Enforcing Local Properties in Online Learning First Order TS-fuzzy Systems by Incremental Regularization

Nils Rosemann Werner Brockmann Bastian Neumann
University of Osnabrück
Albrechtstraße 28
D–49069 Osnabrück

Abstract— *Embedded systems deseminate more and more. Because their complexity increases and their design time has to be reduced, they have to be increasingly equipped with self-tuning properties. One form is self-adaption of the system behavior, which can potentially lead the system into safety critical states. In order to avoid this and to speed up the self-tuning process, we apply a specific form of regularization, incremental regularization. The SILKE approach has been developed as an incremental regularization scheme for a special class of online learning Takagi-Sugeno fuzzy systems. Its aim is to control the process of self-tuning by guiding the online learning process towards local meta-level characteristics such as a smooth system behavior without outliers. This ability has been investigated experimentally and formally for zero order systems before. This paper now analyzes the regularization ability of the SILKE approach to enforce local smoothness in first order TS-fuzzy systems in order to enlarge the methodological basis for more complex applications.*

Keywords— First order Takagi Sugeno fuzzy systems, incremental learning, self-optimization, incremental regularization, fuzzy control

1 Introduction

1.1 Background

Due to the growing dissemination of embedded technical systems and their higher integration into critical environments, the requirements for the next generation of those technical systems will change. The systems will have to adapt intelligently to different environments, different aims and changes within the environment in order to open up new application areas. This greatly increases the complexity of such new systems. At the same time, it will become necessary to decrease the design time. We tackle these problems within the context of Organic Computing, There one tries to face these challenges by introducing self-x properties into the systems, e.g., self-organization and self-optimization [1]. This means that technical systems are equipped with a much higher flexibility. They have to adapt their behavior based on changing demands and prior experience.

But although such a technical system adapts its behavior at run time, safety and trustworthiness are a must. This is especially problematic as the higher integration into natural, i.e., not specially prepared, environments comes with many sources of uncertainty. Thus, any form of adaption based on uncertain interaction with the environment will be uncertain itself. Accordingly, a self-tuning system has to be both flexible enough to react dynamically to new conditions, and stable enough to protect already learned knowledge from transient disturbances, noise and other uncertainties (*stability-plasticity-dilemma*). And, what is more, at the same time it has to be guaranteed that the system does not enter critical states by learning. This means that the self-tuning process has to be controlled, but without limiting its flexibility.

Within this paper, we address issues of self-tuning the control behavior of an embedded system, i.e., a form of self-adaptive control. This online changing of the behavior is done by *incremental online learning*. It has a strong influence on the stability of the system: In addition to the feedback loop between the control system and the environment into which it is embedded, online learning creates another, indirect feedback loop [2]. Whatever is learned in a certain situation changes the system's (future) behavior, and thus also what is learned in the future. Because of this, small deviations of (initial) conditions or the self-optimization process as such might have arbitrarily high impact and can potentially lead to a chaotic system behavior.

Different candidate function approximators have been used in adaptive control by online learning, each of them fullfilling the different requirements to a varying degree. An overview over the different families of function approximators is given in [3]. Accordingly, one has to choose an appropriate approximator to represent the control behavior in a way which suits the application specific needs, e.g., concerning expressiveness, complexity, interpretability and, what is more in our context, controllability of the learning process itself.

For adaptive nonlinear modeling and control, Takagi-Sugeno systems are a favoured trade-off [4, 5]. We hence build upon a specific variant. Then, in order to employ incremental online learning for adaptive control in uncertain environments, one has to address the stability-flexibility-dilemma. In the next section it is argued that this can be done by an appropriate form of regularization.

1.2 Related Work

In data-driven function approximation, regularization techniques have been successfully applied in many different areas, especially in nonlinear system identification [6] and nonlinear adaptive control [7]. Typical for these applications is that the distribution of data points is dense in one region of the input space and sparse in another. For practical reasons it might be impossible to obtain enough data in a sparse region, e.g., because it is related to critical states of a controlled process. Hence regularization techniques have found widespread use in data-driven fields. The reason is that in many cases data-driven problems are ill-posed because there is no unique solution [8]. This is usually due to a misbalance between the degrees of freedom of the function approximator and the amount of data available.

