

Nonlinear context adaptation in the calibration of fuzzy sets

W. Pedrycz^{a,*}, R.R. Gudwin^b, F.A.C. Gomide^b

^a Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada R3T 2N2

^b Unicamp/FEE/DCA, Cx. P. 6101, 13081 - 970 Campinas - SP, Brazil

Received October 1995

Abstract

In this note we elaborate on the concept and use of context adaptation. The underlying idea hinges upon a nonlinear transformation of an actual reference unit universe of discourse into a subset of reals, say $[a, b]$, that is implied by actually available data (current context). Assuming a collection of fuzzy sets $\mathcal{A} = \{A_1, A_2, \dots, A_n\}$ defined over $[0, 1]$, the context adaptation gives rise to a new frame of cognition $\mathcal{A}' = \{A'_1, A'_2, \dots, A'_n\}$ expressed over $[a, b]$. Owing to the inherent nonlinearity of the developed mapping, different elements (fuzzy sets) of \mathcal{A} can be “stretched” or “expanded” according to the given experimental data. Proposed is a neural network as a relevant optimization tool. ©1997 Elsevier Science B.V.

Keywords: Context adaptation; Frame of cognition; Knowledge representation; Information granularity; Neurocomputing

1. Introduction

The frame of cognition [4] constitutes a fundamental concept of fuzzy information processing. It can be readily encountered in almost if not all basic constructs involving fuzzy sets. Fuzzy controllers, fuzzy modeling, and fuzzy clustering are just a few representative examples that are dwelled upon the fundamental concept of set membership.

As formed by a family of fuzzy sets, the frame of cognition provides us with a way of linguistic space quantization. The linguistic terms play an instrumental role in encoding both numerical and nonnumerical information that takes place prior to its further processing. It is obvious that linguistic terms (fuzzy sets) are not universal. When speaking about *comfortable* speed we confine ourselves to a certain context and interpret this term accordingly. When the context

changes so is the meaning of the term. Nevertheless, an order of the terms forming the frame of cognition is retained. For instance, in the frames

$$\mathcal{A} = \{low\ speed, comfortable\ speed, high\ speed\},$$

$$\mathcal{A}' = \{low' speed, comfortable' speed, high' speed\},$$

the order of the terms standing there is preserved no matter how much the meaning attached to the terms tend to vary. The membership functions of the elements of \mathcal{A} and \mathcal{A}' could be very distinct, though. As illustrated in Fig. 1, the same notion of *comfortable* speed in \mathcal{A} is more specific than its linguistic counterpart encountered in \mathcal{A}' .

The issue we are interested in pursuing in this study addresses a key question about the generality of linguistic terms. By solving this fundamental problem of representation of fuzzy notions we will be at position of constructing relevant membership functions based upon experimental data being currently used.

* Corresponding author. E-mail: pedrycz@ee.umanitoba.ca.

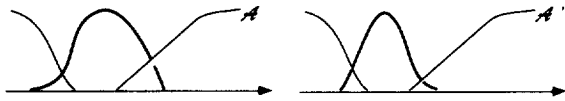


Fig. 1. Examples of two frames of cognition \mathcal{A} and \mathcal{A}' .

In the existing literature one can envision some different attempts aimed at handling this problem:

(1) The earliest approach emerges in the form of a linear scaling of the numerical variables prior to their linguistic encoding (fuzzification). As outlined in the literature, these changes do not concern directly the universe of discourse. Nevertheless the procedure lends itself to a *linear* scaling of the universe. The methods of this category were characteristic to fuzzy controllers (scaling factors); they were first reported in [5], cf. also [1].

(2) In [3] proposed was an idea of linear and *polynomial-based* modification of the universe of discourse – the discussion was focused on the design of fuzzy controllers.

(3) In [2] discussed was the idea of context adaptation in which the universe of discourse is changed according to the actual values of the variable under consideration; one of the particular alternatives arising therein exploits exponential filtering.

To put the problem into a certain perspective and emphasize again its vital importance, let us briefly review a simple example emerging in the realm of fuzzy controllers that definitely calls for the study in context adaptation. As far as the design of the fuzzy controller is concerned, the key point made therein is that the rules of the controller are expected to be *universal* to a high extent. This means that the controller utilizing the same collection of prudently established control rules need to perform equally well when applied to a broad range of problems (systems) exhibiting a similar pattern of dynamic behavior. What makes the controller adaptive are the mechanisms of context adaptation that apply exclusively to the input and output variables of the controller fully retaining the domain control knowledge, see Fig. 2.

