



ELSEVIER

Mathematics and Computers in Simulation 46 (1998) 507–516



MATHEMATICS
AND
COMPUTERS
IN SIMULATION

Complexity reduction in fuzzy modeling

M. Setnes^{1,*}, R. Babuška, H.B. Verbruggen

*Control Laboratory, Department of Electrical Engineering, Delft University of Technology, P.O.Box 5031,
2600 GA Delft, The Netherlands*

Abstract

The interest in data driven approaches to the acquisition of fuzzy systems is increasing. Most of the approaches in the literature emphasize the global quantitative accuracy and not the transparency and interpretability of the resulting model. This paper discusses methods based on similarity analysis that, without performing additional knowledge or data acquisition, allow for the generation of fuzzy models of varying complexity. While models for simulation emphasize numerical accuracy, models for understanding the system and for operator interface are required to be transparent and interpretable. An application of the presented fuzzy modeling techniques to an air-conditioning system is described. © 1998 IMACS/Elsevier Science B.V.

1. Introduction

Computational Intelligence techniques, such as fuzzy and neural systems, have proven to be useful in the modeling of complex nonlinear systems. Both fuzzy and neural systems are recognized as universal approximators. Traditionally, a fuzzy model is built by using expert knowledge in the form of linguistic rules. Recently, there is an increasing interest in obtaining fuzzy models from measured numerical data. Different approaches have been proposed for this purpose, like fuzzy relations [13], neural network training techniques [8], and product-space clustering [2]. However, most of these approaches emphasize the global quantitative accuracy of the resulting model, and little attention is paid to linguistic and qualitative aspects, see, e.g., [10] for an example.

In this paper, we discuss methods based on similarity analysis that can be applied to fuzzy models in order to obtain models of varying complexity and qualitative properties depending on the purpose of the modeling exercise. Three approaches are considered: (1) iterative compatibility analysis [1,11], (2) similarity relations, and (3) linguistic approximation. These approaches do not require additional knowledge or data acquisition. The user can fine-tune the numerical accuracy and transparency in order to obtain a suitable model.

* Corresponding author. Fax: +31 15 2786679; e-mail: m.setnes@et.tudelft.nl

¹This work is partially supported by the Research Council of Norway.

The presented techniques are generally applicable to fuzzy rule-based models, and are illustrated on a fuzzy model of an air-conditioning system obtained from numerical data by means of product-space clustering. In the following sections, the used model structure and modeling method are first shortly described. Then the three approaches to model simplification are presented. Finally, the approaches are applied to the fuzzy model of the air-conditioning system, and the results are discussed with respect to accuracy, interpretability and computational load.

2. The Takagi–Sugeno fuzzy model

A rule-based model of the Takagi–Sugeno (TS) type [12] is considered. It consists of a set of fuzzy implications, or rules, which each describe a local input–output relation, typically in a linear form:

$$R_i : w_i(\text{If } x_1 \text{ is } A_{i1} \text{ and...and } x_n \text{ is } A_{in} \text{ then } y_i = \mathbf{a}_i \mathbf{x} + b_i), \quad i = 1, 2, \dots, K \quad (1)$$

Here R_i is the i th rule, $\mathbf{x}=[x_1, \dots, x_n]^T$ is the input (antecedent) variable, A_{i1}, \dots, A_{in} are fuzzy sets defined in the antecedent space, y_i is the rule output variable, and w_i is the rule weight. Typically, $w_i=1, \forall i$, but these weights can be adjusted during the model reduction. K denotes the number of rules in the rule base, and the aggregated output of the model, \hat{y} , is calculated by taking the weighted average of the rule consequents:

$$\hat{y} = \frac{\sum_{i=1}^K w_i \beta_i y_i}{\sum_{i=1}^K w_i \beta_i} \quad (2)$$

where β_i is the degree of activation of the i th rule:

$$\beta_i = \prod_{j=1}^n \mu_{A_{ij}}(x_j), \quad i = 1, 2, \dots, K \quad (3)$$

and $\mu_{A_{ij}}(x_j) : \mathbb{R} \rightarrow [0, 1]$ is the membership function of the fuzzy set A_{ij} in the antecedent of R_i .

