Proc. Estonian Acad. Sci. Eng., 2000, **6**, 2, 76–95
https://doi.org/10.3176/eng.2000.2.01

INTERPRETABILITY VERSUS ADAPTABILITY IN FUZZY SYSTEMS

Andri RIID and Ennu RÜSTERN

Department of Computer Control, Tallinn Technical University, Ehitajate tee 5, 19086 Tallinn, Estonia; andri@dcc.ttu.ee

Received 14 October 1999, in revised form 8 February 2000

Abstract. This paper considers the most common adaptation techniques for standard and Sugeno fuzzy systems with special regard to post-adaptation linguistic interpretability. The level of adaptation is characterized by the approximation error, but there is no similarly accepted measure for the validity of linguistic interpretation (transparency) and the detection of the latter relies much on empirical observation. Such observation is hardly possible beyond three-dimensional space. The general purpose of this paper is to find out how adaptation algorithms for antecedent and/or consequent parameters, such as gradient descent method, least squares estimation, and clustering techniques, act on the interpretability of the system besides their approximation properties. The comparison of the algorithms is based on the modelling of a simple single-input single-output system. The conclusion that the transparency of the observed system depends on the degree of overlap of neighbouring input fuzzy sets can, however, be generalized for multivariate systems.

Key words: modelling, fuzzy systems, Sugeno systems, fuzzy clustering, gradient descent, least squares estimation, accuracy, interpretability, adaptability.

1. INTRODUCTION

The most attractive property of fuzzy systems lies in their ability to process the information both linguistically and numerically. The universal approximation property of fuzzy systems has been thoroughly investigated [1–3] and successfully applied. Besides the approximation capabilities of fuzzy systems, few authors have given proper importance to the linguistic interpretation problem. It can be said that the full potential of fuzzy systems is not used yet, because linguistic interpretation (when valid) is a rather powerful tool for analysing the numerical data and can give useful information about the modelled unknown system.

The majority of fuzzy systems employed in modelling and control belong to the Sugeno family for which the adaptation rules are most easily derived and efficient. At the same time, the issue of linguistic interpretation is somewhat shadowed, partly due to the semi-linguistic nature of the Sugeno rulebase. For standard (Mamdani) fuzzy systems with clear linguistic rules, adaptation algorithms giving satisfactory accuracy of the model are rare. Possible exceptions are genetic algorithms [4] that often are not usable in practical situations due to the amount of computations and/or large adaptation time.

There is a trade-off between interpretability and adaptability and it depends on the particular control/modelling problem to which the preference should be given. It is usually possible to reach a satisfactory solution by selecting a suitable fuzzy system and adaptation algorithm. For that reason, general understanding of the adaptation techniques and features of different fuzzy systems is useful. In this paper the most common adaptation techniques are described and analysed from the viewpoint of transparency.

2. FUZZY SYSTEMS

A multi-input multi-output fuzzy system is given by the rulebase the rth ($r=1,\dots,R$) rule of which defines the linguistic relationship between the inputs U_i ($i=1,\dots,N$) and outputs V_j ($j=1,\dots,M$) of the system via their linguistic labels A_{ir}, B_{ir}:

$$\text{if } U_1 \text{ is } A_{1r} \text{ and } U_2 \text{ is } A_{2r} \dots \text{ and } U_i \text{ is } A_{ir} \dots \text{ and } U_N \text{ is } A_{Nr},$$
$$\text{then } V_1 \text{ is } B_{1r} \text{ and } V_2 \text{ is } B_{2r} \dots \text{ and } V_j \text{ is } B_{jr} \dots \text{ and } V_M \text{ is } B_{Mr}.$$

The numerical interpretation is given by normal convex fuzzy sets defined by standard membership functions (such as triangular, trapezoidal, Gaussian, etc.) having one-to-one correspondence with linguistic labels, and by a five-step inference algorithm consisting of fuzzification, premise conjunction, implication, aggregation, and defuzzification. First four steps of this algorithm result in the expression of the jth fuzzy output $F(y_j)$ of the system:

$$F(y_j)=\bigcup_{r=1}^{R}\left(\left(\bigcap_{i=1}^{N}\mu_{ir}(x_i)\right)\cap\gamma_{jr}\right), \tag{1}$$

where μ_{ir} and γ_{jr} denote the membership functions of the ith input variable and jth output associated with the rth rule, respectively, x_i denotes the numerical value of the ith input variable, and $\cap$, $\cup$ denote the operators called t-norm and t-conorm, respectively.

The inference algorithm can be specified to meet one's needs by the selection of a suitable t-norm (minimum, product, etc.), t-conorm (maximum, probabilistic sum, etc.), and defuzzification method (centroid, mean of maxima, etc.). An overview of different t-norms and t-conorms as well as of defuzzification algorithms is given in [5,6].

Although the number of different types of fuzzy systems that can be obtained by combining different t-norms, t-conorms, and defuzzificaton methods is large, a separate family of fuzzy systems is well distinguished, being obtained by specifying the output membership functions as the functions of inputs (usually as a linear combination)

$$\gamma_r(x_1,\ldots,x_N) = a_{r0} + \sum_{i=1}^{N} a_{ri}x_i. \quad (2)$$

The output membership functions of Sugeno systems represent a compromise between standard fuzzy and mathematical systems and such systems are usually regarded as piecewise linear input-output mappers. The center-of-gravity defuzzification in case of Sugeno systems reduces to the weighted average (fuzzy-mean) method.

Finally, as can be easily seen from Eq. (1), a fuzzy system having several outputs can be decomposed into M multi-input single-output (MISO) systems, that in many cases relieves the implementation and development of fuzzy systems/algorithms.

