Data-Driven Fuzzy Modeling and Control with Kernel Density Based Clustering Technique

P. A. Kowalski[1*], S. Łukasik[1*], M. Charytanowicz[2**], P. Kulczycki[1*]

[1] Department of Automatic Control, Cracow University of Technology, Warszawska 24, 31-155 Krakow, Poland
[2] Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warszawa, Poland

Abstract

Deriving parameters and structure of fuzzy model for a dynamical system by means of a clustering procedure is a very popular and frequently applied technique in fuzzy identification. The aim of the paper is to present a novel method of fuzzy model formulation based on this approach. Introduced algorithm is based on clustering method employing nonparametric kernel density gradient estimation. Proposed technique is automatic and attains functionality free from arbitrary assumptions concerning „shapes" of data samples, the number of rules and any other user defined parameters. Illustrative results of computer simulations using MATLAB scientific environment are enclosed. The outcome of such experimental verification demonstrates high efficiency of proposed technique in fuzzy controllers synthesis and nonlinear systems modeling.

Keywords: fuzzy modeling, fuzzy control, clustering, kernel density estimation.

Introduction

Fuzzy modeling introduced by Takagi and Sugeno [1] is a powerful identification method based on fuzzy sets theory. It has been successfully applied in several research areas i.e. control engineering [2], prediction tasks [3], systems analysis [4]. Fuzzy reasoning in above-mentioned methodology, in its I order variant, is accomplished by using a set of fuzzy rules in form: *if* x_1 *and* x_2 *is* A_{i2} *then out* $= C^T_i x + D_i$ *for* $i = 1,2,K,R$, where A_{ij} represents membership functions, C_i and D_i are the local linear function parameters and R denotes the number of rules.

The problem of extracting fuzzy rules from experimental data is solved using different techniques i.e. evolutionary algorithms [5], neuro-fuzzy methods [6] or statistical approaches [7]. The most popular clustering technique applied in this area of research is the subtractive clustering procedure proposed by Chiu [8]. The general concept of this density based algorithm has been successfully applied as well in some other fuzzy modeling techniques. In this paper the novel alternative approach of obtaining rules prototypes by means of clustering based on nonparametric kernel estimation is proposed and positively evaluated. Preliminary results in this field were presented in cited reference [9].

The paper is organized as follows. In the next section some methodological preliminaries of the introduced technique are given. This part of the contribution contains a short description of statistical kernel density estimation (KDE) and the gradient clustering algorithm based on KDE. Subsequent part of the paper covers the details of fuzzy model construction based on obtained clusters parameters. Section *Numerical simulations* includes the results of numerical experiments involving the comparison of method under investigation with state-of-art fuzzy modeling techniques. Finally some summarizing remarks are given in the last section.

[1]*e-mail: {pkowal, szymonl, kulczycki}@pk.edu.pl
**e-mail: mchmat@ibspan.waw.pl

Methodological Preliminaries

Statistical Kernel Density Estimation

In most practical data exploration tasks the probability density function f of given sample is multimodal and can be hardly mapped into a function of any typical distribution. Therefore methods of nonparametric density estimation which do not need any assumptions on the distribution type are used commonly. The kernel density estimation is one of the classical techniques with such property [10, 11].

Consider a n-dimensional random variable, with a distribution having the density f. Its kernel estimator $\hat{f}: \mathrm{R}^n \to [0,\infty)$ based on the m-elements data sample x_1, x_2,K ,x_m, can be defined as:

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where the positive coefficient h is called a smoothing parameter, while the measurable function , of unit integral, symmetrical with respect to zero, and having a weak global maximum at this point is called a kernel.

