Neurocomputing 149 (2015) 160–170

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-dimensional extreme learning machine

Wentao Mao [a,*], Shengjie Zhao [a], Xiaoxia Mu [a], Haicheng Wang [b]

[a] College of Computer and Information Engineering, Henan Normal University, Henan, Xinxiang 453007, China
[b] Department of Engineering Physics, TsingHua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 3 August 2013
Received in revised form
28 January 2014
Accepted 3 February 2014
Available online 8 September 2014

Keywords:
Extreme learning machine
Multi-dimensional regression
Mixed integer programming
Loss function

ABSTRACT

As an important branch of neural network, extreme learning machines (ELMs) have attracted wide interests in the fields of pattern classification and regression estimation. However, when facing learning problems with multi-dimensional outputs, named multi-dimensional regression, the conventional ELMs could not generally get satisfactory results because it is incapable of exploiting the relatedness among outputs efficiently. To solve this problem, a new regularized ELM is firstly proposed in this paper by introducing a hyper-spherical loss function as regularizer. As the regularization form with this loss function cannot be solved directly, an solution with iterative procedure is presented. For improving the learning performance, the algorithm proposed above is further reformulated to identify the inner grouping structure hidden in outputs by assuming that the grouping structure is determined by different linear combinations of a small number of latent basis neurons. This is achieved as a mixed integer programming, and finally an alternating minimization method is presented to solve this problem. Experiments on two multi-dimensional data sets, a toy problem and a real-life dynamical cylindrical vibration data set, are conducted, and the results demonstrate the effectiveness of the proposed algorithm.

1. Introduction

Multi-dimensional regression problem, also called the multi-input multi-output (MIMO) regression, is very common to confront for many different engineering fields, for example, system identification and state estimation [1,2]. In this problem, there are two variables $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^m (m > 1)$ which are connected by a functional dependency or a probability density function. Applying multi-dimensional regression to nonlinear black-box modeling helps us to predict multiple outputs efficiently in many situations where the output variables are coupled [3].

Traditionally, the multi-dimensional regression problems are separately divided into total k single-dimensional problems, and then integrate every individual regressor to obtain a final MIMO model. This method is simple and effective. However, if there exist common or related parameters for various output variables, establishing a MIMO regressor simultaneously on all outputs may get more robust and explicable results. To simultaneously predict multi-dimensional output, hierarchical least squares algorithm [4] and multi-innovation stochastic gradient optimization [5] are proposed. Many classic methods, like support vector machine (SVM), have also been developed from single dimensional form to multi-dimensional regression scenario [6]. Specifically, SVM and other optimization algorithms can be integrated to optimize MIMO model based on the RBF network [7]. However, the corresponding price is increasing model complexity and human intervention. Due to its innate structure, multi-layer neural network can predict multiple outputs simultaneously where each output servers as an output node. In this sense, extreme learning machine (ELM) is a competitively good solution for such tasks on account of better generalization performance as well as faster learning speed and least human intervention. ELM provides a unified learning platform for single-hidden-layer feedforward neural network (SLFN) [8]. Its main idea is that the hidden layer of SLFNs need not be tuned. Unlike conventional neural networks, ELMs randomly initialize the input weights and hidden layer biases, and finally determine the output weights via a simple matrix inversion procedure [9]. Obviously, it is important to improve the precision of ELM for multi-dimensional regression.