3. ADAPTATION IN SUGENO SYSTEMS

3.1. Gradient descent

Most fuzzy system adaptation approaches rely on gradient descent optimization by minimizing the objective function

$$\varepsilon(k) = \frac{1}{2}\left[y(k) - \tilde{y}(k)\right]^2, \quad (3)$$

where $y(k)$ denotes the measured fuzzy system output and $\tilde{y}(k)$ is the target value for $y(k)$ at the moment k.

The output of MISO Sugeno systems with product operator for conjunction and implication, sum operator for aggregation, and Gaussian input membership functions μ_{ir} given by

$$\mu_{ir}(x_i) = e^{-\frac{(x_i - c_{ir})^2}{2\sigma_{ir}^2}}, \quad (4)$$

is calculated by making appropriate replacements in Eq. (1) and applying weighted average defuzzification

$$y(k) = \frac{\sum_{r=1}^{R} \tau_r (a_{0r} + a_{1r}x_1(k) + \ldots + a_{Nr}x_N(k))}{\sum_{r=1}^{R} \tau_r}, \quad (5)$$

where τ_r is activation degree of the rth rule, given by

$$\tau_r = \prod_{i=1}^{N} \mu_{ir}(x_i(k)). \tag{6}$$

The usage of above-mentioned operators makes the system differentiable and the following laws for updating the parameters of the fuzzy system are used [7]:

$$c_{ir}(k+1) = c_{ir}(k) - \alpha_2 \frac{\partial \varepsilon(k)}{\partial c_{ir}(k)}, \tag{7}$$

$$\sigma_{ir}(k+1) = \sigma_{ir}(k) - \alpha_3 \frac{\partial \varepsilon(k)}{\partial \sigma_{ir}(k)}, \tag{8}$$

$$a_{0r}(k+1) = a_{0r}(k) - \alpha_4 \frac{\partial \varepsilon(k)}{\partial a_{0r}(k)}, \tag{9}$$

$$a_{ir}(k+1) = a_{ir}(k) - \alpha_5 \frac{\partial \varepsilon(k)}{\partial a_{ir}(k)}. \tag{10}$$

These are obtained by applying the chain rule

$$a_{0r}(k+1) = a_{0r}(k) - \alpha_1 (y(k) - \tilde{y}(k)) \frac{\tau_r(k)}{\sum_{r=1}^{R} \tau_r(k)}, \tag{11}$$

$$a_{ir}(k+1) = a_{ir}(k) - \alpha_2 (y(k) - \tilde{y}(k)) \frac{x_i \tau_r(k)}{\sum_{r=1}^{R} \tau_r(k)}, \tag{12}$$

$$c_{ir}(k+1) = c_{ir}(k) - \alpha_2 (y(k) - \tilde{y}(k))(b_r(k) - y(k)) \frac{\tau_r(k)}{\sum_{r=1}^{R} \tau_r(k)} \frac{\tilde{x}_i(k) - c_{ir}(k)}{\sigma_{ir}^{\,2}(k)}, \tag{13}$$

$$\sigma_{ir}(k+1) = \sigma_{ir}(k) - \alpha_3 (y(k) - \tilde{y}(k))(b_r(k) - y(k)) \frac{\tau_r(k)}{\sum_{r=1}^{R} \tau_r(k)} \frac{(\tilde{x}_i(k) - c_{ir}(k))^2}{\sigma_{ir}^{\,3}(k)}. \tag{14}$$

Application of the criterion (3) guarantees only minimization of the local error, not necessarily the minimum value of the error calculated over the whole measured space. That may be improved by a different minimization criterion ensuring the minimization of the overall error if the learning rates are properly selected:

$$\varepsilon = \sum_{k=1}^{K} \varepsilon(k) = \sum_{k=1}^{K} \frac{1}{2} [y(k) - \tilde{y}(k)]^2, \quad (15)$$

leading to update laws with

$$\Delta p = -\alpha_p \frac{\partial \varepsilon}{\partial p} = \alpha_p \sum_{k=1}^{K} \frac{\partial \varepsilon(k)}{\partial p}, \quad (16)$$

where p is the parameter to be updated.

3.2. Least squares

The consequent parameters of a fuzzy system can be adapted by other means than gradient descent, notably by the least squares procedure. Denoting

$$\phi_r = \frac{\tau_r}{\sum_{r=1}^{R} \tau_r}, \quad (17)$$

and combining ϕ_r into a matrix

$$J = [\phi_1, \phi_2, ..., \phi_R, \phi_1 x_1, \phi_2 x_1, ..., \phi_R x_1, ..., \phi_1 x_N, ..., \phi_R x_N], \quad (18)$$

we obtain that Eq. (5) is equivalent to

$$y = J\theta, \quad (19)$$

where

$$\theta = [a_{01}, a_{02}, ..., a_{0R}, a_{11}, a_{12}, ..., a_{1R}, ..., a_{N1}, ..., a_{NR}]^{\mathrm{T}}. \quad (20)$$

Therefore the output parameters can be estimated as

$$\theta = [J^{\mathrm{T}} J]^{-1} J^{\mathrm{T}} y. \quad (21)$$

3.3. Clustering techniques

Cluster is a group of objects that are mathematically more similar to one another than to members of other clusters. Clustering is basically detection of subspaces of the data space. The potential of clustering algorithms to reveal the underlying structures in the data can be exploited for partitioning the input space of fuzzy systems.

Fuzzy clustering methods allow objects to belong to several clusters simultaneously, with a different membership degree. A large family of fuzzy clustering algorithms is based on the minimization of the fuzzy c-means objective function

$$J=\sum_{k=1}^{K}\sum_{h=1}^{H}(\mu_{hk})^{m}\left|x(k)-v_{h}\right|^{2}, \tag{22}$$

where H is the number of clusters, to find grades of membership μ_{hk} and cluster centres v_h [8].

