FUZZY LOCAL REGRESSION MODELS WITH FUZZY CLUSTERING *

A. F. Gómez-Skarmeta, H. Martínez, J. Botía Blaya
Dept. Informática, I.A. y Electrónica,
Facultad de Informática, Univ. de Murcia,
e-mail:skarmeta@dif.um.es

M. Delgado
Dept. Ciencias de la Computación e I.A.,
E.T.S.I. Informática, Univ. de Granada,
mdelgado@ugr.es

Summary

In this work we present an alternative approach to generate the fuzzy rules with multidimensional fuzzy sets in the antecedent and functional consequent associated to the TSK fuzzy model. In order to obtain the fuzzy rules we use a fuzzy clustering algorithm that look for linear behaviours in the product space of the input-output data.

Keywords: fuzzy clustering; unsupervised learning; fuzzy modeling.

1 INTRODUCTION

The TSK model introduced by Takagi, Sugeno and Kang (TSK fuzzy reasoning) [10] is associated with fuzzy rules that have a special format with a functional type consequent instead of the fuzzy consequent that normally appears in the Mamdami Model. In this way the TSK approach tries to decompose the input space into subspaces and then approximate the system in each subspace by a simple linear regression model. This characteristic provides efficient models to deal with complex system, although the generation of the corresponding fuzzy rules, specially the premise structure, is technically difficult and may lead to a nonlinear programming problem. Several alternative approaches have been proposed to reduce the complexity of this building process.

In the area of fuzzy modeling the procedure of fuzzy clustering has been utilized in different ways. In one of them fuzzy clusters give rise to "local" regression models - this is in fact the essence of the idea introduced originally in [10] and that have been used in differents works like [9][12]. The overall model is then structured into a series of IF-THEN statements. The condition part of the statement might includes linguistic labels of the input variables, or directly fuzzy relations on the product space of the input variables, while the action part contains a linear (or generally nonlinear) numerical relationship between input and output variables.

In this work we present an alternative approach to generate the fuzzy rules with functional consequent associated to the TSK fuzzy model. In our case using fuzzy clustering algorithms that look for linear behaviours in the product space of the input-output data, we develop an approach that detect multidimensional fuzzy sets in the product space of the input variables and in this way trying to identificate the premise of the fuzzy rules and then assigning a linear consequent to each rule.

The identification of fuzzy models can be improved in this way using these multidimensional reference fuzzy sets . Other authors work with fuzzy clusters in the product space of the input-output space using in this way the clustering in order to detect the interaction between the input and output variables, hence fuzzy clusters give rise to "local" regression models . The model is then structured into a set of IF-THEN statements. The condition part of each rule is obtained by projecting the fuzzy clusters in the input and output spaces as in [5]. Several of this authors have proposed the use of the fuzzy clustering techniques that tend to find clusters that are more adequate to the geometry of fuzzy rules than the classical fuzzy c-means [3] [7]. The most significant difference with other works is our use of a Gustafson-Kessel clustering algorithm in the product space of the input-output space in order to identify the fuzzy rules and then use it to generate the fuzzy rules with functional consequent within the TSK fuzzy model.

The paper is organized as follows. In section 2 we formulate the fuzzy model with which we are going to

*This work has been partially supported by CICYT project TIC97-1343-C002-02 and FEDER and EU project 1FD97-0255-C03-01

work. In section 3 we define the methods that we propose here to use the fuzzy clusters in order to generate the fuzzy rules of the TSK model. Numerical examples are provided in section 4 to illustrate the performance of the presented methods. Finally, section 5 contains some conclusions and comments about future trends.

2 FUZZY MODELS AND FUZZY CLUSTERING

Usually the basic knowledge to model a (fuzzy) system is assumed to be given in terms of a (usually large) set of sample I/O pairs, say $\Omega = \{(x_{t1}, x_{t2}, \ldots, x_{tp}), y_t)\} = \{(x_t, y_t), t = 1, \ldots, N\}$, and obviously the selected rules are required to approximate the function $\varphi : X^p \rightarrow Y$ that theoretically underlies the system behaviour.