In other words, context adaptation provides us with a useful tool of calibration of all the variables thus making the rules sensitive to the variety of the conditions formulated by the control environment. It is really what makes a human being so successful in ad-

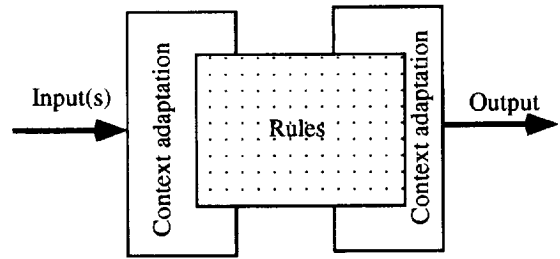


Fig. 2. Fuzzy controller and context adaptation.

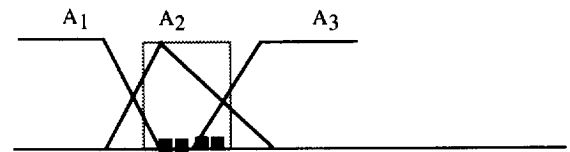


Fig. 3. An effect of underutilization of linguistic terms (and rules) in the control protocol – only A_2 fully exploited.

justing himself to a variety of situations – we do not learn from scratch but adapt to the changeable environment. To an average driver driving a different make of a car may require some adaptation but does not involve excessive learning.

This need for a prudent calibration is additionally illustrated in Fig. 3; note that the lack of the adaptation mechanism could easily make the controller idle by restricting its activity to a fairly narrow range of the control protocol thus making only a few rules to become fully responsible for the performance of the entire controller.

2. Problem statement

Let us consider a collection of generic fuzzy sets (linguistic terms)

$$\mathcal{A} = \{A_1, A_2, \dots, A_c\}$$

defined in $[0, 1]$. As usual, cf. [4], we require that \mathcal{A} satisfies some obvious requirements of semantic integrity. Essentially, we insist on unimodality and normality of the membership functions of the generic fuzzy sets. Moreover, we request that A_i 's do not fully overlap.

Given a data set of experimental outcomes (arising e.g., as a result of expert polling), they can be arranged

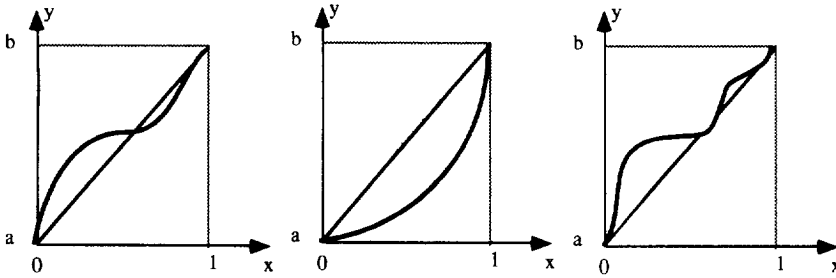


Fig. 4. Examples of nonlinear mappings between the universes of discourse.

in the form of $(c + 1)$ -tuples, namely

$$\begin{aligned} &(d_1, (\mu_{11}, \mu_{12}, \dots, \mu_{1n})), \\ &(d_2, (\mu_{21}, \mu_{22}, \dots, \mu_{2n})), \\ &\quad \vdots \\ &(d_N, (\mu_{N1}, \mu_{N2}, \dots, \mu_{Nn})), \end{aligned}$$

where d_k denotes a given element of the universe of discourse whose membership grades to some linguistic categories under discussion are equal to $\mu_{k1}, \mu_{k2}, \dots, \mu_{kn}$, respectively. Our intent is to accommodate these data to the highest extent by adapting the context of \mathcal{A} . The essence of this process is to nonlinearly map the unit interval of the generic universe of discourse (unit interval) onto the current one being an interval $[a, b]$ in \mathbf{R} whose bounds are specified as

$$a = \min_{1 \leq k \leq N} d_k, \quad b = \max_{1 \leq k \leq N} d_k.$$

This makes the generic membership functions adjusted to the current situation conveyed by the available data – thus the context in which the frame of cognition \mathcal{A} has been originally developed becomes modified (adapted) to the new environment. Due to the inherently nonlinear character of this mapping, see Fig. 4, context adaptation expands some subregions of the unit interval while contracts the others – this feature is definitely not accessible through a straightforward linear mapping (linear scaling).