The idea of regularization is to incorporate additional in-

ISBN: 978-989-95079-6-8

formation about the solution into the problem statement, i.e., to reformulate it [9]. Usually this is some kind of structural or meta-information. A well known example is to assign a penalty to non-zero parameter values [10]. Another example is to assign a penalty to solutions with a high overall curvature [11]. They are called ridge regression or Tikhonov regularization. Despite missing data, sparse regions and large input spaces, regularization is able to improve the generalization performance of the identified models.

Basically, these approaches require all data points to be accessible at once in order to find a globally optimal solution, i.e., they are offline approaches. But for many online applications, there is a change of requirements while new data points are generated (continuously), for example because the system which is to be modeled changes or is modelled from scratch.

Recent approaches for these kind of applications have so far neglected the problem of requiring regularization [4]. In order to tackle the above challenges online, these approaches learn incrementally, i.e., by looking at every data point only once. But without access to the whole dataset, it is not *a priori* clear how to apply the conventional regularization techniques. Other approaches use regularization techniques for ill-posed matrix inversions [12] which is suitable for certain modeling applications.

Because of our background of adaptive nonlinear control by direct adaptation [2], we are developing schemes for *incremental regularization*, i.e., which perform special forms of regularization online, for appropriate function approximators. The control tasks which we address typically have only a low number of input dimensions (less than 10). In the case of more input dimensions, it is usually possible and advisable to decompose the overall problem into subproblems. Thus, the first one of the schemes has been developed for special Takagi-Sugeno fuzzy systems which are especially suited for adaptive nonlinear modeling and control [13, 5]. They employ an all-coverage approach for generating the rule base which allows for a high interpretability albeit limiting the maximal possible number of input dimensions. These fuzzy systems and the incremental regularization, called SILKE approach, will be reviewed in section 2. Up to now, we have developed incremental regularization schemes for zero order Takagi-Sugeno fuzzy systems [14]. They have been investigated experimentally and formally [15, 16]. The purpose of this paper is to extend the SILKE approach to first order Takagi-Sugeno fuzzy systems. In section 3 it is then shown that each incremental application of our approach to a first order Takagi-Sugeno system increases its smoothness. Because of this, the SILKE approach can be seen as a kind of incremental Tikhonov regularization which is also applicable to first order TS-fuzzy systems.

2 Controlled Self-optimization by the SILKE Approach

2.1 Basic Concept

The basis of the SILKE approach is a special class of Takagi-Sugeno fuzzy systems (hereinafter called sTS fuzzy systems) with polynomial functions as rule conclusions. They use linear B-spline input membership functions, such that each input dimension is split by a partition of unity. This allows for an intuitive handling by the system designer and an implementation which requires a very low computational effort. The rule inference is done by the *sum-prod* technique. These sTS-systems are ideally suited for application to online learning within embedded systems [3].

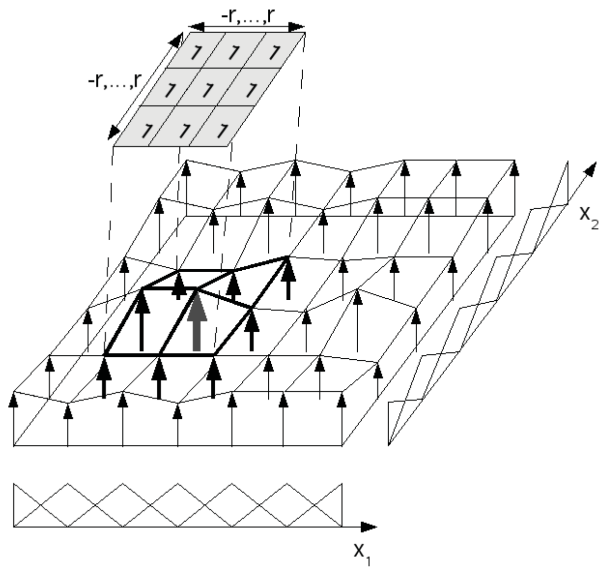

Figure 1: Schematic correspondence of a SILKE-template (shaded in grey) to the conclusions of a part of an sTS-fuzzy rule base in a two-dimensional system. The membership functions determine the lattice of rule conclusions. The thick arrow indicates the active rule under consideration. The numbers within the template are the coefficients of the averaging template mask [15].

