Neural Networks 77 (2016) 14–28

Contents lists available at ScienceDirect

Neural Networks

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

A Fast SVD-Hidden-nodes based Extreme Learning Machine for Large-Scale Data Analytics

Wan-Yu Deng [a,b,*], Zuo Bai [c], Guang-Bin Huang [c], Qing-Hua Zheng [d]

[a] School of Computer, Xian University of Posts & Telecommunications, Shaanxi, China
[b] School of Computer Engineering, Nanyang Technological University, Singapore
[c] School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore
[d] Department of Computer Science and Technology, Xi'an Jiaotong University, China

ARTICLE INFO

Article history:
Received 4 February 2015
Received in revised form 3 August 2015
Accepted 7 September 2015
Available online 21 October 2015

Keywords:
Extreme Learning Machine
Singular value decomposition
Big data
Big dimensional data
Fast approximation method

ABSTRACT

Big dimensional data is a growing trend that is emerging in many real world contexts, extending from web mining, gene expression analysis, protein–protein interaction to high-frequency financial data. Nowadays, there is a growing consensus that the increasing dimensionality poses impeding effects on the performances of classifiers, which is termed as the "peaking phenomenon" in the field of machine intelligence. To address the issue, dimensionality reduction is commonly employed as a preprocessing step on the Big dimensional data before building the classifiers. In this paper, we propose an Extreme Learning Machine (ELM) approach for large-scale data analytic. In contrast to existing approaches, we embed hidden nodes that are designed using singular value decomposition (SVD) into the classical ELM. These SVD nodes in the hidden layer are shown to capture the underlying characteristics of the Big dimensional data well, exhibiting excellent generalization performances. The drawback of using SVD on the entire dataset, however, is the high computational complexity involved. To address this, a fast divide and conquer approximation scheme is introduced to maintain computational tractability on high volume data. The resultant algorithm proposed is labeled here as Fast Singular Value Decomposition-Hidden-nodes based Extreme Learning Machine or FSVD-H-ELM in short. In FSVD-H-ELM, instead of identifying the SVD hidden nodes directly from the entire dataset, SVD hidden nodes are derived from multiple random subsets of data sampled from the original dataset. Comprehensive experiments and comparisons are conducted to assess the FSVD-H-ELM against other *state-of-the-art* algorithms. The results obtained demonstrated the superior generalization performance and efficiency of the FSVD-H-ELM.

1. Introduction

The notion of "Big Dimensionality" was established recently in Zhai, Ong, and Tsang (2014) to place emphasis on the growing phenomenon of problem dimensionality that is observed across diverse scenarios and applications in the real world. When addressing Big Data, volume often comes to mind naturally and take immediate precedence, since it presents the challenges pertaining to the scalability issue. Also, this is what directly comes to our mind when we refer to the term "Big". However, it is worth highlighting that researchers have largely considered the "Big Instance Size" factor of "Volume" in the Big Data analytics community, which refers to the massive amounts of data that we continue to produce daily. In the context of data analytics, volume can be defined as a product of *instance size* and *dimensionality* of the data. The issue of "Big Dimensionality", however, has received much lesser attention. For more coverage on this relatively under-explored topic of "Big Dimensionality" wherein the explosion of features (variables) brings about new challenges to computational intelligence and Big data, the reader is referred to Zhai et al. (2014). *Big dimensional* data is a growing trend that is emerging in many real world contexts, extending from web mining, gene expression analysis, protein–protein interaction to high-frequency financial data Today, there is a growing consensus that the increasing dimensionality poses impeding effects on the performances of classifiers, which is termed as the "peaking phenomenon" in the field of machine intelligence (Sarunas & Vitalijus, 1980). Previous methods (Hanchuan, Fulmi, & Chris, 2005) to address the high dimensionality characteristics of data have mainly concentrated

* Corresponding author at: School of Computer, Xian University of Posts & Telecommunications, Shaanxi, China.
E-mail address: wanyu.deng@gmail.com (W.-Y. Deng).

http://dx.doi.org/10.1016/j.neunet.2015.09.003
0893-6080/

on preprocessing the data with dimensionality reduction methods, keeping only the salient features and discarding the non-informative ones before building the classifier. In this paper, we present an attempt to address both *Big dimensionality* and *Big instance size* simultaneously in the context of Big Data analytic by embedding Singular Value Decomposition Hidden nodes into the classical Extreme Learning Machine.