A parallel with the membership functions of fuzzy systems is obvious and, indeed, input membership functions can be approximated from the projections of μ_{hk} onto the space of input variables x_i by suitable parametric membership functions. Of greatest interest of such algorithms is the Gustafson–Kessel (GK) clustering [9].

Fuzzy clustering appears more natural and gives more information about fuzzy systems than hard clustering based on the classical set theory. However, in addition to fuzzy c-means clustering, a hard clustering method, subtractive clustering, is implemented in Fuzzy Logic Toolbox of MATLAB and is found more suitable for system approximation in combination with the least squares method.

4. FUZZY SYSTEMS WITH SIMPLIFIED INFERENCE

This section of the paper should be properly titled "Adaptation in standard fuzzy systems". However, deriving the adaptation algorithms for standard fuzzy systems is an extremely complicated process. A reasonable compromise is obtained by employing a product operator for conjunction and implication and sum operator for aggregation; in this case centre-of-gravity defuzzification transforms to the weighted average defuzzification. Output fuzzy sets are defined as fuzzy singletons, characterized by a real number b_r.

Defuzzified output of such a system is computed as

$$y(k)=\frac{\sum_{r=1}^{R}b_{r}\tau_{r}(k)}{\sum_{r=1}^{R}\tau_{r}(k)}. \tag{23}$$

Despite all simplifications, Eq. (23) remains a standard fuzzy system, although it can also be regarded as a special case of Sugeno systems and can be derived from Eq. (5), if

$$\forall a_{ir}=0,\quad i=1,\ldots,N,\qquad r=1,\ldots,R. \tag{24}$$

While it is a common practice to employ smooth membership functions in association with Sugeno systems, triangular membership functions are preferred in the standard case:

$$\mu_{ir}(x_i)=\begin{cases}\dfrac{x_i-a_{ir}}{b_{ir}-a_{ir}}, & a_{ir}<x_i<b_{ir},\\[2mm] \dfrac{c_{ir}-x_i}{c_{ir}-b_{ir}}, & b_{ir}<x_i<c_{ir},\\[2mm] 0, & c_{ir}<x_i<a_{ir}.\end{cases} \quad (25)$$

4.1. Gradient descent

Updating law for the consequent parameters b_r in Eq. (23) is equivalent to Eq. (11):

$$b_r(k+1)=b_r(k)-\alpha_1(y(k)-\tilde{y}(k))\frac{\tau_r(k)}{\sum_{r=1}^{R}\tau_r(k)}. \quad (26)$$

Note that triangular membership function given by Eq. (25) is not continuous and therefore we obtain different updating laws for each continuous part of the function:

(1) if $a_{ir}<x_i<b_{ir}$, then

$$a_{ir}(k+1)=a_{ir}(k)-\alpha_2(y(k)-\tilde{y}(k))(b_r(k)-y(k)) \times\frac{\tau_r(k)}{\sum_{r=1}^{R}\tau_r(k)}\frac{(x_i(k)-b_{ir}(k))}{(x_i(k)-a_{ir}(k))(b_{ir}(k)-a_{ir}(k))}, \quad (27)$$

$$b_{ir}(k+1)=b_{ir}(k) +\alpha_3(y(k)-\tilde{y}(k))(b_r(k)-y(k))\frac{\tau_r(k)}{\sum_{r=1}^{R}\tau_r(k)}\frac{1}{(b_{ir}(k)-a_{ir}(k))}, \quad (28)$$

$$c_{ir}(k+1)=c_{ir}(k), \quad (29)$$

(2) if $b_{ir}<x_i<c_{ir}$, then

$$a_{ir}(k+1)=a_{ir}(k), \quad (30)$$

$$b_{ir}(k+1)=b_{ir}(k) +\alpha_3(y(k)-\tilde{y}(k))(b_r(k)-y(k))\frac{\tau_r(k)}{\sum_{r=1}^{R}\tau_r(k)}\frac{1}{(c_{ir}(k)-b_{ir}(k))}, \quad (31)$$

$$c_{ir}(k+1)=c_{ir}(k)-\alpha_4(y(k)-\tilde{y}(k))(b_r(k)-y(k))\frac{\tau_r(k)}{\sum_{r=1}^{R}\tau_r(k)}$$

$$\times\frac{(x_i(k)-b_{ir}(k))}{(c_{ir}(k)-x_i(k))(c_{ir}(k)-a_{ir}(k))}. \tag{32}$$

The derived algorithm, however, is not directly applicable. A set of restrictions must be applied, part of which come from the definition of fuzzy sets, e.g., b_{ir} cannot be bigger than c_{ir}, etc. Also the problem of "blank spots" may cause difficulties, particularly in input space, with the inference engine. Other restrictions are often required in order to retain transparency, e.g., to avoid the occurrence of strongly overlapping neighbouring sets.

Another, more elegant way to reduce the need for learning restrictions, proposed by Jager [6], makes use of the following definition of input membership functions (note that this is not a rule-oriented notation as in all the previous cases, but "neighbour-oriented", resulting in a different training algorithm):

$$\mu_i^j(x_i(k))=\begin{cases}\dfrac{x-a_i^{j-1}}{a_i^j-a_i^{j-1}}, & a_i^{j-1}<x<a_i^j,\\[2ex] \dfrac{a_i^{j+1}-x}{a_i^{j+1}-a_i^j}, & a_i^j<x<a_i^{j+1},\\[2ex] 0, & a_i^{j+1}<x<a_i^{j-1}.\end{cases} \tag{33}$$

Each fuzzy set is defined through the neighbouring fuzzy sets so that its edge parameters b_i^j and c_i^j are determined by the centres of the neighbouring sets, a_i^{j-1} and a_i^{j+1}, respectively (Fig. 1). Thus, the overlap height of 0.5 is always maintained.