To identify the TS fuzzy model, the input and the output variables must first be determined. The regression matrix \mathbf{X} and an output vector \mathbf{y} are constructed from data measurements:

$$\mathbf{X}^T = [\mathbf{x}_1, \dots, \mathbf{x}_N], \quad \mathbf{y}^T = [y_1, \dots, y_N] \quad (4)$$

Here $N \gg n$ is the number of samples used for identification. The objective of identification is to construct the unknown nonlinear function $\mathbf{y}=f(\mathbf{X})$ from the data, where f is the TS model (1).

The number of rules, K , the antecedent fuzzy sets, A_{ij} , and the consequent parameters, \mathbf{a}_i, b_i are determined by means of fuzzy clustering in the product space of $\mathcal{X} \times \mathcal{Y}$ [2]. Hence, the data set \mathbf{Z} to be clustered is composed from \mathbf{X} and \mathbf{y} :

$$\mathbf{Z}^T = [\mathbf{X}; \mathbf{y}] \quad (5)$$

Given \mathbf{Z} and an estimated number of clusters K , a fuzzy clustering algorithm [6] is applied to compute the fuzzy partition matrix \mathbf{U} . This provides a description of the system in terms of its local characteristic behavior in regions of the data identified by the clustering algorithm, and each cluster defines a rule. Cluster validity measures can be applied to select K and a suitable fuzzy partition of \mathbf{Z} [5].

The fuzzy sets in the antecedent of the rules are obtained from the partition matrix U , whose ik th element $\mu_{ik} \in [0,1]$ is the membership degree of the data object z_k in cluster i . One-dimensional fuzzy sets A_{ij} are obtained from the multidimensional fuzzy sets defined point-wise in the i th row of the partition matrix by projections onto the space of the input variables x_j :

$$\mu_{A_{ij}}(x_{jk}) = \text{proj}_j^{\mathbb{N}_{n+1}}(\mu_{ik}) \tag{6}$$

where proj is the point-wise projection operator [9]. The point-wise defined fuzzy sets A_{ij} are approximated by suitable parametric functions in order to compute $\mu_{A_{ij}}(x_j)$ for any value of x_j .

The consequent parameters for each rule are obtained as a weighted ordinary least-square estimate. Let $\theta_i^T = [a_i^T; b_i]$, let X_e denote the matrix $[X; \mathbf{1}]$ and let W_i denote a diagonal matrix in $\mathbb{R}^{N \times N}$ having the weighted degree of activation, $w_i \beta_i(x_k)$, as its k th diagonal element. If the columns of X_e are linearly independent and $w_i \beta_i(x_k) > 0$ for $1 \leq k \leq N$, then the weighted least-squares solution of $y = X_e \theta + \epsilon$ becomes

$$\theta_i = [X_e^T W_i X_e]^{-1} X_e^T W_i y \tag{7}$$

3. Simplification and reduction

The transparency of fuzzy rule-based models obtained from data is often hampered by redundancy present in the form of many overlapping (compatible) fuzzy sets. In [1,11] we proposed to use a similarity measure to assess the compatibility (pair-wise similarity) of fuzzy sets in the rule base, in order to identify fuzzy sets that can be merged. Fuzzy sets estimated from data can also be similar to the universal set, adding no information to the model. Such sets can be removed from the antecedent of a rule. These operations reduce the number of fuzzy sets in the model. Reduction of the rule base follows when the antecedents of some rules become equal. Such rules are combined into one rule. In the following, we describe three approaches to model simplification and reduction. To assess the compatibility of fuzzy sets we apply the fuzzy analog to the Jaccard index [3]:

$$c_{jlm} = \frac{|A_{lj} \cap A_{mj}|}{|A_{lj} \cup A_{mj}|} \tag{8}$$

where $l, m = 1, 2, \dots, K$, and $c_{jlm} \in [0,1]$. The \cap and \cup operators are the intersection and the union, respectively, and $|\cdot|$ denotes the cardinality of a fuzzy set [3,7]. The measure c_{jlm} is computed for discretized domains and quantifies the compatibility between the fuzzy sets A_{lj} and A_{mj} in the rules R_l and R_m , respectively.

