the type of fuzzy rules generated by the Chiu method is of the form: if "response time is average" and "collisions number is average" then "network charge is a linear function of the response time and

collisions number", in other words, the output of this last rule is not as significant as the output obtained by our method.

VI. 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 $C= \{C_1, C_2,...., C_K\}$ 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 $C'=\{C'_1, C'_2,...., C'_K\}$ be the set of K fuzzy values of the predicted output classes.

Our goal is to find K predicted classes closest to the fuzzy classes predefined by the expert. And as the choice of these K classes among the set of the other classes found by the first part of our approach is based on K predefined classes, we will propose to use a measure of conditional fuzzy entropy which compares the classes predicted by the fuzzy model and the classes predefined by the expert. We can possibly use the other measures of similarity, based at the distance between the objects, to make this comparison.

The entropy of an event is a measure of the mean information amount brought by this event. In other words, it is a measure of the unpredictability mean that the observer has on the sought event. The quantity of information and the entropy vary in contrary direction. If an event is sure, its entropy is null, on the other hand the quantity of information is maximum. In our case, the entropy measures the quantity of information corresponding to ambiguity between the predicted classes and those predefined. Thus we seek a set of fuzzy classes knowing the predefined classes, which lead us to use the conditional fuzzy entropy.

For every value of r used in our approach, we compute the fuzzy entropy $E^*(C'/C)$ defined [19] by:

$$E^*(C/C)=\sum_{i=1}^{K}p^*(C_i)E^*(C/C_i) \qquad (6)$$

with

$$E^*(C'/C_i)=-\sum_{j=1}^{K}p^*(C'_j/C_i)\log p^*((C'_j/C_i)) \qquad (7)$$

where $p^*(C_i)$ is the fuzzy probability of C_i which measures the probability that an element belongs to the set C_i. It is defined by [17]:

$$p*(C_i) = \sum_{y \in Y} p*(y) = \sum_{y \in Y} f_{C_i}(y)p(y) \qquad (8)$$

f_{C_i} 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:

$$p*(C_j'/C_i) = \frac{p*(C_j' \cap C_i)}{p*(C_i)} \qquad (9)$$

where:

$$p*(C_i \cap C_j') = \sum_{y \in Y} \min(f_{C_i}(y), f_{C_j'}(y))p(y) \qquad (10)$$

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

$$p*(C_i) = \frac{|C_i|}{N} \text{ , where } |C_i| = \sum_{y \in Y} f_{C_i}(y) \qquad (11)$$

We use the genetic algorithm in order to minimize the fuzzy entropy of predicted outputs regard with to the outputs suggested by the user. The minimization is computed according to the radiuses.

Optimization of fuzzy model by GAs

For the choice of the predicted fuzzy outputs, closest to the outputs suggested by the user, we must minimize the fuzzy entropy [18] which is our objective function. For this minimization, we use the genetic algorithms [9], [10]. We want to adjust the predicted output, then we look for the cluster centers and the corresponding spreads for which the objective function is minimal. We take then an individual, as matrix M of K lines and 2 columns, where the first column contains the centers and the second column represents the corresponding spreads. Let:

$$M = \begin{bmatrix} y_1^* & s_1 \\ y_2^* & s_2 \\ . & . \\ . & . \\ . & . \\ y_k^* & s_k \end{bmatrix}$$

where each (y_i^*, s_i) $(i=1...k)$ represents the center and the corresponding spread, which determines us the membership function of a data point to the predicted class C_i'.

We code each individual by size binary chain. In our case we use real coding.

For each r previously selected, used in the Mamdani fuzzy rules extraction method, we will have K fuzzy rules. These outputs rules represent an individual and the whole of the individuals forms 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.

Application on network management example

We applied our method to the same example of management traffic in a local network [1], [3]. The training set is constituted by 100 examples, with two input variables: the collisions number and the response time, the network charge is the output variable.

