

Structure identification for complex system and inference interpretable rules

Cherif Remache
Computer science departement
University of Oum El Bouaghi
Oum El Bouaghi, Algeria
e-mail: remachec@yahoo.fr

Ramdane Maamri
LIRE laboratory
University of Constantine
Constantine, Algeria
e-mail: rmaamri@yahoo.fr

Zaidi Sahnoun
LIRE laboratory
University of Constantine
Constantine, Algeria
e-mail: sahnounz@yahoo.fr

Abstract. The large and complex systems are characterized by the difficulty of formalizing and the lack of expertise. They are built from data representing instances of the system. The drawback of these structure selection methods is that pay particular attention to the numerical accuracy of the resulting model and little attention to the qualitative and semantic aspect. The use of a large number of input variables results in an introduction of redundant elements, poor transparency and an excessive complexity of the model obtained. To solve these problems, a particular interest should be given for selecting relevant input variables that can provide a reasonable compromise between the quality of approximation, the complexity and the transparency of the model. The proposed approach for structure selection assumes constraints on the number of variables to be selected in the initial combination to describe the model. Relevant inputs are found from input-output data and variables with the highest correlation coefficient with the output variable are selected in the initial combination

Keywords: *fuzzy inference systems, complex systems, extraction rules, learning, induction rules, interpretive rules, Partitioning.*

I. INTRODUCTION

The majority of fuzzy inference systems are built on expertise and knowledge often known by an expert of the domain. Such knowledge is expressed in the form of fuzzy rules. A fuzzy rule is defined as a conditional statement in the form: IF x is A THEN y is B , where x and y are linguistic variables; A and B are linguistic values determined by fuzzy sets on the universe of discourse X and Y , respectively.

These systems have achieved good performance; however, their use was limited to small systems. When the system to be modeled is complex, the expertise is difficult to obtain and to develop or completely absent. Generally the only information available is data representing instances of the behavior of the system; they are an important source of knowledge for its understanding. The existing relationship between the inputs and outputs of the system are formalized by using learning techniques from data. Induction of fuzzy rules is that a particular form of learning. The rule induction process contains in general two phases: rules generation and optimization phase of the model.

Different methods of generating rules are developed, they can be divided into three classes.

- Partitioning methods: They consist in dividing the domain of variation of each variable in a fixed number of intervals. All rules corresponding to all possible combinations of inputs are implemented. The choice of the number of fuzzy subsets for each input will directly affects the structure of the rule base. Two approaches are possible to perform the partitioning: an incremental approach which, from a rough initial partition, increases progressively the number of fuzzy subsets by adding in each step a fuzzy subset on one of the inputs. Various criteria are available to make the choice of the part of the input space to cut more finely. For the other approach, the number of fuzzy subsets for each input is initially fixed. All the rules corresponding to all possible combinations are implemented.

- Clustering methods: Clustering methods allocate data in clusters; characteristics of each cluster are described in the premise of a rule. The fuzzy sets obtained are not shared by all the rules; their interpretation in linguistic terms is greatly reduced. Different techniques are used to determine the number of clusters and evaluate the quality of the partition obtained.

- Hybrid methods: In this category, we cite genetic algorithms; they have the advantage of finding the global optimum and simultaneously optimize the parameters and the structure of the system taking in account the number of rules in the objective function; however, they are guided by numerical performance; the semantic aspect is neglected.

Actually, neuro-fuzzy models are widely used, these models combine the learning ability of neural networks and the formalism of fuzzy systems giving them the ability to acquire knowledge from data; unfortunately, the interpretability of knowledge is, in most cases, compromised during the learning phase: usually numerical accuracy is favored over semantics [1].

The rest of this paper is organized as follow. In section 2, we briefly describe some methods of structure selection for complex systems. Section 3, we present our method for selecting relevant input variables that can provide a reasonable compromise between the quality of approximation, the complexity and the transparency of the model. In section 4, we present an evaluation of the performance of our proposed method. Finally, some concluding remarks are presented in section 5.