We assume that the collection of data samples represent the system behavior in the product space $(X^p \times Y)$, where $X^p = (X_1 \times X_2 \times \ldots \times X_p)$, with $X_1, X_2, ..., X_p$ being the domains of discourse of the inputs and Y is the domain of the output.

It is clear that the approximation of the function $\varphi : X^p \rightarrow Y$ it is equivalent to the establishment of a relation between the data in X^p and Y, which in the fuzzy modeling context will be a fuzzy relation in the product space of the input-output data In this way, the clustering of the data in $(X^p \times Y)$ seems to be an adequate technique to disclose this relation and several authors, like Yoshinari et al.[12] Yager and Filev[11], Babuska and Verbruggen[1] among others, have use it in different ways to identify fuzzy rules. Our contribution is in the use of the information obtained from the clustering, in a direct way in order to generate the fuzzy modeling.

Let us remark that in the TSK model we need to specify fuzzy sets only in the input universe of discourse, so that the antecedent of each rule define a fuzzy region in the product space of the input variables X^p. Several authors (Sugeno et al.[9], Yager et al.[11], Babuska et al.[1] etc.) have made use of fuzzy clustering in fuzzy modeling by projecting the clusters in the domains of the variables. We have adopted a different point of view: instead of specifying fuzzy sets in each universe of discourse, we directly estimate the fuzzy region of the input space by using the fuzzy relation in the input-output product space obtained from the fuzzy clustering. In general, we will directly characterize fuzzy sets in X^p, thus trying to capture most of the information about the system behaviour. The centroids of the clusters (which describe a local behaviour) and their relation with the data (the fuzzy sets they generate), play a key role in our process.

Hence, what we need is a mean to establish a relation between fuzzy clusters in $(X^p \times Y)$ and associated fuzzy sets in X^p, as will be explained in the next section.

Let suppose that by using the fuzzy clustering we have already disclosed k local behaviour of the data (see [3] in order to detect the optimal number of fuzzy clusters) around the obtained centroids as captured by the fuzzy relations. By using the information the clusters in $(X^p \times Y)$ represent (centroids and distribution of the data around them), we are going to assign to any data x of X^p a value in Y, so we approximate the function $\varphi : X^p \rightarrow Y$, being able to generate k fuzzy rules associated with the fuzzy clusters:

$$\begin{array}{l} R_h: \text{ If } x \text{ is } A_h \text{ then } y \text{ is } f_h(x) \\ \text{where} \\ f_h(x) = a_{0h} + a_{1h}x_1 + ... + a_{ph}x_p \\ \text{with } h = 1, ..., k \end{array} \quad (1)$$

where $x = (x_1, ..., x_p)$ is the whole input variable and A_h fuzzy sets in X^p to be obtained from the hth fuzzy cluster C_h of $(X^p \times Y)$.

3 FUZZY LOCAL REGRESSION MODELS

Instead of using directly a classical fuzzy clustering algorithm like the Fuzzy C-Means [2], as is usual in the literature - an algorithm that is one of the most popular - we have adopted an alternative approach. We have used the Gustafson-Kessel fuzzy clustering algorithm [4] starting from the results of a previous application of a FCM method. This algorithm is based on the detection of hyperplane fuzzy clustering and thus it is adequate to detect the fuzzy partitions that better fulfill the assumption of fuzzy linear models [1]. The objective function of this algorithm is similar to the FCM, although it admits different shape and size in the fuzzy clusters.

$$J(U, V : X) = \sum_{i=1}^{N} \sum_{h=1}^{k} \mu_{hi}^m D_{hi,A_h}^2 \quad (2)$$

where N is the number of samples, k the number of fuzzy clusters that are established using a cluster validity measure, m is a smoothing parameter which controls the "fuzziness" of the clusters (normally between $1.75 - 2.5$), and $J(\cdot)$ the objective function that must be minimized with respect to the fuzzy partition U the centroids V and the data set X, and whose solution will be determined by solving by means of a Picard iteration the expressions:

$$v_h = \frac{\sum_{i=1}^{N} \mu_{hi}^m x_i}{\sum_{i=1}^{N} \mu_{hi}^m} \text{ with } h = 1, ..., k$$

$$\mu_{hi} = \frac{1}{\left[\sum_{j=1}^{k} \left(\frac{D_{hi,A_h}}{D_{ji,A_j}}\right)^{\frac{2}{m-1}}\right]} \text{ with } i = 1, ..., N \ h = 1, ..., k$$