3. The optimization algorithm

The calibration of the universe of discourse is carried out in two main steps:

- (i) Identification of a position of the collected membership values in the unit interval by locating the available membership vectors with respect to the linguistic labels of the original frame of cognition.
- (ii) Construction of a nonlinear mapping involving the locations derived in (i)

The first step concerns a specification of an element in the unit interval such that the given vector $\mu_1, \mu_2, \dots, \mu_n$, (the first index pertaining to the data point has been suppressed) matches the vector of the membership values in \mathcal{A} to the highest extent. This leads to the optimization task of the form

$$\min_{x \in [0, 1]} \|A(x) - \mu\| = \|A(x_0) - \mu\|,$$

where

$$A(x) = [A_1(x) A_2(x) \dots A_c(x)]$$

and

$$\mu = [\mu_1 \ \mu_2 \ \dots \ \mu_c]$$

while $\|\cdot\|$ is a certain normalized distance function computed between the corresponding membership values.

The result of this processing phase is concisely summarized in the form of the pairs of the corresponding elements defined in $[0, 1]$ and $[a, b]$, respectively,

$$\begin{aligned} &(x_1, d_1, f_1), \\ &(x_2, d_2, f_2), \\ &\quad \vdots \\ &(x_N, d_N, f_N). \end{aligned}$$

Each of these discrete associations, (x_k, d_k) , is equipped with the resulting relevance factor (coefficient) f_k determined as

$$f_k = 1 - \min_{x \in [0,1]} \|A(x) - \mu\|.$$

If $f_k \approx 1$ then the associated correspondence is regarded as highly essential.

The second step of the optimization algorithm departs from the pairs of data summarized in the above form and constructs the nonlinear mapping

$$\varphi: [0, 1] \rightarrow [a, b].$$

To properly address the core issue of context adaptation, we impose several straightforward requirements on the above mapping such as:

- Continuity.
- Monotonicity. We require that φ is nondecreasing (we allow it to remain constant over some regions of the universe of discourse). This requirement assures us that the meaning of the mapped linguistic terms is not changed (the semantics becomes retained). Eventually, we may request that φ is nondecreasing – by imposing this requirement we consistently reverse the meaning of the linguistic terms of \mathcal{A} .

- Boundary conditions. The boundary conditions $\varphi(0) = a$ and $\varphi(1) = b$ allow us to fully accommodate currently available experimental data.

In light of the above properties φ is a one-to-one mapping.

Finally, once this nonlinear transformation has been constructed, we locate the original fuzzy sets of \mathcal{A} in the actual universe of discourse $([a, b])$ by computing

$$A'_i = \varphi(A_i),$$

namely,

$$A'_i(y) = \varphi(A_i(x)),$$

$i = 1, 2, \dots, c$. When collected together these new fuzzy sets form the required frame of cognition \mathcal{A}' .

4. Neural network realization of the nonlinear mapping

The nonlinear mapping is realized through a neural network with its structure shown in Fig. 5.

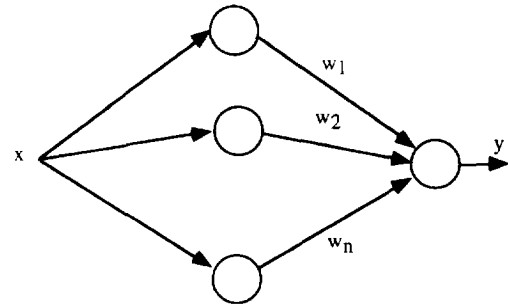


Fig. 5. Architecture of a neural network.

The network is composed of “ n ” nodes situated in the hidden layer and a single node placed at the output layer. The neurons in the hidden layer implement a series of receptive fields equipped with two-parametric sigmoid nonlinearities. The connections of these elements are fixed and equal to 1. Formally, speaking we obtain

$$z_i = \frac{1}{1 + e^{-x_i(x - m_i)}},$$

$i = 1, 2, \dots, n$ where $m_i \in [0, 1]$ $x_i > 0$, are the modal values and spreads of the corresponding fields. The neuron forming the output layer is described as

$$y = a + \frac{b - a}{1 + e^{-z}}$$

with

$$z = \sum_{i=1}^n w_i z_i.$$

Concisely, the network can be written down as a single input–single output mapping of the form

$$y = NN(x).$$

The learning of the network is supervised and guided via a gradient-based optimization of a specified performance index. As the training method is standard to a high degree, the details are not discussed here. Moreover, the proposed method easily generalizes to multidimensional case.

