A way to reduce model complexity for Non Linear System Approximators

Cesare Fantuzzi, Giuseppe Perrica and Cristian Secchi.
Dipartimento di Scienze e Metodi per l'Ingegneria
University of Modena and Reggio Emilia
Via Amendola, 2 42100, Reggio Emilia
E-mail: {cesare.fantuzzi,giuseppe.perrica,cristian.secchi}@unimore.it

March 16, 2009

Abstract

In this paper we exploit the approximation capabilities of non linear function approximator based on fuzzy system structure, to devise an identification procedure for Single-Input Single-Output systems, which minimizes the squared error between the model and the target. The adoption of a model featuring an increased locality allows a substantial reduction in the complexity of the identification phase in which samples are taken into account. Then, a data-independent mapping is devised to translate modified Non-Linear models into conventional ones.

1 Introduction

The fuzzy system's ability of conjugating heuristic knowledge with quantitative and accurate representation has been widely exploited for the identification of complex processes [1]. Many approaches to fuzzy modeling have been investigated and, for many of the possible choices, theoretical results have been established confirming that fuzzy systems are universal approximators, i.e. that they are suitable for the identification of general non-linear systems.

Actually, this is true for classical Mamdani models with constant consequents (e.g. [2][3]) which are also able to reproduce the first-order differential characteristics of the target system. It is also true for Takagi-Sugeno (TS) models [4] which benefit from their enriched consequent structure and provide reproduction of first and second order differential behavior of the target system at least in the Single-Input Single-Output (SISO) case.

Relying on the above results we introduce a simple and efficient procedure for the identification of the consequent parameters of a SISO TS model in the sense of the least squares. It is based on triangular membership functions modified to increase the locality of the resulting fuzzy base functions and granting

a favorable structure to the matrix involved in the classical least square identification. A data independent technique, is also defined to map a modified TS model into a conventional one.

The paper is organized as follows. In Section 2, the definition of Takagi-Sugeno models is briefly recalled along with its cardinal properties which are subsequently discussed in relation with the main steps in identification procedures.

Section 3 introduces the increased-locality fuzzy basis functions and show how this reduces the complexity of the least squares identification of consequents. The general principles on which the translation between modified and conventional TS models is based are discussed in Section 4 which also details the related procedure.

Final version of the paper will present somo more theoretical insight and some practical examples.

2 Non-Linear Models

Let us concentrate on SISO fuzzy systems mapping a real variable $x_m \leq x \leq x_M$ into a single real number and let the output of such systems be given by [4]

$$T(x) = \frac{\sum_{i=1}^{n} \mu_{A_i}(x)(a_i x + b_i)}{\sum_{i=1}^{n} \mu_{A_i}(x)} \tag{1}$$

where n rules of the kind "**if** x **is** A_i **then** y **is** $a_i x + b_i$" are assumed and μ_{A_i} is the membership function of the fuzzy set A_i. We will also assume that n points $x_1 \leq x_2 \leq \ldots \leq x_{n-1} \leq x_n$, with $x_1 = x_m$ and $x_n = x_M$, exist for which

$$\mu_{A_i}(x) = \begin{cases} 0 & \text{if} \quad x < x_{i-1} \\ \frac{x - x_{i-1}}{x_i - x_{i-1}} & \text{if} \quad x_{i-1} \leq x < x_i \\ \frac{x_{i+1} - x}{x_{i+1} - x_i} & \text{if} \quad x_i \leq x < x_{i+1} \\ 0 & \text{if} \quad x_{i+1} \leq x \end{cases} \tag{2}$$

so that the A_i are triangular and form a strict partition of $[x_m, x_M]$, i.e. $\sum_{i=1}^{n} \mu_{A_i}(x) = 1$.

For a more compact development of the following discussion let us also define the generic basis function $\phi_i(x) = \begin{cases} x\mu_{A_i}(x) & \text{if} \quad i = 1, \ldots, n \\ \mu_{A_{i-n}}(x) & \text{if} \quad i = n+1, \ldots, 2n \end{cases}$ and the generic parameter $p_i = \begin{cases} a_i & \text{if} \quad i = 1, \ldots, n \\ b_{i-n} & \text{if} \quad i = n+1, \ldots, 2n \end{cases}$ for $i = 1, \ldots, 2n$, so that (1) actually translates into the linear combination

$$T(x) = \sum_{i=1}^{2n} p_i \phi_i(x) \tag{3}$$

Let us finally state the following

Theorem 1 *For any F with continuous second derivative there are three constants $C_0, C_1, C_2 > 0$ with the following property.*

If $\Delta = \max_{i=1,\ldots,n-1}(x_{i+1} - x_i)$ then a TS model based on the triangular membership functions in (2) exists such that

$$\sup_{x_m \leq x \leq x_M} |F^{(i)}(x) - T^{(i)}(x)| \leq C\Delta^2 \; i = 1, 2, 3 \tag{4}$$

It can be easily seen that this is enough to say that sufficiently dense fuzzy basis functions will reproduce a target behavior also in the least square sense.

