Neurocomputing 128 (2014) 104–112

Contents lists available at ScienceDirect

Neurocomputing

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

A multi-output two-stage locally regularized model construction method using the extreme learning machine

Dajun Du [a,b], Kang Li [b,*], Xue Li [a], Minrui Fei [a], Haikuan Wang [a]

[a] Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronical Engineering and Automation, Shanghai University, Shanghai 200072, China

[b] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT9 5 AH, UK

ARTICLE INFO

Article history:
Received 5 September 2012
Received in revised form
25 February 2013
Accepted 4 March 2013
Available online 24 October 2013

Keywords:
Extreme learning machine
Multi-output linear-in-the-parameters (LITP) model
Regularization
Two-stage stepwise selection

ABSTRACT

This paper investigates the construction of linear-in-the-parameters (LITP) models for multi-output regression problems. Most existing stepwise forward algorithms choose the regressor terms one by one, each time maximizing the model error reduction ratio. The drawback is that such procedures cannot guarantee a sparse model, especially under highly noisy learning conditions. The main objective of this paper is to improve the sparsity and generalization capability of a model for multi-output regression problems, while reducing the computational complexity. This is achieved by proposing a novel multi-output two-stage locally regularized model construction (MTLRMC) method using the extreme learning machine (ELM). In this new algorithm, the nonlinear parameters in each term, such as the width of the Gaussian function and the power of a polynomial term, are firstly determined by the ELM. An initial multi-output LITP model is then generated according to the termination criteria in the first stage. The significance of each selected regressor is checked and the insignificant ones are replaced at the second stage. The proposed method can produce an optimized compact model by using the regularized parameters. Further, to reduce the computational complexity, a proper regression context is used to allow fast implementation of the proposed method. Simulation results confirm the effectiveness of the proposed technique.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The modeling and identification of multi-input multi-output (MIMO) dynamic systems have been used in many industrial applications [1,2]. A conventional approach is to identify a multi-input single-output model for each output separately and then combine every individual model to produce a final MIMO model [2]. However, if there are common or correlated parameters for different output variables, then performing identification on all outputs simultaneously may lead to better and more robust models.

Some research has been reported on the simultaneous identification of the MIMO systems. For example, multi-innovation stochastic gradient [3] and hierarchical least squares algorithms [4] have been proposed for multi-output systems. Gradient-based and least-squares-based iterative estimation algorithms for MIMO systems have also been proposed [5]. Integrating support vector regression and annealing dynamical learning algorithm, a robust approach was developed to optimize a radial basis function (RBF) network for the identification of MIMO systems [1].

A popular alternative approach is to formulate the modeling of MIMO systems as a linear-in-the-parameters (LITP) problem (e.g., support vector machine (SVM) model [1] or RBF neural model [6]), for which some well-known solutions can be applied. For a LITP model, its performance critically depends upon the determination of the nonlinear parameters in each model term, such as the width of a Gaussian function or the fractional power of a polynomial term. A conventional strategy is to randomly select some input data points as the RBF centers [7], which may unfortunately produce a network with poor performance. To tackle this, clustering techniques have been introduced for the center location [8]. In contrast to such traditional computational intelligence techniques, the extreme learning machine (ELM) has been proposed in [9–11]. It applies random computational nodes in the hidden layer that do not need to be tuned. The hidden layer thus has the fixed parameters, allowing the output weights to be solved using the least-squares.

There are some well-known methods for the identification of LITP models. These include the popular forward orthogonal least squares (OLS) [12] and the fast forward recursive algorithm (FRA), which are used to select candidate terms (regressors) based on their contributions to maximizing the model error reduction ratio, and for RBF neural networks, all the training samples are usually used in generating the candidate terms. The OLS algorithm has also been extended for selecting the centers for multi-output RBF neural networks [13]. Further, recursive OLS algorithm is also employed to select the centers for multi-output RBF neural

* Corresponding author.
E-mail addresses: ddj559@hotmail.com, k.li@qub.ac.uk (K. Li).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.03.056

networks [6]. Unlike OLS that uses QR decomposition on the regression matrix, the recently FRA [14,15] proposes a regression context based on which fast selection of the model structure and fast estimation of model parameters are achievable. It has been shown that FRA requires much less computational effort and is also numerically more stable than some of the alternatives. The FRA method has been further extended to construct multi-output RBF neural model [16].

