Paper:

Optimal Size Fuzzy Models

Tamás D. Gedeon*, **László T. Kóczy****, **and Alessandro Zorat*****

*Dept. of Computer Science, Australian National University
Canberra 0200, Australia
E-mail: tom@cs.anu.edu.au
**Dept. of Telecommunications and Media Informatics, Budapest University of Technology and Economics
Sztocek u. 2, H-1111 Budapest, Hungary
E-mail: koczy@tmit.bme.hu
***Dept. of Information and Communication Technologies, University of Trento
Via Sommarive 14, Povo, I-38100 Italy
E-mail: zorat@unitn.it

[Received April 20, 2006; accepted September 6, 2006]

Approximate models using fuzzy rule bases can be made more precise by suitably increasing the size of the rule base and decreasing uncertainty in the rules. A large rule base, however, requires more time for its evaluation and hence the problem arises of determining the size that is good enough for the task at hand, but allows as fast as possible reasoning using the rule base. This trade-off between computation time and precision is significant whenever a prediction is made which can become "out of date" or "too old" if not used in time. The trade off is considered here in the context of tracking a moving target. In this problem, a higher degree of accuracy results in tighter precision of determining target location, but at the cost of longer computation time, during which the target can move further away, thus ultimately requiring a longer search for exact target localisation. This paper examines the problem of determining the optimal rule base size that will yield a minimum total time required to repeatedly re-acquire the moving target, as done by a cat that plays with a mouse. While this problem has no known solution in its general formulation, solutions are shown here for specific contexts.

Keywords: optimisation, model size, fuzzy logic

1. Introduction

Fuzzy control based on If..then rule models has become very successful in many fields of application [3, 6, 8, 9]. Rule based control is usually fast and transparent, even when analytic knowledge about a model of a system to be controlled is not available or is known only partially.

Fuzzy rule based control algorithms are suitable for function approximation. Where the optimal transfer function for a particular control problem is known, a fuzzy controller will provide an approximation of the system. It will generate a transfer function that has a lower absolute difference from the ideal (e.g. in the sup-norm sense) the more (and more precise) rules. There are various opinions in the literature stating that fuzzy controllers are universal approximators with arbitrary fineness, or that low computational complexity can be obtained by a trade-off with rough approximation only (cf. [2]). Obviously, if the size of the model is "too big", the run-time will increase considerably as the evaluation of this "too big" model takes too much time. However, if the size is "too small", the knowledge will be too imprecise and the conclusion will not be exact enough. This trade-off between computation time and precision is significant whenever a prediction is made which can become "out of date" or "too old" if not used in time.

This optimisation problem can be illustrated with a target tracking problem introduced in [5], called the "Cat and Mouse". In this problem the inaccuracy of the conclusion maps into the search time that is assumed to be proportional with the inaccuracy itself, while model complexity obviously maps to calculation time. In this way the total "cost" is expressed in terms of the total problem solution time, so that this sum can be calculated and optimised. Several initial results showing how to optimize the number of rules in the base in some simplified cases will be presented in this paper.

2. The "Cat and Mouse" Problem

2.1. Problem Introduction

Let us suppose that a target point (the "Mouse") is moving in an n-dimensional space ($n \geq 1$). The model of the mouse's behaviour is analytically not known *a priori*, or is known only partly. However, some knowledge about the mouse's movements is available: e.g., *behavioral* information might be given, or *bounds* on some quantities might be known. For instance, for *velocity*, the mouse cannot move with arbitrary velocity in any direction; for *acceleration*, the mouse cannot change its velocity arbitrarily, etc. This knowledge might be contained in a fuzzy rule based model.

Let us suppose that there is a fuzzy rule based model available, based on present observations. Based on this model, a prediction is calculated in the form of a fuzzy conclusion set, which represents the possible position of the mouse at the next period. The model and the prediction algorithm form the inference part of the "Cat" that attempts to locate the mouse exactly in the given time. Clearly, the time necessary for calculating the prediction depends on the model: as the more rules there are, the longer the time needed to obtain the prediction.

2.2. Uncertainty Criteria

There are two ways of interpreting uncertainty in the conclusion here: One is the size of the total area where the mouse might be in the next period, i.e., the size of the support of the normalised conclusion, which is the same as the support of the original conclusion:

$$\begin{aligned} U_{wc} &= \left| supp\left(\bigcup_{i=1}^{r} B_i^*(y)\right)\right| \\ &= \left| supp\left(\max_{i=1}^{r} \min\{w_i, B_i(y)\}\right)\right| \end{aligned}$$

where r is the number of rules in the base, w_i are the degrees of matching between the observations $A^*(x)$ and the rule antecedents $A_i(x)$, and $B_i(y)$ are the corresponding rule consequents. The latter part of the formula refers to the Mamdani-inference. If the prediction refers to the position directly, $Y = \{y\} = X = \{x\}$. This interpretation corresponds exactly to the usual concept of the worst case.