For consequent parameters, Eq. (26) is still valid. However, since

$$\sum_{r=1}^{R}\tau_r(k)=1, \tag{34}$$

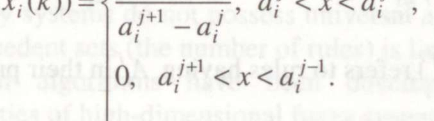

Fig. 1. Neighbour-oriented definition of triangular membership functions.

it can be neglected and we have

$$b_r(k+1) = b_r(k) - \alpha_1 (y(k) - \tilde{y}(k))\tau_r(k). \tag{35}$$

Derivation of update formulas for input parameters leads to the following rules

(1) if $a_i^{j-1} < x_i < a_i^j$, then

$$a_i^j(k+1) = a_i^j(k) - \alpha_2 \frac{(y(k) - \tilde{y}(k))}{a_i^j(k) - a_i^{j-1}(k)} \times \left[\frac{\mu_i^j(x_i(k))}{\mu_i^{j-1}(x_i(k))} \sum_{r'=1}^{R(\mu_i^{j-1})} \tau_{r'}(k)(b_{r'}(k) - y(k)) - \sum_{r'=1}^{R(\mu_i^j)} \tau_{r'}(k)(b_{r'}(k) - y(k)) \right], \tag{36}$$

(2) if $a_i^j < x_i < a_i^{j+1}$, then

$$a_i^j(k+1) = a_i^j(k) - \alpha_2 \frac{(y(k) - \tilde{y}(k))}{a_i^j(k) - a_i^{j+1}(k)} \times \left[\frac{\mu_i^j(x_i(k))}{\mu_i^{j+1}(x_i(k))} \sum_{r'=1}^{R(\mu_i^{j+1})} \tau_{r'}(k)(b_{r'}(k) - y(k)) - \sum_{r'=1}^{R(\mu_i^j)} \tau_{r'}(k)(b_{r'}(k) - y(k)) \right], \tag{37}$$

where $r' = 1, \ldots, R(\mu_i^j)$ refers to rules having A_i^j in their premise.

4.2. Least squares

The least squares algorithm in case of simplified inference is even more straightforward than in the Sugeno case. Equation (23) is equivalent to Eq. (19) if

$$\theta = [b_1, b_2, \ldots, b_R]^{\mathrm{T}}. \tag{38}$$

For computing the estimated consequent singletons, Eq. (21) is applied again.

4.3. Clustering techniques

For standard systems, membership functions cannot be directly obtained from the projections of the fuzzy cluster membership degrees onto input space(s). That is because while Sugeno rules themselves result in linear input-output mapping, in standard systems the linear relation is a consequence of the interpolation between the neighbouring rules [10].

Common practice, therefore, is to place the centres of the triangular membership functions (Fig. 2, solid line) to where the estimated membership function projections μ_{hk} (dashed line) intersect and to add two additional sets at

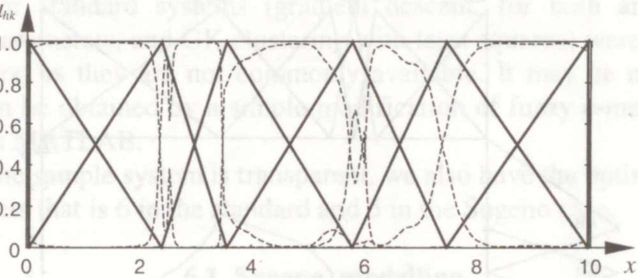

Fig. 2. Partition for standard fuzzy systems derived from GK clusters.

the extreme points of the domain. Similarly to Jager algorithm, overlap height of 0.5 is permanently maintained.

5. A SAMPLE SYSTEM

Perhaps the main area of application of fuzzy systems is approximation of complex systems. The number of rules, however, grows exponentially with the number of fuzzy sets and input variables. This is the reason why in fuzzy applications the number of inputs is usually limited to five and the number of antecedent fuzzy sets per variable does not exceed seven [11]. However, it is shown that most fuzzy systems do not possess universal approximation property if the number of antecedent sets (the number of rules) is limited [12]. A number of complexity reduction algorithms have been developed to preserve the approximation properties of high-dimensional fuzzy systems [13].

On the other hand, popular benchmarks employed in testing of adaptation algorithms such as the Jang "sombrero" function [14]

$$f(x_1, x_2) = \frac{\sin(\sqrt{x_1^2 + x_2^2})}{\sqrt{x_1^2 + x_2^2}}, \tag{39}$$

the Rosenbrock function [15]

$$f(x_1, x_2) = 100(x_1 - x_1^2) + (1 - x_2)^2, \tag{40}$$

or the fuzzy function used by Takagi and Sugeno [16], are two-variable functions. Indeed, the problems that plague the modelling of complex systems do not derive so much from the deficiencies of the modelling algorithms than from the architecture of fuzzy systems. For transparency detection, two-dimensional space is obviously the most suitable environment (particularly for Sugeno systems).

Summing up, the scheme used here for comparison of the different adaptation techniques employs the single-input single-output fuzzy system introduced in [17]. It is a standard fuzzy system with a minimum operator for conjunction and implication, and maximum for aggregation and center-of-gravity defuzzification.