II. II. METHODS OF STRUCTURE SELECTION FOR COMPLEX SYSTEMS.

These methods pay particular attention to large and complex systems. The large and complex systems are characterized by the difficulty of formalizing. In addition, they are characterized by a large number of variables which can be highly correlated with each other and may contain redundancies [2], [3]. The number of input variables must necessarily be reduced. The input structure selection is an essential step for reducing the complexity of the model, the elimination of redundant variables and the noise contained in the data [4]. Different appropriate methods are proposed to perform this selection, However, they pay more attention to the numerical accuracy of the model but the qualitative and semantics aspect are neglected. [1]

Often these methods are based on heuristic criteria, they develop different models (for each possible combination of input variables) compare them in order to retain the combination using the most significant variables. This approach is time consuming and depends on the choice of the model [2], [4]. In addition, these make implicit assumption of independence of input variables while complex systems present always interactions between these variables[3]. An increasing interest is given to the development of methods for selecting the input variables which can provide a reasonable compromise between the quality of approximation, the complexity and the transparency. The neuro-fuzzy systems with different learning techniques are commonly used: the model of Takagi-Sugeno [5] is the most commonly used to represent a nonlinear dynamic system. The system output is a linear combination of inputs; it guarantees a better numerical approximation [6].

JM Zurada proposes a model incorporating the interactions between the input variables in the learning process [7], [8], [9]; it uses binary relations representing linear dependencies between two variables.

The other method presented uses an index of "sensitivity" input-output that will iteratively determine the least significant input which will be removed from the model until the satisfaction of a criterion for dimensionality reduction [10].

In [2], the significant variables are selected directly from the input-output data by performing an exhaustive search from all possible combinations of inputs. The combination of variables which gives a minimal total variance for the model will be selected for analysis. The final selection structure of the model is left to the user who must find a compromise between performance and complexity.

Another method proposed by Babuska [4] selects the model structure by performing the choice of variables directly from the data, this method is based on geometric criteria and an elementary property of a function:

Giving a function:

$$f: x \rightarrow f(x) ; \quad \text{if } x_i = x_j \\ \text{then } f(x_i) = f(x_j) \quad \text{for all } i \text{ and } j$$

After clustering the data, a similarity measure is used to quantify the degree to which "the data do not verify this property of the function". In this way, potential entries are ordered in relation to their relative importance. The final selection of inputs of the model is left to the user.

To save the knowledge acquired by the neuro-fuzzy system in the learning phase and in order to produce interpretable rules; a number of constraints are taken into account by the learning algorithm presented by M.Y. Chow [11].

Starting from the same principle, G. Castellano proposes a neuro-fuzzy model working on a parameter space of reduced size by formulating a set of properties that must verify the fuzzy sets in the partition entry in order to be semantically interpretable [12], [13]. Other heuristic criteria are suggested to guide the generation of membership functions [12], [13], [14], [15], [16], [17].

We propose a new approach: the relevant input variables are found from input-output data; the variables with the highest correlation coefficient with the output variable are selected

III. PROPOSED METHOD

The proposed approach is intended to ensure a better compromise between transparency and the complexity of the model. It is based on the following remarks and observations:

- The use of a large number of variables affects the capabilities of interpretability of the model [18]
- A large number of rules usually results in low generalization and loss of perspicacity of knowledge included in the data [19].
- A large number of variables with several linguistic terms makes the model unnecessarily complex and reduces its transparency; therefore, a certain number of constraints will be considered:
 - Minimizing the number of input variables and select only the most relevant.

- Reducing as much as possible the number of linguistic terms per variables.
- The transparency of data partitioning in the learning phase should be guaranteed.

These constraints will be set by the user depending on the level of complexity and precision required.

Selection of input variables.

We shall select the combination of variables to be used in the model structure as follows:
Construct the matrix of the initial data: $Z = [X | Y]$;
 X is a matrix of dimension $N \times N$ where N is the number of instances and n the number of variables and Y : a vector of dimension $N \times 1$ is the corresponding output.