The other way of interpreting the uncertainty in the conclusion is to consider the possibility distribution $B^*(y)$ as proportional to the subjective probabilities of the mouse being at a certain location. In this case it is necessary that the search be carried out according to decreasing membership values in the conclusion (thus, it must be started at one of the maximal points). In this case, search time will be considered proportional on the average with the average size of the area that has to be searched until the mouse is found. Let us suppose there are n discrete units Δy within $supp(B^*(y))$ that have to be checked. Each unit is represented by a certain y_i, and so, the subjective probability of finding the mouse in Δy is

$$p_i^* = \frac{B^*(y_i)}{\sum_{j=1}^{n} B^*(y_j)}.$$

Let us order these subjective probabilities so that $p_1^* \geq p_2^* \geq \cdots \geq p_n^*$. Then the best search strategy is that we first look for the mouse in y_1, then in y_2, etc. Treating p_i^* as real probabilities (which they are not, as the search according to a concrete prediction cannot be repeated, consequently it would be senseless to consider p_i^* as the limit of any relative frequencies), the expected number of "looks" is estimated by:

$$E(i) = \sum_{i=1}^{n} i p_i^* = \frac{\sum_{i=1}^{n} iB^*(y_i)}{\sum_{j=1}^{n} B^*(y_j)}$$

where the duration of the "look" is proportional to the area that is represented by a single y_i (i.e. to the nth part of the total support size). Consequently the uncertainty is:

$$U_{ev} = \frac{|supp(B^*(y))|}{n} \frac{\sum_{i=1}^{n} iB^*(y_i)}{\sum_{i=1}^{n} B^*(y_i)} = \Delta y_i \frac{\sum_{i=1}^{n} iB^*(y_i)}{\sum_{i=1}^{n} B^*(y_i)}$$

if Δy_i denotes the ith part of the support size. This type of uncertainty corresponds to the usual concept of expected value.

It is reasonable to construct a continuous form of this expression. In order to do so, we slightly rearrange the above equation:

$$U_{ev} = |supp(B^*(y))| \frac{\sum_{i=1}^{n} \frac{i}{n} B^*(y_i)}{\sum_{i=1}^{n} B^*(y_i)}.$$

In this equation $\frac{i}{n}$ represents the relative extension of the area where the membership degree is at least $B^*(y_i)$ so we introduce the concept of the "relative α-cut function" $\alpha_{B^*}(y)$ of fuzzy membership functions $B^*(y)$ with the bounded support $\alpha_{B^*} : [0,1] \to [0,1]$ and α represents the relative size of the α-cut of $B^*(y)$ compared to the total size of the support. Obviously α is monotone decreasing, as the α-cuts are nested into each other. Moreover, its limit for $\alpha = 0$ is always 1. The value $\alpha_{B^*}(|supp(B^*(y))|)$ expresses the ratio

$$\frac{|core(B^*(y))|}{|supp(B^*(y))|}.$$

(Where *core* denotes the core or kernel of the fuzzy set.) For crisp membership functions $C(y)$, $\alpha_C(y) = 1$ holds. For crisp singletons α is by the definition also identically 1. With the help of this new function, the limit form of U_{ev} is:

$$U_{ev} = U_{wc} \frac{\int y\alpha_{B^*}(y)dy}{\int B^*(y)dy} = \frac{\int y\alpha_{B^*}(y)dy}{\int \alpha_{B^*}(y)dy}$$

from which it is clear that for crisp conclusions (singleton or not), the two uncertainties are identical, and that in any other case $U_{ev} < U_{wc}$. It is difficult to decide which definition of U is more suitable for describing the uncertainty in the prediction. The difficulty lies mainly in the problem that the interpretation of possibilities as subjective probabilities has weak justification, unless there is additional knowledge that describes the statistical behaviour of the rule base. Suppose that the conclusion is the following (verbally expressed): "The mouse is possibly in cell 1, but it might with a low possibility also be in cells 2, 3 or 4". It is quite obvious here to check cell 1 first, but we cannot be sure that the mouse will be there. Maybe it will be found only in the last cell to be checked. It is little help for us that probably, if the same conclusion could be obtained many times, and if it could be guaranteed that those cases were independent from each other, our experience would prove that in most cases the mouse would be found in cell 1. The reason is that there is no possibility for a repeat of the experiment, and the probability is high that it will never occur in the future. On the other hand, it can also not be assumed that the individual conclusions will behave as if they were manifestations of a single event space, namely, that the distribution of the actual positions of the mouse relative to the subjective probability distributions represented by the conclusions should be e.g. uniform, as there is no justification of assuming any statistical hypotheses on the behaviour of the rule base – at least not without additional statistical knowledge about them. Thus, if there is no additional information available concerning the distributions, the calculation will be treated as if we had interval valued conclusions only, without the finer knowledge of the membership functions $B_i(y)$. (However, this information is neglected in the estimation of uncertainty, can be used in the actual search or control task, e.g. when calculating the defuzzified conclusion.)