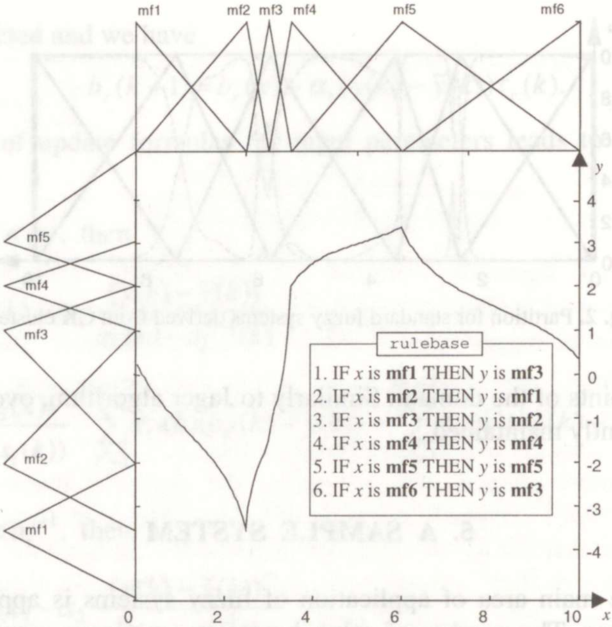

Fig. 3. A sample system: **mfi** denotes input membership functions.

Input x and output y have been partitioned into 6 and 5 fuzzy sets (denoted by linguistic labels **mfi**), respectively (Fig. 3). The parameters of these fuzzy sets and the system rulebase ($R = 6$) have been chosen so that the resulting curve (Fig. 3) includes a steep rise, making achievement of a good approximation difficult. The input signal has been discretized uniformly with the step of 0.05, so that the resulting training data set consists of 201 samples.

6. MODELLING AND RESULTS

It is a common practice to combine the algorithms for the adaptation of antecedent and consequent parameters to reduce the required human factor in the generation of the model as much as possible. Most popular combinations are clustering with the least squares procedure and gradient descent with least squares, ANFIS [14]. In the result, only the number of the rules of the model must be specified by a human expert since combined algorithms are capable to take care of everything else. In addition, special algorithms, such as compatible cluster merging or clustering with validity measures for the determination of the number of the rules, have been proposed [8] to take over also this task.

ANFIS as well as subtractive clustering with least squares are implemented through the Fuzzy Logic Toolbox of MATLAB; Gustafson–Kessel clustering with least squares is available in FMID Toolbox for MATLAB [8]. Training algorithms for standard systems (gradient descent, for both antecedent and consequent parameters, and GK clustering with least squares) were implemented by the authors, as they are not commonly available. It may be noted that GK clustering can be obtained by a simple modification of fuzzy c-means algorithm supplied with MATLAB.

Because the sample system is transparent, we also have the optimal number of input fuzzy sets that is 6 in the standard and 5 in the Sugeno case.

6.1. Sugeno modelling

6.1.1. ANFIS

ANFIS is initialized with uniformly distributed input membership functions and arbitrary nonzero consequent parameters. Low approximation error is usually achieved, provided that a sufficient number of input fuzzy sets is specified and a sufficient number of training epochs is conducted. Both numbers are input parameters of the ANFIS algorithm and can be specified by the user. It is possible to obtain relatively transparent model (Fig. 4) but the result is almost unpredictable. A transparent model was obtained with a number of input sets that reflects best the underlying nature of the modelled system. That is not final, however, as ANFIS's transparency is subject to overtraining.

System's transparency is obviously related to overlapping of the input membership functions. If those overlap strongly (Fig. 5), the rules are not local linear approximators of the global nonlinear function; that is most common with ANFIS, although the model is numerically adequate.

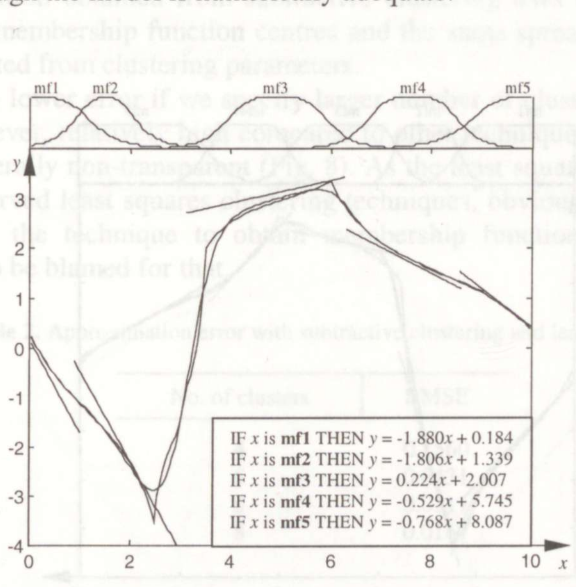

Fig. 4. ANFIS approximation (RMSE = 0.0180).

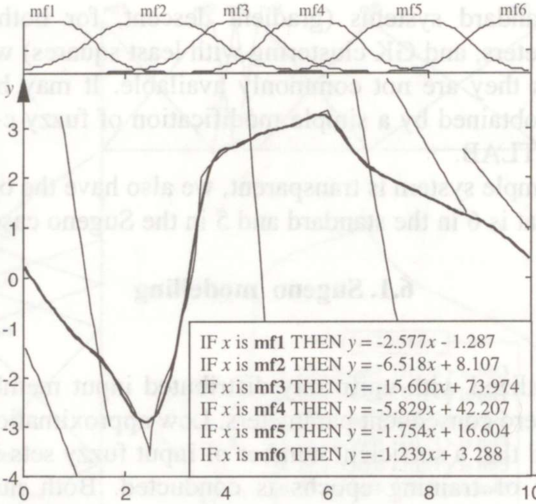

Fig. 5. ANFIS approximation (RMSE = 0.0128).

6.1.2. GK clustering and least squares

Low approximation error and good transparency of the model (Fig. 6) are universal properties of the models obtained by this method, even in cases where the number of clusters (Fig. 7) does not directly reflect the nature of the system (Table 1).