$$D_{hi,A_h}^2 = (X_i - V_h)^T M_h (X_i - V_h)$$

$$M_h = (\rho_h \det (C_{fh}))^{\frac{1}{p}} C_{fh}^{-1}$$

$$C_{fh} = \frac{\sum_{i=1}^{N} \mu_{hi}^m (X_i - V_h)(X_i - V_h)^T}{\sum_{i=1}^{N} \mu_{hi}^m}$$

where v_h are the centroids of the fuzzy clusters, μ_{hi} is the membership degree of (x_i, y_i) to the $h-$th cluster C_h, C_{fh} is the fuzzy covariance matrix of this C_h. The optimal value would be such that $\det \left(M_{fh}^*\right) = \rho_h$

Once we have found the fuzzy clusters in $(X^p \times Y)$ by the Gustafson-Kessel algorithm, we need to determine A_h and $f_h(\cdot)$ of expression 1.

The fuzzy clusters in the product space $(X^p \times Y)$ give information about how the data are structured in X^p captured in the centroids and the metric found by the clustering on the product space, which is to be used to induce fuzzy regions in X^p. We define fuzzy sets with exponential membership functions in the domain X^p, using the components in this space of the centroids found in $(X^p \times Y)$ and the fuzzy covariances of the sample data components in X^p. The centroids of the clusters found in $(X^p \times Y)$ are noted as c_{XY}^h, and the components of these centroids in X^p as c_X^h.

Once we have found the fuzzy clusters in $(X^p \times Y)$ using the Gustafson-Kessel algorithm, one important difference of our approach from the classical TSK model, is that we will use directly the grade of membership of the data to the fuzzy clusters found in $(X^p \times Y)$, in order to obtain by means of the recursive least-squares algorithm (or a stable-state Kalman Filter), the coefficients of the linear regression model that will be associated to the fuzzy region corresponding to each fuzzy cluster.

This way we obtain a method:

$$GMBD \equiv \hat{y} = \frac{\sum_{h=1}^{k} \exp\left(-\frac{\left\|x - c_X^h\right\|_{C_{fh}}^2}{2}\right) * f_h(x)}{\sum_{h=1}^{k} \exp\left(-\frac{\left\|x - c_X^h\right\|_{C_{fh}}^2}{2}\right)} \quad (3)$$

where C_{fh} is the covariance matrix over the components in X^p associated with the hth cluster, and $f_h(\cdot)$ is the linear function with coefficient obtained by means of the recursive least-squares algorithm using the grade of membership of the data to the fuzzy clusters found in $(X^p \times Y)$.

The motivations for using exponential membership functions are that these membership functions are symmetric under rotation in the multidimensional space and therefore can be easily projected onto each domain of the variables. These projections can be used to generate fuzzy rules with fuzzy sets in each domain of discourse, this kind of fuzzy rules being more appropiate for linguistic interpretations.

4 NUMERICAL EXAMPLES

In this section we present the result of the different methods applied to two different examples: A) the Box-Jenkins gas furnace and B) the Zimmermann and Zysno model based on empirical data [13].

The performance index considered to evaluate the obtained fuzzy models will be the root mean square error:

$$E = \sqrt{\frac{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{N}} \quad (4)$$

where N is the number of data, y_t is the actual output, and $\hat{y}_t$ is the model output.

4.1 BOX-JENKINS GAS FURNANCE

Box and Jenkins's gas Furnace is a famous example of system identification. The well-known Box-Jenkins data set consists of 296 input-output observations, where the input $u(t)$ is the rate of gas flow into a furnace and the output $y(t)$ is the CO_2 concentration in the outlet gases. Since the process is dynamical, there are differents values of the variables that can be consider as candidates to affect the present output $y(t)$. In our case we have considered as inputs $y(t-1)$ and $u(t-4)$ as different authors have done [12], etc.

In table(1) we present the comparison of our method (GMBD) with several fuzzy models proposed in the literature, showing the number of variables chosen in each case, the number of rules and the performance of the models. In order to compare our results we carry out a fuzzy clustering on the 296 data points, considering 2 clusters.