The above forward selection methods only provide an efficient pathway for the identification of MIMO systems. However, these methods do not consider how to control the model complexity. In general, a model with too many parameters will tend to overfit the training set and therefore fail to generalize to the test set. Conversely, a model with too few parameters will underfit the training data and hence achieve poor predictive power on both the training data and the test set. Ideally, a sparse learner balances model complexity against training set size, with the goal of balancing between under- and over-fitting. The benefits of the resulting sparse model include improved generalization capability and robustness to new test data and greater efficiency [17,18]. The regularization approach [19–21] is a useful technique to enforce the sparsity of MIMO model and to overcome the over-fitting problem. However, the regularization parameters have to be tuned to obtain satisfactory performance [22]. According to the Bayesian learning theory [19,20], a regularization parameter is equivalent to the ratio of the related hyperparameter to a noise parameter. Compared with traditional regularization methods, the Bayesian approach provides a rigorous framework for automatic adjustment of the regularization parameters to their near-optimal values. This is achieved by marginalizing the hyperparameters when making inferences, and no validation data set is needed. The Bayesian evidence procedure has also been incorporated into multi-output OLS (MOLS) [24].

In this paper, the extreme learning machine and regularized technique are introduced into the recently proposed two-stage stepwise selection algorithm [15], leading to a novel multi-output two-stage locally regularized model construction (MTLRMC) method. In this new algorithm, the nonlinear parameters in each term, such as the width of the Gaussian function and the power of a polynomial term, are firstly determined by the ELM. An initial multi-output LITP model is generated according to the termination criteria in the first stage. The significance of each selected regressor is then checked and the insignificant ones are replaced at the second stage. The proposed method can produce an optimized compact model by the regularization parameters. Further, to reduce the computational complexity, a proper regression context is defined which allows fast implementation of the proposed method.

The paper is organized as follows. Section 2 gives some preliminaries on multi-output linear-in-the-parameters model, determining the centers and widths using the ELM and the parameter estimation of multi-output linear-in-the-parameters models. Section 3 presents the proposed multi-output two-stage locally regularized model construction method, including the net error reduction to the regularized cost function, stage 1—forward model selection, stage 2—backward model refinement, complete algorithm, and computational complexity analysis. Simulation results are presented in Section 4, followed by the concluding remarks in Section 5.

2. Preliminaries

2.1. Multi-output linear-in-the-parameters model

Consider a discrete-time multivariable nonlinear system with m inputs and g outputs:

$$y_i(t)=f(y_1(t-1),\ldots,y_1(t-n_y^1),\ldots,y_g(t-1),\ldots,y_g(t-n_y^g),\ldots,u_1(t-1),\ldots,u_1(t-n_u^1),\ldots,u_m(t-1),\ldots,u_m(t-n_u^m))=f(x(t)), \quad (1)$$

where $y_i(t),\ i=1,\ldots,g$ and $u_j(t),\ j=1,\ldots,m$ are the system output and input; n_y^i and n_u^j are the corresponding maximal lags of the ith output and jth input, respectively; $x(t)=[y_1(t-1),\ldots,u_m(t-n_u^m)]^T$ is model "input" vector; $f(\cdot)$ is some unknown nonlinear function, g and m are the number of system outputs and inputs, respectively.