Compute the correlation matrix $C (N + 1) \times (N + 1)$. Variables are sorted in descending order of correlation coefficient with the output variable. First, we select the variable which is the most strongly correlated with the output then step by step we add another variable; this one will be selected if it does not present a strong correlation with one of the variables already included in the combination and we continue this process until the number of variables to be used is reached.

B. Identification of the structure of the model.

We adopt the model of Takagi-Sugeno [5]; this model is oriented to numerical accuracy and we consider a transparent and interpretable partitioning of the input space.

We use the Gaussian membership functions as they meet the required properties to ensure the interpretability of the model [13].

For a Takagi-Sugeno model, a rule is of the form:

$$\mathbf{R}_k : \text{If } x_1 \text{ is } A_1^k \text{ and } x_2 \text{ is } A_2^k \text{ and } \dots \text{ and } x_n \text{ is } A_n^k \\ \text{Then } y_k \text{ is } b_k, \quad k = 1 \dots K \quad (1)$$

K is the number of rules is given by $K = \prod K_i$ where n is the number of variables and K_i is the number of terms of the variable x_i . The value inferred by the model is calculated by:

$$y_c = \frac{\sum_{k=1}^K \beta_k y_k}{\sum_{k=1}^K \beta_k} \quad (2)$$

β_k is the degree of activation of the k th rule, and y_k is the contribution of this rule.

The degree of activation is calculated by:

$$\beta_k = \prod_{i=1}^n \mu_{A_i^k}(x_i), \quad k = 1, 2, \dots, K \quad (3)$$

$\mu_{A_i^k}$ is the membership function of the input x_i to fuzzy subset A_i^k in the premise of the k^{th} rule.

The input x_i is defined in the interval $[a_i, b_i]$, $a_i = \inf x_i$ and $b_i = \sup x_i$

Each variable x_i , is defined by A_{ij} ; $j = 1, 2, \dots, K_i$ linguistic terms defined by the membership functions $\mu_{A_{ij}}$ which are Gaussian with center ω_{ij} and standard deviation σ_{ij} .

$$\mu_{A_{ij}}(x_i) = \exp\left[-\frac{(x_i - \omega_{ij})^2}{2\sigma_{ij}^2}\right]; \\ i = 1, 2, \dots, n \quad \text{et } j = 1, 2, \dots, K_i \quad (4)$$

Averages are uniformly distributed in the interval $[a_i, b_i]$ and calculated as follows:

$$\omega_{ij} = (j-1) \frac{b_i - a_i}{K_i - 1}, \quad j = 1, 2, \dots, K_i \quad (5)$$

The standard deviations σ_{ij} are calculated by:

$$\sigma_{ij} = \frac{b_i - a_i}{2(K_i - 1)\sqrt{-2 \ln \varepsilon}}, \quad j = 1, 2, \dots, K_i \quad (6)$$

$\varepsilon \in [0, 1]$ to provide good coverage of the input space.

K_i is the number of linguistic terms retained for the input x_i is determined by the user.

The performance of the model is measured by the mean square error:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - y_{c_i})^2 \quad (7)$$

IV. EVALUATION OF THE PERFORMANCE.

To examine the performance of the proposed method; we consider the following non linear function often used in different publications to illustrate the performance of methods selecting structure of the model [20]; [21].

$$y = 10 \sin(\pi x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 \quad (8)$$

A set of 1200 samples is constituted as following:

variables x_1 to x_5 are randomly generated according to the uniform distribution $U(0, 1)$, the corresponding y is calculated using (8). In addition; five other variables are added to the inputs, they are obtained as follows:

$x_6 = x_1 + 2x_2$; $x_7 = 2x_1$; x_8 ; x_9 et x_{10} are randomly generated according to the uniform $U(0, 1)$. We obtain 10 variables that are candidates for selection to build the model structure. The number of input variables in the structure of the model and the number of linguistic terms for each input will be set by the user according to the desired accuracy. These inputs will be sorted according to the criterion for their selection in the structure of the model.

