Neurocomputing 123 (2014) 281–291

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Incremental smooth support vector regression for Takagi–Sugeno fuzzy modeling

Rui Ji [a,b,*], Yupu Yang [a,b], Weidong Zhang [a,b]

[a] Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 21 October 2012
Received in revised form
12 July 2013
Accepted 12 July 2013
Communicated by W.S. Hong
Available online 16 August 2013

Keywords:
Takagi–Sugeno fuzzy systems
Smooth support vector regression
Reference functions
Fuzzy modeling
ε-insensitive learning
Incremental learning

ABSTRACT

We propose an architecture for Takagi–Sugeno (TS) fuzzy system and develop an incremental smooth support vector regression (ISSVR) algorithm to build the TS fuzzy system. ISSVR is based on the ε-insensitive smooth support vector regression (ε-SSVR), a smoothing strategy for solving ε-SVR, and incremental reduced support vector machine (RSVM). The ISSVR incrementally selects representative samples from the given dataset as support vectors. We show that TS fuzzy modeling is equivalent to the ISSVR problem under certain assumptions. A TS fuzzy system can be generated from the given training data based on the ISSVR learning with each fuzzy rule given by a support vector. Compared with other fuzzy modeling methods, more forms of membership functions can be used in our model, and the number of fuzzy rules of our model is much smaller. The performance of our model is illustrated by extensive experiments and comparisons.

1. Introduction

Support vector machine (SVM) [1] is one of the most promising learning algorithms for pattern classification. It is based on the structural risk minimization (SRM) principle. Vapnik introduced an ε-insensitive loss function and applied SVM to regression problems. The ε-insensitive loss function sets an ε-insensitive tube around the data, within which errors are disregarded. This problem is referred to as ε-insensitive support vector regression (ε-SVR) [2]. ε-SVR is formulated as a constrained minimization problem, and is extended to the nonlinear case by using the kernel technique. A smoothing strategy for solving ε-SVR, named ε-insensitive smooth support vector regression (ε-SSVR), was proposed in [3]. The ε-SSVR approximates the ε-insensitive loss by a smooth function and converts ε-SVR to an unconstrained minimization problem. The objective function of ε-SSVR is strongly convex and infinitely differentiable for any arbitrary kernel. It is always solvable using a fast Newton–Armijo method.

For the last decade, there has been an increasing interest in incorporating support vector learning into fuzzy modeling. Chen [4,5] proposed a positive definite fuzzy system (PDFS). The membership functions for the same input variable were generated from location transformation of a reference function [6]. The fuzzy rules were determined by support vectors (SVs) of an SVM, where the kernel was constructed from the reference functions. The kernel was proven to be an admissible Mercer kernel if the reference functions were positive definite functions [7]. Chiang [8] proposed an SVM-based modeling framework for fuzzy basis function inference system [9]. Fuzzy rules were extracted from the training data based on the SVs. In these two models, the number of fuzzy rules equaled the number of SVs. As the number of SVs in an SVM was usually large, the number of fuzzy rules was equally large. Lin proposed an SVR-based fuzzy neural network (SVRFNN) [10]. The number of fuzzy rules in the SVRFNN was reduced by removing irrelevant fuzzy rules, but this rule reduction approach degraded the generalization performance. In all these SVM-based models, the form of the membership functions was restricted by the Mercer condition [11], i.e., the positive definiteness of the membership functions was required. All of the above models use fuzzy rules with singletons in the consequent. A fuzzy system with Takagi–Sugeno (TS)-type consequent, i.e., a linear combination of the input variables, has better performance than that with singleton consequent. Researchers have also proposed methods for TS fuzzy modeling [12–21]. Leski [20] introduced Vapnik's ε-insensitive loss function to Takagi–Sugeno–Kang (TSK) fuzzy modeling. The parameters of the membership functions were determined by fuzzy c-means clustering (FCM). The number of fuzzy rules equaled the number of clusters. The consequent parameters were obtained by solving a minimization problem. Juang proposed a self-organizing TS-type fuzzy network

