UNIVERSITY OF OSTRAVA

Institute for Research and Applications of Fuzzy Modeling

Construction and Errors Estimation of Locally Linear Fuzzy Models

Jiří Močkoř

Research report No. 7

November 6, 1997

Submitted/to appear:

–

Supported by:

Grant 201/96/0985 of GA ČR

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
Bráfova 7, 701 03 Ostrava 1, Czech Republic

tel.: +420-69-622 2808 fax: +420-69-22 28 28
e-mail: Mockor@osu.cz

Construction and Errors Estimation of Locally Linear Fuzzy Models *

Jiří Močkoř

January 29, 1998

Abstract

A method is presented for a construction of a Takagi-Sugeno fuzzy model of a process described by a discrete system in a form $\mathcal{D} = \{(x_{t,1}, \dots, x_{t,m}, y_t) : t \in T\}$, where $(x_1, \dots, x_m)$ are inputs and y is an output. A method is based on some fuzzy partition of this discrete system. Various error functions are estimated for global output function of this fuzzy model.

1 Introduction

In most studies of identification of processes by using its input-output data, it is assumed that there exists a global functional structure between the input and the output. In many real models this global functional structure is representated by linear relation and statistical methods are used to identify the paramethers of this linear regression.

It is, however, very difficult and unrealistic to substitute a nonlinear process by a single linear regression and , hence, very frequently a more sophisticated approaches are considered. All these approaches are well known under the common name *fuzzy controllers*. The fuzzy controller consists of series of fuzzy control rules each of which takes the form of a "*if...then...*" sentence. The quality of performance of this fuzzy controller depends greatly on the reasonable structure of the control rules and to find corresponding control rules represents one of the most important problem in this theory.

Basically there are three applicable methods for deriving these control rules.

1. Methods based on human operator's experience,
2. Methods based on a modelling of human operator's controlling actions, and
3. Methods based on a model of a process.

Let us recall very roughly some principal characterizations and descriptions of all these methods. We will suppose for simplicity that a process under the control consists of one state variable X (like temperature, speed, etc.) and assume that m numerical information may be derived (like value of this state variable, speed of changes of values of X, etc.). By $x(t)$ we denote the process output of the state variable X at time t. Moreover, from this process output $x(t)$ another m values can be derived, say $x_1(t), \dots, x_m(t)$, which then describe a complex behaviour of a process. All these vectors

$$\vec{x}(t) = \big(x_1(t), \dots, x_m(t)\big)$$

then represent input information for fuzzy controller. Values of $x_i(t)$ are not considered to be a continuous functions, their values are mostly calculated at some time $t = \tau, 2.\tau, \dots$, where τ is a sampling period of the process. It should be observed that among values $x_i(t)$ some target values of state variable X are contained frequently, mostly in a form

$$x_j(t) = \frac{d^{j-1}(x(t) - x_0)}{dt^{j-1}}$$

*The author was partially supported by the grant 201/96/0985 of GA ČR.

where x_0 is the target value of X, or sometimes we use simply

$$x_j(t) = x\big(t - (j-1)\cdot\tau\big).$$

For simplicity we may assume that values of $x_i(t)$ are in some technological interval $[A_i, B_i]$. Further, there are r input state variables $Y_1, \dots, Y_r$ in the process, controlling the process, the values of which are defined by the control system. These values will be denoted by $y_1(t), \dots, y_r(t)$ and we have a vector $\vec{y}(t) = (y_1(t), \dots, y_r(t))$. Analogously, there are some technological intervals $[C_j, D_j]$ for values of these functions.

The functions of fuzzy controller can be then described functionally in a form

$$\vec{y}(n.\tau) = \big(y_1(n\cdot\tau), \dots, y_r(n\cdot\tau)\big) = \mathbb{F}\Bigg(\Big(\vec{x}\big(n\cdot\tau\big), \dots, \vec{x}\big((n-k)\cdot\tau\big)\Big), \dots$$
$$\dots, \Big(\vec{y}\big((n-1)\cdot\tau\big), \dots, \vec{y}\big((n-k)\cdot\tau)\big)\Big)\Bigg).$$

