Neurocomputing 168 (2015) 823–828

Contents lists available at ScienceDirect

Neurocomputing

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

A novel extreme learning machine using privileged information

Wenbo Zhang*, Hongbing Ji, Guisheng Liao, Yongquan Zhang

School of Electronic Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 11 December 2014
Received in revised form
6 May 2015
Accepted 12 May 2015
Communicated by G.-B. Huang
Available online 22 May 2015

Keywords:
Extreme learning machine (ELM)
ELM+
Privileged information
Hidden information
Radar emitter recognition

ABSTRACT

Extreme learning machine (ELM) is a competitive machine learning technique, which is much more efficient and usually lead to better generalization performance compared to the traditional classifiers. In order to further improve its performance, we proposed a novel ELM called ELM+ which introduces the privileged information to the traditional ELM method. This privileged information, which is ignored by the classical ELM but often exists in human teaching and learning, will optimize the training stage by constructing a set of correcting functions. We demonstrate the performance of ELM+ on datasets from UCI machine learning repository, Mackey–Glass time series and radar emitter recognition and also present the comparison with SVM, ELM and SVM+. The experimental results indicate the validity and advantage of our method.

1. Introduction

Extreme learning machine (ELM) [1–5] was originally proposed for the single-hidden-layer feedforward neural networks (SLFNs) and then extended to the generalized SLFNs where the hidden layer need not be neuron alike. In ELM, the input weights of the SLFNs are randomly chosen without iterative tuning, and the output weights are analytically determined. Thus, the training speed of ELM can be thousand times faster than that of the traditional iterative implementations of SLFNs. In addition, different from the traditional learning algorithms for a neural type of SLFNs, ELM aims to reach not only the smallest training error but also the smallest norm of output weights. Bartlett's theory [6] shows that for feedforward neural networks reaching smaller training error the smaller the norm of weights is, the better generalization performance the networks tend to have. Because of its good performance, ELM has been attracting the attentions from more and more researchers [6–12], and various extensions have been made to make ELM more efficient and more suitable for the real-world problems, such as ELM for imbalanced data [13], ELM for online sequential data [14–16], ELM for noisy data [17,18], and so on. However, further improvement on the performance of ELM has gradually entered the bottleneck.

In a data-rich world, there often exists privileged information about training samples, which is not reflected directly in the training set. For example, in radar emitter recognition, traditional approaches [19] separate the received pulses into individual emitter group, such as passenger or cargo (civil), Model A or Model B (military). However, the state-of-the-art approaches [20] utilize individual parameters to ascertain specific emitters through the precise measurement of intercepted signals. Then, the groups of radar emitters can be considered as the privileged information which can be used to improve recognition performance. However, this privileged information can be easily ignored by the traditional learning machines. Recently, Vapnik and Vashist [21] proposed a general approach for solving such problems, known as Learning Using Privileged Information (LUPI), where at the training stage some additional information $\mathbf{x}^*$ about training example $\mathbf{x}$ is given. An SVM-based optimization formulation under LUPI setting is called SVM+ which can effectively utilize this privileged information to improve performance. Recent empirical comparisons [22,23] show that SVM+ provides improved generalization accuracy for handwritten digit recognition data and landmine detection, respectively. Liang and Cherkassky [24,25] showed the empirical validation of SVM+ for classification including medical diagnosis data. Zhu and Zhong [26] proved that privileged information can also improve the performance of one-class SVM. Feyereisl and Aickelin [27] utilize the notion of privileged information to the unsupervised learning in order to improve clustering performance. Sharmanska et al. [28] proposed Rank Transfer method based on privileged information and applied it to visual object classification tasks. In fact, for almost all machine learning problems there exists some sort of privileged information. Currently, nearly all learning models based on privileged information focus on SVM. However, there is no published work about ELM using the advantages of privileged information.

In this paper, ELM+ based on privileged information is proposed by embedding the additional information into the corresponding

* Corresponding author. Tel./fax: +86 29 88201658.
E-mail address: zwbsoul@163.com (W. Zhang).

http://dx.doi.org/10.1016/j.neucom.2015.05.042
0925-2312/© 2015 Published by Elsevier B.V.

optimization problem. This so-called privileged information can be seen as group information in a way. We suppose that the available training data can be partitioned into several groups in a meaningful way. In order to be useful for learning, this group information is related to the slack variables and the additional constraints on the slack variables are introduced for samples from different groups. That is, we introduce different constraints on slacks from different groups. The slack variables for each group are modeled by the correcting functions which are defined in the correcting space. Thus, the main difference between ELM+ and the standard ELM is that the standard ELM projects inputs into one space whereas ELM+ projects inputs into two different spaces: decision space and correcting space. The experimental results show that the performance can be improved by introducing privileged information in the correcting space.

The remainder of the paper is organized as follows. Section 2 briefly summarizes the principles of ELM, and illustrates privileged information by two specific examples. The proposed algorithm is described in detail in Section 3. In Section 4, the experiments and results analysis are presented. The conclusions are drawn in Section 5.

2. Related work

The proposed ELM+ is based on ELM. This section provides a brief review of ELM. In addition, the description of the privileged information can be found in this section as well.