Iterative compatibility analysis. This approach is based on iterative merging of compatible fuzzy sets [1,11]. It requires two thresholds from the user, $\lambda, \gamma \in (0,1)$ for merging compatible fuzzy sets, and removing fuzzy sets compatible with the universal set, respectively. In each iteration, the compatibility between all fuzzy sets in each antecedent dimension j is analyzed. The pair of fuzzy sets having the highest compatibility $c > \lambda$ are merged. A new fuzzy set is created by merging, and the rule base is updated by substituting this fuzzy set for the ones merged. The algorithm again evaluates the updated

Table 1
Two algorithms for fuzzy rule base simplification

Given a rule base $\mathbf{R} = \{R_i \mid i=1, \dots, K\}$, where R_i is given by (1). Select the thresholds $\lambda, \gamma \in (0,1)$:	
<p>Repeat for $j=1, 2, \dots, n$: Step 1: Select most compatible fuzzy sets:</p> $A_{L_j} = \left\{ A_{ij} \mid c_{jlm} = \max_{\substack{i \neq p \\ i,p=1, \dots, K}} (c_{jip}) \right\}$ <p>Step 2: Merge selected fuzzy sets: If $c_{jlm} > \lambda$:</p> $A_{c_j} = \text{Merge}(A_{L_j})$ $\forall A_{ij} \in A_{L_j}, \text{ set } A_{ij} = A_{c_j}$ <p>Until: $c_{jlm} < \lambda$</p> <p>Step 3: Remove fuzzy sets similar to universal set:</p> $c_{ij} = \frac{ A_{ij} \cap U_j }{ A_{ij} \cup U_j }, \quad i = 1, 2, \dots, K$ <p>where $\mu_{U_j} = 1, \forall x_j$ If $c_{ij} > \gamma$, remove A_{ij} from the antecedent of R_i. (a) <i>Iterative compatibility analysis.</i></p>	<p>Repeat for $j=1, 2, \dots, n$: Step 1: Calculate similarity relation:</p> $C_j = [c_{jlm}], \quad l, m = 1, 2, \dots, K$ $S_j = [s_{jlm}] = C_{T_j}$ <p>Step 2: Merge similar fuzzy sets:</p> $A_{L_j} = \{A_{ij} \mid s_{jlm} > \lambda, l \neq m\}$ $A_{c_j} = \text{Merge}(A_{L_j})$ <p>$\forall A_{ij} \in A_{L_j}, \text{ set } A_{ij} = A_{c_j}$</p> <p>Step 3: Remove fuzzy sets similar to universal set:</p> $c_{ij} = \frac{ A_{ij} \cap U_j }{ A_{ij} \cup U_j }, \quad i = 1, 2, \dots, K$ <p>where $\mu_{U_j} = 1, \forall x_j$ If $c_{ij} > \gamma$, remove A_{ij} from the antecedent of R_i. (b) <i>Similarity relations.</i></p>

rule base, until there are no more fuzzy sets for which $c > \lambda$. Fuzzy sets compatible with the universal set ($c > \gamma$) are removed from the rules in which they occur. The algorithm is given in Table 1(a).

Similarity relations. Also this approach requires two thresholds λ and γ . For each antecedent dimension, $j=1, \dots, n$, a similarity relation between the fuzzy sets is obtained in two steps: First, a $K \times K$ binary fuzzy compatibility relation $C_j = [c_{jlm}]$ is calculated, whose elements are obtained by (8). C_j is reflexive and symmetric. Second, a similarity relation, S_j , is calculated as the max–min transitive closure, C_{T_j} , of C_j [7]:

1. $C'_j = \max[C_j, (C_j \circ C_j)]$
2. If $C'_j \neq C_j$, set $C_j = C'_j$ and go to 1
3. Stop : $C_{T_j} = C'_j$, set $S_j = C_{T_j}$

Here \circ is the max–min composition. The lm th element of S_j , $[s_{jlm}]$, gives the similarity between A_{ji} and A_{jm} . For each antecedent dimension, the fuzzy sets having similarity $s_{jlm} > \lambda$ are merged. Fuzzy sets compatible with the universal set are removed. This algorithm is given in Table 1(b).