The control rules "if...then..." can be expressed as follows. Let us suppose that in any interval $[A_i, B_i]$, $i = 1, \dots, m$, (or $[C_j, D_j]$, analogously) a system of fuzzy sets $\{\alpha_{ij} : \alpha_{ij} \underset{\sim}{\subsetneq} [A_i, B_i], j = 1, \dots, m_i\}$ is given. Than any such fuzzy control rule R may be defined as a system

$$\begin{aligned}
R &= (\vec{\alpha}_n, \dots, \vec{\alpha}_{n-k_1}, \vec{\beta}_{n-1}, \dots, \vec{\beta}_{n-k_2}, \vec{\beta}) \\
\vec{\alpha}_p &= (\alpha^{(p)}_{1,j_1}, \dots, \alpha^{(p)}_{m,j_m}), \quad p = 1, \dots, n-k_1 \\
\vec{\beta}_s &= (\beta^{(s)}_{1,i_1}, \dots, \beta^{(s)}_{r,i_r}; \beta), \quad s = 1, \dots, n-k_2 \\
\vec{\beta} &= (\beta_1, \dots, \beta_r) \\
\beta^{(s)}_{k,i_k}, \beta_i &\underset{\sim}{\subsetneq} [C_k, D_k]
\end{aligned}$$

which can be interpreted as a fuzzy implication

$$\begin{gathered}
R = \textbf{if } \Big(\vec{x}\big(n\cdot\tau\big) = \vec{\alpha}_n \textbf{ and } \dots \textbf{ and } \vec{x}\big((n-k_1)\cdot\tau\big) = \vec{\alpha}_{n-k_1} \\
\textbf{and } \vec{y}\big((n-1)\cdot\tau\big) = \vec{\beta}_{n-1} \textbf{ and } \dots \textbf{and } \vec{y}\big((n-k_2)\cdot\tau\big) = \vec{\beta}_{n-k_2}\Big) \\
\textbf{then } \vec{y}(n\cdot\tau) = \vec{\beta}.
\end{gathered}$$

The complexity and a type of fuzzy controller depends significantly on a form of these fuzzy sets β_j. Sugeno and Takagi [4] introduced a well known linear fuzzy controller which (in our notation) has the following form.

$$\begin{gathered}
R = \textbf{ if } \Big(\vec{x}\big(n\cdot\tau\big) = \vec{\alpha}_n \textbf{ and } \dots\textbf{and } \vec{x}\big((n-k_1)\cdot\tau\big) = \vec{\alpha}_{n-k_1} \\
\textbf{and } \vec{y}\big((n-1)\cdot\tau\big) = \vec{\beta}_{n-1} \textbf{ and } \dots \textbf{ and } \vec{y}\big((n-k_2)\cdot\tau\big) = \vec{\beta}_{n-k_2}\Big) \\
\textbf{then } \mathbb{Y} = \mathbb{A}\cdot\mathbb{X},
\end{gathered}$$

where

$$\begin{aligned}
\mathbb{Y}^T &= \big(y_1(n\cdot\tau), \dots, y_r(n\cdot\tau)\big) \\
\mathbb{X}^T &= \Big(\vec{x}\big(n\cdot\tau\big), \dots, \vec{x}\big((n-k_1)\cdot\tau\big), \vec{y}\big((n-1)\cdot\tau\big), \dots, \vec{y}\big((n-k_2)\cdot\tau\big)\Big)
\end{aligned}$$

and $\mathbb{A}$ is a corresponding coefficients matrix.

Now, the three above mentioned methods of deriving control rules may be characterized as follows.

1. This method represents a classical fuzzy controller, where fuzzy sets α, β from fuzzy implications are mostly represented as fuzzy terms of some linguistic variable. This method requires that human operator can explain linguistically what control actions he takes. Hence this method requires knowledge of the process.
2. The second method frequently results in fuzzy control rules of a linear type and it is based on human skill rather than knowledge. The coefficients of a matrix $\mathbb{A}$ can be derived using a suitable database of previous human operator's actions and corresponding responses of the process or from dates describing a previous behaviour of the system.
3. The third method also frequently results in fuzzy control rules of a linear type but these rules are derived purely from a database describing previous behaviour of the system.

For applying of the last two methods it is supposed that a database describing previous input-output behaviour of a system is available. Instead of deriving control rules directly from this database $\mathcal{D}$, we derive at first a *functional model* of a process and then use this model in a control process.

In this paper we want to deal with this third method of deriving control rules, namely with deriving a model of a process under the control. We present a method for deriving a system of fuzzy implications with linear functions as consequences (i.e. a modified version of Sugeno-Takagi model) from a given data set, describing a behaviour of the system and, moreover, we estimate error functions of such models.

2 Locally linear fuzzy models

We will start this section with a more traditional description and notation frequently used in fuzzy models of a system.

Let us suppose that a system has $\vec{x} = (x_1, \ldots, x_m)$ as inputs and y as output. Then a database describing process behaviour can be written in the form

$$\mathcal{D} = \{(x_{t,1}, \ldots, x_{t,m}, y_t) : t \in T\}$$

where T is some discrete set. The technological intervals of x_i will be denoted by $\mathcal{U}_i \subseteq \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers. Let $\mathcal{U} = \prod_{i=1}^{m} \mathcal{U}_i$ and $\mathcal{D}_x = \{(x_{t,1}, \ldots, x_{t,m}) : t \in T\} \subseteq \mathcal{U}$.