* Corresponding author at: Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China. Tel.: +86 21 34204261; fax: +86 21 34204038.
E-mail addresses: jirui@sjtu.edu.cn, rui.ji@hotmail.com (R. Ji).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.07.017

with support vector learning (SOTFN-SV) [22]. The antecedent of SOTFN-SV was generated by fuzzy clustering of the data, and then the consequent parameters were tuned by SVM. A TS fuzzy system-based SVR (TSFS-SVR) model was also proposed by Juang [21]. The parameters of TSFS-SVR were learned by a combination of fuzzy clustering and linear SVR. Based on the TSFS-SVR, Juang proposed two other TS fuzzy modeling methods [18,19]. Cai proposed a Gaussian kernel-based high order fuzzy system (KHFIS) [13]. That study extended Leski's model to high order fuzzy systems. In all these TS fuzzy modeling methods, the antecedent fuzzy sets were estimated
by fuzzy clustering and they were only conceived for Gaussian membership function.

The learning of SVMs can be very costly in terms of time and memory consumption, especially on large datasets as we have to deal with a large kernel matrix. In some cases, the data cannot be collected in advance, they come sequentially. To address these problems, incremental SVMs were proposed [23–26]. Juang proposed an incremental SVM-trained TS-type fuzzy classifier (ISVM-FC) [27]. That study was the first design of a fuzzy classifier using incremental SVM and can be applied to online classification problems. A fuzzy modeling method via online SVM (FSVM) was proposed in [28]. The structure identification was performed by using an online SVM, then fuzzy rules were extracted and membership functions were updated. An incremental reduced support vector machine (IRSVM) was proposed in [29]. It combined incremental learning and reduced support vector machine (RSVM). The RSVM, proposed by Lee [30,31], selected a small random portion of the data to generate a reduced kernel. This reduced kernel technique has been applied to ε-SSVR [3]. Instead of purely random selection, IRSVM selected representative samples incrementally from the dataset in forming the reduced kernel. IRSVM achieved comparable accuracy with RSVM, while with a smaller number of SVs.

In this paper, we first apply the concept of IRSVM to ε-SSVR and propose incremental smooth support vector regression (ISSVR). Then we establish a connection between ISSVR and TS fuzzy systems. An ISSVR-based TS fuzzy modeling method is proposed. TS-type fuzzy rules are automatically generated from the given training data based on the ISSVR learning. Since ε-SSVR puts no restrictions on the kernel, our model relaxes the positive definiteness requirement on membership functions. Any arbitrary form of membership functions can be used. Numerical results show that our model has good generalization ability with small number of fuzzy rules.

A brief description of our notation is given as follows. All vectors will be column vectors unless transposed to a row vector by a prime superscript $'$. $x'y$ will denote the inner product of two vectors x and y in $\mathbb{R}^n$. The p-norm of x will be denoted by $\|x\|_p$. For a vector x in $\mathbb{R}^n$, the plus function x_+ is defined as $(x_+)_i = \max\{0, x_i\}$, and the ε-insensitive loss is $(|x|_\varepsilon)_i = \max\{0, |x| - \varepsilon\}$, $i = 1, \ldots, n$. For a matrix $A \in \mathbb{R}^{m\times n}$, A_i will denote the ith row of A. A column vector of ones of arbitrary dimension will be denoted by $\mathbf{1}$. A training dataset is $\{(x^1, y_1), \ldots, (x^m, y_m)\}$, where $x^i \in \mathbb{R}^n$ is the ith sample and $y_i \in \mathbb{R}$ is the observation of real value associated with x^i. For notational convenience, the training dataset will be rearranged as an $m\times n$ matrix A and $y \in \mathbb{R}^m$. For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times l}$, the kernel $K(A,B)$ maps $\mathbb{R}^{m\times n} \times \mathbb{R}^{n\times l}$ into $\mathbb{R}^{m\times l}$. In particular, if x and y are column vectors in $\mathbb{R}^n$, then $K(x', y)$ is a real number, $K(A,x) = K(x', A')'$ is a column vector in $\mathbb{R}^m$ and $K(A, A')$ is an $m\times m$ matrix.

The rest of this paper is organized as follows. Section 2 introduces the incremental smooth support vector regression. In Section 3, we introduce TS fuzzy modeling based on the ε-insensitive learning. Section 4 describes our ISSVR-based TS fuzzy modeling method in detail. Section 5 presents experimental results and comparisons. Section 6 is the conclusion.