The first approach merges only one pair of fuzzy sets per iteration and the rule base is updated between the iterations. The second approach merges all similar fuzzy sets per dimension simultaneously. Hence, the use of the transitive similarity relation may give different results than the

iterative approach. Merging of fuzzy sets is accomplished by letting the support of the union of the sets in \mathbf{A}_{L_j} be the support of the new fuzzy set A_{c_j} . This guarantees preservation of the coverage of the antecedent space. The kernel of A_{c_j} is given by averaging the kernels of the sets in \mathbf{A}_{L_j} .

If the antecedents of $p \geq 2$ rules become equal, the p rules can be replaced by *one* common rule R_c . The consequent parameters of the reduced rule base can be re-estimated from training data (7), or one can calculate the parameters of R_c from the parameters of the p removed rules. The latter method does not depend on the availability of data. This approach is now described: Let $Q \subset \{1, 2, \dots, K\}$ be a subset of rule indices such that $A_{lj} = A_{mj}, \forall j \in \{1, 2, \dots, n\}, \forall l, m \in Q$. \mathbf{R}_Q then denotes the set of rules with equal antecedents. The rule R_c replaces the rules in \mathbf{R}_Q , and its antecedent part equals that of the rules \mathbf{R}_Q , i.e., $A_{c_j} = A_{l_j}, j = 1, 2, \dots, n, l \in Q$. The one common rule R_c is made to account for all the rules \mathbf{R}_Q by weighting it with the total weight of the rules $\mathbf{R}_Q, w_c = \sum_{i \in Q} w_i$, and by letting its consequent y_c be an average of the consequents of \mathbf{R}_Q . Thus, the set of rules \mathbf{R}_Q is represented by a single rule R_c with weight w_c and consequent parameters

$$\theta_c = \frac{1}{w_c} \sum_{i \in Q} w_i \theta_i \tag{9}$$

Let $\bar{Q} = \{1, \dots, K\} - Q$, the model output (2) now becomes

$$\hat{y} = \frac{\sum_{i \in \bar{Q}} w_i \beta_i y_i + w_c \beta_c y_c}{\sum_{i \in \bar{Q}} w_i \beta_i + w_c \beta_c} \tag{10}$$

This substitution of \mathbf{R}_Q by R_c does not alter the input–output mapping of the TS-model (1).

Linguistic approximation. A fuzzy model can be interpreted by means of linguistic approximation [4]. Using (8), the fuzzy sets in the model are compared to some reference fuzzy sets and their modifications by selected linguistic hedges. The model is described in terms of the labels of the reference fuzzy sets. By substituting the reference fuzzy sets for the original fuzzy sets, the model can be directly interpreted linguistically as well as verified in simulation. Three reference fuzzy sets are used in our example, ‘Small’, ‘Medium’ and ‘Big’, shown in Fig. 1(a), together with the linguistic hedges in Table 2.

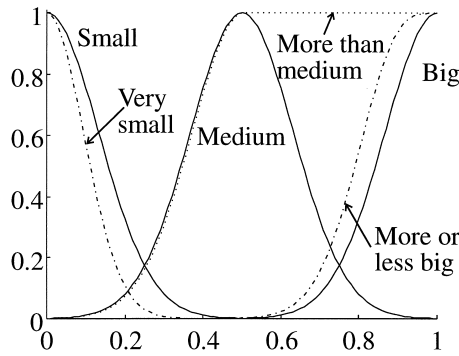
4. Application to a model of an air-conditioning system

