

A New Approach for Online Fuzzy Identification by Potential Clustering Including Rule Reduction

A. karimodдини, K. Salahshoor, A. Fatehi and M. Karimadini

Abstract— This paper uses potential clustering approach to perform online fuzzy clustering. This method is an improvement of the subtractive clustering which is a noniterative clustering algorithm and so is suitable for online applications. In spite of all capabilities of the potential clustering, this method suffers from a major disadvantage. The number of clusters grows fast when the sensitivity of the algorithm is increased. In this article an innovative technique has been proposed to reduce the number of clusters. The proposed method is applied to the Macky-Glass benchmark. It is shown although the number of clusters is reduced; the resulting performance will not be affected.

I. INTRODUCTION

THIS paper introduces a new method for online fuzzy clustering to be used in fuzzy identification. The conventional techniques are difficult to implement and sometimes impracticable for modeling and identification of nonlinear dynamic systems, some of the main reasons are as follows:

- They are model-based approaches and hence require an appropriate dynamic structure to be pre-defined.
- Some of their dynamic information will be neglected due to usual simplification of the defined model structure.
- They are based on a prior knowledge to set some effective parameters.

Therefore techniques based on fuzzy logic are used widely [1]. Among the different fuzzy methods, the Takagi-Sugeno model (TS) has attracted most attention [2].

The methods for learning TS models from data are based on the idea of consecutive structure and parameter identification. Identification of the parameters of the consequent part in TS model is handled by RLS (Recursive Least Squared). Structure identification includes estimation of the focal points of the rules (antecedent parameters) by fuzzy clustering [3]. Identification of the antecedent part of the TS model could be solved by the methods such as fuzzy C-means [4] and Gustafson-Kessel (GK) [5] which are objective based and Mountain Method [6],[7] and Subtractive clustering method

[3],[8] which are nonobjective base. These iterative methods, however, have some shortages and due to their iterative feature can not be used for online methods. In on-line mode the training data are collected continuously, rather than being available as a fixed batch set. Some of the new data reinforce and confirm the information contained in the previous data. While, other data bring new information, which could indicate a change in the operating conditions, development of a fault or simply a more significant change in the dynamic of the process [9],[10],[11]. They may possess enough new information to form a new cluster or to modify an existing one [3]. Real-time on-line applications are hampered by the need to recursive calculation of the model parameters. In this paper, an improved version of the online identification approach based on evolving TS fuzzy models [3] is proposed. The method includes an online clustering methodology to indicate the system dynamic variation in the form of a gradual change of the rule-base and a modified (weighted) RLS algorithm to estimate the free parameters in the consequent part of the TS model. The proposed method is often called as potential fuzzy identification or evolving rule-based (eVR) method in the literature.

II. POTENTIAL FUZZY IDENTIFICATION

A. Online potential clustering

The on-line clustering procedure starts with the first data point established as the focal point of the first cluster. Its coordinates are used to form the antecedent part of the fuzzy rule using Gaussian membership functions. Any other type of membership functions could also be used instead. Its potential is assumed to be equal to 1. Following with the next data points onwards, the potentials of the new data points are calculated using a Cauchy type function of first order [11],[12]:

$$P_k(z_i) = \frac{1}{1 + \frac{1}{(k-1)} \sum_{i=1}^{k-1} \sum_{j=1}^{n+1} (d_{ik}^j)^2} \quad (1)$$

for $k = 2, 3, \dots$ where $P_k(z_i)$ denotes the potential of the data point $(z_i = [x_i^T; y]^T)$ calculated at time k ; $d_{ik}^j = z_i^j - z_k^j$, denotes projection of the distance between two data points, z_i^j and z_k^j , on the axis x^j (x^j for $j = 1, 2, \dots, n$ and on the axis y for $j = n + 1$). This function is monotonic and inversely proportional to the distance and enables recursive calculation, which is important for on-line implementation of the learning algorithm. The potential of

Manuscript received October 12, 2006

A. karimodдини is with the Automation and Instrumentation Department, Petroleum University of Technology, Tehran, Iran (e-mail: akarimodдини@gmail.com).