5. Numerical experiments

As a numerical illustration of the algorithm we consider a data set summarized in Table 1.

Table 1
Experimental data (d_k, μ_k)

| d | Membership | | | | |
|-------|------------|------|------|------|------|
| 2.00 | 1.00 | 0.30 | 0.10 | 0.00 | 0.00 |
| 4.30 | 0.80 | 0.50 | 0.40 | 0.04 | 0.00 |
| 4.70 | 0.70 | 0.65 | 0.42 | 0.06 | 0.00 |
| 5.60 | 0.30 | 0.98 | 0.60 | 0.20 | 0.05 |
| 7.90 | 0.10 | 1.00 | 0.40 | 0.00 | 0.00 |
| 8.20 | 0.00 | 0.00 | 0.70 | 0.60 | 0.20 |
| 9.80 | 0.00 | 0.00 | 0.96 | 0.50 | 0.30 |
| 12.60 | 0.00 | 0.00 | 1.00 | 0.30 | 0.20 |
| 13.60 | 0.00 | 0.10 | 1.00 | 0.20 | 0.05 |
| 14.10 | 0.12 | 0.56 | 0.87 | 0.67 | 0.20 |
| 16.30 | 0.10 | 0.10 | 0.25 | 0.80 | 0.92 |
| 18.10 | 0.00 | 0.05 | 0.20 | 0.40 | 1.00 |

Table 2
A list of associations of elements in $[0, 1]$ and $[a, b]$

| x | y |
|--------|---------|
| 0.1260 | 2.0000 |
| 0.3200 | 4.3000 |
| 0.3700 | 4.7000 |
| 0.4500 | 5.6000 |
| 0.5100 | 7.9000 |
| 0.5800 | 8.2000 |
| 0.6700 | 9.8000 |
| 0.8400 | 12.6000 |
| 0.8600 | 13.6000 |
| 0.9300 | 14.1000 |
| 0.9900 | 16.3000 |
| 1.0000 | 18.1000 |

This family of data consists of the elements situated in a segment of real numbers $[2, 18.1]$ that are assigned to five linguistic categories. The fuzzy sets of \mathcal{A} defined in $[0, 1]$ are defined using Gaussian membership functions

$$G(x; m, \sigma) = \exp(-(x - m)^2/\sigma).$$

More specifically,

$$A_1(x) = G(x; 0, 0.02),$$

$$A_2(x) = G(x; 0.25, 0.02),$$

$$A_3(x) = G(x; 0.50, 0.02),$$

$$A_4(x) = G(x; 0.75, 0.02),$$

$$A_5(x) = G(x; 1.00, 0.02).$$

First the experimental data are associated with the corresponding elements of the unit interval so that (1) becomes minimized – these results are given in Table 2.

The learning of the mapping has been completed using standard gradient-based technique assuming the form

$$\mathbf{param} = \mathbf{param} - \alpha \frac{\partial Q}{\partial \mathbf{param}}.$$

$\alpha \in (0, 1)$, where Q is a sum of squared errors observed between the experimental data and the outputs produced by the network. The confidence factors standing in (2) are not involved in the learning procedure and all data are treated uniformly. The vector of

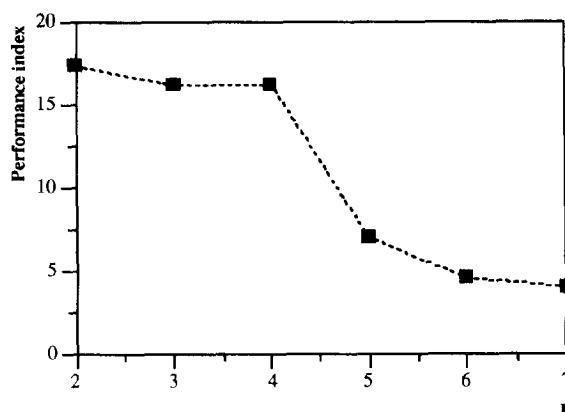


Fig. 6. Performance index Q versus “ n ”.

parameters, **param**, to be adjusted concerns the connections between the hidden and output layer as well as the parameters of the sigmoid functions (spreads and modal values). As the learning was highly sensitive to the changes of the latter, the learning rate used in the training was kept at a low level – the experiments were completed for $\alpha = 0.0005$. The values of Q for several sizes of the hidden layer are visualized in Fig. 6.

As clearly visible, a significant improvement occurs at $n = 5$; subsequently this case is discussed in detail. To visualize the character of learning Fig. 7 shows the changes of Q as they occur in the first 400 learning epochs.