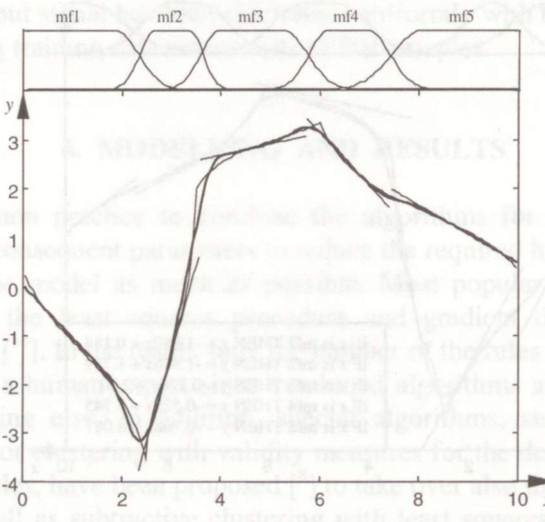

Fig. 6. Approximation with GK clustering and least squares.

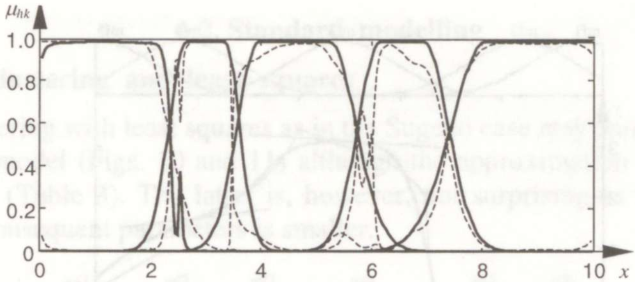

Fig. 7. Input approximation from GK clusters.

Table 1. Approximation error with GK clustering and least squares

No. of clusters	RMSE
4	0.0197
5	0.0164
6	0.0121
8	0.0114

6.1.3. Subtractive clustering and least squares

Input partition obtained from subtractive clustering uses cluster centres for defining the membership function centres and the same spread for all functions that is estimated from clustering parameters.

We obtain lower error if we specify larger number of clusters (Table 2). The error is, however, relatively high compared to other techniques and the obtained model is generally non-transparent (Fig. 8). As the least squares part is identical for both observed least squares clustering techniques, obviously, the subtractive clustering or the technique to obtain membership functions from identified clusters are to be blamed for that.

Table 2. Approximation error with subtractive clustering and least squares

No. of clusters	RMSE
4	0.1560
5	0.0431
6	0.0311
8	0.0161

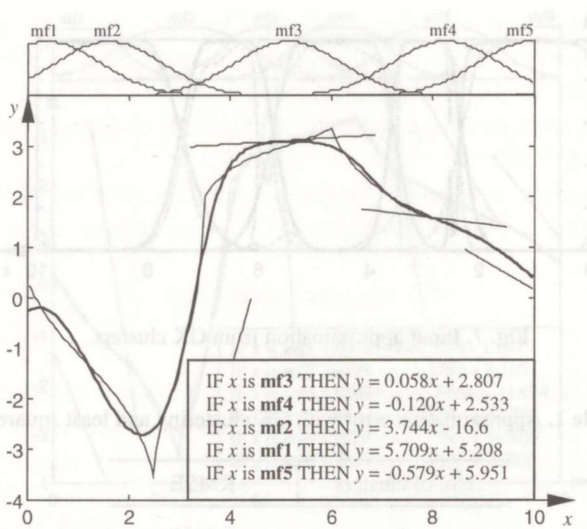

Fig. 8. Approximation with subtractive clustering and least squares.

6.1.4. GK clustering and ANFIS

ANFIS was initialized with the input partition shown in Fig. 7 and arbitrary nonzero consequent parameters and is able to improve the model both numerically (RMSE = 0.0039) and from the point of view of interpretation (Fig. 9).

It is, however, quite clear that improved transparency is a direct result of separation of the input fuzzy sets that is not generally desired.

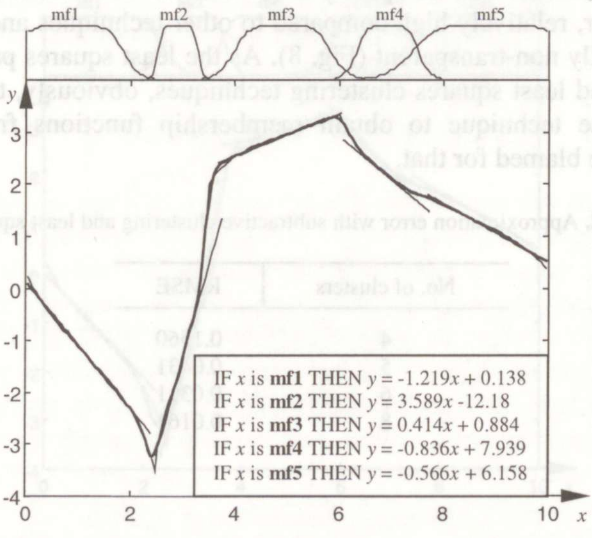

Fig. 9. Function approximation with ANFIS using GK clustering and least squares model.

6.2. Standard modelling

6.2.1. GK clustering and least squares

GK clustering with least squares as in the Sugeno case may extract a perfectly transparent model (Figs. 10 and 11) although the approximation error is somewhat bigger (Table 3). The latter is, however, not surprising as the number of adjustable consequent parameters is smaller.