Let K be an m-dimensional cube in $\mathcal{U}$ such that $K \cap \mathcal{D}_x \neq \emptyset$. Then by y_K we denote a linear function

$$y_K(\vec{x}) = y_K(x_1, \ldots, x_m) = a_{K,0} + \sum_{i=1}^{m} a_{K,i} x_i$$

such that

$$\sum_{\vec{x}_t \in K \cap \mathcal{D}_x} \left(y_K(\vec{x}_t) - y_t\right)^2 \leq \sum_{\vec{x}_t \in K \cap \mathcal{D}_x} \left(y(\vec{x}) - y_t\right)^2 \quad (1)$$

for any other linear function $y(\vec{x})$. It is clear that such a function can be derived by using e.g. the least square method. Moreover, if $K \cap \mathcal{D}_x$ is an one point set $\{\vec{x}_t\}$, then $y_K(\vec{x}_t) = y_t$.

Definition 1 *A cube K in $\mathcal{U}$ is called an ϵ-cluster in $\mathcal{D}$ (where $\epsilon \in \mathbb{R}_+$), if*

1. $K \cap \mathcal{D}_x \neq \emptyset$

2. $\frac{1}{|K \cap \mathcal{D}_x|} \sum_{\vec{x}_t \in K \cap \mathcal{D}_x} |y_K(\vec{x}_t) - y_t| < \epsilon$,

where $|X|$ is a number of elements of a set X.

Definition 2 *A cube $K \subseteq \mathcal{U}$ is called a maximal ϵ-cluster in $\mathcal{D}$, if it is an ϵ-cluster in $\mathcal{D}$ and there does not exist any ϵ-cluster L in $\mathcal{D}$ such that $K \cap \mathcal{D}_x \subset L \cap \mathcal{D}_x$.*

The following lemma is a trivial consequence of the definition of an ϵ-cluster.

Lemma 1 *Let K be an ϵ-cluster in $\mathcal{D}$ and let L be a cube in $\mathcal{U}$ such that $K \cap \mathcal{D}_x = L \cap \mathcal{D}_x$. Then L is an ϵ-cluster in $\mathcal{D}$.*

In the following lemma we prove the existence of maximal ϵ-clusters covering the domain $\mathcal{D}_x$.

Lemma 2 *For any $\epsilon \in \mathbb{R}_+$ and any $t \in T$ there exists a maximal ϵ-cluster K_t in $\mathcal{D}$ containing $\vec{x}_t$.*

Proof. For $t \in T$ we define the set

$$\mathcal{C}_t = \{K : K \text{ is an } \epsilon - \text{cluster in } \mathcal{D}, \vec{x}_t \in K\}.$$

Since any m-dimensional cube K in $\mathcal{U}$ such that $K \cap \mathcal{D}_x = \{\vec{x}_t\}$ is contained in $\mathcal{C}_t$, we have $\mathcal{C}_t \neq \emptyset$. We show that $(\mathcal{C}_t, \subseteq)$ satisfies the conditions of Zorn's lemma. In fact, let $(\{K_i : i \in I\}, \subseteq)$ be an increasing chain in $\mathcal{C}_t$, $K_i = \prod_{s=1}^{m} a_{i,s}$, where $a_{i,s}$ is a subset in $\mathcal{U}_s$. Then for any $s, s = 1, \ldots, m$, a system $(\{a_{i,s} : i \in I\}, \subseteq)$ is a chain. Let $a_s = \bigcup_{i \in I} a_{i,s} \subseteq \mathbb{R}$ and let $K = \prod_{s=1}^{m} a_s$. Clearly $\vec{x}_t \in K$ and $K_i \cap \mathcal{D}_x \subseteq K \cap \mathcal{D}_x$. We show that the opposite inclusion holds for some i, as well. Let $\vec{x} = (x_1, \ldots, x_m) \in K \cap \mathcal{D}_x$. It follows that for any s,$s = 1, \ldots, m$, there exists $i_s \in I$ such that $x_s \in a_{i_s,s}$. Let $i(\vec{x}) = \max\{i_s : s = 1, \ldots, m\}$. Then $a_{i_s,s} \subseteq a_{i(\vec{x}),s}$ for all $s = 1, \ldots, m$. Hence, $x_s \in a_{i(\vec{x}),s}$ and $\vec{x} \in \prod_s a_{i(\vec{x}),s} = K_{i(\vec{x})}$. Since $\mathcal{D}_x$ is a finite set, the subset $\{i(\vec{x}) \in I : \vec{x} \in K \cap \mathcal{D}_x\} \subseteq I$ is finite as well. Let i_0 be the maximal element of this subset. Then since $\{K_j : j \in I\}$ is a chain, we have $\vec{x} \in K_{i_0} \cap \mathcal{D}_x$ for all $\vec{x} \in K \cap \mathcal{D}_x$ and it follows that $K \cap \mathcal{D}_x = K_{i_0} \cap \mathcal{D}_x$. According to Lemma 1, K is an ϵ-cluster and hence it is an upper bound of the chain $\{K_i : i \in I\}$ in $\mathcal{C}_t$. Hence, according to the Zorn's lemma in $\mathcal{C}_t$ there exists a maximal element K_t. Clearly this element is then a maximal ϵ-cluster. □