While ELMs have proved their success in solving regression and multi-class classification problems [10], ELMs are also improved to solve multi-dimensional regression problems. To reach this goal, some researches try to develop ELM according to MIMO's requirements. Du et al. [11] try to enforce the sparsity of ELM in MIMO model in order to overcome the over-fitting problem. Specifically, he proposed a two-stage locally regularized method to establish MIMO model. Wang et al. [12] utilized an improved multi-response sparse regression method to construct an efficient MIMO ELM from a pre-generated model pool, and then proposed an

* Corresponding author.
E-mail address: maowt.mail@gmail.com (W. Mao).

http://dx.doi.org/10.1016/j.neucom.2014.02.073
0925-2312/

constructive model selection method for multi-output ELM [13]. On the other hand, because of ELM's intrinsic structure for MIMO modeling, one way is to improve the generalization performance of ELM. Considering the existence of the acceptable minimal training error, regularization on output weights is important in particular if the observations do not comprise many training samples [14]. Huang et al. [14] also proved that ELM can be linearly extended to SVM with less optimization constraints and simpler random kernel. Following this basic idea, various regularization forms such as LASSO and Tikhonov regularization [15–17] have recently been introduced to improve the generalization ability and decrease the complexity of ELM model. Note that although regularization technique could help us to establish more effective regressor for the available data, it will cause bias. Therefore, for MIMO problems, regularization technique should be used according to the specific requirement of applications.

In many practical applications, MIMO modeling problems have a large number of outputs. For examples, dynamical load identification generally needs to predict dozens of load simultaneously. In our experiment which will be shown in Section 5, we need to establish more than 100 regressors on the total of 144 measuring points. In this scenario, the knowledge among outputs could help us to improve the generalization performance of MIMO model besides the information among input variables. Therefore, the relatedness between outputs should be detected, and exploiting the inner structure among outputs will play a part vital role. Although the classical ELM can work in MIMO mode, it is incapable to draw support from outputs. Similarly, if regularization of ELM are directly applied to solving multi-dimensional regression problems with high output dimension, the results tend to be unsatisfactory. The main reason is that in these formulations, the loss functions with insensitive zone [14] or equation type [18,16] do not have multi-dimensional form, and will not suffer an equal penalty for each output. As the MIMO form of ELM [11–13] mentioned above is dedicated to choosing effective ELM model from the sparsity perspective, they are not involved in exploiting the inner structure among outputs, especially when facing a large number of outputs.

In this paper, aiming at multi-dimensional regression paradigm which has high output dimension, we focus on two important issues. First, we need to find an efficient method to establish regressor for each output using the domain knowledge among all outputs. Second, as multiple outputs have inner structure, e.g., some outputs are more related than others, we need techniques to determine this structure for better generalization performance. Guided by the above idea, a new regularized ELM for multi-dimensional regression is firstly proposed in this paper by introducing a hyper-spherical loss function as regularizer. This loss function estimates the errors in hyper-spherical form which is applicable to multi-dimensional regression. So it will allow us to equally treat every output. On account of the adopted loss function, this new ELM algorithm cannot be solved like SVMs or other conventional methods. So an solution with iterative procedure is also developed. In order to determine the inner structure, a new grouping learning algorithm of this regularized ELM is proposed. This algorithm starts from the assumption that the grouping structure is determined by different linear combinations of a small number of latent basis neurons. After reformulating as a mixed integer programming problem, an alternating minimization method is presented to solve this problem. This grouping learning algorithm can also be performed as a model selection method which could determine the compact network structure, and remarkably reduce computational complexity than the traditional regularized ELM and improve the numerical stability for multi-dimensional regression. Experiments on a toy regression data as well as a real-life dynamical cylindrical vibration data set show the benefit of the proposed algorithm. To our best knowledge, this research serves as the first attempt to study the inner grouping structure theoretically in generalization of ELM for multi-dimensional regression problem.

The rest of this paper is organized as follows. In Section 2, a brief review to ELM is given. In Section 3, both a new regularized form of ELM and its training algorithm are provided. Section 4 further proposes a grouping learning algorithm for the proposed ELM. Section 5 is devoted to computer experiments, followed by a conclusion of the paper in the last section.

