Neurocomputing 149 (2015) 340–353

Contents lists available at ScienceDirect

Neurocomputing

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

Discriminative graph regularized extreme learning machine and its application to face recognition

Yong Peng [a], Suhang Wang [b], Xianzhong Long [a], Bao-Liang Lu [a,c,*]

[a] Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
[c] Key Laboratory of Shanghai Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 3 August 2013
Received in revised form
21 November 2013
Accepted 1 December 2013
Available online 8 September 2014

Keywords:
Extreme learning machine
Graph Laplacian
Manifold regularization
Face recognition

ABSTRACT

Extreme Learning Machine (ELM) has been proposed as a new algorithm for training single hidden layer feed forward neural networks. The main merit of ELM lies in the fact that the input weights as well as hidden layer bias are randomly generated and thus the output weights can be obtained analytically, which can overcome the drawbacks incurred by gradient-based training algorithms such as local optima, improper learning rate and low learning speed. Based on the consistency property of data, which enforces similar samples to share similar properties, we propose a discriminative graph regularized Extreme Learning Machine (GELM) for further enhancing its classification performance in this paper. In the proposed GELM model, the label information of training samples are used to construct an adjacent graph and correspondingly the graph regularization term is formulated to constrain the output weights to learn similar outputs for samples from the same class. The proposed GELM model also has a closed form solution as the standard ELM and thus the output weights can be obtained efficiently. Experiments on several widely used face databases show that our proposed GELM can achieve much performance gain over standard ELM and regularized ELM. Moreover, GELM also performs well when compared with the state-of-the-art classification methods for face recognition.

1. Introduction

Extreme Learning Machine (ELM) is an emerging model proposed by Huang [1] as a least square based learning algorithm for single hidden layer neural networks (SLFNNs) [2–5]. In comparison with traditional neural networks which usually employ back propagation (BP) algorithm to train the connection weights, the tedious process of iterative parameter tuning is eliminated and the slow convergence and local minimum problems are avoided.

The consistency among ELM and SVM, least square support vector machine (LS-SVM) and proximal SVM was well studied and analyzed from the optimization point of view [5–7]. In [7], it was found that ELM can provide an unified solution for generalized SLFNNs, which include but not limit to neural network, support vector network and regularized network. That is to say, the feature mapping function can be any type of nonlinear piecewise function as in conventional ELM random nodes; or an unknown function to form a mercer's kernel as in SVMs and other kernel based algorithms.

Recently, much effort has been made on ELM from both theoretical and application aspects. Huang et al. showed that the universal approximation performance of SLFNNs can be implemented in an incremental method, which may simply choose hidden nodes at random and then adjust the output weights (I-ELM) [2]. An enhanced method for I-ELM (referred as EI-ELM) was proposed in [4]. I-ELM was proven to have the ability of approximating any target function in both the real and complex domains [8]. An error minimized ELM (EM-ELM) which can automatically determine the number of hidden nodes in generalized SLFNNs was proposed in [9]. Zong et al. applied ELM to relevance ranking and studied it as a learning-to-rank algorithm from the perspective of both pointwise and pairwise [10]. The impact of random weights between input and hidden layers was investigated in [11]. To alleviate the effect of outliers, robust ELM was proposed in [12]. Zhang et al. proposed a fuzzy ELM (FELM) in which the inputs with different fuzzy matrix can make different contributions to learn the output weights [13]. Shi et al. proposed the elastic net regularized ELM and put it to EEG based vigilance estimation [14]. Wang et al. proposed a parallelized ELM ensemble based on the Min-Max Modular network

* Corresponding author at: Key Laboratory of Shanghai Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.
E-mail address: bllu@sjtu.edu.cn (B.-L. Lu).