A system $\{K_i : i \in I\}$ is called maximal ϵ-covering of $\mathcal{D}$ if $\mathcal{D}_x \subseteq \bigcup_{i \in I} K_i$ and any K_i is an maximal ϵ-cluster of $\mathcal{D}$ for any $i \in I$.

A system $\{K_t : t \in T\}$ from Lemma 2 is an example of such a system. We can even a little reduce this system. In fact, we select an element $t_1 \in T$ such that t_1 is an arbitrary element of T and if elements $t_1, \ldots, t_n$ are selected, then t_{n+1} is an arbitrary element of the set

$$T \setminus \left\{t \in T : (\exists_{1 \le i \le n} i)\vec{x}_t \in K_{t_i}\right\}$$

otherwise the selection process is finished. In this way we obtain a finite set $\{K_{t_i} : i = 1, \ldots, r\}$ which is maximal ϵ-covering of $\mathcal{D}$.

For an ϵ-covering system $\mathcal{K}$ we introduce a notion of index of its redundancy as follows. For $t \in T$ let $n_t = |\{K \in \mathcal{K} : \vec{x}_t \in K\}|$. Then the *index of redundancy of* $\mathcal{K}$ is the number

$$r(\mathcal{K}) = \max_{t \in T} n_t.$$

If elements of $\mathcal{K}$ are mutually disjoint, we have $r(\mathcal{K}) = 1$. Otherwise, $r(\mathcal{K}) > 1$.

For any maximal ϵ-covering $\mathcal{K} = \{K_i : i = 1, \ldots, r\}$ of $\mathcal{D}$ we can define a global function $y_{\mathcal{K}}$ such that

$$y_{\mathcal{K}}(\vec{x}) = \frac{\sum_{i \in I(\vec{x})} y_{K_i}(\vec{x})}{|I(\vec{x})|}$$

where $I(\vec{x}) = \{i : 1 \le i \le r, \vec{x} \in K_i\}$. It should be mentioned that no fuzzy approach is used yet for defining this global function.

For this global function we can estimate its global inconsistency with original dates from $\mathcal{D}$, i.e. we can evaluate the error function

$$\Delta(y_{\mathcal{K}}) = \frac{1}{|\mathcal{D}_x|} \sum_{t \in T} |y_{\mathcal{K}}(\vec{x}_t) - y_t|.$$

Proposition 1 *For any maximal ϵ-covering $\mathcal{K}$ of $\mathcal{D}$ we have*

$$\Delta(y_{\mathcal{K}}) < \epsilon \cdot r(\mathcal{K}).$$

Proof. We have the following inequalities:

$$\Delta(y_{\mathcal{K}}) = \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\left|\frac{\sum_{i\in I(\vec{x}_t)} y_{K_i}(\vec{x}_t)}{|I(\vec{x}_t)|} - y_t\right| =$$

$$= \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\frac{1}{I(\vec{x}_t)}\left|\sum_{i\in I(\vec{x}_t)}(y_{K_i}(\vec{x}_t) - y_t)\right| \leq$$

$$\leq \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\left|\sum_{i\in I(\vec{x}_t)}(y_{K_i}(\vec{x}_t) - y_t)\right| \leq$$

$$\leq \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\sum_{i\in I(\vec{x}_t)}\big|(y_{K_i}(\vec{x}_t) - y_t)\big| = \frac{1}{|\mathcal{D}_x|}\sum_{i=1}^{r}\sum_{\vec{x}_t\in K_i}\big|y_{K_i}(\vec{x}_t) - y_t\big|.$$

Now, since K_i is an ϵ-cluster, we have

$$\frac{1}{|K_i\cap\mathcal{D}_x|}\sum_{\vec{x}_t\in K_i}|y_{K_i}(\vec{x}_t) - y_t| < \epsilon$$

and it follows that

$$\epsilon\cdot|\mathcal{D}_x|\cdot r(\mathcal{K}) \geq \epsilon\cdot\sum_{t\in T} n_t = \epsilon\cdot\sum_{t\in T}\big|\{K\in\mathcal{K} : \vec{x}_t\in K\}\big| =$$

$$= \epsilon\cdot\sum_{K\in\mathcal{K}}\big|\{t\in T : \vec{x}_t\in K\}\big| = \epsilon\cdot\sum_{i=1}^{r}|K_i\cap\mathcal{D}_x > \sum_{i=1}^{r}\sum_{\vec{x}_t\in K_i}|y_{K_i}(\vec{x}_t) - y_t|.$$