The problem of not having a single event space is also similar to when somebody tries to win by playing roulette every day in a different casino. He observes the frequencies of various outcomes for a while and then he has estimates for the most probable result after the next turn and so he puts his stake on the expected winning number. When he fails, he goes on the next day to another casino and starts the whole procedure again. Experience tells us that some people play for very long times and they never win (on the long term). The reason is that the individual play occasions are not manifestations of the same event space.

In most cases when a rule base is constructed, there is no additional information available on how the individual consequent functions are related to the statistics of the real behaviour. If the cat is an "expert", it will anticipate the next steps of the mouse. In some regions of the state space it will have good expertise (e.g. when the mouse is near to a piece of cheese, where it can be assumed with a high probability that it will chose a trajectory that brings it closer to the food), but in some other regions, the cat might have very little knowledge, and its guess (the rule consequents in the neighbourhood) might be quite imprecise. As a real example, Burkhardt and Bonissone [1] apply a rule tuning method in order to obtain a better rule based controller from a rather rough, equidistantly distributed model. They obtain a good controller (with fine rules, with narrow supports) near the set point, and they keep the original very rough model in the distant regions where during the tuning there was no new information. Probably, some of the rules will give better results (those which are based on more experience), and some will not (those which are very subjectively estimated), and there is no guarantee that the kernels or centres of gravities of the functions will always provide the expected values of the "best" output.

Because of these considerations, in this paper, we will restrict investigations to the worst case uncertainty measure. However, it must be stressed that further research on the optimisation with the more sophisticated U_{ev} criterion might produce some interesting and practically important results. This could be the case where the rules are generated by evaluating statistical data connecting the inputs and outputs of the controller to be assembled, such as by Sugeno and Yasukawa [7].

2.3. Problem Delineation

The mouse has to be located, i.e. a conclusion drawn, and its uncertainty determined. The total tracking time τ between two successive exact localisations of the mouse comprises two components: $\tau = T_i + T_s$, where T_i is the inference time and T_s is the action (search) time (equivalent to the uncertainty in the original context).

Let us call a model finer if there are more rules in it, and those rules have smaller supports. Even though these two features are not formally connected, we will always assume their connection, as there is really no reason to increase the number of rules in a model if it does not help with decreasing the sizes of consequents. As the model is made finer, the size of the rule base will increase and consequently T_i will increase, while T_s will decrease. Thus, one would like to determine a degree of fineness / roughness of the model that minimises $\tau = T_i + T_s$.

Finding the optimal roughness with the aim of reducing computational (e.g. time) complexity illustrates that a theoretical universal approximator becomes completely irrelevant as it might be reasonable to reduce the accuracy of approximation in order to have more effective (fast and successful) action.

The "roughness" or "fineness" of the rule base is first of all expressed by the number of rules. It is although true that an arbitrary rule base might be locally fine but rough elsewhere, but in the next sections we will assume that the rule antecedents are distributed uniformly. If this is not the case in real control systems, the number of rules calculated can be considered as locally interpolated number of rules, and the optimisation can be done iteratively for various neighbourhoods.

With the above simplifying assumptions (worst case search, uniform distribution of rules), an optimisation task can be formulated. It is not obvious whether there is any meaningful solution to the problem. So, in the next section we will examine whether a nontrivial optimum exists for the number of rules in the base if the costs of inference and search are set. As optimality will be considered in the worst case, τ has to be calculated for the case where the cat finds the mouse in the last examined corner of the predicted territory. The position is unambiguous in that the n-dimensional space is discretised into small hypercubes (with edges of ε) within which the positions can not be distinguished. In order to avoid any additional difficulty caused by the mouse's movement while the cat is searching, it will be assumed that the cat takes a snapshot every time when it is ready with its prediction and searches on the snapshot.

3. Models of Position Prediction

It is obvious that many different models can be assumed for the cat. If the cat has tractable analytic and deterministic knowledge about the movements of the mouse, and is able to observe the exact parameters of its target, there is no need for any fuzzy model at all. Fuzzy modelling and control gain importance when the model is too complex and intractable for practical purposes, or when

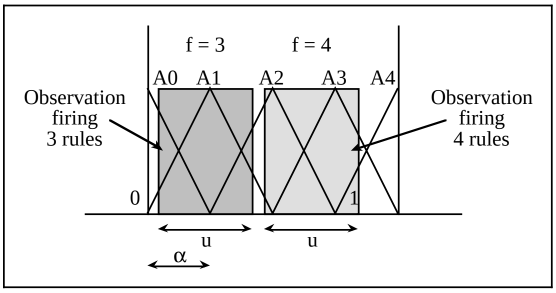

Fig. 1. Rules fired for two fuzzy observations.