K. Salahshoor, is with the Automation and Instrumentation Department, Petroleum University of Technology, Tehran, Iran (e-mail: Dr_Karim_Salahshoor@yahoo.com).

A. Fatehi is with the Electrical Engineering Department, KNTOosi University of Technology, Tehran, Iran (e-mail: fatehi@kntu.ac.ir).

M. Karimadini is with the Electrical and Electronic department, Universiti Putra Malaysia, serdang, Malaysia (e-mail: karimadini@gmail.com)

the new data sample can be recursively calculated as follows:

$$P_k(z_i) = \frac{k-1}{(k-1)(\mathcal{G}_k + 1) + \sigma_k - 2\nu_k} \quad (2)$$

Where :

$$\mathcal{G}_k = \sum_{j=1}^{n+1} (z_k^j)^2 \quad (3)$$

$$\sigma_k = \sum_{i=1}^{k-1} \sum_{j=1}^{n+1} (z_i^j)^2 \quad (4)$$

$$\nu_k = \sum_{j=1}^{n+1} z_k^j \sum_{i=1}^{k-1} z_i^j \quad (5)$$

Parameters \mathcal{G}_k and ν_k are calculated from the current data point z_k , while β_k^j and σ_k are recursively updated as follows:

$$\sigma_k = \sigma_{k-1} + \sum_{j=1}^{n+1} (z_{k-1}^j)^2 \quad (6)$$

$$\beta_k^j = \beta_{k-1}^j + z_{k-1}^j \quad (7)$$

After the new data are available in on-line mode, they influence the potentials of the centers of the clusters (z_i^* , $i = [1, R]$), which are respective to the focal points of the existing rules (x_i^* , $i = [1, R]$). The reason is that by definition the potential depends on the distance to all data points, including the new ones.

The recursive formula to update the potentials of the existing clusters' focal points can easily be derived using the following relation:

$$P_k(z_i^*) = \frac{(k-1)P_{k-1}(z_i^*)}{k-2 + P_{k-1}(z_i^*) + P_{k-1}(z_i^*) \sum_{j=1}^{n+1} (d_{k(k-1)}^j)^2} \quad (8)$$

Where, $P_k(z_i^*)$ is the potential of the cluster at time k , which is a prototype of the i^{th} rule. The proof can be seen in the appendix B.

Potentials of the new data points are compared to the updated potential of the centers of the existing clusters. If the potential of the new data point is higher than the potential of the existing centers then the new data point is accepted as a new center and a new rule is formed with a focal point based on the projection of this center on the x axis ($R = R + 1; x_R^* = x_k$). The condition for creating a new rule by new data is:

$$\text{If } P_K(z_k) > P_K(z_\ell^*); \ell = 1, \dots, R \\ \text{then } (R = R + 1; x_R^* = x_k) \quad (9)$$

The rationale is that in this case the new data point is more descriptive, has more summarization power than all the other previous data points. It should be noted that the condition to

have higher potential is a very strong one. The reason is that with the growing number of data, their concentration is usually decreasing except in the cases some new important region of data space reflecting a new operating regime or new condition appears. In such cases, a new rule is formed, while outlying data are automatically rejected because their potential is significantly lower due to their distance.

If in addition to the previous condition the new data point is close to an old center then the new data point replaces this rule or in the other words:

$$\text{If } P_K(z_k) > P_K(z_\ell^*) \text{ and} \quad (9) \\ \frac{\|z_k - \arg_{\ell=1}^R \min \|z_k - z_\ell^*\|\|}{\text{radii}} + \frac{\max_{\ell=1}^R P_K(z_\ell^*)}{P_K(z_k)} < 1$$

Then ($z_\ell^* = z_k$); $\ell = 1, \dots, R$

This is the mechanism for rule-base adaptation called modification and ensures a replacement of a rule with another one built around the projection of the new data point on the x axis.

It should be noted that using the potential concept instead of the distance to a certain rule center results in rules that are more informative and a more compact rule-base. The reason is that the spatial information and the history are not ignored, but are part of the decision whether to upgrade or modify the rule-base.