We consider a model of an air-conditioning system consisting of a fan-coil unit. Hot or cold water is supplied to the coil through a valve. In the unit, outside (primary) air is mixed with return air from the room, see Fig. 1(b). From systems measurements, a TS fuzzy model has been obtained by clustering in 10 clusters. The model predicts the supply air temperature T_s based on its present and previous value, the mixed air temperature T_m , and the heating valve position u , thus:

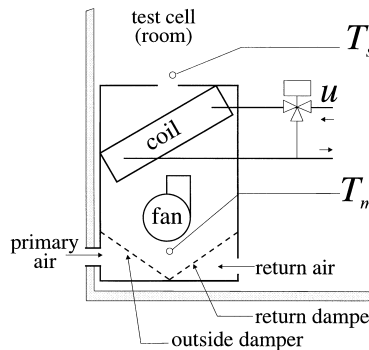
$$\mathbf{x}(k) = [T_s(k), T_s(k - 1), u(k - 1), T_m(k)]^T, \quad y(k) = T_s(k + 1) \tag{11}$$

The model consist of ten rules, each with four antecedent fuzzy sets, of the form:

$$R_i : \mathbf{IF} T_s(k) \text{ is } A_{i1} \text{ and } T_s(k - 1) \text{ is } A_{i2} \text{ and } u(k - 1) \text{ is } A_{i3} \text{ and } T_m(k) \text{ is } A_{i4} \\ \mathbf{THEN} T_s(k + 1) = y_i$$



(a) Reference fuzzy sets.



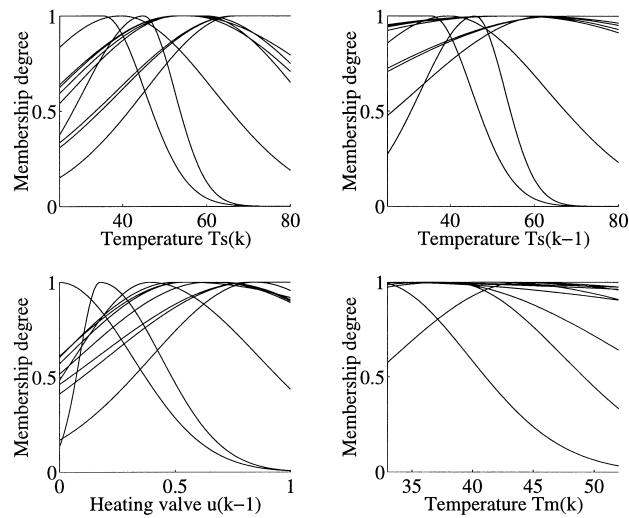
(b) Air-conditioning system.

Fig. 1. Reference fuzzy sets (a) and the system considered in the application (b).

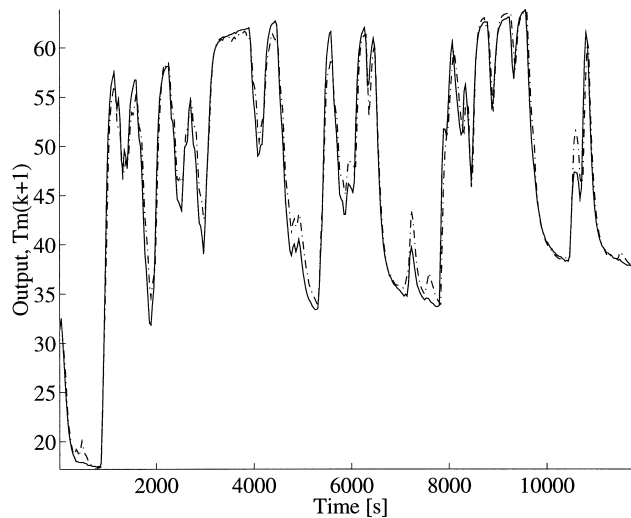
Table 2
Linguistic hedges

Linguistic hedge	Operation	Linguistic hedge	Operation
Very A	μ_A^2	More than A	$\begin{cases} \mu_A(x), & \text{if } x < \min\{x \mid \mu_A(x) = 1\} \\ 1, & \text{otherwise.} \end{cases}$
More or less A	$\sqrt{\mu_A}$	Less than A	$\begin{cases} \mu_A(x), & \text{if } x > \max\{x \mid \mu_A(x) = 1\} \\ 1, & \text{otherwise.} \end{cases}$

