Identifying Fuzzy Inference Models by Means of Possibilistic Clustering: Socio-Economic Applications

Alberto Guillén[1], Federico Montesino[2], Angel Barriga[3],
L.J. Herrera[1], J. González[1], H. Pomares[1] and I. Rojas[1] *

1- Dept of Computer Technology and Architecture
University of Granada - Spain

2- Aalto University SST - Dpt. of Information and Computer Science
P.O. Box 15400, FI-00076 Aalto - Finland

3- University of Seville - Dpt. of Electronics and Electromagnetism
Avda. Reina Mercedes s/n, E-41012 Seville - Spain

Abstract.
This paper analyzes the use of clustering methods for the identification of fuzzy inference models for regression problems which is a problem frequently raised in the socio-economic research field. Traditional, fuzzy and hybrid fuzzy-possibilistic approaches to clustering are compared. In particular, Improved Clustering for Function Approximation (ICFA), and Fuzzy-Possibilistic Clustering for Function Approximation (FPCFA). The methods proposed are applied to three datasets concerning stock prices and housing prices. The experiments show that both ICFA and FPCFA compare favorably against well-established traditional clustering methods for fuzzy inference systems identification.

1 Introduction

The problem of approximating a given function using a model $\mathcal{F}$ can be formulated as, given a set of observations $\{(\vec{x}_k; y_k); k = 1, ..., n\}$ with $y_k = F(\vec{x}_k) \in$ and $\vec{x}_k \in {}^d$, it is desired to obtain a function $\mathcal{F}$ so $\sum_{k=1}^{n} ||y_k - \mathcal{F}(\vec{x}_k)||^2$ is minimum. In order to solve this problem, Fuzzy logic based and neuro-fuzzy modeling techniques are appealing because of their interpretability and potential to address a broad spectrum of problems. In particular, fuzzy inference systems (FIS) exhibit a combined description and prediction capability as a consequence of their rule-based structure [1].

The problem of identifying fuzzy inference rules can be addressed by data-driven techniques for identification of fuzzy systems from numerical examples.

*FM is supported by a Marie Curie Intra-European Fellowship for Career Development (grant agreement PIEF-GA-2009-237450) within the European Community´s Seventh Framework Programme (FP7/2007–2013). We are indebted to CSC-IT Center for Science Ltd. for its support and high-performance computing services. CSC is administered by the Ministry of Education of Finland. This work has been partially supported by the Spanish CICYT Project TIN2007-60587 and Junta Andalucia Project P07-TIC-02768. This work was supported in part by project TEC2008-04920/MICINN from the Spanish Ministry of Science and Innovation, as well as project P08-TIC-03674 from the Andalusian regional Government.

Two approaches are often distinguished in the literature: structure-oriented and clustering-based.

In this paper we analyze the use of clustering-based identification methods for fuzzy inference systems in the context of regression problems. Fuzzy and hybrid fuzzy-possibilistic clustering approaches are followed.

The rest of the paper remains as follow: in the next section, we discuss the structure of FIS identified by means of a fuzzy clustering methods as well as the fuzzy-possibilistic approach used previously. Afterwards, in section 3, we apply clustering-based identification approaches to socio-economic regression datasets. Finally, some conclusion remarks are given.

2 Fuzzy Inference System Identification

In order to build a regression model, a fuzzy inference system can be defined as a mapping between a vector of crisp inputs and a crisp output. Let us denote the inputs for observations $(\vec{x}_k; y_k)$ of a certain regression problem as scalar values $x_1, \ldots, x_M$. This way, assuming all the inputs (M) are used, the fuzzy regression model can be expressed as a set of N fuzzy rules of the following form:

$$R_i : \text{IF } x_1 \text{ is } L_1^i \text{ AND } x_2 \text{ is } L_2^i \text{ AND } \ldots \text{ AND } x_M \text{ is } L_M^i \text{ THEN } \hat{y} \leftarrow \mu_{R_i},$$