Applying the proposed method; the relevant variables are selected according to the number of inputs and linguistic terms by input. The mean square error is computed to measure the performance of the model corresponding to each selected structure. The final model structure is done by the user according to the accuracy obtained. The inputs will be sorted according to the criterion for their selection in structure of the model. Tab1 shows the correlation coefficient of each variable x_j with the output.

TABLE 1

x_j	Cov(x _j ,y)	x_j	Cov(x _j ,y)
x1	0.0067	x6	0.0301
x2	0.0296	x7	0.0329
x3	0.9600	x8	0.0255
x4	0.1565	x9	0.0270
x5	0.0329	x10	0.0528

.correlation coeff. of variable x_i with the output y .

For two input variables, x_3 and x_4 are retained in the structure of the model, they correspond to the variables that have the most contribution to the output y .

After 11 iterations, we obtain a mean square error less than 3%. Fig.1. shows the MSE depending on the number if iteration for two variables.

For three input variables, x_3 , x_4 and x_2 are retained in the Structure of the model; after 13 iterations, we obtain a

mean square error less than 3%. Fig.2. shows the MSE depending on the number if iteration for two variables.

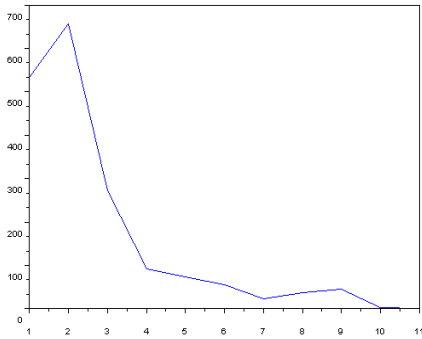


Fig.1. Error depending on the number of iterations (2 variables, 3linguistic terms per variable)

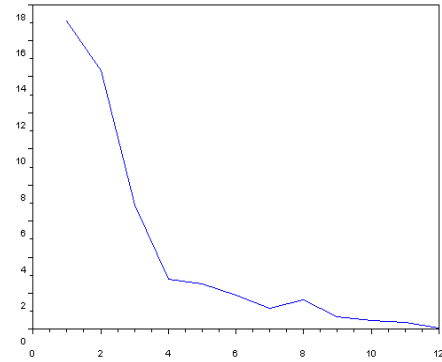


Fig. 1. Error depending on the number of iterations (3 input variables; 3 linguistic terms by variable)

We can note that the results attained with our approach are comparable to those developed in literature. [2]; [4]; [14].

A low error was achieved and a reasonable compromise between the quality of approximation and the interpretability of the model.

Fig.3. shows the first 50 values for y computed by (10) and y_c inferred by the model.

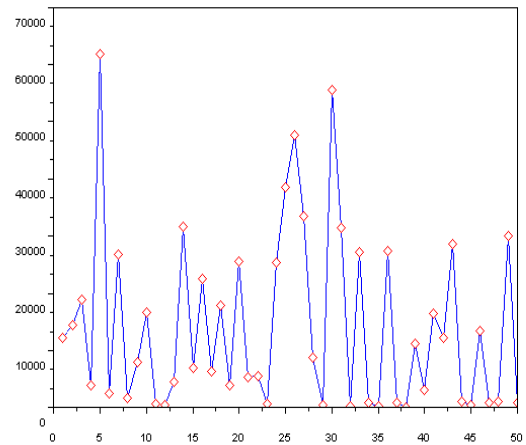


Figure 3: 50 first values of y and y_c

V. CONCLUSION AND PERSPECTIVE

In this paper, we have presented a new method which is intended to simplify a complex system and to provide a better compromise between transparency accuracy and complexity of the model by taking account constraints on the structure of the model: limiting the number of variables and the number of linguistic terms per variable. The transparency of partitioning the input space will ensure the interpretability of the system. As a continuation to this work, we plan to work on a benchmark that will help to evaluate the performance of our proposed method.