Here the Gaussian kernel $K(x) = 1/(2\pi)^{1/2} \exp(-x^2/2)$ will be used. In the multidimensional case it will be generalized to the product kernel notation:

$$\begin{aligned} K(x) &= K([\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_n]^{\mathrm{T}}) = \\ &= \mathrm{K}(\mathrm{x}_1)\cdot\mathrm{K}(\mathrm{x}_2)\cdot\ldots\cdot\mathrm{K}(\mathrm{x}_n) \end{aligned} \tag{2}$$

where K constitutes the one-dimensional Gaussian kernel given above. As a result smoothing parameter takes a form of a vector $h = [h_1, h_2, ..., h_n]$. It can be easily obtained using automatic smoothing selection procedures i.e. plug-in method [12]. More detailed information about the practical issues of KDE methods and usage examples can be found in cited references [13].

Kernel Density Estimation Clustering

KDE clustering belongs to the class of density gradient-based methods [14]. In the introduced algorithm following natural assumptions are made. First, each cluster center should be represented by a local maximum of KDE calculated for given dataset. Furthermore, argument corresponding to the first non-zero minimum of probability density function obtained for distances between data elements ought to describe cluster's radius. The algorithm relying on above-mentioned proposals consists of two phases.

In the initial stage of the clustering procedure every point xi of the considered sample is moved according to a search direction indicated by gradient $\nabla\hat{f}$:

$$x_i^{(k+1)} = x_i^{(k)} + b\frac{\nabla\hat{f}(x_i^{(k)})}{\hat{f}(x_i^{(k)})} \quad for\ i=1,2,\ldots,m\ , \tag{3}$$

$$x_i^{(0)} = x_i$$

The parameter $b = [b_1, b_2,$K$, b_n]^T$ defines a speed of data relocation according to the following formula:

$$b_i = \frac{h_j^2}{3} \tag{4}$$

This part of the clustering algorithm ends when the next condition is fulfilled:

$$\frac{\left\| D^{(k)} - D^{(k-1)} \right\|}{D^{(0)}} \le 0.001 \tag{5}$$

where $D^{(k)}$ is a sum of distances between sample elements in each of k algorithm's iterations.

The subsequent part of KDE clustering algorithm consists of calculation of distances between all input data elements $x_i^{(k)}$ and afterwards - obtaining the probability estimation function (1) for such distance-based sample. The smallest argument of this function ensuring its local minimum (except zero) is determined and such distance $d_{\min}$ serves then as a half of distance between clusters. Mapping elements to target clusters is performed in the following manner: each pair of points belong to the same cluster when the distance between them is lower then $d_{\min}$. This stage of the clustering procedure is definitely most computationally exhausting, especially for large sample sizes, as one have to obtain distance-based density estimates from $(m-1) \cdot m/2$ elements. To deal with this problem it is proposed to use either parallel processing [15] to speed up associated computation time or condense the sample using effective „pruning" algorithm [16].

Detailed description of the clustering algorithm, its properties and usage guidelines can be found in the referenced bibliography [17].

Fuzzy Identification with KDE Clustering

To extract the rules of fuzzy model at first the data set representing input u_1, u_2,K ,u_m and output y_1, y_2,K ,y_m values of some system under convideration is separated into clusters using introduced clustering technique. Consider a set of c cluster centers $\{x_1, x_2,$K$, x_c\}$ where $x_i = \{u_{ci1}, u_{ci2},K, u_{cin_x}, y_{ci1}, y_{ci2},K, y_{cin_y}\}$. Each vector u_{c_i} derived from cluster centre can be put into a fuzzy rule i, with a given degree to which this rule is fulfilled:

$$\mu_i(u) = e^{-\frac{4}{r_a^2}\left\|u - u_{c_i}\right\|^2} \tag{6}$$

where positive parameter r_a allows to control the generalization ability of resulting fuzzy inference system. During method's experimental evaluation it was established that as a „rule of thumb" the value $r_a = 1/c$ could be used.

The output vector y of modeling system is given as:

$$y = \frac{\sum_{i=1}^{c} \mu_i(u) y_{c_i}}{\sum_{i=1}^{c} \mu_i(u)} \tag{7}$$

with y_{ci} being a linear function of input variables as shown in the Introduction. The consequents of this function can be easily established as finding their optimal values constitutes a simple least-squares estimation problem [8].