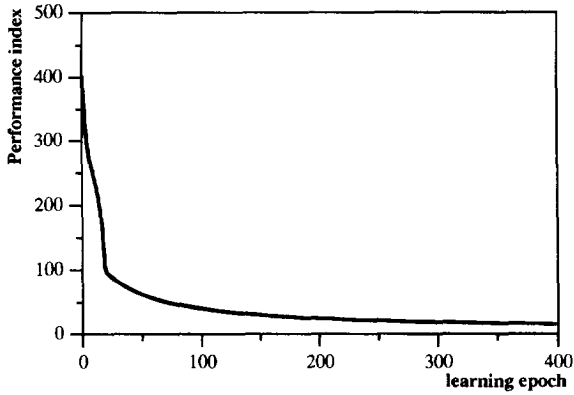


Fig. 7. The values of the performance index Q in successive learning epochs.

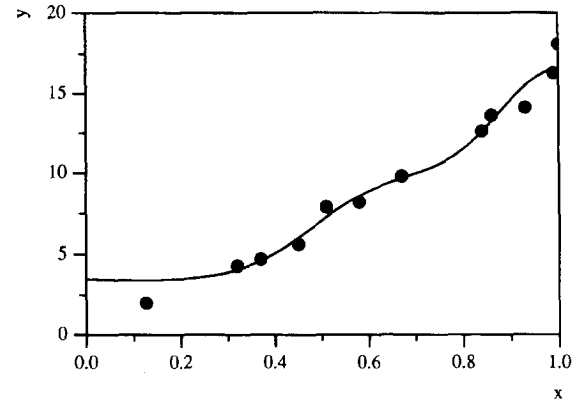


Fig. 8. Results of nonlinear mapping provided by the network.

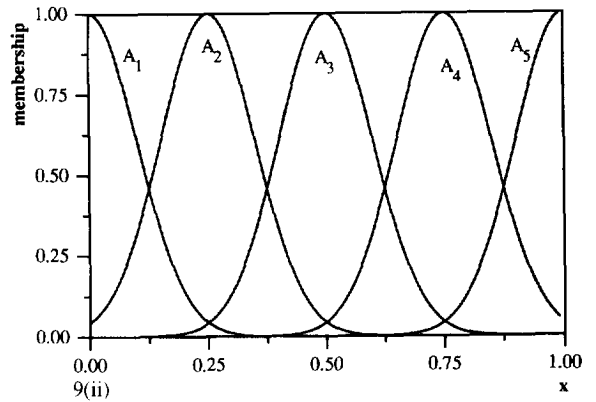
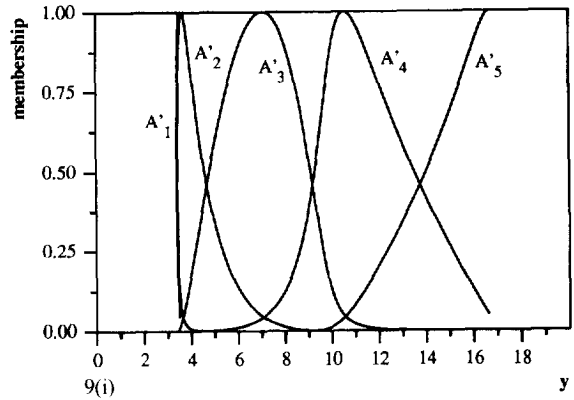


Fig. 9. Frames of cognition and membership functions of their elements (i) \mathcal{A}' and (ii) \mathcal{A} .

The form of the nonlinear mapping produced by the network is illustrated in Fig. 8 – this is shown vis at vis the experimental data (Table 2). Subsequently, Fig. 9 displays the membership functions resulting from the process of context adaptation – to ease a comparison the original membership functions defined in the unit interval are included as well, Fig. 9(ii).

As a continuation of this example, we proceed with the same linguistic terms and adapt them by considering the data contained in Table 3.