The idea of the SILKE approach is then to enhance learning in sTS-systems by a correction step which introduces an additional feedback loop at a higher, functional level in order to rate the learning process itself online [17]. This is done by implementing so called *SILKE templates* which determine whether the rule base of the sTS-system complies with predefined meta-level properties. If this *monitoring* detects too strong deviations, then the rule base should be corrected, which closes the feedback loop at a higher level. It should be noted that such a template acts only on those rules which apply to the current system state and which form a local neighborhood (*active rules*), i.e., a template acts only locally and incrementally on each new data point.

Many different meta-level properties can be expressed, e.g., steadiness, monotonicity, local linearity or gradient ratios of the underlying function. One of the most important examples of meta-properties in the field of adaptive control is *smoothness* of the approximated function, i.e., the learned knowledge, because smoothness means that small changes of the inputs do not cause strong changes at the output. So given a process which is smooth, then one wants a rule base which is also smooth. If online learning changes the rule base in a way that it is smooth, then the process of learning harmonizes with its environment. Hence, robustness of learning is increased when faced with learning stimuli which would otherwise violate the smoothness property and smoothness is enforced, for example the effect of outliers is reduced.

In the next section, it is formally shown that this smoothness property can be enforced for first order sTS-systems by

ISBN: 978-989-95079-6-8

using the SILKE approach. For this purpose, the formalism of the SILKE approach is presented in its general form. Afterwards, a measure for local smoothness in zero and first order sTS-systems is developed. This measure is then used to investigate the effect of the SILKE approach in a simplified, one-dimensional scenario for clarity reasons.

2.2 Formalization

Let n be the number of inputs of a first order sTS-fuzzy system. By compilation of an sTS rule base, it becomes normalized, so that one can imagine it as a lattice [17]. This is illustrated by the schematic in Fig. 1. It shows a two-dimensional lattice of a rule base with two input variables, x_1 and x_2. Both inputs are partitioned by triangular membership functions. For every combination of two membership functions, one from each dimension, there is a single rule. Due to normalization, the rules thus form a lattice given by these membership functions. The membership functions can be numbered sequentially in each input dimension. As there is a single rule for each combination of membership functions, one can assign a vector of the individual membership function indixes to each rule. For example, in a two-dimensional rule base, the index vector $(2, 1)$ is assigned to the rule which has the second membership function of the first input dimension and the first membership function of the second input dimension as its antecedents.

The rule conclusions (which are affine functions) can be viewed as being placed upon the nodes of this lattice. It is useful here to identify the affine conclusions by vectors of their parameters, i.e., a function $g : \mathbb{R}^n \to \mathbb{R}$ with

$$g(\mathbf{x}) = c_0 + c_1x_1 + \ldots + c_nx_n, \quad \mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}, \tag{1}$$

can be identified by the vector of the parameters

$$\mathbf{c} = (c_0, \ldots, c_n)^T \in \mathbb{R}^{n+1}. \tag{2}$$

With this, the rule base of a first-order sTS fuzzy system can be seen as a map $o : \mathbb{N}^n \to \mathbb{R}^{n+1}$ from the space $\mathbb{N}^n$ of rule index vectors to the space $\mathbb{R}^{n+1}$ of parameter vectors. A single node $p \in \mathbb{N}^n$ of the lattice corresponds to a rule, and $o(p)$ to the parameters of an affine function, i.e., the rule conclusion.

A so called SILKE template T is than a map

$$T : \mathbb{N}^n \to \mathbb{R}^{n+1}. \tag{3}$$

The value $\|T(p)\|$ is called *violation degree* of the rule p with respect to T. It indicates how much a rule p violates the meta-level property which is given by T with respect to the neighboring rules. $\|T(p)\| = 0$ means a complete compliance with this property. The higher $\|T(p)\|$, the further rule p is violating the property.