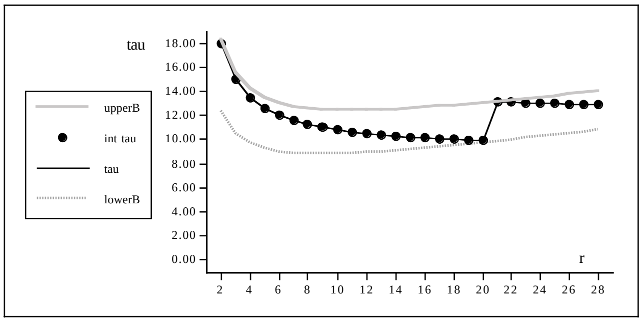

Fig. 2. Comparison between lower bound, upper bound and τ.

the knowledge is uncertain and the behaviour of the target appears to be non-deterministic from the point of view of the tracker, and finally, when it is not possible to observe the exact position. In realistic situations all these difficulties might occur simultaneously (including in the original control application). Because of these, we assume that no kinematic model of the mouse can be formulated.

Let us assume that we have a rule based mouse model, which uses one or more observable parameters of the mouse. In the following sections we investigate the following models:

- Model 2: one input, one output, fuzzy observation (1/1/f)

to introduce our terminology;

- Model 4: k inputs, one output, fuzzy observations (k/1/f)

to extend our treatment to deal with multiple inputs;

- Model 6: one input, n outputs, fuzzy observation (1/n/f)

to extend our treatment to multiple outputs instead. We omit discussion of the intermediate Models 1,3,5,7 which use crisp observations, respectively 1/1/c, k/1/c, 1/n/c, and k/n/c.

This now leaves the reader with the tools to derive the last model:

- Model 8: k inputs, n outputs, fuzzy observations (k/n/f).

Clearly Model 8 is the general model of which all of the presented models are subsets. Instead of providing the lengthy derivation, for clarity of exposition we chose to start with the very simplest models to introduce our terminology, and then discuss the different increasingly complex models which then require less lengthy exposition. This we believe will optimize the length of exposition on the fuzzy models we describe and minimise the reader's time in understanding the concepts and derivations in our paper.

3.1. Model 2: One Input, One Output, Fuzzy Observation

In a realistic situation, the Cat is often unable to identify the *exact* value of the observed parameter (e.g. position) and hence the observation is in the form of a fuzzy set, that is, a *fuzzy number*. This has the effect of increasing the uncertainty in the reasoning procedure, as it increases the number of rules fired. As the uncertainty in the fuzzy conclusion (total united support of the fired consequents) becomes larger, this in turn will increase the search time as well.

Let the support of fuzzy observation A^* be $|supp(A^*)| = u$. The value of u is a fraction indicating the *uncertainty* of the observation. The number of rules fired f is $2 + \lfloor u/d \rfloor \le f \le \lceil u/d \rceil$, where $d = 1/(r-1)$ is the distance measured for example between the two peaks of two neighbouring antecedents, that are assumed to be equidistantly located. This is true almost everywhere, except for the special case when there is coincidence of the endpoints of the observation and the antecedents' membership functions. In the example of **Fig. 2**, $r = 5$, $d = \frac{1}{r-1} = \frac{1}{4} = 0.25$, $u = 0.30$ and hence $u/d = 0.30/0.25 = 6/5$, $\lfloor u/d \rfloor = 1$ and $\lceil u/d \rceil = 2$. As shown in the figure, while both observations have uncertainty $u = 0.30$, the left observation fires 3 rules, while the one on the right fires 4 rules.

In the worst case

$$T_i(r) = c_0 r + c_1 f = c_0 r + c_1 \left(2 + \left\lceil \frac{u}{d} \right\rceil \right),$$

$$T_s(r) = c_2 f \beta = c_2 \frac{2f}{r} = 2c_2 \frac{2 + \lceil \frac{u}{d} \rceil}{r}.$$

From which:

$$\tau(r) = c_0 r + c_1 \left(2 + \left\lceil \frac{u}{d} \right\rceil \right) + 2c_2 \frac{2 + \lceil \frac{u}{d} \rceil}{r}.$$

Substitution of $1/(r-1)$ for d yields:

$$\tau(r) = c_0 r + c_1 (2 + \lceil u(r-1) \rceil) + 2c_2 \frac{2 + \lceil u(r-1) \rceil}{r}.$$