where $i = 1, \ldots, N$, and the fuzzy sets $L_j^i \in \{L_{j,k}\}, k = 1 \ldots, n_j, j = 1, \ldots, M$, with n_j being the number of linguistic labels defined for the jth input variable. L_j^i are the fuzzy sets representing the linguistic terms used for the jth input in the ith rule of the fuzzy model. μ_{R_i} are the consequents of the rules and can take different forms. For example, in a system with two inputs, if L_1^i is renamed LOW_1 and L_2^i is renamed $HIGH_2$, the ith rule R_i, would have the following form:

$$R_i : \text{IF } x_1 \text{ is } LOW_1 \text{ AND } x_2 \text{ is } HIGH_2 \text{ THEN } \hat{y} \leftarrow \mu_{R_i}.$$

Depending on the fuzzy operators, inference model and type of membership functions (MFs) employed, the mapping between inputs and outputs can have different formulations. In principle, the methods proposed in this paper can be applied for any combination of types of MFs, operators and inference model, but the selection can have a significant impact on practical results.

As a concrete implementation for this paper, we use the minimum as T-norm for conjunction operations, Gaussian MFs for inputs, singleton outputs, and product inference of rules. Defuzzification is performed using the fuzzy mean method, i.e., zero-order Takagi-Sugeno systems [2] are defined. Thus, the result of the inference process is a weighted average of the singleton consequents. This inference scheme was chosen in order to keep systems as simple and interpretable as possible. In particular, the use of singleton outputs simplifies both the interpretation of rules and its local optimization.

Therefore, in this particular case a fuzzy regressor can be formulated as follows:

$$\mathcal{F}(\bar{x}) = \frac{\sum_{i=1}^{N} \left(\mu_{R_i} \cdot \min_{1 \le j \le M} \mu_{L_j^i}(x_j) \right)}{\sum_{i=1}^{N} \min_{1 \le j \le M} \mu_{L_j^i}(x_j)},$$

where N is the number of rules in the rule base, μ_{R_i} are singleton output values, and $\mu_{L_j^i}$ are Gaussian MFs for the inputs. Thus, each fuzzy set defined for the input linguistic terms, $L_{j,k}$ (for the kth linguistic term defined for the jth input), is characterized by an MF having the following form:

$$\mu_{L_{j,k}} = exp\left[-(y_j - c_{k,j})^2/2\sigma_{k,j}^2\right],\ k = 1 \ldots, n_j,\ j = 1, \ldots, M, \qquad (1)$$

where $c_{k,j}$ and $\sigma_{k,j}$ are scalar values and represent the centers and widths of the inputs MFs, respectively.

Fuzzy inference systems of the class being designed here are universal approximators [3, 4]. Thus, for a sufficiently large number of rules and MFs, any input-output mapping should be approximated with arbitrary accuracy.

2.1 Clustering-Based Identification of Fuzzy Inference Systems

Different approaches to the identification of fuzzy inference systems from numeric data have been proposed in the literature [5, 6]. Roughly, two classes of methods can be distinguished: structure-oriented and clustering-based.

In this paper we focus on the clustering-based class of methods and specially on those methods that follow an offline approach. The following clustering algorithms are compared for the purposes of identifying fuzzy inference systems: The Hard and Fuzzy C-means (HCM and FCM, respectively) [7] clustering algorithms, the Improved Clustering for Function Approximation (ICFA) algorithm [8], and the hybrid Fuzzy-Possibilistic Clustering for Function Approximation [9]. The latter two algorithms were originally proposed for initializing Radial Basis Function Neural Networks (RBFNNs) for regression problems. In this paper, the $ICFA_f$ variant, tailored for fuzzy inference systems identification, is used [10].