This on-line clustering [11],[12] ensures an evolving fuzzy rule-based system by dynamically upgrading and modifying it while inheriting the bulk of the rules ($R-1$ of the rules are preserved even when a modification or an upgrade take place).

A. Online fuzzy identification of TS model

The problem of fuzzy TS model identification could be divided into the following two steps:

1) Identification of the antecedent part of the model consists of determination of the centers and spreads of the membership functions.

2) Identification of the parameters of the consequent part in TS model.

The antecedent part of the model is identified using the potential clustering approach. The free parameters of the consequent part of the TS model (part 2) are identified by the RLS algorithm [13].

First, according appendix C, online prediction of the output is expressed in terms of the past input-output as follows:

$$\hat{y}_{k+1} = \psi_k^T \hat{\theta}_k \quad k = 2, 3, \dots \quad (10)$$

The proof of the above formula can be seen in the appendix C.

Then the following RLS procedure (called also the Kalman filter) is applied to estimate the unknown free parameters of the TS model:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + C_k \psi_k (y_k - \psi_k^T \hat{\theta}_{k-1}) \quad k=2,3,\dots \quad (12)$$

$$C_k = C_{k-1} - \frac{C_{k-1} \psi_k \psi_k^T C_{k-1}}{1 + \psi_k^T C_{k-1} \psi_k} \quad (13)$$

with initial conditions:

$$\hat{\theta}_1 = [\hat{\pi}_1^T, \hat{\pi}_2^T, \dots, \hat{\pi}_R^T]^T = 0; \quad C_1 = \Omega I \quad (11)$$

Hence when a new rule is added to the rule-base, the RLS algorithm is reseted in the following way:

1) Parameters of the new rule are determined by weighted average of the parameters of the other rules. The weights are the normalized firing levels of the existing rules λ_i . The idea is to use the existing centers as a rule-base to approximate the initialization of the parameters of the new rule by a weighted sum. Parameters of the other rules are inherited from the previous step:

$$\hat{\theta}_k = [\hat{\pi}_{1(k-1)}^T, \hat{\pi}_{2(k-1)}^T, \dots, \hat{\pi}_{R(k-1)}^T, \hat{\pi}_{R+1(k-1)}^T]^T \quad (12)$$

Where:

$$\hat{\pi}_{R+1(k)} = \sum_{i=1}^R \lambda_i \hat{\pi}_i(k-1) \quad (13)$$

2) Covariance matrices are reseted as:

$$C_k = \begin{bmatrix} \rho \zeta_{1,1} & \dots & \rho \zeta_{1,R(n+1)} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho \zeta_{R(n+1),1} & \dots & \rho \zeta_{R(n+1),R(n+1)} & 0 & \dots & 0 \\ 0 & 0 & 0 & \Omega & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \Omega \end{bmatrix} \quad (14)$$

Where ζ_{ij} is an element of the covariance matrix for R rule base. In this way, the part of the covariance matrix associated with the new $(R+1)^{th}$ rule (last $n+1$ columns and last $n+1$ rows) is initialized as usual (with a large number Ω) in its main diagonal. In addition:

$$\rho = \frac{R^2 + 1}{R^2} \quad (15)$$

Where ρ is a coefficient, respective for the rest of the rules (from 1 to R) which are updated by multiplication of ρ .

When a rule is replaced by another one, which has antecedent parameters close to the rule being replaced, then parameters and co-variance matrices are inherited from the previous time

step. Now, the potential clustering algorithm can be summarized as follows [13]:

Step1: The rule-base structure including antecedent part of the rules for the first new data as follows:

$$k = 1; R = 1; x_1^* = x_1; P_1(z_1^*) = 1 \quad (16)$$

$$\theta_1 = \pi_1 = 0; C_1 = \Omega I \quad (17)$$

where z_1^* is the first cluster center; x_1^* is a focal point of the first rule being a projection of z_1^* on the x axis.