Lower ($\check{\tau}$) and upper ($\hat{\tau}$) bounds on τ can be easily obtained as:

$$\tau(r) \ge \check{\tau}(r) = c_0 r + c_1 (2 + u(r-1)) + 2c_2 \frac{2 + u(r-1)}{r}$$

and

$$\tau(r) \le \hat{\tau}(r) = c_0 r + c_1 (3 + (r-1)) + 2c_2 \frac{3 + u(r-1)}{r}.$$

Both $\check{\tau}(r)$ and $\hat{\tau}(r)$ have a minimum that can be obtained by differentiation:

$$\frac{d\check{\tau}(r)}{dr} = c_0 + c_1 u - 2c_2 \frac{(2-u)}{r^2}$$

and

$$\frac{d\hat{\tau}(r)}{dr} = c_0 + c_1 u - 2c_2 \frac{(3-u)}{r^2}.$$

Both have a nontrivial minimum (at a somewhat different value of r):

$$\check{r} = \sqrt{\frac{2c_2(2-u)}{c_0 + c_1 u}} \quad \text{and} \quad \hat{r} = \sqrt{\frac{2c_2(3-u)}{c_0 + c_1 u}}.$$

While neither $\check{r}$ nor $\hat{r}$ is necessarily identical to the real minimum, they provide a relatively good estimate as the two functions $\check{\tau}$ and $\hat{\tau}$ differ by at most $c_1 + c_2$ (in height) and τ itself is *between* these two.

Let ρ be the desired *integer* value for which $\tau(\overline{r})$ is minimal. The value of $\tau(\check{r})$ or $\tau(\hat{r})$ do not differ from the minimal value of $\tau(\rho)$ by more than $c_1 + c_2$, i.e. $\tau(r) \leq \tau_{min} + c_1 + c_2$ for both approximate minima.

As an example, consider the following case: $u = 0.05$, $c_0 = 0.0$, $c_1 = c_2 = 3$. **Fig. 2** shows the graphs of $\check{\tau}(r)$ (lower grey line), and $\hat{\tau}(r)$ (upper grey line), and $\tau(r)$ (middle black line). The values of $\tau(r)$ for *integer* values of r are shown as solid bullets in the graph.

For this case, $\check{r} = 8.83$ and $\hat{r} = 10.86$. However, function $\tau(r)$ of this example is fairly flat and hence it keeps on decreasing for $r = 10, 11, 12, \ldots$ – while both $\check{\tau}(r)$ and $\hat{\tau}(r)$ have already started to increase again.

Actually, $\tau(\rho) = 21$, which seems quite “far away” from the computed values of $\check{r}$ and $\hat{r}$. Thus, at first it might appear that the values of $\check{r}$ and $\hat{r}$ are not very useful, as what is needed is really the value of ρ and that can be quite distant from $\check{r}$ and $\hat{r}$. However, this is not so, as shown next.

Let r_j be a value for which $u(r_j - 1)$ is an integer and hence $\lceil u(r_j - 1) \rceil = u(r_j - 1)$. Notice that $u(r_j - 1) + 1 = u(r_j - 1 + 1/u) = u(r_{j+1} - 1)$ is also an integer and then:

$$\begin{aligned}
\tau(r_j) &= c_0 r + c_1(2 + \lceil u(r_j - 1) \rceil) + 2c_2 \frac{2 + \lceil u(r_j - 1) \rceil}{r} \\
&= c_0 r + c_1(2 + u(r_j - 1)) + 2c_2 \frac{2 + u(r_j - 1)}{r} \\
&= \check{\tau}(r_j)
\end{aligned}$$

and

$$\begin{aligned}
\tau(r_{j+1}) &= c_0 r + c_1(2 + \lceil u(r_{j+1} - 1) \rceil) \\
&\quad + 2c_2 \frac{2 + \lceil u(r_{j+1} - 1) \rceil}{r} \\
&= c_0 r + c_1(2 + u(r_{j+1} - 1)) \\
&\quad + 2c_2 \frac{2 + u(r_{j+1} - 1)}{r} \\
&= \check{\tau}(r_{j+1}).
\end{aligned}$$

That is, at $r_j, r_{j+1}, r_{j+2}, \ldots$, the functions τ and $\check{\tau}$ coincide. Notice that $r_{j+1} - r_j = 1/u$ and recall that $u \leq 1$ (so $1/u \geq 1$). Thus, if one were to start from as initial r and move to the right (left) by $1/u$, it will *certainly* find a position where τ and $\check{\tau}$ coincide.