Observe that the experimental results call for a decreasing characteristics of the mapping to be generated by the neural networks. The learning rate is the same as before, the first 400 learning epochs are included in Fig. 10 while Fig. 11 illustrates the relationship $y = NN(x)$. Fig. 12 illustrates the nonlinear

Table 3
A list of associations of elements in $[0, 1]$ and $[a, b]$

| x | y |
|-------|------|
| 0.126 | 18.1 |
| 0.32 | 16.3 |
| 0.37 | 14.1 |
| 0.45 | 13.6 |
| 0.51 | 12.6 |
| 0.58 | 9.8 |
| 0.67 | 8.2 |
| 0.84 | 7.9 |
| 0.86 | 5.6 |
| 0.93 | 4.7 |
| 0.99 | 4.3 |
| 1.0 | 2 |

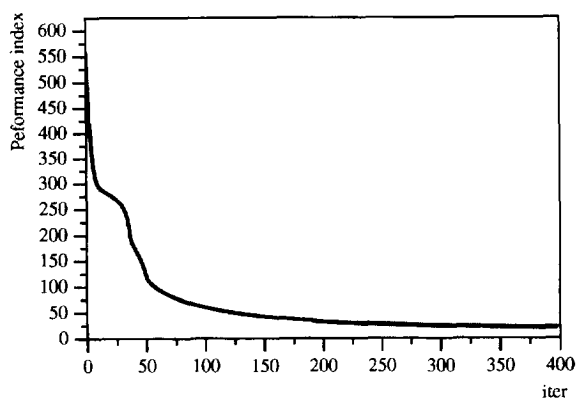
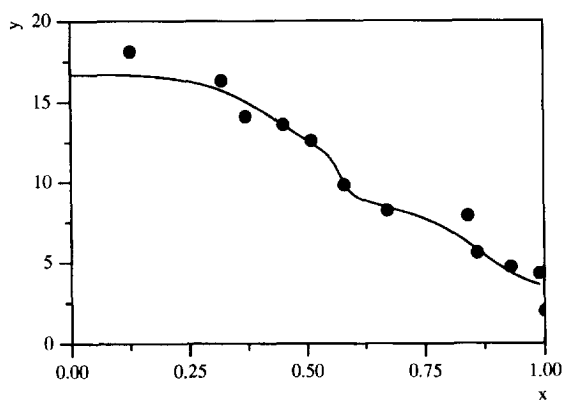
Fig. 10. Performance index Q in successive learning epochs.

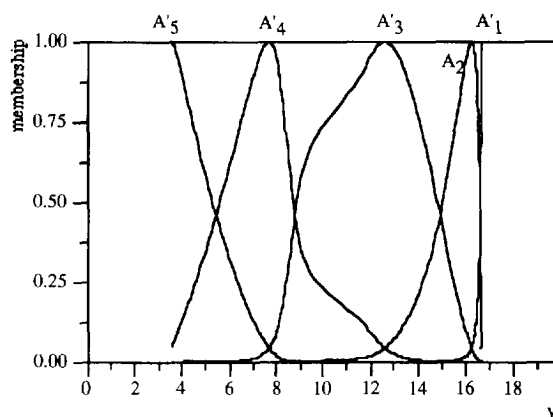
Fig. 11. Nonlinear mapping produced by the neural network.

adaptation effect by showing how much the original Gaussian membership functions have been affected by some local deformation of the original universe of discourse.

6. Conclusions

The primary intent of this note was to pose the problem of nonlinear context adaptation and come up with its relevant optimization framework.

The importance of the addressed idea is evident at the conceptual level. By raising the problem of

Fig. 12. Frame of cognition \mathcal{A}' .

calibrating fuzzy sets we have tackled the issue of attaching meaning of generic linguistic terms whose semantics should be sustained across a broad range of cases.

The direct application aspects have not been discussed. They could be quite easily envisioned in adaptation of rule-based control schemes (in fact, the one example along this line has been proposed in group elevator traffic control [2]) and fuzzy modeling. These will be thoroughly analyzed in a separate study.

References

- [1] M. Braae and D.A. Rutherford, Selection of parameters for a fuzzy logic controller, *Fuzzy Sets and Systems* 2 (1979) 185–199.
- [2] R.R. Gudwin and F.A.C. Gomide, Context adaptation in fuzzy processing, in: *Proc. 1st Brazil–Japan Joint Symp. in Fuzzy Systems*, Campinas, SP, Brazil (July 1994) 15–20.
- [3] W. Pedrycz, *Fuzzy Control and Fuzzy Systems* (Research Studies Press, Wiley, Taunton/New York, 1989).
- [4] W. Pedrycz, *Fuzzy Sets Engineering* (CRC Press, Boca Raton, FL, 1995).
- [5] J.C.T. van der Veen, Fuzzy sets, theoretical reflections, application to ship steering, M.Sc., Dept. of Electrical Eng., Delft University of Technology, Delft, the Netherlands, 1976.