Hence, we obtain $\Delta(y_{\mathcal{K}}) < \epsilon\cdot r(\mathcal{K})$. □

Definition 3 *A system* $\mathbf{P} = \{(K_i, L_i) : i = 1,\dots,r\}$ *is called an* (ϵ,τ)*-fuzzy partition of* $\mathcal{D}$ *(where* $\epsilon < \tau$*), if*

1. $\{L_i : i = 1,\dots,r\}$ *is a maximal* τ*-covering of* $\mathcal{D}$.
2. *For any* i, K_i *is a maximal* ϵ*-cluster in* $\mathcal{D}$ *and* $K_i \subseteq L_i$.
3. *For any* i, $K_i\cap\mathcal{D}_x \neq \emptyset$.

It is a very natural interpretation that any pair $(K, L)\in\mathbf{P}$ then represents a fuzzy set (denoted by $[K, L]$) in $\mathcal{U}$ such that supp $[K, L] = L$ and core $[K, L] = K$. In fact, let $K = \prod_{i=1}^m a_{K,i}, L = \prod_{i=1}^m a_{L,i}$, where $a_{K,i}\subseteq a_{L,i}\subseteq\mathcal{U}_i$ are intervals, i.e.

$$a_{K,i} = [a_{K,i}^{\text{begin}}, a_{K,i}^{\text{end}}],\quad a_{L,i} = [a_{L,i}^{\text{begin}}, a_{K,i}^{\text{end}}].$$

Then any pair $(a_{K,i}, a_{L,i})$ represents a fuzzy set $[a_{K,i}, a_{L,i}]$ such that

$$[a_{K,i}, a_{L,i}](x) = \begin{cases} 1, & \text{if } x\in a_{K,i},\\ 0, & \text{if } x\notin a_{L,i},\\ \frac{x - a_{L,i}^{\text{begin}}}{a_{K,i}^{\text{begin}} - a_{L,i}^{\text{begin}}}, & \text{if } a_{L,i}^{\text{begin}}\leq x\leq a_{K,i}^{\text{begin}},\\ \frac{a_{L,i}^{\text{end}} - x}{a_{L,i}^{\text{end}} - a_{K,i}^{\text{end}}}, & \text{if } a_{K,i}^{\text{end}}\leq x\leq a_{L,i}^{\text{end}}.\end{cases}$$

Then a fuzzy set $[K, L]$ is defined as follows.

$$[K, L](\vec{x}) = \begin{cases} 1, & \text{if } \vec{x}\in K,\\ 0, & \text{if } \vec{x}\notin L,\\ [a_{K,1}, a_{L,1}](x_1) & \text{if } \vec{x}\in L\setminus K.\end{cases}$$

Proposition 2 *For any pair $\tau > \epsilon$ there exists an (ϵ, τ)-fuzzy partition of $\mathcal{D}$.*

Proof. Let $\{L_i : i = 1, \ldots, r\}$ be a maximal τ-covering of $\mathcal{D}$. Such a system exists according to Lemma 2 For any $i, 1 \leq i \leq r$, we put

$$\mathcal{R}_i = \{K : K \text{ is a } \epsilon - \text{ cluster in } \mathcal{D}, \vec{x}_{t_0} \in K \subseteq L_i, \}$$

where x_{t_0} is an (fixed) element of $L_i \cap \mathcal{D}_x$. Then analogously as in the Lemma 2 we can prove that $\mathcal{R}_i$ satisfies the conditions of Zorn's lemma and it follows that there exists a maximal element K_i in $\mathcal{R}_i$. Then $\{(K_i, L_i) : i = 1, \ldots, r\}$ is a required fuzzy partition. □

For any element (K, L) of a fuzzy partition $\mathbf{P}$ we can analogously define a local function $y_{(K,L)}$ such that

$$y_{(K,L)}(\vec{x}) = \begin{cases} y_K(\vec{x}), & \text{if } \vec{x} \in K, \\ y_L(\vec{x}), & \text{if } \vec{x} \in K \setminus L. \end{cases} \tag{2}$$

For these local functions $y_{(K,L)}$ we can define two local error functions such that

$$\Delta_1\big(y_{(K,L)}\big) = \frac{1}{|L \cap \mathcal{D}_x|} \sum_{\vec{x}_t \in L \cap \mathcal{D}_x} |y_{(K,L)}(\vec{x}_t) - y_t|,$$

$$\Delta_2\big(y_{(K,L)}\big) = \frac{1}{|L \cap \mathcal{D}_x|^2} \sum_{\vec{x}_t \in L \cap \mathcal{D}_x} (y_{(K,L)}(\vec{x}_t) - y_t)^2.$$

In the following proposition we estimate values of these local error functions.