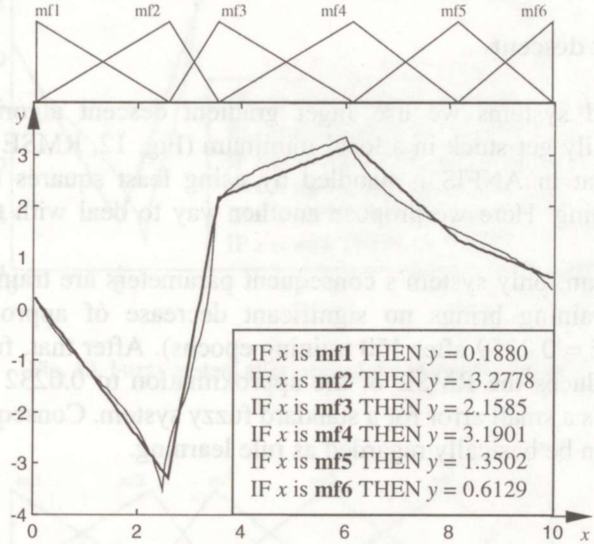

Fig. 10. Approximation with 6 GK clusters and least squares.

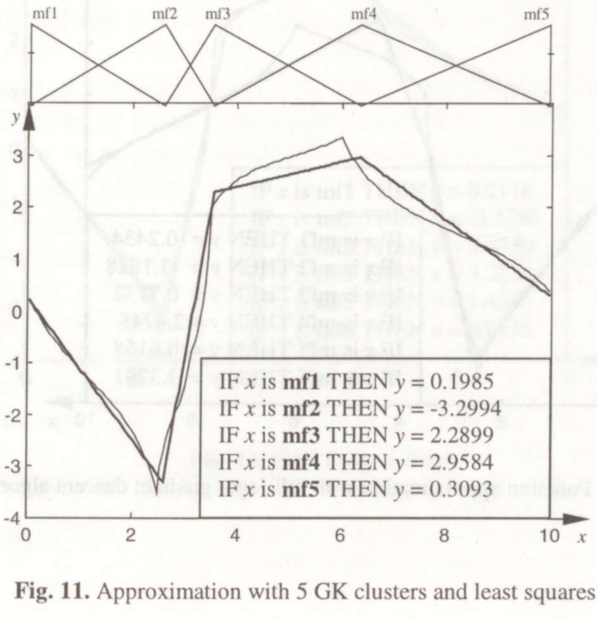

Fig. 11. Approximation with 5 GK clusters and least squares.

Table 3. Approximation error with GK clustering and least squares

No. of clusters	RMSE
5	0.0836
6	0.0586
7	0.0386
8	0.0380

6.2.2. Gradient descent

For standard systems we use Jager gradient descent algorithm. Gradient descent can easily get stuck in a local minimum (Fig. 12, RMSE = 0.0779 after 250 epochs) that in ANFIS is handled by using least squares for consequent parameter learning. Here we propose another way to deal with that unpleasant symptom.

In the first run, only system's consequent parameters are trained by Eq. (35) until further training brings no significant decrease of approximation error (Fig. 13, RMSE = 0.2159 after 150 training epochs). After that, full algorithm is applied that reduces the RMSE of the approximation to 0.0232 in 100 epochs (Fig. 14). This is a small error for a standard fuzzy system. Consequent parameter training here can be basically regarded as rule learning.

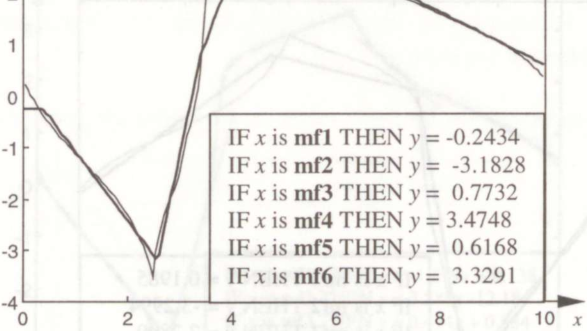

Fig. 12. Function approximation with full Jager gradient descent algorithm.

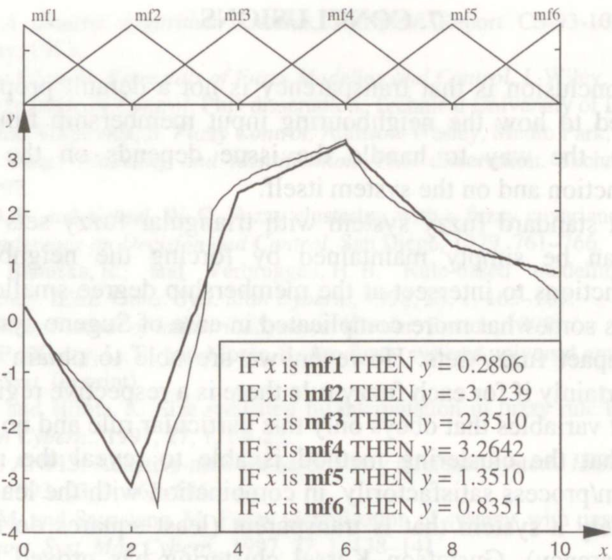

Fig. 13. Fuzzy system after consequent parameter training.

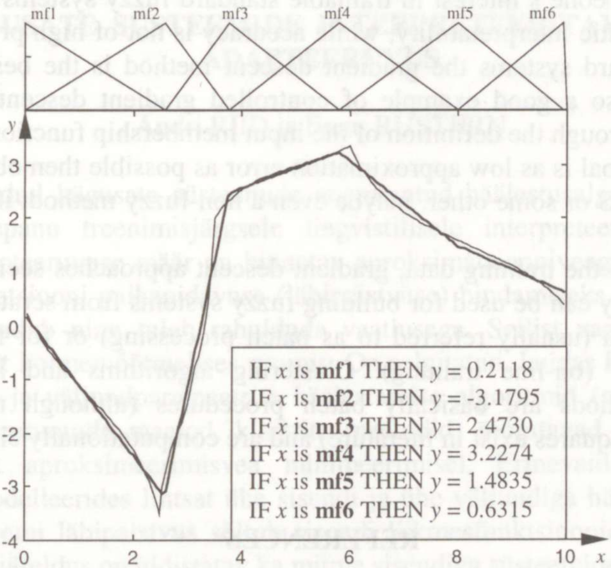

Fig. 14. Final fuzzy system.