The core of the SILKE approach is then to correct the given rule p according to

$$o(p) \leftarrow (1-\alpha)\cdot o(p) + \alpha\cdot(o(p) - T(p)). \tag{4}$$

Obviously, if $\|T(p)\| = 0$, the rule is not changed. The *adjustment rate* $\alpha \in [0; 1]$ determines how much the meta-level property is enforced, i.e., $\alpha = 0$ results in no correction of the rule base, whereas $\alpha = 1$ makes it fully compliant with the meta-level property in one step. As said before, such a correction step is done directly after each learning step. Every rule which has been changed by learning is subject to the SILKE template(s). This means that $\alpha = 1$ ignores all changes of the rules by learning. Instead, the meta-level property is fully enforced. Thus, high values of α limit the influence of learning severely but increase the stability.

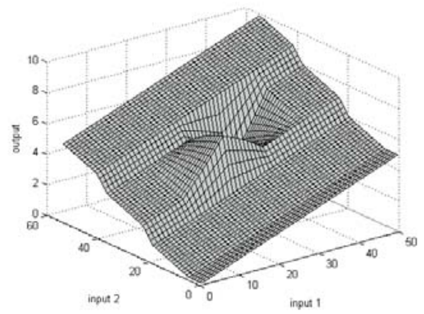

Figure 2: Original graph of a sample two-dimensional sTS-system which is smoothed by the SILKE approach, see Fig. 3 and Fig. 4.

In the following, only SILKE templates are considered which can be represented by a convolution of the rule base:

$$T(p) = o(p) - \frac{1}{(2r+1)^n}\sum_{u\in U} o(p-u)m(u). \tag{5}$$

Here, $r \in \mathbb{N}^+$ is a radius. All neighbor rules with a L_1-distance to the given rule p of less than or equal to r are considered by the template (For illustration, in Fig. 1 an exemplary template is shown with radius $r = 1$). So the sum indices u are in

$$U = [-r, r]^n. \tag{6}$$

The map $m : U \to \mathbb{R}$ is called a *mask* and determines the effect of the SILKE template, i. e., the represented meta-level property, by specific coefficients. Because of this, the mask is the most important part of the SILKE approach concerning design. The designer has to use the right coefficients in the mask in order to represent the desired meta-level properties. The mask thus gives the SILKE approach the necessary flexibility to express a wide range of local meta-level properties. The example in Fig. 1 shows the mask for an averaging template which is given by $m(u) = 1 \forall u \in U$. This mask can be used to guide zero order sTS systems towards a higher smoothness [15]. Another example is the one-dimensional averaging template which only use one of the input dimensions. It is given by $m(u) = 1 \forall u \in \{0\}^{a-1} \times [-r, r] \times \{0\}^{n-a}$ with a denoting the dimension to be used.

3 Smoothness in First Order sTS Fuzzy Systems

In order to extend the formal investigations of the SILKE approach to first order sTS fuzzy systems, they are investigated concerning their smoothness properties. In this section, based on the definition of a measure of smoothness for the individual rules, it is proven that the application of an averaging SILKE

ISBN: 978-989-95079-6-8

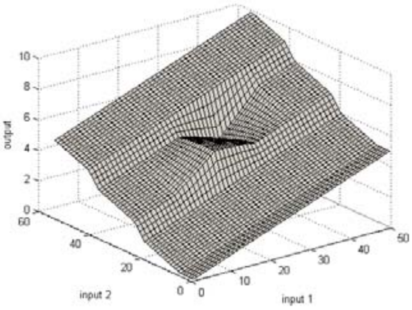

Figure 3: Graph of the sample two-dimensional sTS-system from Fig. 2 after the first application of an averaging template ($\alpha = 0.5$).

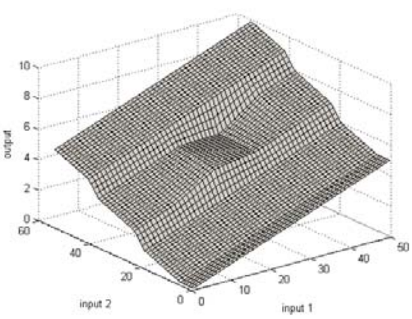

Figure 4: Graph of the sample two-dimensional sTS-system from Fig. 2 after applying an averaging template for two times ($\alpha = 0.5$).

template to a rule increases its smoothness. This can be seen as a form of incremental regularization.