From here the derivatives are

$$\begin{aligned}
\frac{d\check{\tau}}{dr} &= c_0 + c_1 u + 2c_2 n(u-2) \frac{(2 + u(r-1))^{n-1}}{r^{n+1}} \\
\frac{d\hat{\tau}}{dr} &= c_0 + c_1 u + 2c_2 n(u-3) \frac{(3 + u(r-1))^{n-1}}{r^{n+1}}.
\end{aligned}$$

3.2. Model 4: *k* Inputs, One Output, Fuzzy Observation

The model and rule type are the same as in the above model, and the support of the observation is $|supp(A^*)| = a^k$. (Similarly to the assumption concerning t, we consider uniform uncertainty in every input variable.) Accordingly, the number of fired rules in one dimension is

$$f_1 = \left(2 + \left\lceil \frac{a}{\alpha} \right\rceil\right), \text{ and } f = f_1^k.$$

As before, α is the distance of two neighbouring antecedents in each dimension. The value of α depends on t: $\alpha = \frac{|X|}{i-1}$ (assuming that for every dimension i, $|X_i|$ is constant). As in Model 2:

$$2 + \frac{a}{\alpha} \leq f \leq 3 + \frac{a}{\alpha}.$$

Using similar formulas as above we obtain

$$T_i = c_0 k t + c_1 \left(2 + \left\lceil \frac{a}{\alpha} \right\rceil\right)^k$$

while

$$T_s = 2c_2 \frac{\left(2 + \left\lceil \frac{a}{\alpha} \right\rceil\right)^k}{r} = 2c_2 \left(\frac{2 + \left\lceil \frac{a}{\alpha} \right\rceil}{t}\right)^k$$

from which substitution of $1/(r-1)$ for α yields

$$\tau = c_0 k t + c_1(2 + \lceil a(t-1) \rceil)^k + 2c_2 \left(\frac{2 + \lceil a(t-1) \rceil}{t-1}\right)^k.$$

Because of the existence of the upper and lower bound functions for f the following two continuously differentiable bounds can be considered for τ:

$$\begin{aligned}
\check{\tau} &= c_0 k t + c_1[2 + a(t-1)]^k + 2c_2 \left(\frac{2 + a(t-1)}{t-1}\right)^k \leq \tau \\
&\leq \hat{\tau} = c_0 k t + c_1[3 + a(t-1)]^k + 2c_2 \left(\frac{3 + a(t-1)}{t-1}\right)^k.
\end{aligned}$$

The first derivatives are

$$\begin{aligned}
\frac{d\check{\tau}}{dt} &= c_0 k + c_1 k a[2 + a(t-1)]^k \\
&\quad - 4c_2 k \left(\frac{2}{t-1} + a\right)^{k-1} \frac{1}{(t-1)^2}
\end{aligned}$$

and

$$\begin{aligned}
\frac{d\hat{\tau}}{dt} &= c_0 k + c_1 k a[3 + a(t-1)]^k \\
&\quad - 6c_2 k \left(\frac{3}{t-1} + a\right)^{k-1} \frac{1}{(t-1)^2}.
\end{aligned}$$

Unfortunately, it is not possible to give the explicit solutions of $\frac{d\check{\tau}}{dt} = 0$, and $\frac{d\hat{\tau}}{dt} = 0$. We show, however, that such solutions necessarily do exist.

The "increasing parts" of the functions ($g_{i1} = c_0k + c_1ka[2+a(t-1)]^{k-1}$ and $g_{i2} = c_0k + c_1ka[3+a(t-1)]^{k-1}$ respectively) assume $g_{i1}(2) = c_0k + c_1ka(2+a)^{k-1}$ and $g_{i2}(2) = c_0k + c_1ka(3+a)^{k-1}$ at $t = 2$ which is the minimal value for t, while the "decreasing parts" ($g_{d1} = 4c_2k\left(\frac{2}{t-1}+a\right)^{k-1}\frac{1}{(t-1)^2}$ and $g_{d2} = 6c_2k\left(\frac{3}{t-1}+a\right)^{k-1}\frac{1}{(t-1)^2}$) assume at the same locations $g_{d1}(2) = 4c_2k(2+a)^{k-1}$ and $g_{d2}(2) = 6c_2k(3+a)^{k-1}$.

The expression $g_{i1} - g_{i2} < 0$ iff $c_0 < (4c_2 - c_1a)(2+a)^{k-1}$ is satisfied if c_0 is small compared to c_2. Otherwise the function has no zero to the right of t, so the minimum is the extremal $t = 2$. If this difference is negative in $t = 2$, there must be a zero on the right of 2 as the increasing part goes to infinity, and the decreasing part to zero when t goes to infinity, i.e., the difference itself goes to infinity. So in every case there is a minimum for $\check{\tau}$, and in a similar manner for $\hat{\tau}$ (even though the two may not be identical). The two values give a rough estimate for the minimum of τ itself.

