Procedia Computer Science
Volume 53, 2015, Pages 420–427
2015 INNS Conference on Big Data

# Regularized Extreme Learning Machine for large-scale media content analysis

Alexandros Iosifidis, Anastasios Tefas and Ioannis Pitas

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
{tefas,pitas}@aiia.csd.auth.gr

**Abstract**

In this paper, we propose a new regularization approach for Extreme Learning Machine-based Single-hidden Layer Feedforward Neural network training. We show that the proposed regularizer is able to weight the dimensions of the ELM space according to the importance of the network's hidden layer weights, without imposing additional computational and memory costs in the network learning process. This enhances the network's performance and makes the proposed approach suitable for learning non-linear decision surfaces in large-scale classification problems. We test our approach in medium- and large-scale face recognition problems, where we observe its superiority when compared to the existing regularized Extreme Learning Machine classifier in both constrained and unconstrained problems, thus making our approach applicable in demanding media analysis applications such as those appearing in digital cinema production.

*Keywords:* Extreme Learning Machine, Regularization, Face Recognition, Large-scale learning

## 1 Introduction

Extreme Learning Machine has been proposed as an alternative algorithm for Single-hidden Layer Feed-forward Neural (SLFN) networks training [8], towards overcoming the computational bottleneck of related SLFN network training algorithms which are based on gradient descend optimization, e.g. the Backpropagation [16] algorithm. The main idea of ELM is that the network hidden layer weights need not to be learned, but they can be randomly assigned instead. A learning process is applied only for the determination of the network output weights by solving an optimization problem that has a closed form solution. Such an approach has also been found to be efficient in earlier attempts on neural networks training of several topologies [1, 18, 21, 2]. By using a very large number of hidden layer neurons, ELM networks can achieve satisfactory performance in many classification problems [15]. It has been also proven that ELMs have the properties of global approximators in the case where the number of hidden layer neurons is equal to the cardinality of the training set [5, 22]. Recently, it has been shown that ELM networks can achieve state-of-the-art performance in many small- and medium-scale classification problems related to media content analysis, since for such problems the realization of networks having a number of hidden layer neurons comparable to the training set cardinality is possible [7].


Selection and peer-review under responsibility of the Scientific Programme Committee of INNS-BigData2015


doi:10.1016/j.procs.2015.07.319

In order to achieve satisfactory performance in large-scale classification problems involving high-dimensional data, ELM networks need to exploit regularized solutions for the calculation of the network output weights [7]. Regularized ELM networks minimizing both the network training error and the (Frobenius) norm of the network output weights have been proposed [7, 9]. Regularized ELM networks have been shown to outperform standard ELM networks, while not requiring additional computational cost. In this paper, we propose a regularized solution for the network output weights calculation of ELM networks. When compared to the standard (Frobenius) norm regularization, which leads to uniform network output weights regularization, the proposed solution can appropriately regularize the dimensions of the obtained network output weights, while not requiring additional training computational cost.

We test the proposed classifier in a media content analysis application, i.e. human face recognition. This problem has received much attention during the last two decades, since it is the first processing step towards semantic image/video analysis and visual content analytics [10]. However, much of the research conducted until now has been focused on a restricted application scenario, that involves lab-generated visual data having small to medium scale in both resolution and size. Resent advances in technological equipment (e.g. cameras and smartphones), as well as the accessibility of social and image/video sharing web applications in our daily lives have rejuvenated research interest in this area, since the problem to be solved has been extended in three directions, i.e. visual content resolution, data size and difficulty. In order to highlight these differences between the two application scenarios, recent works characterize the new face recognition problem as an open-universe problem (when compared to the restricted application scenario noted as closed-universe face recognition problem) [14]. Experimental analysis on five publicly available databases shows that the proposed classifier can achieve almost perfect classification performance in the closed-universe face recognition problem by exploiting very simple facial image representations (i.e. vectorized image intensity values). In the open-universe face recognition problem, the proposed classifier outperforms both standard and regularized ELM networks, while both its training and test complexities are the same with that of the ELM algorithm. These observations indicate that the proposed approach is appropriate for demanding large-scale media analysis applications such as digital cinema production where a large amount of video streams from multiple cameras is captured every shooting day and it should be analysed and described for the post-production. In such an application scenario, the proposed approach can be used for automatic actor recognition, something that can facilitate subsequent post-processing steps.

## 2 Previous Work

In this Section, we introduce notation that will be used throughout the paper and we briefly describe the ELM and regularized ELM algorithms. Let us assume that an annotated visual database contains facial images depicting $C$ persons. By applying face detection and tracking techniques [23], facial images depicting the persons in the database can be extracted and pre-processed. This process leads to the determination of facial image vectors $\mathbf{x}_i \in \mathbb{R}^D,\ i = 1, \dots, N$, which are accompanied by the corresponding person ID labels $c_i$. We would like to employ the data $\{\mathbf{x}_i, c_i\}_{i=1,\dots,N}$ in order to train a SLFN network. In such classification problems, the SLFN network consists of $D$ input, $L$ hidden and $C$ output neurons. The number of hidden layer neurons $L$ is a parameter of any neural network training algorithm. We employ the person ID labels $c_i$ in order to form target vectors $\mathbf{t}_i \in \mathbb{R}^C$. The elements of the target vectors are set equal to $t_{ik} = 1$, when $c_i = k$, and $t_{ik} = -1$, otherwise.

ELM assigns randomly the network hidden layer weights $\mathbf{W}_{in} \in \mathbb{R}^{L \times D}$. By exploiting an activation function $\phi(\cdot)$, the training data $\mathbf{x}_i$ are mapped to the so-called ELM space $\mathbb{R}^L$, i.e. $\mathbf{x}_i \in \mathbb{R}^D \overset{\phi(\cdot)}{\rightarrow} \boldsymbol{\phi}_i \in \mathbb{R}^L$. It has been shown that almost any nonlinear piecewise continuous activation functions $\Phi(\cdot)$ can be used for the calculation of the network hidden layer outputs, e.g. the sigmoid, sine,