ELSEVIER

Fuzzy Sets and Systems 118 (2001) 281–296

FUZZY sets and systems

www.elsevier.com/locate/fss

Autostructuration of fuzzy systems by rules sensitivity analysis

Antonio Fiordaliso *

Faculté Polytechnique de Mons, Department of Mathematics and Operational Research, 9, rue de Houdain, 7000 Mons, Belgium

Received May 1998; received in revised form October 1998

Abstract

We present a destructive (pruning) method aiming at gradually finding the appropriate number of rules in the case of fuzzy models. A particular attention has been paid to Takagi–Sugeno fuzzy systems (TS) for the problem of functions approximation. The proposed system can be seen as a generalization of the conventional TS system and allows to evaluate the importance of one particular rule in the inference process. The advantage of our approach has been put in light on two well-known benchmarks related to the field of chaotic time series forecasting. We also study and compare possible local and global learning strategies for these systems in terms of readability and performance.

Keywords: Takagi–Sugeno systems; Fuzzy rules learning; Membership functions; Chaotic time series forecasting; Empirical research

1. Introduction

The process of identifying a fuzzy inference system (or fuzzy model) generally requires two types of tuning designated as structural and parametric tunings. The first one concerns the structure of the rules and deals with problems such as the partition of the universe of discourse, the number of fuzzy if–then rules and the number of membership functions for each input.

Though several methods have been proposed to automatize this task (see for example [8,17,24,30,41]), the identification of rules structure is still an extremely difficult process where human intervention is generally required. Once a satisfactory structure is available, it is relatively easy to automatically adjust the membership functions (parametric tuning) and different techniques have been designed including gradient-based algorithms [7,35], Karr's genetic algorithm [24] or Lin's reinforcement learning method [30].

This paper is mainly motivated by two important questions that can be stated as follows: (1) we know several fuzzy sytems (including TS systems) are universal approximators [6], but how many rules are needed if the system is intented to be used for generalization tasks such as forecasting or patterns recognition? (2) what is the influence of the parametric tuning strategy on the readability and the performance of the resulting system?

From a learning-generalization point of view, it is well known that overparametrized structures often fail to generalize on new data; this is the overfitting phenomenon. Hence, for many tasks related to

* Tel.: +32 065 374686; fax: +32 065 374689.

E-mail address: fiordal@mathro.fpms.ac.be (A. Fiordaliso).

0165-0114/01/$ - see front matter
PII: S0165-0114(98)00430-8

approximation or forecasting, the structure of the fuzzy model to discover must be *rich enough* to learn the inputs–outputs associations of the training set, but *not too rich* in order to avoid the system to model the noise inherent to the data set. This preoccupation also exists in other disciplines such as statistics or neural networks: the use of information criteria [39], complexity penalization techniques (weight decay [21], weight elimination [47]) or regularization methods [37] are some illustrations of this parcimony principle.

The decremental algorithm presented in this paper for TS fuzzy systems autostructuration is clearly in the spirit of this philosophy. The key idea is to associate a parameter with each rule corresponding to the rule importance in the inference process. Then, the algorithm attempts to find redundant or least sensitive rules in order to remove them until a stable structure is found.

The paper is organized as follows. Takagi–Sugeno fuzzy systems are briefly described in Section 2 and some classical parametric adjustment strategies are discussed and compared in Section 3. It appears that global learning methods are superior in term of approximation error criteria while local learning allows for a better interpretation of the generated rules. Section 4 presents the decremental approach (DEC) and illustrates the ability of the algorithm to detect the redundancies of a TS fuzzy system. This characteristic is fully exploited in Section 5 where DEC has been successfully used to forecast two chaotic time series. The last section provides the conclusions.

2. Takagi–Sugeno fuzzy systems

Takagi–Sugeno fuzzy systems [42] form a very special class of fuzzy systems because the conclusion of each rule is crisp (not a fuzzy set). A typical single antecedent fuzzy rule in a Takagi–Sugeno model of order d (TS, in the sequel) has the form

R_k: If $\boldsymbol{x}_t$ is A_k then $\hat{y}_{t,k} = P_k^{(d)}(\boldsymbol{x}_t)$, $k = 1, 2, \ldots, c$

where $\boldsymbol{x}_t$ is the input variable ($\boldsymbol{x}_t \in \mathbb{R}^n$), A_k is a fuzzy set of $\mathbb{R}^n$ and $P_k^{(d)}(\boldsymbol{x}_t)$ is a polynomial of order d in the components $\boldsymbol{x}_{t,i}$ of $\boldsymbol{x}_t$. In the sequel, we will suppose $d = 1$.

For convenience, we will write the conclusion $\hat{y}_{t,k}$ of rule R_k relatively to input $\boldsymbol{x}_t$ as

$$\hat{y}_{t,k} = \boldsymbol{x}_t' \boldsymbol{\beta}_k \tag{1}$$

where $\boldsymbol{\beta}_k = (\beta_1, \ldots, \beta_n)'$. An intercept is allowed in the conclusion $\hat{y}_{t,k}$ if we suppose $x_{t,1} = 1$ (bias term).

Output $\hat{y}_t$ relative to input $\boldsymbol{x}_t$ obtained after aggregating a set of c TS-rules can be written as a weighted sum of the individual conclusions:

$$\hat{y}_t = \sum_{k=1}^{c} \pi_k(\boldsymbol{x}_t) \hat{y}_{t,k} \tag{2}$$

with

$$\pi_k(\boldsymbol{x}_t) = \frac{\mu_{A_k}(\boldsymbol{x}_t)}{\sum_{j=1}^{c} \mu_{A_j}(\boldsymbol{x}_t)} \tag{3}$$

where μ_{A_k} is the membership function related to the fuzzy set A_k.

Several links can be drawn between TS systems and other models such as *radial basis functions neural networks* [37], *switching regression models* [38], *mixture of distributions* [33], *fuzzy c-regression models* [20] or *mixture of experts* [23].

It has been shown that TS systems (along with several other fuzzy models) are universal approximators (see [19] for a review). This mainly motivates their use in fields related to functions approximation such as time series forecasting or control. Several proofs of the universal approximation property are based on the Stone–Weierstrass theorem (see, e.g. [45,46]). Other papers show the uniform approximation capability of specific families of fuzzy systems not fulfilling the hypotheses of the Stone–Weierstrass theorem. This mainly concerns fuzzy systems that do not form an algebra (see, e.g. [25,34,48]). Other existing proofs are adaptations of some universal approximation results known in the field of neural networks (see, e.g. [5,6]).

A universal approximation theorem is an existence theorem. As such, it does not help in specifying the fuzzy model that will achieve a satisfactory approximation. For this reason, the results on the universal approximation property of various models (including fuzzy systems) are rather academic. Note also that condition for arbitrary accuracy is the exponential growth of the size of the system (and consequently, the computational time) (see, e.g. [26–28,34]).