The exact minimum can be found by applying an algorithm of repeated check, as τ consists of several sections of the "reference functions"

$$\tau_r = c_0kt + c_1const^k + 2c_2\left(\frac{const}{t}\right)^k$$

of which the minimum can be found by solving

$$\frac{d\tau}{dt} = c_0k + c_1k\,const^{k-1} - 2c_2k\frac{const^k}{t^{k+1}} = 0$$

which is easily done. This is the first candidate for the exact minimum, and next the left breakpoints must be examined in a similar manner to Model 2. Unfortunately, here there is also a chance that some further minima on the right must be checked.

Example: Let us use the same parameters as before: $c_0 = 1$ ms, $c_1 = 5$ ms, $c_2 = 5$ s, $k = 2$, and $const = 3$. The solution of the equation

$$2\times10^{-3} + 5\times10^{-3}\times2\times3 - \frac{2\times5\times3^2}{t^3} = 0$$

gives the number of terms per dimension. This produces $t = 18$, and $r = 324$. As $a = 10^{-3}$, $\lceil(t-1)a\rceil = \lceil17\times10^{-3}\rceil = 1$, this is the exact minimum.

3.3. Model 6: One Input, *n* Outputs, Fuzzy Observation

The expression of τ is

$$\begin{aligned}\tau &= c_0r + c_1f + c_2f\beta^n \\ &= c_0r + c_1\lceil2+a(r-1)\rceil + c_2\lceil2+a(r-1)\rceil\left(\frac{2}{r}\right)^n\end{aligned}$$

which can be similarly bounded from both above and below as in the previous models. The approximate optimum is calculated then by solving

$$\frac{d\check{\tau}}{dt} = c_0 + ac_1 + 2^nc_2\frac{(n(a-2) - ar(n-1))}{r^{n+1}}$$

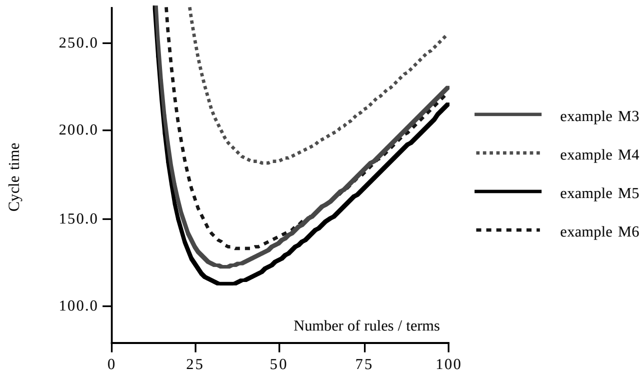

Fig. 3. Comparison of Model 4 (2/1/f) and Model 6 (1/2/1) examples.

and its pair for $\frac{d\hat{\tau}}{dt}$, which is a rather similar problem to the previous ones.

Figure 3 shows the comparison of the major models discussed, using the examples from the respective sections. Note that as the number of rules becomes very large, for all models the c_0r term dominates. We are, however, mostly interested in the minimum of the curve. Comparing models 3 and 4, we notice that the fuzzy output model has its minimum at a higher cycle time and also higher number of rules. The same can be noticed in comparing models 5 and 6. It is interesting to notice that the difference between the crisp and fuzzy versions appears to be greater in model 3&4 than in model 5&6, that is, the difference appears greater when the fuzziness is at the input, rather than the output.

The models where both the inputs and outputs are multidimensional can be treated in a similar manner, on the analogy of Models 3 and 5, and Models 4 and 6, respectively.

Finally we would like to point out the fact that once a model is given, it is possible to reduce it to a "rougher" one using the interpolation model as proposed in [4].

4. Conclusions

The use of linguistic variables and terms for values is an attempt to mimic human ways of solving control problems using experience expressed as rules. By analyzing the clear mathematical formulas behind these intuitive methods, some further conclusions for the applicability of fuzzy control can be obtained. The analysis of real controllers is especially useful in showing their limitations, that are in distinct contradiction with the rather optimistic conclusions coming from results on idealistic fuzzy controllers that have some nice analytic properties.

We have shown that it is possible to investigate the optimal size of the model if a certain – rather practical – optimality criterion is set, by assuming cost factors for inference time and uncertainty in the fuzzy conclusion. In summary, in many concrete models there exists an optimal trade-off between accuracy and complexity of the model. Even though the cases shown here do not represent the whole range of practically important types of controllers, they give concrete answer for several important simple and fundamental cases. Further research should follow with the extension of the class of controllers examined, and also with the use of the expected value uncertainty which may add to the full understanding of the problem.

Acknowledgements

The work of T. D. G. was supported by the Australian Research Council. Part of work of L. T. K. was done during a visiting appointment at the Department of Computer and Management Science, University of Trento, Italy, and supported by the (Hungarian) National Science Research Foundation projects T19671 and T16480. The work of A. Z. was supported by Special Project "Fuzzy Logic: Applicazioni ai modelli multiattributo di supporto alle decisioni ed al controllo di sistemi" of the University of Trento.