2.1. Extreme learning machine

ELM [2] was originally proposed for the single-hidden layer feedforward neural networks and was then extended to the "generalized" single-hidden layer feedforward networks (SLFNs) where the hidden layer need not be neuron alike. The output of an ELM with $\tilde{N}$ hidden nodes can be represented by

$$f_{\tilde{N}}(\mathbf{x}) = \sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i G(\mathbf{a}_i, b_i, \mathbf{x}), \quad \mathbf{a}_i \in \mathbf{R}^n,\ \mathbf{x} \in \mathbf{R}^n \tag{1}$$

where $\mathbf{a}_i$ and b_i are the connection weights between inputs and hidden nodes, $\boldsymbol{\beta}_i$ is the weight connecting the ith hidden node to the output node, and $G(\mathbf{a}_i, b_i, \mathbf{x})$ is the output of the ith hidden node with respect to the input $\mathbf{x}$. For N arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k)$, if ELM can classify them accurately, it implies that there exist $\mathbf{a}_i$, b_i and $\boldsymbol{\beta}_i$ such that

$$\sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i G(\mathbf{a}_i, b_i, \mathbf{x}) = \mathbf{t}_k, \quad k = 1, \ldots, N. \tag{2}$$

Eq. (2) can be written compactly as

$$H\boldsymbol{\beta} = \mathbf{T}, \tag{3}$$

where

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

and $\mathbf{T} = [t_1, t_2, \ldots, t_N]^T$. $\mathbf{H}$ is called the hidden layer output matrix of the network, and the parameters $(\mathbf{a}_i, b_i)$ of $\mathbf{H}$ are randomly chosen. Then, the classification problem for ELM can be formulated as

$$\text{Minimize}: L_{ELM} = \tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + C\tfrac{1}{2}\sum_{i=1}^{N}\|\boldsymbol{\xi}_i\|^2$$

$$\text{Subject to}: \mathbf{H}(\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{t}_i^T - \boldsymbol{\xi}_i^T, \quad i = 1, \ldots, N \tag{4}$$

where $\boldsymbol{\xi}_i$ is the training error vector for the training sample $\mathbf{x}_i$, and C is the regularization parameter which represents the trade-off between the minimization of training errors and the maximization of the marginal distance. According to Karush–Kuhn–Tucker (KKT) theorem [29], to train ELM is equivalent to solving the following dual optimization problem:

$$L_{ELM} = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\frac{1}{2}\sum_{i=1}^{N}\|\boldsymbol{\xi}_i\|^2 - \sum_{i=1}^{N}\sum_{j=1}^{m}\alpha_{i,j}\left(\mathbf{h}(\mathbf{x}_i)\beta_j - t_{i,j} + \xi_{i,j}\right). \tag{5}$$

The KKT corresponding optimality conditions can be obtained as:

$$\frac{\partial L_{ELM}}{\partial \beta_j} = 0 \rightarrow \beta_j = \sum_{i=1}^{N}\alpha_{i,j}\mathbf{h}(x_i)^T \rightarrow \beta = \mathbf{H}^{\mathbf{T}}\boldsymbol{\alpha}$$

$$\frac{\partial L_{ELM}}{\partial \boldsymbol{\xi}_i} = 0 \rightarrow \boldsymbol{\alpha}_i = C\boldsymbol{\xi}_i, \quad i = 1, \ldots, N$$

$$\frac{\partial L_{ELM}}{\partial \boldsymbol{\alpha}_i} = 0 \rightarrow \mathbf{h}(x_i)\beta - \mathbf{t}_i^T + \xi_i^T = 0, \quad i = 1, \ldots, N \tag{6}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$. From (6), we have

$$\boldsymbol{\beta} = \mathbf{H}^T\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{T}. \tag{7}$$

Then, the output function of ELM classifier is

$$\mathbf{f}(x) = \mathbf{h}(x)\boldsymbol{\beta} = \mathbf{h}(x)\mathbf{H}^T\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{T}. \tag{8}$$

In summary, the input weights in ELM are randomly chosen without iterative tuning, and the output weights are analytically determined. Thus, ELM can achieve similar or much better generalization performance at much faster learning speed than the traditional learning machines, such as SVM.

2.2. Privileged information

Different from human learning, a teacher does not play an important role in the traditional machine learning paradigm. However, a teacher is very important in the process of teaching and learning, because along with examples a teacher can provide students with explanations, comments, comparisons, and so on. Vapnik and Vashist [21] proposed an advanced learning paradigm called LUPI, where at the training stage a teacher gives some additional information $\mathbf{x}^*$ about training example $\mathbf{x}$ to improve generalization. Since the additional information is available only at the training stage but it is not available for the test set, it is called privileged information.

Let us consider several examples where a teacher provides privileged information during the training stage.

1) In medical diagnosis, the lung cancer predictive model is estimated by using a training set of male and female patients. The gender can impact the medical tests results to a certain degree, and men and women have different lung cancer risk. Thus, the gender can be considered as the privileged information, and this information can be used to improve generalization.
2) In radar emitter recognition, our goal is to find a rule to classify radar emitters into specific types. At the training stage, apart from the radar signal datasets, a teacher can also give the additional information about the uses of radar emitters, such as civil and military. This additional information is available at the training stage but it is not available at the testing stage, and it can be used to improve the classification performance.

3. ELM+ using privileged information

As the description above, ELM is a competitive learning method, which achieves excellent performance both in accuracy rate and run time. Recently, many researchers devote to further