2. Brief introduction of ELM

As studied by [19], the theoretical foundations of ELM are that SLFN with at most N hidden neurons can learn N distinct samples with zero error by adopting any bounded nonlinear activation function. Following this concept, Huang et al. [9] proposed ELM algorithm whose main procedure is determining the output weights by a matrix pseudo-inversion computation after initializing the input weights and hidden layer biases randomly. As proved empirically by many researchers [20], ELM has very high learning speed, simple network structure and good generalization performance. Here a brief summary of ELM is provided.

Given a set of *i.i.d* training samples $\{(\mathbf{x}_1,\mathbf{y}_1),\ldots,(\mathbf{x}_N,\mathbf{y}_N)\}\subset \mathbb{R}^d\times\mathbb{R}^m$, standard SLFNs with $\tilde{N}$ hidden nodes are mathematically formulated as

$$\sum_{i=1}^{\tilde{N}}\beta_i g_i(\mathbf{x}_j)=\sum_{i=1}^{\tilde{N}}\beta_i g_i(\mathbf{w}_i\cdot\mathbf{x}_j+b_i)=\mathbf{o}_j,\quad j=1,\ldots,N \tag{1}$$

where $g(x)$ is the activation function, $\mathbf{w}_i=[w_{i1},w_{i2},\ldots,w_{id}]^T$ is the input weight vector connecting input nodes and the ith hidden node, $\boldsymbol{\beta}_i=[\beta_{i1},\beta_{i2},\ldots,\beta_{im}]^T$ is the output weight vector connecting output nodes and the ith hidden node, b_i is the bias of the ith hidden node. Huang et al. [19] have rigorously proved that then for N arbitrary distinct samples and any $(\mathbf{w}_i,b_i)$ randomly chosen from $\mathbb{R}^d\times\mathbb{R}$ according to any continuous probability distribution, the hidden layer output matrix $\mathbf{H}$ of a standard SLFN with N hidden nodes is invertible and $\|\mathbf{H}\boldsymbol{\beta}-\mathbf{T}\|=0$ with probability one if the activation function $g:\mathbb{R}\mapsto\mathbb{R}$ is infinitely differentiable in any interval. Then given $(\mathbf{w}_i,b_i)$, training a SLFN equals finding a least-squares solution of the following equation [9]:

$$\mathbf{H}\boldsymbol{\beta}=\mathbf{Y} \tag{2}$$

where

$$\mathbf{H}(\mathbf{w}_1,\ldots,\mathbf{w}_{\tilde{N}},b_1,\ldots,b_{\tilde{N}},\mathbf{x}_1,\ldots,\mathbf{x}_{\tilde{N}})=\begin{bmatrix} g(\mathbf{w}_1\cdot\mathbf{x}_1+b_1) & \cdots & g(\mathbf{w}_{\tilde{N}}\cdot\mathbf{x}_1+b_{\tilde{N}})\\ \vdots & \cdots & \vdots\\ g(\mathbf{w}_1\cdot\mathbf{x}_N+b_1) & \cdots & g(\mathbf{w}_{\tilde{N}}\cdot\mathbf{x}_N+b_{\tilde{N}})\end{bmatrix}_{N\times\tilde{N}}$$

$$\boldsymbol{\beta}=[\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_{\tilde{N}}]^T$$

$$\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_N]^T$$

Considering most cases that $\tilde{N}\ll N$, $\boldsymbol{\beta}$ cannot be computed through the direct matrix inversion. Therefore, Huang et al. [9] calculated the *smallest norm* least-squares solution of Eq. (2):

$$\hat{\boldsymbol{\beta}}=\mathbf{H}^{\dagger}\mathbf{T} \tag{3}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$ [2]. Based on the above analysis, Huang [3] proposed ELM whose framework can be stated as follows [9]:

Step 1 : Randomly generate input weight and bias $(\mathbf{w}_i,b_i)$, $i=1,\ldots,\tilde{N}$.
Step 2 : Compute the hidden layer output matrix $\mathbf{H}$.
Step 3 : Compute the output weight $\hat{\boldsymbol{\beta}}=\mathbf{H}^{\dagger}\mathbf{T}$.