References:

[1] D. G. Burkhardt and P. P. Bonissone, "Automated fuzzy knowledge base generation and tuning," IEEE Int. Conf. on Fuzzy Systems, San Diego, pp. 179-196, 1992.

[2] J. L. Castro, "Fuzzy logic controllers are universal approximators," IEEE Tr. on SMC 25, pp. 629-635, 1995.

[3] D. Drinkov, H. Hellendoorn, and M. Reinfrank, "An Introduction to Fuzzy Control," Springer, Berlin, etc., 1993.

[4] L. T. Kóczy, "Compression of fuzzy rule bases by interpolation," Proceedings of the First Asian Fuzzy Systems Symposium, Singapore, pp. 500-507, 1993.

[5] L. T. Kóczy, A. Zorat, and T. D. Gedeon, "The Cat and Mouse Problem – optimising the size of fuzzy rule bases," Proceedings of CIFT '95, Trento, pp. 139-151, 1995.

[6] W. Pedrycz, "Fuzzy Control and Fuzzy Systems," 2nd edition, J. Wiley, Chichester, 1992.

[7] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modelling," IEEE Tr. on Fuzzy Systems 1, pp. 7-31, 1993.

[8] L. X. Wang, "Adaptive Fuzzy Systems and Control," Prentice Hall, Englewood Cliffs, N.J., 1994.

[9] R. Yager and D. Filev, "Essentials of Fuzzy Modelling and Control," J. Wiley, Chichester, 2004.

Name:
Tamás (Tom) D. Gedeon

Affiliation:
Chair Professor of Computer Science, Australian National University

Address:
Dept of Comp Sci, Bld 108, Acton ACT 0200 Australia

Brief Biographical History:
1991- UNSW, Lecturer, Snr Lect, Assoc Prof, Head Dept of Info Eng
1999- Murdoch Uni, Head of School of IT
2003- ANU, Head of Dept, later Dean/Assoc Dir Educ

Main Works:
• Please see http://cs.anu.edu.au/˜Tom.Gedeon

Membership in Academic Societies:
• Australian Computer Society
• Asia Pacific Neural Network Assembly (Board member)
• IEEE

Name:
László T. Kóczy

Affiliation:
Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics

Address:
Magyar Tudósok Krt. 2, Budapest, H-1117, Hungary

Brief Biographical History:
1975, 1976, 1977 M.Sc., M. Phil., Ph.D. in Electrical Engineering from Technical University of Budapest
1998- Doctor of the Hungarian Academy of Science (the highest earnable postdoctoral degree in Hungary)
Professor (Telecommunication and Information Technology) at Budapest University of Technology and Economics, and Professor and Dean at the Széchenyi István University, Győr

Main Works:
• Y. Yam and L. T. Kóczy, "Representing membership functions as points in high-dimensional spaces for fuzzy interpolation and extrapolation," IEEE Tr. on Fuzzy Systems, Vol.8, No.6, pp. 761-772, 2000.
• L. T. Kóczy, "Fuzzy if...then rule models and their transformation into one another," IEEE Tr. SMC A, Vol.26, No.5, pp. 621-637, 1996.
• L. T. Kóczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," International Journal of Approximate Reasoning, Vol.9, pp. 197-225, 1993.

Membership in Academic Societies:
• 2001-2003 President of the International Fuzzy Systems Association
• 1990- Hungarian Fuzzy Society Founding & Life Honorary President
• 2004- IEEE Computational Intelligence Society AdCom Member
• 2005- President of CIS Hungarian Chapter

Name:
Alessandro Zorat

Affiliation:
Department of Information and Communication Technology, University of Trento

Address:
Via Sommarive 14, I-38100 Povo (Trento), Italy

Brief Biographical History:
1973 Graduated in Computer Science, Univ. of Pisa, Italy
1975- M.S. in Computer Science, USC, California, USA
1979- Ph.D. in Computer Science, USC, California, USA
1979- Dept. of Computer Science, SUNY at Stony Brook, NY, USA
1986- Dept. of Computation and Management, Univ. of Trento, Italy
2001- Dept. of Information and Communication Technology, Univ. of Trento, Italy

Main Works:
• A. Boni and A. Zorat, "FPGA Implementation of Support Vector Machines with Pseudo-Logarithmic Number Representation," Proceeding of the IEEE 2006 International Joint Conference on Neural Networks [IJCNN 2006], Vancouver, Canada, July, 2006.
• A. Zorat, A. Sartori, G. Tecchiolli, and L. Koczy, "A Flexible VLSI Processor for Fast Neural Network and Fuzzy Control Implementation," Proceedings of the Conference of Soft Computing [Iizuka'96], Iizuka, Japan, October, 1996.

Membership in Academic Societies:
• IEEE