We obtain a better performance with a number of rules remarkably smaller in comparison with other "classical" approaches. The Sugeno and Tanaka model, Sugeno (91) provide smaller error, although with more variables and more computational complexity.

4.2 ZIMMERMANN AND ZYSNO EMPIRICAL DATA

In this section we present the results of our method on the empirical data presented in [13] and that other

Table 1: Errors in Literature

Model	N. Inputs	N. Rules	Error
Tong(77)	2	19	0.684
Pedrycz(84)	2	81	0.565
Xu(87)	2	25	0.572
Peng(88)	2	49	0.548
Sugeno(91)	6	2	0.261
Sugeno(93)	3	6	0.435
Wang(96)	2	5	0.397
GMBD	2	2	0.396

authors like Dyckhoff and Pedrycz, Krishnapuram and Lee [6], or Langari and Wang [8] have also used to test their algorithms.

Table(2) shows the performance of the method suggested in the literature and our methods (with four rules as suggested by Langari et al.[8].)

Table 2: Errors in Literature

Model	PI/Error
Zimmermann and Zysno(80)	0.057
Dyckhoff and Pedrycz(84)	0.052
Krishnapuram and Lee(92)	0.051
Langari and Wang(96)	0.042
GMBD	0.034

Again in this example it is clear the good performance of our method.

5 CONCLUSIONS AND FUTURE TRENDS

The use of a fuzzy clustering algorithm like fuzzy C-means(FCM) in combination with the Gustafson-Kessel algorithm, results into an efficient method to identify a collection of fuzzy rules. That suggest us the future analysis of other kind of combinations between fuzzy clustering algorithms. Like all unsupervised techniques, clustering suffers from the presence of noise in the data. While this is not really a drawback when nothing is known about the nature and number of subgroups in the data set, it can be a serious problem in situations where one whishes to generate membership functions from training data. In order to treat this problem, the use of fuzzy clustering algorithm like the Possibilistic C Means (PCM) or its variants could be an interesting alternative.

References

[1] R. Babuska and H. B. Verbruggen, Applied fuzzy modeling, *Proceedings IFAC congress AIRTC,* Valencia, Spain, 1994.

[2] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms, Plenun, New York, 1981.

[3] M. Delgado, A. F. Gómez-Skarmeta and F. Martín Using fuzzy clusters to model fuzzy systems in a descriptive approach , *Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU'96,* , Granada, Spain, 1996.

[4] D. E. Gustafson and W. C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, *Proc. of the IEEE Int. Conf. on Fuzzy Systems,* San Diego, pp. 761-766, 1979.

[5] F. Klawonn and R. Kruse, Constructing a fuzzy controller from data, *Fuzzy Sets and Systems,* vol. 85, pp. 177-193, 1997.

[6] R. Krishnapuram and J. Lee, Fuzzy connective-based hierarchical aggregation networks for decision making, *Fuzzy Sets and Systems*, vol 46, pp. 11-27, 1992

[7] A. Kroll, Identification of functional fuzzy models using multidimensional reference fuzzy sets, *Fuzzy Sets and Systems*, vol. 80, pp. 149-158, 1996

[8] R. Langari and L. Wang, Fuzzy models, modular networks, and hybrid learning, *Fuzzy Sets Syst.*, vol. 79, pp. 141-150, 1996

[9] M. Sugeno and T. Yasukawa, A Fuzzy-logic-based approach to qualitative modeling, *IEEE Trans. Fuzzy Systems*, vol 1, n.1, pp. 7-31, 1993.

[10] T. Takagi and M Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man, and Cybernetics,* vol. SMC-15, n. 1, pp. 116-132, 1985.

[11] R. R. Yager and D. P. Filev, Unified structure and parameter identification of fuzzy models, *IEEE Trans. on System Man and Cybernetics*, vol. 23, pp. 1198-1205, 1993

[12] Y. Yoshinari, W. Pedrycz and K. Hirota, Construction of fuzzy models through clustering techniques, *Fuzzy Sets and Systems,* 54, pp. 157-165, 1993.

[13] H-J. Zimmermann and P. Zysno, Latent connective in human decision making, *Fuzzy Sets and Systems*, vol 4, pp. 37-51, 1980