For clarity reasons, we first consider a one dimensional first order sTS-fuzzy system f with triangular membership functions which sum up to 1. Then the input space is covered completely by a set of neighboring membership functions. Due to the normalization of the rule base, there is a single rule for each membership function. Thus, the rules can be ordered according to the ordinality of the membership functions so that they can be addressed by an index $i \in \mathbb{N}$, such that the indices i and $i+1$ address neighboring rules. The rule conclusion of rule number i is given by

$$f_i(x) = c_{0,i} + c_{1,i} \cdot x \tag{7}$$

with x as the input to the fuzzy system.

In the following, let $p_i \in \mathbb{R}$ denote the core of the membership function of rule i. As the membership functions are triangular, it is convenient here to distinguish between the left and the right part, $\mu_{i,l}(x)$ and $\mu_{i,r}(x)$, respectively, of each of them:

$$\begin{aligned} \mu_{i,l}(x) &= \frac{x - p_{i-1}}{p_i - p_{i-1}}, \quad p_{i-1} \leq x < p_i \\ \mu_{i,r}(x) &= \frac{p_{i+1} - x}{p_{i+1} - p_i}, \quad p_i \leq x < p_{i+1}. \end{aligned} \tag{8}$$

We are now considering the smoothness of the sTS fuzzy system in the environment of p_i. The functional behavior in this environment is determined by the neighboring rules, i.e., the rules $i-1$, i and $i+1$. To analyze the smoothness, we first calculate the first derivative of the sTS fuzzy system on the left and on the right of p_i. Between p_{i-1} and p_i, the output of the fuzzy system is then given by

$$f(x) = \mu_{i-1,r}(x) f_{i-1}(x) + \mu_{i,l}(x) f_i(x) \tag{9}$$

because of the linear interpolation due to sum-prod inference. Similarly, between p_i and p_{i+1}, the output is

$$f(x) = \mu_{i,r}(x) f_i(x) + \mu_{i+1,l}(x) f_{i+1}(x). \tag{10}$$

With these preparations, one can calculate the first derivative f' of f between the cores. For $p_{i-1} < x < p_i$ one gets

$$\begin{aligned} f'(x) \quad = \quad & \mu'_{i-1,r}(x) f_{i-1}(x) + \mu_{i-1,r}(x) f'_{i-1}(x) \\ & + \mu'_{i,l}(x) f_i(x) + \mu_{i,l}(x) f'_i(x) \end{aligned} \tag{11}$$

and for $p_i < x < p_{i+1}$

$$\begin{aligned} f'(x) \quad = \quad & \mu'_{i,r}(x) f_i(x) + \mu_{i,r}(x) f'_i(x) + \\ & \mu'_{i+1,l}(x) f_{i+1}(x) + \mu_{i+1,l}(x) f'_{i+1}(x). \end{aligned} \tag{12}$$

We can now determine the difference between the right and the left limit of $f'(x)$ with $x \to p_i$ and simplify it as the membership functions are either 0 or 1 at p_i:

$$\begin{aligned} \Delta_i \quad &= \quad \lim_{x \searrow p_i} f'(x) - \lim_{x \nearrow p_i} f'(x) \\ &= \quad (\mu'_{i,r}(p_i) - \mu'_{i,l}(p_i)) f_i(p_i) \\ & \qquad + \mu_{i+1,l}(p_i) f_{i+1}(p_i) - \mu'_{i-1,r}(p_i) f_{i-1}(p_i) \\ &= \quad \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) (c_{0,i} + c_{1,i} p_i) \\ & \qquad + \frac{1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ & \qquad + \frac{1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i). \end{aligned} \tag{13}$$

ISBN: 978-989-95079-6-8

This expression means that in general one obtains a different result for the first derivative when approaching p_i from left than from right. To put it otherwise, Δ_i is a measure of the violation of the smoothness property at rule i.