where $y_i = \theta_i^T [x(k)^T; 1]^T$. The total of 40 antecedent fuzzy sets used by the model are shown in Fig. 2. Applying the similarity relation approach of Table 1(b) with $\lambda=0.8$ and $\gamma=0.9$ gives an interesting result for the two antecedent variables $T_s(k)$ and $T_s(k-1)$. The partitionings of their domains are equal. This result is supported by knowledge about sampling time and the system's dynamics. This suggests that one of the two variables could be removed from the model in further analysis. The simplified and reduced model consist of only four rules, given in Table 3, and nine fuzzy sets, shown in Fig. 3. The consequents of the rules in the new model has been reestimated using the original training data. The result is a much more transparent and less computational intensive model with comparable numeric



(a) Fuzzy sets.



(b) Recursive simulation.

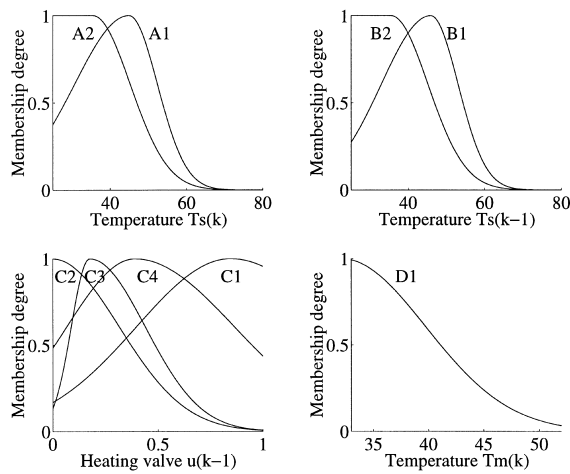
Fig. 2. Original model: fuzzy sets (a), recursive simulation (b).

accuracy. In a recursive simulation consisting of 397 predictions of unseen data, the original model uses 1.4 Mflops, while the simplified model uses 0.4 Mflops. In this simulation, the root mean square (RMS) error of the original model is 1.89, while the RMS error of the new model is 1.91.

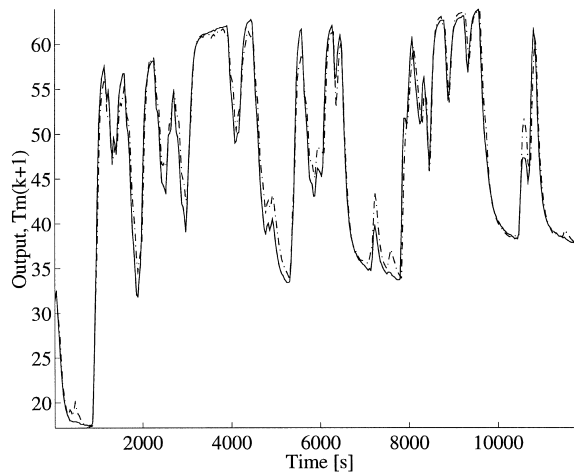
The application of linguistic approximation to the original model also gives a highly reduced rule base with 10 qualitative linguistic terms in 5 rules with the following antecedent parts:

Table 3
Simplified model

IF	$T_s(k)$ is,	$T_s(k-1)$ is,	$u(k-1)$ is,	$T_m(k)$ is,	THEN $T_s(k+1)=$
$R'_1:$	–	–	C_1	–	y_1
$R'_2:$	A_1	B_1	C_2	–	y_2
$R'_3:$	A_2	B_2	C_3	D_1	y_3
$R'_4:$	–	–	C_4	–	y_4



(a) Fuzzy sets.



(b) Recursive simulation.

Fig. 3. Original model: fuzzy sets (a), recursive simulation (b).