Of course the rule-base can be initialized by using an expert knowledge or based on off-line identification approaches. In this case:

$$R = R^{ini}; P_1(z_i^*) = 1; i = 1, \dots, R^{ini} \quad (18)$$

Where R^{ini} denotes the number of rules defined initially off-line.

Step2: At the next time step reading the next data sample z_k is collected;

Step3: The potential of each new data sample is calculated recursively to influence the structure of the rule-base according to (2).

Step4: The potentials of old centers are updated recursively taking into account the influence of the new data sample according to (12).

Step5: The potential of the new data sample is compared to the updated potential of existing centers and a decision whether to modify or upgrade the rule-base is taken.

(a) **If**

The potential of the new data point is higher than the potential of the existing centers:

$$P_k(z_k) > P_k(z_i^*); \quad i = 1, \dots, R \quad (19)$$

AND

The new data point is close to an old center according to (9)

THEN

The new data point (z_k) replaces it.

In this case, the new data point is used as a prototype of a focal point. Let us suppose that it has index j , indeed:

$$z_j^* = \arg \min_{i=1}^R \|z_k - z_i^*\| \quad (20)$$

Then x_j^* is equaled to x_k and the potential of the regarded focal point, $P_k(z_j^*)$, is as $P_k(z_k)$.

Consequence parameters and covariance matrices are inherited from the rule to be replaced:

$$\hat{\pi}_k = \hat{\pi}_k^j, \quad C_k = C_k^j \quad (21)$$

Since the new center is close to the replaced one by definition, this change is marginal. The disturbance caused by this change could be ignored, because the Kalman filter is able to cope with using the existing estimations of the parameters and covariance matrices.

(b) ELSE IF

The potential of the new data point is higher than the potential of the existing centers:

$$P_k(z_k) > P_k(z_i^*) ; \quad i = 1, \dots, R \quad (22)$$

THEN

it is added to the rule-base as a new rule's center.

In this case, the new data point becomes a prototype of a focal point of a new rule:

$$R = R + 1 \quad (23)$$

$$x_{R+1}^* = x_k$$

$$P_k(z_{R+1}^*) = P_k(z_k)$$

Consequence parameters and co-variance matrices are reset by $C_{(R+1,k)} = \Omega I$.

END IF.

Step6: The consequent parameters are estimated recursively by RLS, for globally optimal parameters.

Step7: Prediction of the model output for the next time step (y_{k+1}).

$$\hat{y}_{k+1} = \psi_k^T \hat{\theta}_k ; \quad k = 2, 3, \dots \quad (24)$$

The algorithm continues from stage 2 by reading the next data sample at the next time-step. All steps are non-iterative. Some of the advantageous of this approach can be summarized as follows:

1. Using the approach a transparent, compact and accurate model can be found by rule base evolution based on experimental data with the simultaneous recursive estimation of the fuzzy set parameters.

2. The proposed rule base uses non-iterative training, which is very appropriate for real-time applications (calculations for each time-step take only in a fraction of a second).

3. It is interesting to note that the rule addition with new data does not lead to an excessively large rule base. The reason for this is that the condition for the new data point to have higher potential than the focal points of all existing rules is a hard requirement. Additionally, the possible proximity of a candidate center to the already existing focal points leads to just a replacement of the existing point, i.e. modification of its coordinates without enlarging the rule-base size.

III. A NEW APPROACH FOR RULE REDUCTION IN POTENTIAL FUZZY IDENTIFICATION

A new online approach is proposed to controls the number of rules. In the beginning of the identification, it is natural that some rules may be produced due to the fast changes in the dynamics of the system. These erratic initial changes can be exogenous by the specific algorithm initialization conditions. After some while, when the algorithm is more tuned to the actual situation, the rule-base may be ended up with some inactivity in number of the identified rules.

Therefore, an improvement to the potential fuzzy identification can be achieved by recognizing and deleting the non-efficient rules. The rule-production mechanism is active in all times and the deleted rules can be created if necessary.