Hence, we may adopt a TS model and go through the two phases of clustering and consequent estimation needed for a complete identification process.

To this end, note that Theorem 1 ensures that no complex membership structure is needed to achieve universal approximation capabilities so that any well-posed clustering scheme can be applied to compute the points x_i.

3 Consequent identification

From (3) we easily get that $T(x)$ is linear in every parameter p_i. Hence, least square identification reduces to a classical quadratic problem.

To see this assume that $m > 2n$ samples $(\bar{x}_1, \bar{y}_1 = F(\bar{x}_1)), \ldots, (\bar{x}_m, \bar{y}_m = F(\bar{x}_m))$ are given and define the rectangular regression matrix Φ setting $\Phi_{ij} = \phi_i(\bar{x}_j)$. Indicate also with $y = (\bar{y}_1, \ldots, \bar{y}_m)$ the vector of the observed system outputs and with $p = (p_1, \ldots, p_{2n})$ the parameter vector.

If the $\cdot^T$ operator transpose its argument, the square error between the model and the sampled behavior is $(\Phi p - y)^T(\Phi p - y)$ which is trivially minimized by

$$p = (\Phi^T \Phi)^* \Phi^T y \tag{5}$$

where the $\cdot^*$ operator gives the pseudo-inverse of its argument. The computational complexity of such a calculation is dominated by the number of multiplications needed for 1 matrix by matrix product, 1 matrix inversion and 2 matrix by vector products. In our case the total number of multiplications cannot be less than $(2n)^{\log_2 7} + 8n^2 + 2nm$, where the $(2n)^{\log_2 7}$ term comes from the inversion [8] of the sample-dependent matrix $\Phi^T\Phi$.

Assume, now,that $x_{i+1} - x_i = (x_M - x_m)/(n-1)$ and note that, as it can be also proved that the bound in Theorem 1 is strict (i.e. that functions F exist for which (4) holds with the equal sign) the asymptotic decay of the error is like $O(1/n^2)$ while the asymptotic increase in identification complexity is $O(n^{\log_2 7})$. A different technique is therefore needed to make precision computationally affordable.

To this end, assume that the point $x_1 < x_2 < \ldots < x_n$ are given. Define another sequence $x'_1, \ldots, x'_{2n-2}$ as

$$x'_i = \begin{cases} x_1 & \text{if} \quad i = 1 \\ x_{i/2+1} & \text{if} \quad i < 2n-2 \quad \text{even} \\ x_{(i+1)/2} & \text{if} \quad i > 1 \quad \text{odd} \\ x_n & \text{if} \quad i = 2n-2 \end{cases} \tag{6}$$

and indicate with A'_i, $\phi'_i(x)$, Φ', p' the related quantities. This new TS model is defined on the same points x_i but relies on basis functions with a more local behavior.

As it can be expected, this increased locality, eases the least square identification. To see how let us now prove the following

Theorem 2 *Let the m samples be sorted so that $\bar{x}_1 \leq \bar{x}_2 \leq \ldots \leq \bar{x}_m$ and indicate with m_i the number of samples falling within $[x'_i, x'_{i+1}]$.*

The columns of Φ' (and thus of the parameter vector p') can be appropriately permuted to obtain the following structure

$$\Phi' = \begin{pmatrix} \Phi'_1 & 0 & \ldots & 0 \\ 0 & \Phi'_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Phi'_{n-1} \end{pmatrix} \tag{7}$$

where each Φ'_i is a $m_i \times 4$ matrix.

Proof: Note first that from (6) we get that $x'_i = x'_{i+1}$ for any even $i < 2n-2$ and $x'_i < x'_{i+1}$ for any odd i. Thus, if i is odd and $x \in]x'_i, x'_{i+1}[$ the only non-vanishing basis functions are $\phi'_i(x)$, $\phi'_{i+1}(x)$, $\phi'_{i+2n-2}(x)$ and $\phi'_{i+2n-1}(x)$.

From the definition of Φ' we get that the structure in (7) can be easily obtained if the columns $i, i+1, i+2n-2, i+2n-1$ are made adjacent. ∎

From Theorem 2 we easily get that, once that the parameter vector is appropriately permuted, the quadratic minimization problem Φ' can be decomposed into $n-1$ smaller problems involving Φ'_i.

As Φ'_i is a $m_i \times 4$ matrix, $\Phi'^T\Phi$ is a 4×4 matrix and solving each of these problems entails $4^{\log_2 7} + 2 \times 4^2 + 4m_i$ for a total of $\sum_{i=1}^{n-1} 4^{\log_2 7} + 2 \times 4^2 + 4m_i = (n-1)(4^{\log_2 7} + 32) + 4m$.