7. CONCLUSIONS

The main conclusion is that transparency is not a default property of fuzzy systems. Related to how the neighbouring input membership functions of the system overlap, the way to handle the issue depends on the type of the membership function and on the system itself.

In case of a standard fuzzy system with triangular fuzzy sets the system's transparency can be simply maintained by forcing the neighbouring input membership functions to intersect at the membership degree smaller or equal to 0.5. The issue is somewhat more complicated in case of Sugeno systems because of the non-compact fuzzy sets. However, we are able to obtain a transparent system more certainly if for each fuzzy rule there is a respective region within the domain of input variables that obeys only this particular rule and no other rules.

Assuming that the clustering method is able to reveal the nature of the modelled system/process satisfactorily, in combination with the least squares we obtain most likely a system that is transparent (least squares do not distort or improve transparency). Gustafson–Kessel clustering has proved to be such a method. Not less important is the fact that being one of the fuzzy clustering methods, the cluster membership degrees can be directly used to determine membership functions. This approach works both for standard and Sugeno systems.

Usually, someone's interest in trainable standard fuzzy systems is due to the system's linguistic interpretability, while accuracy is not of high priority. In this case, for standard systems the gradient descent method is the best. The Jager algorithm is also a good example of controlled gradient descent to maintain transparency through the definition of the input membership functions.

If the only goal is as low approximation error as possible then obvious choice would be ANFIS or some other, maybe even a non-fuzzy method, if it serves the goal better.

In respect of the training data, gradient descent approaches seem to be more universal as they can be used for building fuzzy systems from scratch with large amounts of data (usually referred to as batch processing) or for improving an existing model (on-line training). Clustering algorithms and least squares estimation methods are basically batch procedures (although references to recursive least squares exist in literature) and are computationally cheaper.

REFERENCES

1. Kosko, B. Fuzzy systems as universal approximators. In *Proc. IEEE International Conference on Fuzzy Systems*. San Diego, 1992, 1153–1162.
2. Castro, J. L. Fuzzy logic controllers are universal approximators. *IEEE Trans. Syst. Man Cybern.*, 1995, **25**, 4, 629–635.
3. Wang, L.-X. Fuzzy systems are universal approximators. In *Proc. IEEE International Conference on Fuzzy Systems*. San Diego, 1992, 1163–1170.

4. Whitley, D. *A Genetic Algorithm Tutorial.* Technical Report CS-93-103, Colorado State University, 1993.
5. Yager, R. and Filev, D. *Essentials of Fuzzy Modeling and Control.* J. Wiley, New York, 1994.
6. Jager, R. *Fuzzy Logic in Control.* PhD dissertation. Technical University of Delft, 1995.
7. Passino, K. and Yurkovich, S. *Fuzzy Control.* Addison-Wesley, Menlo Park, 1998.
8. Babuška, R. *Fuzzy Modeling and Identification.* PhD dissertation. Technical University of Delft, 1997.
9. Gustafson, D. E. and Kessel, W. C. Fuzzy clustering with a fuzzy covariance matrix. In *Proc. IEEE Conference on Decision and Control.* San Diego, 1979, 761–766.
10. Setnes, M., Babuška, R., and Verbruggen, H. B. Rule-based modeling: Precision and transparency. *IEEE Trans. Syst. Man Cybern.*, 1998, **28**, 1, 165–169.
11. Shaw, I. S. *Fuzzy Control of Industrial Systems.* Kluwer, Boston, 1998.
12. Klement, E. P., Koczy, L. T., and Moser, B. Are fuzzy systems universal approximators? *Int. J. General Syst.* (in print).
13. Koczy, L. T. and Hirota, K. Size reduction by interpolation in fuzzy rule bases. *IEEE Trans. Syst. Man Cybern.*, 1997, **27**, 1, 14–25.
14. Jang, J.-S. R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.*, 1993, **23**, 3, 665–685.
15. Luciano, A. M. and Savastano, M. Fuzzy identification of systems with unsupervised learning. *IEEE Trans. Syst. Man Cybern.*, 1997, **27**, 1, 138–141.
16. Takagi, T. and Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.*, 1985, **15**, 1, 116–132.
17. Riid, A. and Rüstern, E. Comparison of fuzzy function approximators. In *Proc. 6th Biennal Baltic Electronics Conference.* Tallinn, 1998, 139–142.

HÄGUSATE SÜSTEEMIDE INTERPRETEERITAVUS JA ADAPTEERUVUS

Andri RIID ja Ennu RÜSTERN

On võrreldud hägusate süsteemide enamtuntud häälestusalgoritme pöörates erilist tähelepanu treenimisjärgsele lingvistilisele interpreteeritavusele. Kui süsteemi adapteeruvuse määr on hinnatav aproksimatsiooniveaga, siis lingvistilise interpretatsiooni paikapidavuse (läbipaistvuse) hindamiseks üheselt mõistetav mõõt puudub ning tuleb rahulduda vaatlusega. Sellist vaatlust saab teha maksimaalselt kolmemõõtmelises ruumis. On selgitatud, kuidas hägusate süsteemide sisend- ja väljundparameetrite häälestamise algoritmid (suurima languse meetod, vähimruutude meetod, klastrite meetodid) mõjutavad süsteemi interpreteeritavust aproksimeerimisvea minimeerimisel. Erinevaid algoritme on võrreldud modelleerides lihtsat ühe sisendi ja ühe väljundiga hägusat süsteemi. Hägusa süsteemi läbipaistvus sõltub sisendi liikmesfunktsioonide ülekattumise määrast. See järeldus on üldistatav ka mitme sisendiga süsteemidele.