The first step for clustering-based identification of fuzzy inference systems is to apply a clustering algorithm on the input-output patterns. Once this process finishes, Q clusters have been identified. Then, the structure of the corresponding fuzzy inference systems has to be defined. In general, fuzzy rules can be interpreted as joint constraints [6] rather than implication rules. Thus, it is sensible to define a fuzzy rule from each cluster identified. This is the most frequent approach in the literature. This way, the clusters and their corresponding rules are considered as prototypes or models of the whole input pattern sequence.

Let us consider as above a multiple scalar input, single scalar output case where the input patterns to the clustering algorithm consist of M inputs and one output. Let us denote the clusters identified by $\vec{c}_k$, $i = 1, \ldots, Q$. Let every cluster have the following general form:

$$\vec{c}_k : (c_{k,1}, \ldots, c_{k,M+1}), \text{ with } k = 1, \ldots, Q,$$

where the $c_{k,M+1}$ correspond to the outputs of the fuzzy inference model, whereas the $c_{k,1}, \ldots, c_{k,M}$ correspond to the inputs $(x_1, \ldots, x_M)$ to the fuzzy model. For each cluster, a matching rule is generated with the following form:

$$R_k : \text{IF } x_1 \text{ is } L_{1,k} \text{ AND } x_2 \text{ is } L_{2,k} \text{ AND } \ldots \text{ AND } x_M \text{ is } L_{M,k} \text{ THEN } \hat{y} \leftarrow c_{k,M+1}, \quad k = 1, \ldots, Q, \; Q = N = n_j,$$

where a set of input linguistic terms is created $\{L_{j,k}\}, k = 1 \ldots, n_j, j = 1, \ldots, M,$. These linguistic terms are defined by Gaussian MFs, $\mu_{L_{j,k}}$, as in equation 1. The output membership functions are defined as singleton functions centered at the corresponding element of the cluster centers, $c_{k,M+1}$. The centers of the input Gaussian MFs for the jth input and kth rule ($c_{k,j}$ in equation 1) are set to the jth elements of the corresponding clusters $\vec{c}_k$.

When inference systems are identified with clustering methods following this approach, the number of linguistic terms defined for every input variable, $n_j, j = 1, \ldots, M$, is equal to the number of clusters identified, Q, which in turn is equal to the number of rules identified, N. Hence, Q different membership functions are generated for each input and output variable, and Q rules are generated for horizon h.

The way the widths of the input Gaussian MFs ($\sigma_{k,j}$ in equation 1) are set depends on the clustering algorithm used. For the Hard C-means and Fuzzy C-means algorithms the widths are set as a function of the membership degrees of the input patterns to the clusters.

Recently, an adaptation of the ICFA [8] algorithm for identification of FIS was proposed [10]. This adaptation, ICFA$_f$, is a simple generalization of the original ICFA proposal where all the widths for a certain rule are set to a value inversely proportional to the average weighting parameter w.

The ICFA algorithm performs an initialization of the centers of the clusters, taking into account the output of the function to be approximated. The way in which the output is considered is by defining a value for each center in the output space. This value is named expected output (o_i) of a center i and allows the algorithm to weight the distance computed between the input vectors and each center.

2.2 Fuzzy-Possibilistic approach

As was shown in [9], the combination of possibilistic and fuzzy membership functions could lead to a better center initialization for RBFNNs.

The development of the FPCFA algorithm relied in the approach presented in [11] where a combination of a fuzzy partition and a possibilistic partition is used. The authors assert that the membership value of the fuzzy partition is important to be able to assign a hard label to classify an input vector, but at the same time, it is very useful to use the typicality (possibility) value to move the centers properly in presence of outliers. Let $U^p = [u^p_{ik}]$ be the matrix containing all the possibilistic memberships, $U^f = [u^f_{ik}]$ the matrix containing the fuzzy memberships, and $C = [\vec{c_i}]$ the matrix containing the center positions for $i = 1...m$ and $k = 1...n$. The distortion function to be minimized is:

$$J_{h_f,h_p}(U^f, C, U^p; X) = \sum_{k=1}^{n}\sum_{i=1}^{m}((u^f_{ik})^{h_f} + (u^p_{ik})^{h_p})D^2_{ik} \tag{2}$$

with the following constraints:

$$\sum_{i=1}^{m} u^f_{ik} = 1 \ \forall k = 1...n \tag{3}$$

$$\sum_{k=1}^{n} u^p_{ik} = 1 \ \forall i = 1...m \tag{4}$$

Let $U^p = [u^p_{ik}]$, then, the constraint shown above requires each row of U^p to sum up to 1 but its columns are free up to the requirement that each column contains at least one non-zero entry. Therefore, there is a possibility of input vectors not belonging to any cluster. The design of the FPCFA algorithm weighted the similarity criteria used in the computation of the distances and defining an expected output for each center, so the distortion function to be optimized remained as:

$$J_{h_f,h_p}(U^f, C, U^p, W; X) = \sum_{k=1}^{n}\sum_{i=1}^{m}((u^f_{ik})^{h_f} + (u^p_{ik})^{h_p})D^2_{ikW} \tag{5}$$

restricted to the same constraints than the FPCM algorithm.

The iteration method to minimize considered the following equations to compute the membership and expected output:

$$u^f_{ik} = \left(\sum_{j=1}^{m}\left(\frac{D_{ikW}}{D_{jkW}}\right)^{\frac{2}{h_f-1}}\right)^{-1} \tag{6}$$

$$u^p_{ik} = \left(\sum_{j=1}^{n}\left(\frac{D_{ikW}}{D_{ijW}}\right)^{\frac{2}{h_p-1}}\right)^{-1} \tag{7}$$

$$\vec{c_i} = \frac{\sum_{k=1}^{n}((u^f_{ik})^{h_f} + (u^p_{ik})^{h_p})\vec{x}_k w^2_{ki}}{\sum_{k=1}^{n}((u^f_{ik})^{h_f} + (u^p_{ik})^{h_p})w^2_{ki}} \tag{8}$$

$$o_i = \frac{\sum_{k=1}^{n}((u^f_{ik})^{h_f} + (u^p_{ik})^{h_p})y_k d^2_{ik}}{\sum_{k=1}^{n}((u^f_{ik})^{h_f} + (u^p_{ik})^{h_p})d^2_{ik}} \tag{9}$$

The algorithm iterates until the centers have not moved significantly.

3 Experiments

The identification methods described above have been applied to three regression benchmarks related to socio-economic applications: The Stocks Domain, Boston Housing, and California Housing datasets. The following approach was followed in order to test the performance of different identification methods. The observations of the datasets are randomly randomly rearranged and two thirds of the complete set of observations is selected for training models, while the remaining third is selected for testing purposes.

The overall properties of the three problems are shown in table 1. Overall, FCM yields better results than HCM. Both ICFA and FPCFA can improve the results obtained with FCM in general, with exceptions for certain numbers of clusters. The most accurate results are obtained using FPCFA for a sufficiently large number of clusters.

Dataset	Inputs	Training length	Test length
Stocks Domain	9	634	316
Boston Housing	13	338	168
California Housing	8	13760	6880

Table 1: Characteristics of Datasets.

In what follows we summarize the results obtained for the different clustering methods described above. The experiments have been performed using the Xfuzzy environment for the design of fuzzy inference systems [12, 13]. For the regression approximation, errors are given as normalized root mean square error (RMSE) values, where the normalization is performed against the variance of the output variable.

3.1 Stocks Domain

The Stocks Domain dataset consists of daily stock prices for ten aerospace companies recorded from January 1988 through October 1991, and can be obtained from the StatLib repository [14].

Table 2 reports the results obtained for this dataset. The lowest test error achieved is highlighted in bold face. In this case, both HCM and FCM exhibit a higher sensitivity to the number of clusters selected. These limitations are overcome by both ICFA and PCFA. In particular, FPCFA can provide the best accuracy for a relatively small number of clusters.