Proposition 3 *Let $\mathbf{P}$ be an (ϵ, τ)-fuzzy partition of $\mathcal{D}$ and let $(K, L) \in \mathbf{P}$. Then*

$$\Delta_1\big(y_{(K,L)}\big) < \epsilon + \tau,$$
$$\Delta_2\big(y_{(K,L)}\big) < \tau^2.$$

Proof. According to the definition of a function $y_{(K,L)}$ we have

$$\Delta_1(y_{(K,L)}) = \frac{1}{|L \cap \mathcal{D}_x|} \sum_{\vec{x}_t \in L \cap \mathcal{D}_x} |y_{(K,L)}(\vec{x}_t) - y_t| =$$

$$= \frac{1}{|L \cap \mathcal{D}_x|} \left(\sum_{\vec{x}_t \in K \cap \mathcal{D}_x} |y_K(\vec{x}_t) - y_t| + \sum_{\vec{x}_t \in (L \setminus K) \cap \mathcal{D}_x} |y_L(\vec{x}_t) - y_t| \right) <$$

$$< \frac{1}{|L \cap \mathcal{D}_x|} \big(\epsilon \cdot |K \cap \mathcal{D}_x| + \tau \cdot |L \cap \mathcal{D}_x|\big) \leq \epsilon + \tau.$$

For the other local error function we have

$$\Delta_2(y_{(K,L)}) = \frac{1}{|L \cap \mathcal{D}_x|^2} \sum_{\vec{x}_t \in L \cap \mathcal{D}_x} (y_{(K,L)}(\vec{x}_t) - y_t)^2 =$$

$$= \frac{1}{|L \cap \mathcal{D}_x|^2} \left(\sum_{\vec{x}_t \in K \cap \mathcal{D}_x} (y_K(\vec{x}_t) - y_t)^2 + \sum_{\vec{x}_t \in (L \setminus K) \cap \mathcal{D}_x} (y_L(\vec{x}_t) - y_t)^2 \right).$$

On the other hand, according to the property (1) of a local function y_K we have

$$\tau^2 \cdot |L \cap \mathcal{D}_x|^2 > \left(\sum_{\vec{x}_t \in L \cap \mathcal{D}_x} \big|y_L(\vec{x}_t) - y_t\big| \right)^2 \geq \sum_{\vec{x}_t \in L \cap \mathcal{D}_x} \big(y_L(\vec{x}_t) - y_t\big)^2 =$$

$$= \sum_{\vec{x}_t \in (L \setminus K) \cap \mathcal{D}_x} \big(y_L(\vec{x}_t) - y_t\big)^2 + \sum_{\vec{x}_t \in K \cap \mathcal{D}_x} \big(y_L(\vec{x}_t) - y_t\big)^2 \geq$$

$$\geq \sum_{\vec{x}_t \in (L \setminus K) \cap \mathcal{D}_x} \big(y_L(\vec{x}_t) - y_t\big)^2 + \sum_{\vec{x}_t \in K \cap \mathcal{D}_x} \big(y_K(\vec{x}_t) - y_t\big)^2.$$

Hence, we obtain

$$\Delta_2\big(y_{(K,L)}\big) \leq \frac{1}{|L\cap \mathcal{D}_x|^2}\Big[\sum_{\vec{x}_t\in K\cap\mathcal{D}_x}\big(y_K(\vec{x}_t)-y_t\big)^2+\tau^2\cdot|L\cap\mathcal{D}_x|^2 - \\ - \sum_{\vec{x}_t\in K\cap\mathcal{D}_x}\big(y_K(\vec{x}_t)-y_t\big)^2\Big] = \tau^2.$$

□

A local function $y_{(K,L)}$ can be defined also in a little different way. In fact, we can put

$$y_{(K,L)}(\vec{x}) := [K,L](\vec{x}).y_K(\vec{x}) + \big(1-[K,L](\vec{x})\big)\cdot y_L(\vec{x}) \tag{3}$$

For this another version of a local function we receive a little worser estimation of its error function.