Suppose a multi-output linear-in-the-parameters (LITP) model is used to represent (1) such that

$$y_i(t)=\hat{y}_i(t)+\varepsilon_i(t)=\sum_{j=1}^{M}\theta_{j,i}\varphi_j(x(t))+\varepsilon_i(t), \quad i=1,\ldots,g, \quad (2)$$

where $\theta_{j,i}$ are the model weights, $\varepsilon_i(t)$ is the error between $y_i(t)$ and the ith model output $\hat{y}_i(t)$ and $\{\varepsilon_i(t)\}$ is assumed to be a white sequence, M is the number of basis functions, $\varphi_j(\cdot)$ is a known nonlinear basis function, such as Gaussian, polynomial or B-spline functions, and so on. If a Gaussian kernel function $\phi(x,c_j,\sigma_j)=\exp(-\|x-c_j\|/\sigma_j^2)$ is used, then (2) can be re-written as

$$y_i(t)=\sum_{j=1}^{M}\theta_{j,i}\varphi_j(\|x(t)-c_j\|;\sigma_j)+\varepsilon_i(t), \quad i=1,\ldots,g. \quad (3)$$

Suppose N data samples $\{x(t),Y(t)\}_{t=1}^{N}$ are used for model identification, (3) can be re-written in matrix form as

$$Y=\Phi\Theta+\Xi, \quad (4)$$

where $Y=[y_1,y_2,\ldots,y_g]\in\Re^{N\times g}$ with column vectors $y_i=[y_i(1),y_i(2),\ldots,y_i(N)]^T,\ i=1,2,\ldots,g$; $\Phi=[\phi_1,\phi_2,\ldots,\phi_M]\in\Re^{N\times M}$ with column vectors $\phi_i=[\phi_i(\|x(1)-c_i\|,\sigma_i),\ldots,\phi_i(\|x(N)-c_i\|,\sigma_i)]^T,\ i=1,2,\ldots,M$; $\Theta=[\theta_1,\ldots,\theta_g]\in\Re^{M\times g}$ with column vectors $\theta_i=[\theta_{1,i},\ldots,\theta_{M,i}]^T,\ i=1,\ldots,g$; $\Xi=[\varepsilon_1,\ldots,\varepsilon_g]\in\Re^{N\times g}$ with column vectors $\varepsilon_i=[\varepsilon_i(1),\ldots,\varepsilon_i(N)]^T,\ i=1,\ldots,g$.

2.2. Determining the centers and widths using the ELM

It is noted that each Gaussian basis function in (3) contains two adjustable parameters, the center c_j and the width σ_j. The suitable parameters can help to improve the modeling performance. Compared to the conventional conjugate gradient and exhaustive search methods [25], the extreme learning machine (ELM) [9–11] assigns random values to these parameters and is claimed to produce better generalization performance at a much faster learning speed and with least human intervene. Here, the ELM approach is employed to determine these parameters. It should be noted that the idea of choosing arbitrary data samples as the centers of candidate RBF hidden node have been proposed early in [26]. The ELM further extended the idea by assigning randomly values to the nonlinear parameters, and Huang et al. [10] further provided a theoretic framework to justify the efficacy of such approaches.

The ELM [9–11] works for the generalized single-hidden layer feedforward networks (SLFNs). The essence of the ELM is that the nonlinear parameters in the hidden layer of SLFNs need not be tuned. From the interpolation capability point of view, it has been proved that if the activation function (i.e., nonlinear basis function) $\varphi_j(\cdot)$ is infinitely differentiable in any interval, the hidden layer parameters (i.e., c_j, σ_j) can be randomly generated. The corresponding theorem is as follows.

Theorem 1 (*Huang et al. [10]*). *Given any small positive value $\varepsilon>0$ and activation function $\varphi(\cdot)\ g:R\to R$ which is infinitely differentiable in any interval, and N arbitrary distinct sample $(x(t),y(t))\in\Re^m\times\Re^g$, there exists $M\le N$ such that for any $\{c_j,\sigma_j\}_{j=1}^{M}$ randomly generated from any intervals of $\Re^m\times\Re$, according to any continuous probability distribution, then with probability one, the regression matrix Φ is invertible and $\|\Phi W-Y\|<\varepsilon$.*

Download English Version:

https://daneshyari.com/en/article/408134

Download Persian Version:

https://daneshyari.com/article/408134

Daneshyari.com