3.2 Boston Housing

The Boston Housing contains various parameters about housing values in suburbs of Boston. The inputs variables include the per capita crime rate by town,

Clusters	HCM	FCM	$ICFA_f$	FPCFA
2	0.398±0.260	0.544±0.356	0.309±0.184	0.346±0.211
3	0.327±0.191	0.538±0.352	0.263±0.166	0.295±0.180
4	0.541±0.425	0.278±0.170	0.302±0.194	0.225±0.130
5	0.288±0.187	0.248±0.149	0.254±0.160	0.243±0.143
6	0.324±0.189	0.215±0.136	0.215±0.138	0.231±0.152
7	0.238±0.152	0.188±0.110	0.218±0.132	0.216±0.120
8	0.246±0.159	0.213±0.136	0.184±0.113	0.175±0.110
9	0.223±0.145	0.242±0.156	0.188±0.117	0.192±0.116
10	0.311±0.225	0.177±0.115	0.214±0.136	0.204±0.125
15	0.180±0.122	0.186±0.129	0.175±0.110	**0.140±0.087**

Table 2: Results for Stocks Domain. Average and standard deviation of the test RMSE (normalized)

index of accessibility to radial highways, pupil-teacher ratio by town, and % lower status of the population, among others. The output variable to be modeled is the Median value of owner-occupied homes. The dataset can be obtained from the UCI Machine Learning Repository [15].

The results obtained are shown in table 2. Both ICFA and FPCFA provide overall better results than HCM and FCM. FPCFA consistently provides the most accurate results for a wide range of number of clusters.

Clusters	HCM	FCM	$ICFA_f$	FPCFA
2	0.779±0.535	0.775±0.539	0.487±0.319	0.492±0.328
3	0.611±0.407	0.585±0.445	0.526±0.327	0.397±0.248
4	0.497±0.343	0.792±0.516	0.597±0.418	0.417±0.278
5	0.607±0.398	0.628±0.445	0.631±0.433	0.435±0.298
6	0.643±0.450	0.703±0.540	0.792±0.543	0.374±0.268
7	0.668±0.504	0.585±0.417	0.401±0.262	**0.343±0.236**
8	0.615±0.440	0.527±0.378	0.428±0.291	0.359±0.244
9	0.546±0.378	0.610±0.424	0.435±0.302	0.405±0.289
10	0.606±0.446	0.645±0.482	0.405±0.280	0.387±0.269
15	0.565±0.407	0.622±0.468	0.391±0.272	0.426±0.319

Table 3: Results for Boston Housing. Average and standard deviation of the test RMSE (normalized).

3.3 California Housing

The California Housing Datasets contains information collected from the 1990 Census. The data describes all the block groups in California, where a block group on average includes 1425.5 individuals living in a geographically compact area. The dataset can be obtained from the StatLib repository [14]. The inputs variable include information about the location, median income, median age, number of rooms, and number of householders, among others.

Table 4 summarizes the results obtained. Both HCM and FCM provide a slightly lower test error for sufficiently high numbers of clusters. Overall, errors are similar for the 4 clustering methods, with only two exceptions (FCM for 2 clusters and FPCFA for 4 clusters).

Clusters	HCM	FCM	$ICFA_f$	FPCFA
2	0.666±0.448	1.031±0.612	0.686±0.452	0.637±0.432
3	0.612±0.426	0.625±0.426	0.616±0.422	0.624±0.435
4	0.592±0.414	0.604±0.417	0.665±0.448	0.762±0.499
5	0.595±0.413	0.594±0.412	0.600±0.423	0.597±0.415
6	0.596±0.413	0.600±0.412	0.634±0.435	0.585±0.414
7	0.589±0.410	0.585±0.405	0.584±0.412	0.585±0.414
8	0.573±0.397	0.566±0.398	0.581±0.407	0.580±0.408
9	0.561±0.394	0.570±0.401	0.597±0.422	0.534±0.381
10	0.567±0.391	0.551±0.391	0.583±0.407	0.566±0.400
15	0.522±0.372	**0.514±0.366**	0.593±0.414	0.552±0.388

Table 4: Results for California Housing. Average and standard deviation of the test RMSE (normalized).