Numerical Simulations

Modeling of Nonlinear Dynamical Systems

For testing purposes as an example of nonlinear plant modeling a benchmark problem, taken from Narendra et al. paper [3] was used. The plant is given by the second order highly nonlinear differential equation:

$$y(k)=\frac{y(k-1)y(k-2)[y(k-1)+2.5]}{1+y^2(k-1)+y^2(k-2)}+u(k) \quad (8)$$

The task is to use past values of control signal u and output y up to time step k to predict the value of $y(k)$. The model with three inputs $\{u(k),y(k-2),y(k-1)\}$ and one output $y(k)$ was under consideration. As a training data 500 elements $\{u(k),y(k-2),y(k-1),y(k)\}$ with $u(k)$ uniformly distributed over the interval $[-2, 2]$ were used. Testing data was generated for $u(k) = \sin(2\pi\, k/25)$ (testing set 1) and $u(k) = 1{,}6 \cos(2\pi\, k/30)$ (testing set 2).

Introduced fuzzy modeling technique was compared with subtractive clustering method and with Adaptive-Network-Based Fuzzy Inference System (ANFIS) [18] procedure. For KDE clustering algorithm $h = 1.6h_{plug-in}$ was assumed. Subtractive clustering was executed with $r_a = 0.4$ and ANFIS with 100 epochs and two membership functions for each input. Those values provided the possibility to compare the efficiency of the algorithms at the same level of the rules number. The outcome of such comparison is presented in Table 1. It includes as well some previously published results [18].

Analysis of methods' performance leads to the conclusion that the presented method ensures high modeling efficiency with compact number of rules incorporated in a fuzzy system. In the course of performed numerical tests KDE clustering procedure achieved high generalization ability similar to the one shown by GA technique [18]. Yet it was attained with significantly reduced rule set. At the fixed level of fuzzy system complexity the introduced technique offers better efficiency than state-of-art subtractive clustering and ANFIS methods.

HDD Servo-Motor Fuzzy Control

The method under investigation was tested as well for a hard-drive servo motor control design [19]. The following continuous state-space rigid model of the servo system was used:

$$\begin{bmatrix}\dot{y}(t)\\ \dot{v}(t)\end{bmatrix}=\begin{bmatrix}1 & 1.664\\ 0 & 1\end{bmatrix}\begin{bmatrix}y(t)\\ v(t)\end{bmatrix}+\begin{bmatrix}1.384\\ 1.664\end{bmatrix}u(t)$$
$$z(t)=\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}y(t)\\ v(t)\end{bmatrix} \quad (9)$$

where u is the actuator input (in volts), y and v are the position (in tracks) and velocity of the disk drive's head. The problem of accurate positioning was under consideration with the output $z(t) = y(t)$.

First the standard 49-rules PD fuzzy logic controller (FLC) [20] was properly tuned for quick response in time domain with the step reference signal. Obtained 101 input-output data elements $\{e,\Delta e,u\}$ were used for the clustering procedure. As a result new FLC with reduced rule base was synthesized. For comparison the performance of classical PID controller [20]:

$$G(s)=\frac{0.13s^2+221.8s+6.052\cdot 10^{-10}}{s^2+5545s} \quad (10)$$

was under experimental verification as well. The results of numerical simulations for all above-mentioned controllers are enclosed in Table 2. The number of fuzzy system's rules (if appropriate) was reported, as well as the root-mean square error versus the reference signal, settling time $T_{s,\,2\%}$ and the percentage overshoot of the output's response.

Results of such simulations in the time domain are also illustrated in Fig. 1. It can be seen that even though the application of the introduced KDE-based clustering technique does not ensure the minimal value of the root-mean square error, the fuzzy logic controller synthesized using this method is the quickest one and it achieves significantly smallest overshoot.

Table 1. Comparison of fuzzy modeling algorithms in a nonlinear plant identification problem.