Now suppose that our sTS fuzzy system contains a rule i with $\Delta_i \neq 0$. We will now prove that we can use a special SILKE template to reduce the violation of the smoothness property. This special SILKE template is given by the radius $r = 1$ and the mask

$$\begin{aligned} m(-1) &= \frac{3(p_{i+1} - p_i)}{p_{i+1} - p_{i-1}} \\ m(0) &= 0 \\ m(1) &= \frac{3(p_i - p_{i-1})}{p_{i+1} - p_{i-1}}. \end{aligned} \tag{14}$$

The application of this mask according to (4) and (5) changes $c_{0,i}$ and $c_{1,i}$ to $\hat{c}_{0,i}$ and $\hat{c}_{1,i}$, respectively:

$$\begin{aligned} \hat{c}_{a,i} &= c_{a,i} + \alpha \left(\frac{p_{i+1} - p_i}{p_{i+1} - p_{i-1}} c_{a,i-1} \right. \\ &\quad \left. + \frac{p_i - p_{i-1}}{p_{i+1} - p_{i-1}} c_{a,i+1} - c_{a,i} \right), \; a \in \{1, 2\}. \end{aligned} \tag{15}$$

This yields an updated value $\hat{\Delta}_i$ for the violation of the smoothness property. By comparing the updated value $\hat{\Delta}_i$ with the old value Δ_i, we can prove that the SILKE template (14) has the desired effect:

$$\begin{aligned} \hat{\Delta}_i &= \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) (\hat{c}_{0,i} + \hat{c}_{1,i} p_i) \\ &\quad + \frac{1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad + \frac{1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \\ &= \Delta_i + \alpha \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) \\ &\quad \cdot \left(\frac{p_{i+1} - p_i}{p_{i+1} - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \right. \\ &\quad + \frac{p_i - p_{i-1}}{p_{i+1} - p_{i-1}} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad \left. + \frac{-1}{1} (c_{0,i} + c_{1,i} p_i) \right) \\ &= \Delta_i + \alpha \left(\left(\frac{1}{p_{i+1} - p_i} + \frac{1}{p_i - p_{i-1}} \right) (c_{0,1} + c_{1,i} p_i) \right. \\ &\quad + \frac{-1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad \left. + \frac{-1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \right) \\ &= \Delta_i - \alpha \Delta_i \\ &= (1 - \alpha) \Delta_i. \end{aligned} \tag{16}$$

From this, it can be concluded that

$$|\hat{\Delta}_i| \leq |\Delta_i|. \tag{17}$$

The effect of this parameter update on a sample two-dimensional sTS-system is illustrated in Fig. 2 to 4. The figures demonstrate that each application of (4) on the central rule (in this case with $\alpha = 0.5$) increases the smoothness.

4 Discussion

The above investigations have shown that first order sTS-fuzzy systems with triangular membership functions have discontinuities of their first derivative precisely at the cores of the input membership functions, i.e., at the nodes of the normalized rule base. These discontinuities correspond to violating the *local smoothness* of the system behavior. It was shown for the one dimensional case that the SILKE approach reduces the discontinuity locally by a special averaging template, i.e., it increases the local smoothness. The effect depends on the adjustment rate α. For $\alpha = 0$, the rule base is not changed at all. But for higher values of α, the discontinuity of the first derivative is more and more reduced. Thus, the learning process is guided towards the meta-property of global smoothness expressed by the template function. Up to now, the formal proof is for the one dimensional case, although the SILKE algorithm is of course applicable to arbitrary dimensionality. We have shown by examplary investigations that the smoothing property also holds for higher dimensions. The extension of the formal proof is currently in progress.

In contrast to prior work on regularization, the SILKE approach has the advantage of working incrementally. It is thus well suited for incremental online learning which is required for self-adapting systems or building models online. During online operation, it is applied periodically and affects all currently active rules of the fuzzy system. This way, SILKE templates get a global influence although they work locally, which results in a low computational effort and good real-time capabilities. But what is more, sufficient plasticity of learning is still given due to locality, although the learning process is guided towards global meta-level properties.

Incremental online learning in a closed-loop setting is potentially chaotic. But with the SILKE approach, the influence of learning on the behavior of an embedded system in its environment can be guided by correcting the learned knowledge locally towards higher smoothness or towards compliance with other local meta-level properties when other templates are used. Depending on the adjustment rate, this yields a system which is very flexible in the face of learning stimuli that comply the meta-properties, but which is also very robust against learning stimuli which violate these meta-properties. So for this case, self-adaptive systems can fight outliers and overcome the stability-plasticity-dilemma to a certain extent. Based on this it is clear that the focus of this work is not to achieve the highest possible modeling quality, but to stabilize adaptive control systems.