We studied the case where the manager wants to classify the output variables in four fuzzy classes. For that, we varied the ray r in the method of Mamdani fuzzy rules extraction and we chose the unit R of radius which we produced four cluster centers and four corresponding spreads in the data space.

An individual corresponds to a matrix of four lines and two columns obtained after each use of a radius r from the unit R in the method of Mamdani fuzzy rules extraction.

We chose an initial population formed by 34 individuals, and apply the wheel of lottery in order to create a new population.

For the genetic operators we used the crossover at a point with a rate $p_c= 0.8$, and two criteria of stopping: the stagnation of fitness and the maximum number of generations. The latest is set to 20. The research stop of optimum intervenes since one condition is satisfied.

We tested our method on several cases of fuzzy classes predicted by the user. We present in figure 5 cases of these results where we compare the predicted and predefined outputs.
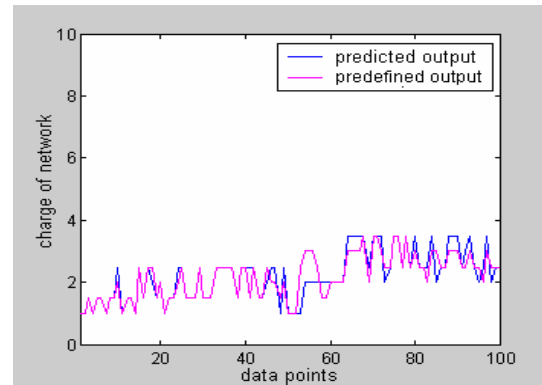

Fig. 5 Predefined and predicted outputs after adjustment with GA

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. The RMSE in this case is 0.0425.

VII. 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 in consideration the suggested classes or the second approach that allows for the predefined classes.

In the perspective of data mining application with very large databases, we shall study the scalability of the proposed method with 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, M. Ramdani, C. Marsala. Inductive learning and fuzziness, Scientia Iranica, International Journal of Science and Technology, volume 2, Number 4, pp. 289-298, 1996.

[3] 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.

[4] 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.

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

[6] A. Elmzabi, M. Bellafkih, M. Ramdani, K. Zeitouni. Conditional Fuzzy Clustering with Adaptive Method. IPMU´04, 4-9 July 2004, Perugia – Italie.

[7] A. Elmzabi, M. Bellafkih, M. Ramdani, F. Elhebil. Méthode adaptative pour générer des règles floues: Application à la gestion du réseau. Journal Marocain d'Automatique, d'Informatique et de Traitement de Signal. Pages 37-45, Numéro Spécial CoPSTIC'03, November 2004.

[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] J. S. R. Jang, C. T. Sun, E. Mizutani. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. 1997 by Prentice-Hall, Inc. Upper Saddle River, NJ 07458.

[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] 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.

[16] 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.

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

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

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

Amal Elmzabi was born in Taounate, Morocco. She received the PhD thesis in Computer Science from the Mohammedia Faculty of Sciences and Technologies (FSTM), Morocco, in July 2005. Her research interests include the network management, artificial intelligence (A.I), Data mining and time series. She is temporary teacher in Mohammedia Faculty of Sciences and Technologies (FSTM), Morocco since 2003.

Mostafa Bellafkih was born in Oujda, Morocco. He received the PhD thesis in Computer Science from the University of Paris 6, France, in June 1994 and Doctorat Es Science in Computer Science (option networks) from the University of Mohammed V in Rabat, Morocco, in May 2001. His research interests include the network management, knowledge management, A.I., Data mining and Database. He is Professor in The National Institute of Posts and Telecommunications (INPT) in Rabat, Morocco since 1995.

Mohammed Ramdani was born in Fes. He received the PhD thesis in Computer Science from the University of Paris 6, France, in February 1994 and Habilitation in Computer Science from the University of Paris 6, France, in June 2001. His research interests include the, knowledge management, A.I., Data mining and Database. He is Professor in Mohammedia Faculty of Sciences and Technologies (FSTM), Morocco since 1995.