To identify the inactive rules, the following procedure performed when any data sample enters in the identification algorithm, the potential of the centers should be updated according to (12). Indeed the more a rule is active the more the center potential of its regarded cluster is increased. If we exert a forgetting factor in the potential formula as bellow:

$$MP_K(z_k^*) = (1 - \gamma)MP_{k-1}^* + \gamma\Delta \quad (25)$$

Where MP is a variable for monitoring of the cluster center potential and γ is a forgetting factor and Δ is differential between the cluster center potential in two steps. The amount of stimulation of any rule can be monitored when a new data enters. If MP be lower than a threshold after some time, the corresponding rule will be omitted.

IV. SIMULATION

This method is implemented on a well-known benchmark problem known as Macky-Glass.

The Macky-Glass chaotic time series is generated from the Macky-Glass differential delay equation defined by [15],[3]:

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (26)$$

The aim is to use the past values of x to predict some future value of x . we assume $x(0) = 1.2$, $\tau = 17$ and the value of the signal 85 step ahead $x(t+85)$ is predicted based on the

values of the signal at the current time, 6, 12 and 18 steps back. As a result it can be written:

$$\begin{aligned} \text{output} &= [x(t + 85)] \\ \text{inputs} &= [x(t - 18), x(t - 1), x(y - 6), x(t)] \end{aligned} \quad (27)$$

First the original potential fuzzy identification algorithm is applied to this benchmark problem. The resulting outcome has been demonstrated in Fig. 1.

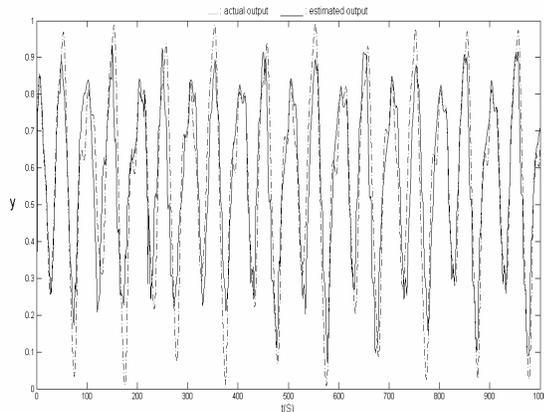


Fig. 1. The output of the clustering method

The validation criteria for the whole data set are as (see appendix A):

$$\text{RMSE}=0.484 \quad \text{NDEI}=0.4681 \quad \text{VAF}=78.0724$$

For the last 500 data set, after identification completes, these criteria will be as:

$$\text{RMSE}=0.1093 \quad \text{NDEI}=0.4313 \quad \text{VAF}=81.3669$$

The number of the created rules is increased as shown in Fig. 2.

As depicted, the algorithm can lead to a large number of rules where some of them might be inactive.

If these inactive rules be modified, identification process will not be influenced while the speed and the amount of calculations can be improved in large extent.

To demonstrate this disadvantage of the original potential clustering algorithm, the levels of thresholds are decreased to obtain more accuracy in the identification. However, this results in a large number of rules and consequently require more computational time, as shown in Table I.

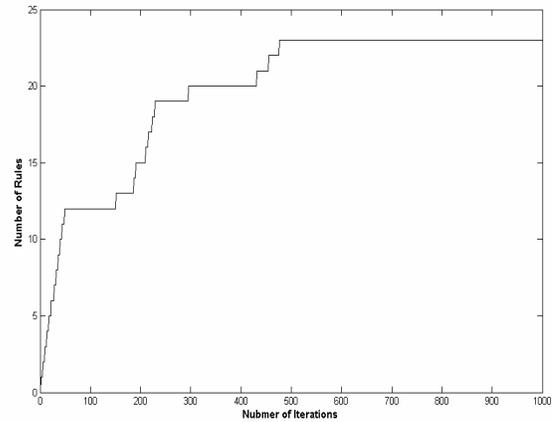


Fig. 2. increasing in the number of rules

TABLE I
RESULTS OF THE ORIGINAL POTENTIAL CLUSTERING ALGORITHM WHEN THE LEVELS OF THRESHOLDS ARE DECREASED

r	RMSE	NDEL	VAF	No. rules
0.5	0.1205	0.4745	77.5730	14
0.45	0.1170	0.4623	78.6428	17
0.4	0.1125	0.4445	80.2229	21
0.35	0.1057	0.4179	82.5533	30
0.3	0.1010	0.3991	84.0973	44
0.25	0.0889	0.3515	87.6606	61
0.2	0.0845	0.3337	88.9141	71

As shown in table I, the number of rules is increased with more precision requirements.