If $T_s(k)$ is More than Low,	$T_s(k-1)$ is More than Low,	$u(k-1)$ is More than Low,	$T_m(k)$ is Less than High
If $T_s(k)$ is More than Low,	$T_s(k-1)$ is Less than High,	$u(k-1)$ is More than Low,	$T_m(k)$ is Less than High
If $T_s(k)$ is Less than High,	$T_s(k-1)$ is Less than High,	$u(k-1)$ is Less than High,	$T_m(k)$ is More than Low
If $T_s(k)$ is Less than High,	$T_s(k-1)$ is Less than High,	$u(k-1)$ is Medium,	$T_m(k)$ is Less than High
If $T_s(k)$ is Less than High,	$T_s(k-1)$ is Less than High,	$u(k-1)$ is More or less Med.,	$T_m(k)$ is Less than High

This linguistic description matches the simplified model in Fig. 3 quite well. For both variables $T_s(k)$ and $T_s(k-1)$, a partition into two fuzzy regions is found, and for variable $u(k-1)$, a partitioning into four fuzzy regions is found. For the input $T_m(k)$, both methods recognize the region ‘Less than High’, but the linguistic models also uses the region ‘More than Low’. This region is removed from the simplified model due to its similarity with the universal set, and is thus implicitly present with a membership 1 in the rules R_1' , R_2' and R_4' in Table 3 where $T_m(k)$ is not used in the premise. The accuracy of the linguistic model is verified in a recursive simulation giving an RMS error of 2.17.

5. Conclusions

Methods for complexity reduction in fuzzy models acquired from numerical data have been presented. The methods are based on similarity analysis of the fuzzy sets used in the antecedent space of the model. Distinction is made between iterative rule base reduction, reduction based on transitive similarity relations, and on linguistic approximation.

The consequent parameters of the reduced fuzzy models can be re-estimated by least-squares techniques or recomputed from the parameters of the original model. The presented techniques have been applied to the fuzzy modeling of a real-world air-conditioning system. It is shown that the originally obtained model can be strongly reduced, allowing for qualitative interpretation, and faster computations, without deteriorating the prediction accuracy.

References

- [1] R. Babuška, M. Setnes, U. Kaymak, H.R. van Nauta Lemke, Rule base simplification with similarity measures, Proceedings FUZZ-IEEE'96, New Orleans, USA, 1996, p. 1642–1647.
- [2] R. Babuška, H.B. Verbruggen, Fuzzy set methods for local modeling and identification, in: R. Murray-Smith, T.A. Johansen (Eds.), Multiple Model Approaches to Nonlinear Modeling and Control, Taylor & Francis, London, UK, 1997.
- [3] D. Dubois, H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1980.
- [4] F. Esmagh, E.H. Mamdani, A general approach to linguistic approximation, Int. J. Man-Machine Studies 11 (1979) 501–519.
- [5] I. Gath, A.B. Geva, Unsupervised optimal fuzzy clustering, IEEE Trans. Pattern Analysis and Machine Intelligence 11(7) (1989) 773–781.
- [6] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, Proceedings IEEE CDC, San Diego, USA, 1979, p. 761–766.
- [7] G.J. Klir, B. Youan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, New Jersey, 1995.
- [8] B. Kosko, Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice Hall, New Jersey, 1992.
- [9] R. Kruse, J. Gebhardt, F. Klawonn, Foundations of Fuzzy Systems, Wiley, Chichester, 1994.
- [10] J.H. Nie, T.H. Lee, Rule-based modeling: Fast construction and optimal manipulation, IEEE Trans. Systems, Man and Cybernetics – Part A: Systems and Humans 26 (1996) 728–738.

- [11] M. Setnes, Fuzzy Rule-Base Simplification Using Similarity Measures, M.Sc. Thesis, Delft University of Technology, Dep. of El. Eng., Control Laboratory, Delft, The Netherlands, 1995, (A.95.023).
- [12] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modelling and control, IEEE Trans. Systems, Man and Cybernetics 15 (1985) 116–132.
- [13] R.M. Tong, The evaluation of fuzzy model derived from experimental data, Fuzzy Sets and Systems 4 (1980) 1–12.