Proposition 4 *Let* $\mathbf{P}$ *be an* (ϵ,τ)*-partition of* $\mathcal{D}$ *and let* $(K,L)\in\mathbf{P}$*. Then for a local function* $y_{(K,L)}$ *defined above we have*

$$\Delta_1(y_{(K,L)}) < \epsilon + 3\cdot\tau.$$

Proof. For $t\in T$ by α_t we denote a number $[K,L](\vec{x}_t)$. Then we have

$$\begin{aligned}\Delta_1(y_{(K,L)}) &= \frac{1}{|L\cap\mathcal{D}_x|}\sum_{\vec{x}_t\in L\cap\mathcal{D}_x}|y_{(K,L)}(\vec{x}_t)-y_t| = \\ &= \frac{1}{|L\cap\mathcal{D}_x|}\Big[\sum_{\vec{x}_t\in K\cap\mathcal{D}_x}\big|y_K(\vec{x}_t)-y_t\big| + \\ &+ \sum_{\vec{x}_t\in(L\setminus K)\cap\mathcal{D}_x}\big|\alpha_t\cdot y_K(\vec{x}_t)+(1-\alpha_t)\cdot y_L(\vec{x}_t)-y_t\big|\Big] < \\ &< \frac{1}{|L\cap\mathcal{D}_x|}\Big[\epsilon\cdot\big|K\cap\mathcal{D}_x\big| + \sum_{\vec{x}_t\in(L\setminus K)\cap\mathcal{D}_x}\alpha_t\big|y_K(\vec{x}_t)-y_t\big| + \\ &+ \sum_{\vec{x}_t\in(L\setminus K)\cap\mathcal{D}_x}(\alpha_t+1)\big|y_L(\vec{x}_t)-y_t\big|\Big] \leq \\ &\leq \frac{1}{|L\cap\mathcal{D}_x|}\big(\epsilon\cdot|K\cap\mathcal{D}_x|+\tau\cdot|L\cap\mathcal{D}_x|+2\cdot\tau\cdot|L\cap\mathcal{D}_x|\big) \leq \epsilon+3\cdot\tau.\end{aligned}$$

□

A system of these local functions (of any of the two mentioned type) then represents naturally a fuzzy model of data set $\mathcal{D}$ (of Sugeno-Takagi type). This fuzzy model $\mathcal{M}(\mathbf{P})$ can be then written in a form

$$\mathcal{M}(\mathbf{P}) = \big\{\textbf{if } \vec{x}\in[K,L] \textbf{ then } y = y_{[K,L]}(\vec{x}) : (K,L)\in\mathbf{P}\big\}.$$

By using an (ϵ,τ)-fuzzy partition $\mathbf{P}$ of $\mathcal{D}$ we can define a global function $y_{\mathbf{P}}$ describing our data structure as follows.

$$y_{\mathbf{P}}(\vec{x}) := y_{(K,L)}(\vec{x}),$$

where $(K,L)\in\mathbf{P}$ is such that $[K,L](\vec{x}) = \max\{[R,S](\vec{x}) : (R,S)\in\mathbf{P}\}$.

The following proposition estimates a global error of this function.

Proposition 5 *Let* $\mathbf{P}$ *be an* (ϵ,τ)*-fuzzy partition of* $\mathcal{D}$ *and let* $\mathcal{L} = \{L : (K,L)\in\mathbf{P}$ *for some* $K\}$ *be a corresponding maximal* τ*-covering. Then*

$$\Delta_1(y_{\mathbf{P}}) < \tau\cdot r(\mathcal{L}).$$

Proof. At first, the following inequality holds for a system $\mathcal{L}$:

$$\sum_{L\in\mathcal{L}} |L\cap\mathcal{D}_x| = \sum_{t\in T} \big|\{L\in\mathcal{L} : \vec{x}_t\in L\}\big| =$$

$$= \sum_{t\in T} n_t \leq r(\mathcal{L})\cdot|\mathcal{D}_x|. \tag{4}$$

For $(K,L)\in\mathbf{P}$ we put

$$\sigma(K,L) = \big\{\vec{x}_t\in\mathcal{D}_x : [K,L](\vec{x}_t) = \max\{[R,S](\vec{x}_t) : (R,S)\in\mathbf{P}\}\big\} \subseteq L\cap\mathcal{D}_x.$$

Let for $t\in T$, (K_t,L_t) be an element of $\mathbf{P}$ such that $y_{\mathbf{P}}(\vec{x}_t) = y_{(K_t,L_t)}(\vec{x}_t)$. Then we have

$$\big\{(K_t,L_t,t) : t\in T\big\} \subseteq \bigcup_{(K,L)\in\mathbf{P}} \big\{(K,L,t) : t\in\sigma(K,L)\big\}$$

and it follows (according to (4)) that

$$\begin{aligned}
\Delta_1(y_{\mathbf{P}}) &= \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}|y_{\mathbf{P}}(\vec{x}_t)-y_t| = \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}|y_{(K_t,L_t)}(\vec{x}_t)-y_t| \leq \\
&\leq \frac{1}{|\mathcal{D}_x|}\sum_{(K,L)\in\mathbf{P}}\ \sum_{t\in\sigma(K,L)} |y_{(K,L)}(\vec{x}_t)-y_t| \leq \\
&\leq \frac{1}{|\mathcal{D}_x|}\sum_{(K,L)\in\mathbf{P}}\left(\sum_{\vec{x}_t\in L\cap\mathcal{D}_x}|y_{(K,L)}(\vec{x}_t)-y_t|\right) \leq \\
&\leq \frac{1}{|\mathcal{D}_x|}\sum_{(K,L)\in\mathbf{P}} \tau\cdot|L\cap\mathcal{D}_x| \leq \frac{\tau}{|\mathcal{D}_x|}\cdot|\mathcal{D}_x|\cdot r(\mathcal{L}) = \tau\cdot r(\mathcal{L}).
\end{aligned}$$