Next, the proposed rule reduction approach is utilized for identification of the Macky-Glass benchmark problem and the obtained results have been shown in Table II:

TABLE II
RESULTS OF RULE REDUCTION APPROACH ON THE MACKY-GLASS BENCHMARK PROBLEM

R	Rmse	NDEL	VAF	No. rules
0.5	0.1201	0.4745	77.5730	14
0.45	0.1170	0.4623	78.6428	17
0.4	0.125	0.4445	80.2225	21
0.35	0.1062	0.4195	82.3920	29
0.3	0.1018	0.4023	83.8147	41
0.25	0.0893	0.3530	87.5945	46
0.2	0.0795	0.3140	90.1469	53

The summarized results in tables I and II demonstrate that more accurate performance has been achieved with less number of created rules.

For radius $r=0.2$, the output and the number of created rules during simulation study, are depicted in Fig. 3 and 4, respectively.

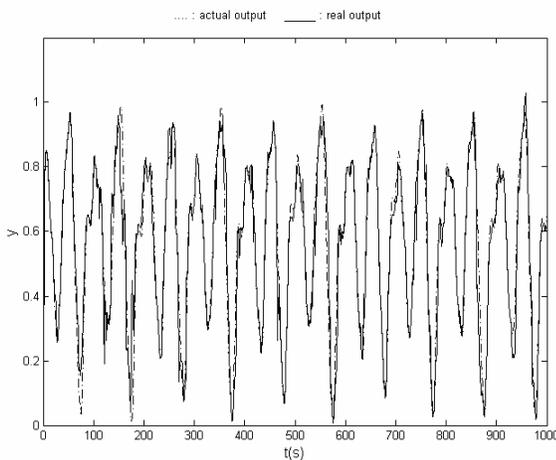


Fig. 3. Potential identification, $r=0.2$

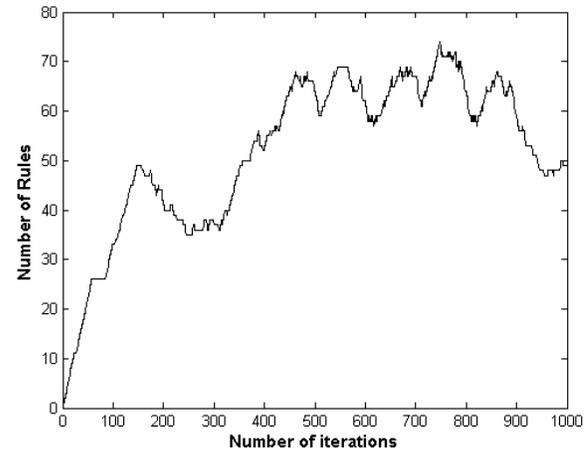


Fig. 4. Rule reduction using forgetting factor scheme

V. CONCLUSION

The basic online potential fuzzy identification algorithm is prone a lot of rules during the initial commissioning phase mainly due to non-exact parameter initialization. As a result, some of the created rules may be non-active during the continuing identification process. An improved algorithm was proposed in order to recognize and delete these non-active rules. Besides, the algorithm can re-create these rules if necessary afterwards. The basic and the improved algorithms were applied to Macky-Glass benchmark. It was demonstrated that improved algorithm is able to produce more accurate results with more less number of rules.