This reduction in the computational burden is not without price. In fact, as the number of parameters increases from $2n$ to $4n-4$, the modified model is more prone to follow noisy fluctuations in the samples and may even be non-continuous at any point x'_i with $i > 1$ odd.

If this is not desired a procedure for the translation of modified TS models into conventional TS models can be applied as described in the following Section.

4 Consequent Reduction

The core idea allowing a smooth transition from local TS models to conventional TS models is contained in the following rephrasing of a well-known Theorem from linear algebra.

Theorem 3 *Let two set of basis functions $\phi_1(x), \ldots, \phi_{2n}(x)$ and $\phi'_1(x), \ldots, \phi'_{4n-4}(x)$ be given along with a vector of coefficients $p' = (p'_1, \ldots, p'_{4n-4})$.*

The vector of coefficients $p = (p_1, \ldots, p_{2n})$ which minimizes the error

$$\int_{x_m}^{x_M} \left[\sum_{i=1}^{2n} p_i \phi_i(x) - \sum_{i=1}^{4n-4} p'_i \phi'_i(x) \right]^2 dx \tag{8}$$

is given by $p = A^* B p'$ where A is a $2n \times 2n$ symmetric matrix defined by $A_{ij} = \int_{x_m}^{x_M} \phi_i(x)\phi_j(x)dx$ and B is a $(4n-4) \times 2n$ matrix defined by $B_{ij} = \int_{x_m}^{x_M} \phi'_i(x)\phi_j(x)dx$.

Proof: We have to require that the derivative of (8) with respect to p_j vanishes for each j, i.e.

$$\int_{x_m}^{x_M} \left[\sum_{i=1}^{2n} p_i \phi_i(x) - \sum_{i=1}^{4n-4} p'_i \phi'_i(x) \right] \phi_j(x)dx = 0$$

or, alternatively,

$$\sum_{i=1}^{2n} \left[\int_{x_m}^{x_M} \phi_i(x)\phi_j(x)dx \right] p_i - \sum_{i=1}^{4n-4} \left[\int_{x_m}^{x_M} \phi'_i(x)\phi_j(x)dx \right] p'_i = 0$$

which, with the definition of A and B, yields $Ap - Bp' = 0$. The property of the pseudo-inverse finally guarantee that $p = A^* B p'$ satisfies this latter equation. ∎

Theorem 3 allows the translation of local TS models as defined and identified in the previous Section into conventional TS models. This translations minimizes the square error between the local and the conventional model under the assumption of uniformly distributed samples and benefits from the noise rejection properties of least square identification. Model smoothing is therefore performed as the number of parameter is brought down to $2n$ and the conventional a_i and b_i are determined.

The computational complexity of this procedure is dominated by the multiplications needed in the inversion of A which are not less than $(2n)^{\log_2 7}$. Nevertheless, the matrix A^*B do not depend on the samples but only on the two sets of basis functions and can be re-applied to every subsequent identification exploiting the same antecedent structure. Moreover, A^*B can be pre-computed in the case that is by far the most common, i.e. when the x_i are evenly distributed from x_m to x_M.

Further on, the proposed technique is of general applicability as it allows the translation from any fuzzy basis to any other fuzzy basis, whatever is the shape and the position of the underlying membership functions.

References

[1] Tsung-Ying Sun and Shang-Jeng Tsai and Cheng-Han Tsai and Chih-Li Huo and Chan-Cheng Liu, "Nonlinear Function Approximation Based

on Least Wilcoxon Takagi-Sugeno Fuzzy Model", *Proceedings of the 2008 Eighth International Conference on Intelligent Systems Design and Applications*, Pages 312-317.

[2] L.-X. Wang, J. M. Mendel, "Fuzzy Basis functions, Universal Approximation, and Orthogonal Least-Squares Learning," *IEEE Transactions on Neural Networks*, vol. 3, pp. 807-814, 1992

[3] B. Kosko, "Fuzzy Systems as Universal Approximators," *IEEE Transactions on Computers*, vol. 43, pp. 1329-133, 1994

[4] T. Takagi, M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, pp. 116-132, 1985

[5] R. Babuška, R. Jager, H.B. Verbruggen, "Interpolation Issues in Sugeno-Takagi Reasoning," Third IEEE International Conference on Fuzzy Systems, Orlando (Florida), March 1994

[6] R. Babuška, C. Fantuzzi, U. Kaymak, H.B. Verbruggen, "Improved Inference for Takagi-Sugeno Models," Fifth IEEE International Conference on Fuzzy Systems, New Orleans (Louisiana), September 1996

[7] R. Babuška, M. Setnes, U. Kaymak, H.R. van Nauta Lemke, "Rule Base Simplification with Similarity Measures," Fifth IEEE International Conference on Fuzzy Systems, New Orleans (Louisiana), September 1996

[8] W.H. Press, S.A. Teukolsky, W. T. Vetterling and B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 2nd ed.,1992