□

Analogously as we did for a local function $y_{(K,L)}$, we can define another version of a global function $y_{\mathbf{P}}$ such that

$$y_{\mathbf{P}}(\vec{x}) = \frac{\sum_{(K,L)\in\mathbf{P}} y_{(K,L)}(\vec{x})\cdot[K,L](\vec{x})}{\sum_{(K,L)\in\mathbf{P}}[K,L](\vec{x})}. \tag{5}$$

Even in this case we can estimate an error function for this global function.

Proposition 6 *Let* $\mathbf{P}$ *be an* (ϵ,τ)*-fuzzy partition of* $\mathcal{D}$ *and let* $\mathcal{L} = \{L : (K,L)\in\mathbf{P}$ *for some* $K\}$ *be a corresponding maximal* τ*-covering. Then for a global function* $y_{\mathbf{P}}$ *defined in (5) we have*

$$\Delta_1(y_{\mathbf{P}}) < r(\mathcal{L})\cdot(\epsilon+\tau), \text{ if local functions are defined by (2),}$$
$$\Delta_1(y_{\mathbf{P}}) < r(\mathcal{L})\cdot(\epsilon+3\cdot\tau), \text{ if local functions are defined by (3).}$$

Proof. Let

$$\alpha = \begin{cases} \epsilon+\tau, & \text{if local functions are defined by (2),}\\ \epsilon+3\cdot\tau, & \text{if local functions are defined by (3).}\end{cases}$$

Then according to Propositions 3 and 4 we have

$$
\begin{aligned}
\Delta_1(y_{\mathbf{P}}) &= \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}|y_{\mathbf{P}}(\vec{x}_t)-y_t| = \\
&= \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\left|\frac{\sum_{(K,L)\in\mathbf{P}} y_{(K,L)}(\vec{x})\cdot[K,L](\vec{x})}{\sum_{(K,L)\in\mathbf{P}}[K,L](\vec{x})}-y_t\right| = \\
&= \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\frac{1}{a_t}\left|\sum_{(K,L)\in\mathbf{P}}(y_{(K,L)}(\vec{x}_t)-y_t)[K,L](\vec{x}_t)\right| \le \\
&\le \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\frac{1}{a_t}\sum_{(K,L)\in\mathbf{P}}|y_{(K,L)}(\vec{x}_t)-y_t|[K,L](\vec{x}_t) \le \\
&\le \frac{1}{|\mathcal{D}_x|}\sum_{t\in T}\sum_{\substack{(K,L)\in\mathbf{P}\\ \vec{x}_t\in L}}|y_{(K,L)}(\vec{x}_t)-y_t| = \\
&= \frac{1}{|\mathcal{D}_x|}\sum_{(K,L)\in\mathbf{P}}\sum_{\vec{x}_t\in L\cap\mathcal{D}_x}|y_{(K,L)}(\vec{x}_t)-y_t| \le \\
&\le \frac{1}{|\mathcal{D}_x|}\sum_{(K,L)\in\mathbf{P}}\alpha\cdot|L\cap\mathcal{D}_x| \le \frac{\alpha}{|\mathcal{D}_x|}\cdot r(\mathcal{L})\cdot|\mathcal{D}_x| = r(\mathcal{L})\cdot\alpha
\end{aligned}
$$

where

$$
a_t = \sum_{(K,L)\in\mathbf{P}}[K,L](\vec{x}).
$$

□

References

[1] Ivaknenko, A.G.,Vysockij, V.N., Ivaknenko, N.A.: *Principal version of the minimum bias criterion for a model and an investigation of their noise immunity*, Soviet Automat. Control, 11, 1978 , 27–45

[2] Mamdani, E.H., Ostergaard, J.J., Lembessis,E.,: *Use of fuzzy logic for implementation rule based control of industrial processes*, Studies in the Management Sciences, 20, 1984, 429–445

[3] Mockor, J.: *Fuzzy models with locally linear output functions*, Universitat Duisburg, SM-DU-274, 1994

[4] Takagi, T., Sugeno, M.: *Fuzzy identification of systems and its applications to modelling and control*, IEEE trans. System Man. Cybernet., 15, 1985, 116–132