http://dx.doi.org/10.1016/j.neucom.2013.12.065
0925-2312/

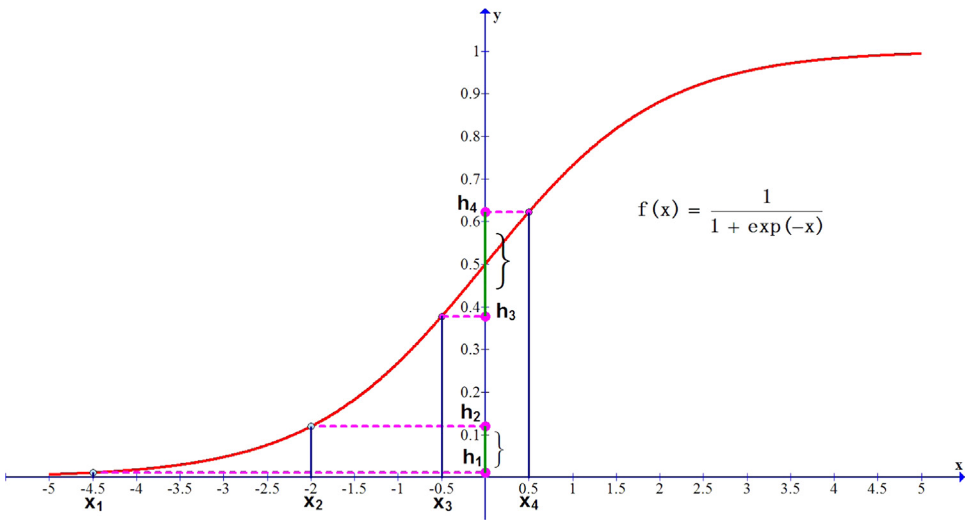

Fig. 1. An example to explain the distance information cannot be preserved by 'sigmoid' mapping.

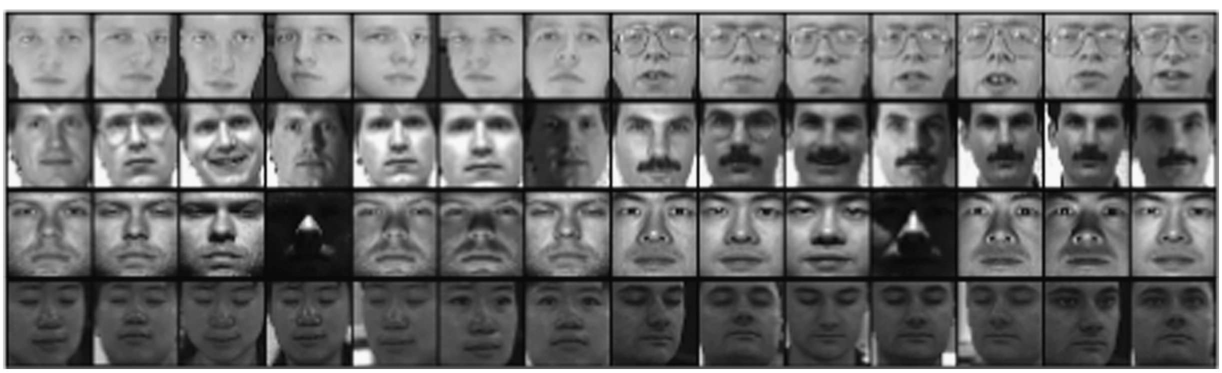

Fig. 2. The four rows show the fourteen image samples from the ORL, Yale, Extended Yale B and CMU PIE databases, respectively.

Table 1
Statistics of the four face databases.

Database	#samples	# classes	# samples/subject	dimension
ORL	400	40	10	32×32
Yale	165	15	11	32×32
Extended Yale B	2414	38	$\sim$64	32×32
CMU PIE	11,544	68	$\sim$170	32×32

(M^3-network) to meet the challenge of the so-called big data [15]. In addition, ELM has been put into diverse applications such as speaker recognition [16], neuroimage data classification [17], security assessment [18], data privacy [19], EEG and seizure detection [20], image quality assessment [21], image super-resolution [22], FPGA [23], face recognition [24], and human action recognition [25].

Recently, learning with local consistency of data has drawn much attention to improve the performance of existing machine learning models. In this paper, based on the idea that similar samples should share similar properties, we propose a discriminative graph regularized Extreme Learning Machine (GELM). In GELM, the constraint imposed on output weights enforces the output of samples from the same class to be similar. The constraint is formulated as a regularization term being added on the objective of basic ELM model, which also makes the output weights be solved analytically. We conduct experiments on four popular face databases to evaluate the performance of GELM. The experimental results demonstrate that GELM can obtain much better performance on most cases in comparison with basic ELM and state-of-the-art models.

The remainder of this paper is organized as follows. Section 2 describes the basic extreme learning machine model as well as its ℓ_2-norm regularized version. Section 3 introduces the proposed discriminative graph regularized ELM (GELM) including its model formulation and optimization method. Section 4 gives the detailed experiments to evaluate the efficiency of applying GELM to face recognition on several widely used data sets. Conclusion is given in Section 5.

2. Extreme Learning Machine

In this section, we review the Extreme Learning Machine algorithm in detail as the preliminary of our work.

Extreme Learning Machine proposed by Huang et al. [1] is an efficient and practical learning mechanism for the single layer feed forward neural networks.

Given a training data set, $L = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbb{R}^d, \mathbf{t}_i \in \mathbb{R}^m, i = 1, 2, \ldots, N\}$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})^T$ and $\mathbf{t}_i = (t_{i1}, t_{i2}, \ldots, t_{im})^T$. An ELM with K