Method	KDE-Based Clustering	Subtractive Clustering	ANFIS	Genetic Algorithm [18]	Least-squares + Gradient Descent [18]	Fuzzy C-Means Clustering [18]
Input variables	u(k), y(k-2), y(k-1)					
No of rules	10	10	8	75	12	8
Learning MSE	0.086	0.041	0.005	0.037	0.507	0.618
Testing MSE (set 1)	0.031	0.112	0.010	0.040	0.245	0.204
Testing MSE (set 2)	0.035	0.306	1.619	0.037	not reported	not reported

Table 2. Comparison of controllers performance.

Controller	Rules	RMSE	$T_{s,2\%}[s]$	OV
KDE-based FLC	38	0.2129	0.0013	11%
Classical PID	-	0.2909	0.0096	78 %
PD FLC	49	0.1981	0.0020	92 %

Conclusion

In this paper a novel approach for modeling of complex-dynamical systems using fuzzy logic based structure was proposed. Applying presented method of estimating cluster centers from numerical data lead to the construction of accurate procedure for fuzzy model identification. Its efficiency is comparable (or even superior) to the one demonstrated by other fuzzy modeling techniques and it was positively verified in prediction and automatic control tasks.

Further work in the subject could concern the improved method of mapping the obtained clusters structure to the rules based set, taking into consideration a shape of each cluster in an input-output space.

References

1. TAKAGI T., SUGENO M., Fuzzy Identification of Systems and its Applications to Modeling and Control, IEEE Transactions on Systems, Man and Cybernetics, **15**:116-132, **1985**.
2. CHOPRA S., MITRA R., KUMAR V., Analysis of PI and PD-Type Controllers Using Subtractive Clustering, International Journal of Computational Cognition, **4**(2):30-34, **2006**.
3. NARENDRA K. S., PARTHASARATHY K.,: Identification and Control of Dynamical Systems Using Neural Networks, IEEE Transactions on Neural Networks, **1**:4-27, **1990**.
4. TRABELSI A., LAFONT F., KAMOUN M., ENEA G., Identification of Nonlinear Multivariable Systems by Adaptive Fuzzy Takagi-Sugeno Model, International Journal of Computational Cognition, **2**:137-153, **2004**.
5. PAPADAKIS S. E., THEOCHARIS I. B., A GA-Based Fuzzy Modeling Approach for Generating TSK Models, Fuzzy Sets and Systems, **131**:121-152, **2002**.
6. JANG J.-S. R., ANFIS: Adaptive-Network-Based Fuzzy Inference System, IEEE Transactions on Systems, Man and Cybernetics, **23**:665-684, **1993**.
7. YEN J., WANG L., Application of Statistical Information Criteria for Optimal Fuzzy Model Construction, IEEE Transactions on Fuzzy Systems, **6**:362-372, **1998**.
8. CHIU S. L., Fuzzy Model Identification Based on Cluster Estimation, Journal of Intelligent and Fuzzy Systems, **2**:267- 278, **1994**.
9. ŁUKASIK S., KOWALSKI P., CHARYTANOWICZ M., KULCZYCKI P., Fuzzy Model Identification Using Kernel-Density-Based Clustering, in Proceedings of the Sixth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets, Warsaw, **2007** [In Print].
10. KULCZYCKI P., Estymatory jądrowe w analizie systemowej, WNT, Warsaw, **2005** [In Polish].
11. WAND M. P., JONES M. C., Kernel Smoothing, Chapman and Hall, London, **1995**.
12. SHEATHER S. J., JONES M. C., A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation, Journal of the Royal Statistical Society. Series B. Methodological **53** (3):683-690, **1991**.
13. KULCZYCKI P., Kernel Estimators in Industrial Applications, in B. Prasad, editor, Soft Computing Ap-