2. Incremental smooth support vector regression

2.1. Basic ε-SSVR concepts

We consider a given dataset $\{(x^1, y_1), \ldots, (x^m, y_m)\}$ which consists of m samples in $\mathbb{R}^n$ represented by $A \in \mathbb{R}^{m\times n}$ and m observations of real value associated with each sample. The goal of a regression problem is to find a function $f(x)$ that tolerates a small error in fitting all the data. By utilizing the ε-insensitive loss function [2], the tiny errors that fall within some tolerance ε are disregarded. Based on the idea of SVMs, the function $f(x)$ is made as flat as possible at the same time. We begin with the case of linear function $f(x)$, taking the form $f(x) = x'w + b$, where w is the normal vector. The problem can be formulated as the following unconstrained minimization problem:

$$\min_{w,b} \ \frac{1}{2}\|w\|_2^2 + C\mathbf{1}'|\xi|_\varepsilon \tag{1}$$

where $(|\xi|_\varepsilon)_i = \max\{0, |A_i w + b - y_i| - \varepsilon\}$, $i = 1, \ldots, m$, is the ε-insensitive loss and C is a positive parameter controlling the tradeoff between the flatness of $f(x)$ and the amount up to which deviations larger than ε are tolerated. Conventionally, this problem is reformulated as a convex quadratic minimization problem called ε-insensitive support vector regression (ε-SVR). Mercer kernels [11] are used to make the algorithm nonlinear.

The ε-SSVR [3] modifies the problem slightly and solves it as an unconstrained minimization problem directly. In ε-SSVR, the square of 2-norm of the ε-insensitive loss is minimized with weight $C/2$ instead of the 1-norm of the ε-insensitive loss as in (1). In addition, $b^2/2$ is added in the objective function, and this induces strong convexity and has little or no effect on the problem. These modifications lead to the following unconstrained minimization problem:

$$\min_{w,b} \ \frac{1}{2}(\|w\|_2^2 + b^2) + \frac{C}{2}\sum_{i=1}^{m} |A_i w + b - y_i|_\varepsilon^2 \tag{2}$$

For all $x \in \mathbb{R}$ and $\varepsilon > 0$, we have $|x|_\varepsilon^2 = (x-\varepsilon)_+^2 + (-x-\varepsilon)_+^2$, where x_+ is a plus function. The following p-function can provide a very accurate smooth approximation to x_+:

$$p(x, \alpha) = x + \frac{1}{\alpha}\log(1 + \exp(-\alpha x)) \tag{3}$$

where $\alpha > 0$ is the smoothing parameter. Therefore, $|x|_\varepsilon^2$ can be accurately approximated by the following p_ε^2-function:

$$p_\varepsilon^2(x, \alpha) = (p(x-\varepsilon, \alpha))^2 + (p(-x-\varepsilon, \alpha))^2 \tag{4}$$

Replacing the square of the ε-insensitive loss in (2) using this p_ε^2-function yields the following ε-SSVR formulation:

$$\min_{w,b} \ \frac{1}{2}(\|w\|_2^2 + b^2) + \frac{C}{2}\mathbf{1}'p_\varepsilon^2(Aw + \mathbf{1}b - y, \alpha) \tag{5}$$

where $p_\varepsilon^2(Aw + \mathbf{1}b - y, \alpha)_i = p_\varepsilon^2(A_i w + b - y_i, \alpha)$, $i = 1, \ldots, m$. The objective function in this problem is strongly convex and infinitely differentiable. Therefore, this problem has a unique solution and can be solved using a fast Newton–Armijo method. The solution of (2) can be obtained by solving (5) with α approaching infinity [3].

For the nonlinear case, the duality theorem in convex minimization problem [32,33] and the kernel technique [1] are applied. The observation $y \in \mathbb{R}^m$ is approximated by a nonlinear function of the form $y \approx K(A, A')u + \mathbf{1}b$, where $K(A, A')$ is a nonlinear kernel with $K(A, A')_{ij} = K(A_i, A_j')$. The regression parameter $u \in \mathbb{R}^m$ and the bias $b \in \mathbb{R}$ are determined by solving the following unconstrained minimization problem.

$$\min_{u,b} \ \frac{1}{2}(\|u\|_2^2 + b^2) + \frac{C}{2}\sum_{i=1}^{m} |K(A_i, A')u + b - y_i|_\varepsilon^2 \tag{6}$$