4 Conclusions

We have discussed the use of fuzzy and hybrid fuzzy-possibilistic clustering methods for identifying fuzzy inference systems. Four clustering methods, HCM, FCM, ICFA and FPCFA, were compared on datasets concerning stock prices and housing statistics. Both ICFA and FPCFA, recently proposed for function approximation problems, compare favorably against well-established traditional clustering methods for fuzzy inference systems identification.

References

[1] Jerry M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall PTR, January 2001. ISBN: 0130409693.

[2] Hung T. Nguyen and Nadipuram R. Prasad, editors. *Fuzzy Modeling and Control: Selected Works of M. Sugeno*. CRC Press, Inc, March 1999.

[3] Hao Ying. Sufficient Conditions on Uniform Approximation of Multivariate Functions by General Takagi-Sugeno Fuzzy Systems with Linear Rule Consequent. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 28(4):515–520, July 1998.

[4] Jyh-Shing Roger Jang, Chuen-Tsai Sun, and Eiji Mizutani. *Neuro-Fuzzy and Soft Computing A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Upper Saddle River, New Jersey, September 1997. ISBN 0-13-261066-3.

[5] Sushmita Mitra and Yoichi Hayashi. Neuro-fuzzy rule generation: survey in soft computing framework. *IEEE Transactions on Neural Networks*, 11(3):748–768, 2000.

[6] Leszek Rutkowski. *Flexible Neuro-Fuzzy Systems. Structures, Learning and Performance Evaluation.* Kluwer Academic Publishers, Boston, MA, USA, 2004. ISBN: 1-4020-8042-5.

[7] J. Valente de Oliveira and W. Pedrycz, editors. *Advances in Fuzzy Clustering and its Applications.* John Wiley & Sons, Ltd., West Sussex, England, 2007. ISBN: 978-0-470-02760-8.

[8] A. Guillén, J. González, I. Rojas, H. Pomares, L. J. Herrera, O. Valenzuela, and A. Prieto. Using fuzzy logic to improve a clustering technique for function approximation. *Neurocomputing*, 70(16–18):2853–2860, October 2007.

[9] A. Guillén, H. Pomares, I. Rojas, J. González, L. J. Herrera, F. Rojas, and O. Valenzuela. Studying possibility in a clustering algorithm for RBFNN design for function approximation. *Neural Computing & Applications*, 17(1):75–89, August 2007.

[10] Federico Montesino Pouzols and Angel Barriga Barros. Automatic clustering-based identification of autoregressive fuzzy inference models for time series. *Neurocomputing*, to appear.

[11] N. R. Pal, K. Pal, and J. C. Bezdek. A mixed c–means clustering model. In *Proceedings of the 6th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'97)*, volume 1, pages 11–21, Barcelona, July 1997.

[12] The Xfuzzy Development Environment: Fuzzy Logic Design Tools. https://forja.rediris.es/projects/xfuzzy, http://www.imse-cnm.csic.es/Xfuzzy, December 2009.

[13] Francisco José Moreno-Velo, Iluminada Baturone, Angel Barriga, and Santiago Sánchez-Solano. Automatic Tuning of Complex Fuzzy Systems with Xfuzzy. *Fuzzy Sets and Systems*, 158(18):2026–2038, September 2007.

[14] Dataset Archive, StatLib Repository. http://lib.stat.cmu.edu, January 2009.

[15] A. Asuncion and D. J. Newman. UCI Machine Learning Repository, January 2010. University of California, Irvine, Center for Machine Learning and Intelligent Systems.