REFERENCES

- [1] R. Babuska, M.Setnes et H.B. Verbruggen . "Rule-Based Modeling : Precision and Transparency" in IEEE Transactions on Systems, Man and Cybernetics, Vol. 28, No 1, February 1998.
- [2] R. Babuska, D. Girimonte et J. Abonyi "Fuzzy Clustering for Selecting Structure of Nonlinear Models with Mixed Discrete and Continuous Inputs" in Proceedings IEEE International Conference on Fuzzy Systems, Budapest, Hungary, July 2004.
- [3] J. Zurada et A.E. Gaweda " Data-Driven Linguistic Modeling Using Relational Fuzzy Rules" in IEEE Transactions on Fuzzy Systems, Vol. 11, No 1, February 2003.
- [4] R. Babuska et R.Sindelar "Input Selection for Nonlinear Regression Models" in IEEE Transactions on Fuzzy Systems, Vol. 12, No 5, October 2004.
- [5] T. Takagi; M.Sugeno"Fuzzy identification systems and its application to modelling and control." IEEE, Transaction on Systems, Man and Cybernetics p116-132, 1985.
- [6]. Abonyi et R. Babuska"Local and Global Identification and Interpretation of Parameters in akgi-Sugeno Fuzzy Models", in Proceedings IEEE International Conference on Fuzzy Systems, San Antonio,USA, May 2000.
- [7]. J. Zurada et A.E. Gaweda "Fuzzy Neural Network with Relational Fuzzy Rules", in Proceedings of the International Joint Conference on Neural Network. IJCNN' 2000.
- [8]. J. Zurada et A.E. Gaweda "Data-Driven of Fuzzy Systems with Relational Input Partition" In Proceedings of the International Conference on Fuzzy Systems. FUZZ-IEEE' 2001.
- [9] J. Zurada, A.E. Gaweda et P.B. Aronhime"Efficient Data-Driven Fuzzy Modeling with Fuzzy Relational Rule Network", Work Sponsored by the Department of the Navy. Office of the Research.
- [10] J. Zurada , A.E. Gaweda et R. Setiono"Input Selection in Data-Driven Fuzzy Modeling" In Proceedings of the International Conference on Fuzzy Systems. FUZZ-IEEE' 2001
- [11] M.Y. Chow et S. Altug et H.J. Trussel " Heuristic Constraints Enforcement for Training of and Knowledge Extraction from a Fuzzy Neural Architecture- Part 1, Foundation, Part 2, Implementation and Application, IEEE transactions on Fuzzy Systems, 1999.
- [12] G.Ccastellano, A.M. Fanelli et C.Mencar "Discovering Interpretable Classification Rules from Neural Processed Data", Cognitive Systems, 2002
- [13] G. Ccastellano, A.M. Fanelli et C. Mencar"A Neuro-Fuzzy Network to Generate Human Understandable Knowledge from Data", Cognitive Systems, 2002
- [14] J.V. de Oliveira"Towards Neuro-Linguistic Modeling, Constraints for Optimization of Membership Functions", Fuzzy Set and Systems 1999
- [15] S-M Zhou " Improving the interpretability of Takagi-Sugeno Fuzzy model by using linguistic modifiers and multiple objective learning scheme " IEEE 2004
- [16] J.V de Olivera" Semantic constraints for membership function optimization «IEEE Trans on systems, Man and cybernetics, Part A, vol 29 p128-138
- [17] Y.Jin "Fuzzy modeling of high-dimensional system: complexity reduction and interpretability improvement " IEEE, Trans and Fuzzy Systems,Vol 8 p 212-221,2000
- [18] J.Abonyi et F. Szeifert"Supervised Fuzzy Clustering for the Identification of Fuzzy Classifiers" in pattern recognition letters 24 (2003)
- [19] J.H.Friedman "Multivariate adaptive regression splines" the annals of statistics ,vol. 19n01 pp 1-141, 1991.
- [20] W. Duch; R. Setiono et J. Zurada "Computational Intelligence Methods for Rule Base Data Understanding " in Proceedings IEEE , Vol. 92, No 5, May 2004