APPENDIX A

The following evaluation measures were used to validate the performance of the identified models :

i. VAF

In [15], a criterion is proposed for validation. VAF is a function which computes the percentile variance accounted for between two signals as follow:

$$VAF = \left[1 - \frac{Var(y_{act} - y_{est})}{Var(y_{act})} \right] . 100\% \tag{A.1}$$

y_{act} is the output of the plant and y_{est} is the output of the model. If the signals differ, VAF is lower.

ii. RMSE

Root mean square is a famous criterion for error measurement which is defined as bellow:

$$Rmse = \sqrt{\frac{\sum (y_{act} - y_{est})^2}{N}} \tag{A.2}$$

iii. NDEI

In [3] NDEI is an appropriate criterion for model validation. To evaluate the performance of the models we used the NDEI (None Dimensional Error Index) defined as the ratio of the root mean square error over the standard deviation of the target data, as bellow:

$$NDEI = \frac{RMSE(y_{act} - y_{est})}{std(y_{act})} \quad A.3$$

APPENDIX B

In (8), a recursive formula was introduced to update the potentials of the existing clusters' focal points. The proof of this formula is as follows:

Assume that the $(k-1)^{th}$ data sample is admitted to be the i^{th} cluster center or $z_{k-1} = z_i^*$.

From (1):

$$P_{k-1}(z_i^*) = \frac{k-2}{(k-2) + \sum_{i=1}^{k-2} \sum_{j=1}^{n+1} (d_{i(k-1)}^j)^2} \quad (B.1)$$

Therefore:

$$(k-2)P_{k-1}(z_i^*) + P_{k-1}(z_i^*) \sum_{i=1}^{k-2} \sum_{j=1}^{n+1} (d_{i(k-1)}^j)^2 = (k-2) \quad (B.2)$$

Consequently:

$$\sum_{i=1}^{k-2} \sum_{j=1}^{n+1} (d_{i(k-1)}^j)^2 = \frac{(k-2)}{P_{k-1}(z_i^*)} - (k-2) \quad (B.3)$$

In k^{th} step:

$$P_k(z_i^*) = \frac{k-1}{(k-1) + \sum_{i=1}^{k-1} \sum_{j=1}^{n+1} (d_{ik}^j)^2} \quad (B.4)$$

or:

$$P_k(z_i^*) = \frac{k-1}{(k-1) + \sum_{i=1}^{k-2} \sum_{j=1}^{n+1} (d_{i(k-1)}^j)^2 + \sum_{j=1}^{n+1} (d_{k(k-1)}^j)^2} \quad (B.5)$$

Substitute (B.3) in (B.5), this results in :

$$P_k(z_i^*) = \frac{(k-1)}{(k-1) + (k-2) \left(\frac{1}{P_{k-1}(z_i^*)} - 1 \right) + \sum_{j=1}^{n+1} (d_{k(k-1)}^j)^2} \quad (B.6)$$

Now Multiply numerator and denominator of (B.6) in $P_{k-1}(z_i^*)$, the equation (8) is obtained. The proof is completed.

APPENDIX C

TS model as a quasi-linear systems

The TS model consists of if-then rules as bellows:

$$\mathfrak{R}^i : IF(x_1 \in \mu_{i1}) \text{ and } (x_2 \in \mu_{i2}) \dots \text{ and } (x_n \in \mu_{in}) \\ \text{Then } (y_i = a_{i0} + a_{i1}x_1 + \dots + a_{in}x_n \quad i = 1, \dots, R$$

Where \mathfrak{R}^i is i^{th} fuzzy rule, R is the number of fuzzy rules and x is the input vector, $x = [x_1, x_2, \dots, x_n]^T$, μ_{ij} denotes the antecedent fuzzy sets, $j = \{1, \dots, n\}$, y_i is the output of the i^{th} linear subsystem and a_{ij} are its parameters, $i = \{0, \dots, n\}$. Membership functions are selected as:

$$\mu_{ij} = e^{-\alpha \|x_j - x_{ij}^*\|} \quad i = \{1, R\}, j = \{1, n\} \quad (C.1)$$

Where $\alpha = 4/r^2$ and r is a positive constant, which define the spread of the antecedent and the zone of influence of the i^{th} rule (radius of the neighborhood of a data point); too large a value of r leads to averaging and too small a value of r leads to overfitting; values of $r \in [0.3, 0.5]$ is recommended [3].

x_i^* is the focal point of the i^{th} rule antecedent.