At system level, the presented approach is an effective and efficient method for controlling incremental online learning without limiting the required flexibility. Such a system can cope with changes online and in a reliable way. It is based on the use of rules. On the one hand this aids engineerability at design time. On the other hand, rules which have been modified by self-tuning can also be analyzed directly. By that good engineerability and high trustworthiness are achieved.

In future work, the interplay between closed-loop incremental online learning and the SILKE approach will be investigated especially for the application of other template functions to zero and first order sTS-fuzzy systems in order to achieve other meta-properties such as monotony and curvature. As before, this will be done in experiments as well as

ISBN: 978-989-95079-6-8

in formal analysis. In addition, the extension of the results to more than one dimension has to be addressed.

Acknowledgement

This project is funded by the German Research Foundation DFG in the framework of the *Organic Computing* priority research program under No. BR 1980/3-1.

References

[1] C. Müller-Schloer, C. von der Malsburg, and R. P. Würtz. Organic Computing. *Informatik-Spektrum*, 27(4):332–336, 2004.

[2] W. Brockmann. Online Machine Learning For Adaptive Control. *IEEE Int. Work. Emerging Technologies and Factory Automation, 1992*, pages 190–195, 1992.

[3] J. Farrell and M. Polycarpou. *Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. Wiley Interscience, Hoboken, New Jersey, 2006.

[4] P. P. Angelov and D. P. Filev. An Approach to Online Identification of Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):484–498, 2004.

[5] T. A. Johansen and R. Murray-Smith. The operating regime approach to nonlinear modelling and control. *Multiple Model Approaches to Modelling and Control*, pages 3–72, 1997.

[6] T. A. Johansen. Robust identification of Takagi-Sugeno-Kang fuzzy models using regularization. In *Fifth IEEE Int. Conf. Fuzzy Systems*, volume 1, pages 180–193, 1996.

[7] J. Farrell and M. Polycarpou. *Adaptive Approximation Based Control*. Wiley–Interscience, 2006.

[8] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.

[9] F. Bauer, S. Pereverzev, and L. Rosasco. On regularization algorithms in learning theory. *Journal of Complexity*, 23(1):52–72, 2007.

[10] A.E. Hoerl and R.W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 42(1):80–86, 2000.

[11] A. Tikhonov. On the stability of inverse problems. *CR (Dokl.) Acad. Sci. URSS, n. Ser.*, 39:176–179, 1943.

[12] E. Lughofer and S. Kindermann. Improving the Robustness of Data-driven Fuzzy Systems with Regularization. In *Proc. Int. Conf. Fuzzy Systems*, pages 703–709. IEEE, 2008.

[13] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

[14] W. Brockmann and A. Horst. Stabilizing the Convergence of Online-Learning in Neuro-Fuzzy Systems by an Immune System-inspired Approach. In *Proc. IEEE Int. Conf. Fuzzy Systems, FUZZ-IEEE*, pages 351–356. IEEE, 2007.

[15] N. Rosemann and W. Brockmann. Concept for Controlled Self-optimization in Online Learning Neuro-fuzzy Systems. In J. Hertzberg, M. Beetz, and R. Englert, editors, *KI 2007: Advances in Artificial Intelligence, LNAI*, volume 4667, pages 498–501. Springer, Berlin, 2007.

[16] N. Rosemann, B. Neumann, and W. Brockmann. Formale Eigenschaften des SILKE-Ansatzes zur Kontrolle selbstoptimierender Systeme. In H.-G. Hegering and A. et al. Lehmann, editors, *Proc. Informatik 2008 Beherrschbare Systeme – dank Informatik, LNI*, volume 134, pages 755–762. Gesellschaft für Informatik, Bonn, 2008.

[17] W. Brockmann and N. Rosemann. Instantaneous Anomaly Detection in Online Learning Fuzzy-Systems. In F. Hoffmann and O. et al. Cordon, editors, *3rd Int. Workshop on Genetic and Evolving Fuzzy Systems*, pages 23–28. IEEE Press, Piscataway, 2008.

ISBN: 978-989-95079-6-8