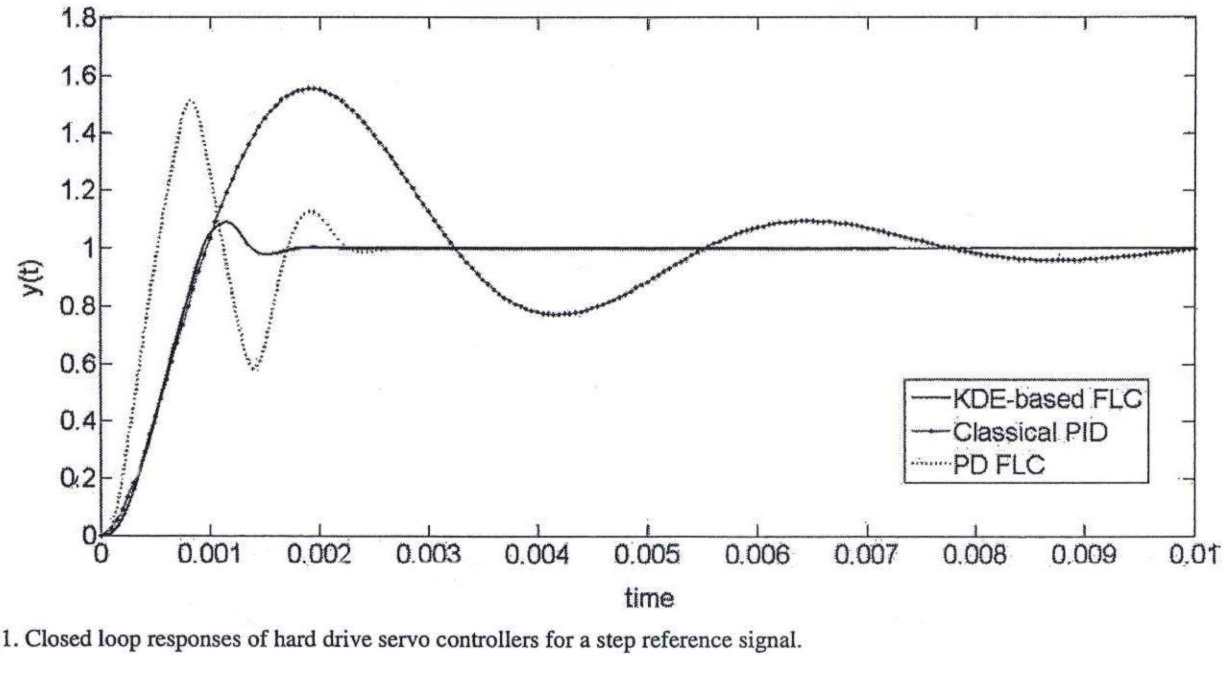

Fig 1. Closed loop responses of hard drive servo controllers for a step reference signal.

plications in Industry, pages 69-93, Springer-Verlag, **2008**.

14. FUKUNAGA K., HOSTETLER L. D., The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition, IEEE Transactions on Information Theory, **21**:32-40, **1975**.
15. ŁUKASIK S., Parallel Computing of Kernel Density Estimates with MPI, Lecture Notes in Computer Science, **4489**:726-734, **2007**.
16. MITRA P., MURTHY C.A., PAL S. K., Density-Based Multiscale Data Condensation, IEEE Transactions on Pattern Analysis and Machine Intelligence, **24** (6): 734-747, **2002**.
17. KULCZYCKI P., CHARYTANOWICZ M., A Complete Gradient Clustering Algorithm by Kernel Estimators, **2008** [In Press].
18. FARAG W. A., QUINTANA V. H., LAMBERT-TORRES G., A Genetic-Based Neuro Fuzzy Approach for Modeling and Control of Dynamic Systems, IEEE Transactions on Neural Networks, **9**: 756-766, **1998**.
19. TAN K. C., SATHIKANNAN R., TAN W.W., LOH A.P., Evolutionary Design and Implementation of a Hard Disk Drive Servo Control System, Soft Computing, **11**:131 -139, **2007**.
20. MUDI R., PAL N. R., A Robust Self-Tuning Scheme for PI and PD Type Fuzzy Controllers, IEEE Transactions on Fuzzy Systems, **7**:2-16, **1999**.