The firing level of the rules is defined as Cartesian product or conjunction of respective fuzzy sets for this rule:

$$\tau_i = \mu_{i1}(x_1) \cdot \mu_{i2}(x_2) \cdot \dots \cdot \mu_{in}(x_n) = \prod_{j=1}^n \mu_{ij}(x_j) \quad (C.2)$$

Where the TS model output is calculated by weighted averaging of individual rules' contributions as below:

$$y = \sum_{i=1}^R \lambda_i y_i = \sum_{i=1}^R \lambda_i x_e^T \pi_i \quad (C.3)$$

Where $\lambda_i = \frac{\tau_i}{\sum_{j=1}^R \tau_j}$ is the normalized firing level of the i^{th}

rule, y_i represents the output of the i^{th} linear model, $\pi_i = [a_{i0}, a_{i1}, \dots, a_{in}]^T$, $i = [1, R]$, is the vector of parameters of the i^{th} linear model, $x_e = [1 \ x^T]^T$ is the expanded data vector.

$$\theta = [\pi_1, \pi_2, \dots, \pi_R]^T \quad (C.4)$$

$$\psi = [\lambda_1 x_e^T, \lambda_n x_e^T, \dots, \lambda_n x_e^T]^T \quad (C.5)$$

$$y = \psi^T \theta \quad (C.6)$$

REFERENCES

- [1] J.Q.Chen, J.Chen, An online identification algorithm for fuzzy systems, *Fuzzy Sets and Systems*, 1994, pp.63-72.
- [2] T.Takagi, and M.Sugeno, "Fuzzy Identification of Systems and its Application to Modeling and Control", *IEEE Transaction on Systems, Man and Cybernetics*, Vol.15, 1985, pp 116-132
- [3] P.P.Anegelev, and D.P.Filev, "An Approach to Online Identification of Takagi-Sugeno Fuzzy Models", *IEEE Transaction on Systems, Man and Cybernetics*, Vol.34, No.1, FEBRUARY 2004.
- [4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [5] I. Gath, A. B. Geva, Unsupervised optimal fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7 (1989), pp 773-781.
- [6] Yager, R. and Filev, D. Generation of fuzzy rules by mountain clustering. *J. of Intelligent and Fuzzy Systems* 2, 209-219, 1994
- [7] J.T.Rickard, R.R.Yager, and W.Miller, "Mountain Clustering on Nonuniform Grids", *Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop (AIPR'04)* 1550-5219/04 \$ 20.00 IEEE
- [8] Stephen 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. Parade, and R. Yager, JohnWiley & Sons, 1997
- [9] P.P. Angelov, *Evolving Rule-based Models: A Tool for Design of Flexible Adaptive Systems*, Springer, Physica-Verlag, Heidelberg, Germany, 2002
- [10] P.P. Angelov, R.A. Buswell, Identification of evolving fuzzy rule-based models, *IEEE Transactions on Fuzzy Systems* 10 (2002) 667-677
- [11] P.P. Angelov, D.P. Filev, An approach to on-line identification of Takagi-Sugeno fuzzy models, *IEEE Transactions on System, Man and Cybernetics, Part B*, vol. 33, December, 2003, to appear.
- [12] P.P. Angelov, D.P. Filev, Flexible models with evolving structure, in: *Proceedings of the IEEE Symposium on Intelligent Systems IS'02*, Varna Bulgaria, 10-12 September, vol. 2, 2002, pp. 28-33
- [13] P.Angelov, "A fuzzy controller with evolving structure", *Information Sciences* 161 (2004) 21-35, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK, accepted 18 March 2003
- [14] S.L. Chiu, "Fuzzy model identification based on cluster estimation", *J. Intel. Fuzzy Syst.*, vol.2, pp.267-278, 1994.
- [15] J.C. Bezdek and S.K. Pal, *Fuzzy Models for Pattern Recognition*, IEEE Press, New York, 1992.