In spite of the extensive works on ELM research, existing efforts still cannot arrive at representations and models that are elegant for Big data. Like all methods known to date, ELM suffers from the effects of 'curse of dimensionality and instance size'. The performance of ELM has been shown to deteriorate substantially and becomes unstable in the face of Big data. Taking this cue, this paper considers an implicit incorporation of dimension reduction mechanism into the classical ELM to achieve high generalization performances. In particular, we present a Fast Singular Value Decomposition-Hidden-nodes based Extreme Learning Machine (or FSVD-H-ELM in short) to cope with *Big dimensional and Big instance* data. We embed SVD hidden nodes into the classical Extreme Learning Machine to arrive at SVD-H-ELM and provide theoretical proofs on the optimality of using SVD hidden nodes over random hidden nodes. Further, instead of identifying the SVD hidden nodes directly from the Big dataset, a fast divide and conquer approximation scheme is also introduced to maintain computational tractability and scalability by deriving SVD hidden nodes from multiple random subsets of data sampled from the original Big dataset, in order to arrive at the Fast SVD-H-ELM or FSVD-H-ELM. Comprehensive experiments conducted on commonly used Big dataset indicate that the SVD nodes in the hidden layer are able to capture the underlying characteristics of the Big dataset well, leading to excellent generalization performances and efficiencies.

The rest of the paper is organized as follows: Section 2 gives a brief overview of the classical ELM. Section 3 details the proposed Fast Singular Value Decomposition-Hidden-nodes based Extreme Learning Machine (or FSVD-H-ELM). In Section 4, A discussion on the feature learning ability of the fast divide-and-conquer scheme is made. In Section 5, we summarize the experimental studies on twelve commonly used benchmark problems. Section 6 summarizes the main conclusions.

2. Review of Extreme Learning Machine and large-scale data analytics

2.1. Extreme Learning Machine

Feedforward neural networks play key roles in data analytic and have been widely applied in many applications for its promising generalization ability (Chaturvedi, Ong, & Arumugam, 2015; Feng, Ong, Lim, & Tsang, 2014; Schaefer, Krawczyk, Celebi, & Iyatomi, 2014; Seah, Ong, & Tsang, 2013; Seah, Tsang, & Ong, 2012, 2013). However, popular learning techniques face some challenging issues such as intensive human intervene and slow learning speed (Bullinaria & AlYahya, 2014). Although many effective algorithms, such as the back-propagation are available, training a neural network with all parameters adjustable is usually of high computational burden (Rummelhart, 1986; Salama, Hassanien, & Revett, 2013). To overcome such issue, a useful learning scheme, the Extreme Learning Machine (ELM), was suggested in Huang, Zhu, and Siew (2006) for single layer feedforward neural networks and subsequently extended to different variations such as local connected structure (Huang, Bai, Kasun, & Vong, 2015) and multi hidden layers structure (Kasun, Zhou, Huang, & Vong, 2013). ELM and its variations (Bueno-Crespo, García-Laencina, & Sancho-Gómez, 2013; Cambria et al., 2013; Chen, Peng, Zhou, Li, & Pan, 2014; Fernandez-Delgado, Cernadas, Barro, Ribeiro, & Neves, 2014; Huang, Chen, & Siew, 2006; Huang, Ding, & Zhou, 2010; Huang, Zhou, Ding, & Zhang, 2012; Rong, Ong, Tan, & Zhu, 2008) have been successfully used in the fields as object recognition (Huang et al., 2015), terrain-based navigation (Kan, Lim, Ong, Tan, & Yeo, 2013) activity recognition (Deng, Zheng, & Wang, 2014), time series (Butcher, Verstraeten, Schrauwen, Day, & Haycock, 2013), security assessment (Xu, Dong, Zhao, Zhang, & Wong, 2012), written character recognition (Chacko, Krishnan, Raju, & Anto, 2012), face recognition (Mohammed, Minhas, Wu, & Sid-Ahmed, 2011), gene selection and cancer classification (Saraswathi, Sundaram, Sundararajan, Zimmermann, & Nilsen-Hamilton, 2011). In essence ELM is a learning scheme whose hidden nodes need not be tuned and can be randomly generated. Then the ELM transforms the training of the neural network into a linear problem where the output weights of ELMs can be analytically determined instead of being tuned (Luo, Vong, & Wong, 2014). For the sake of efficiency, in real applications they may be determined in different ways such as with or without iterations, with or without incremental implementations, etc. Xu etc. (Lin, Liu, Fang, & Xu, 2015; Liu, Lin, Fang, & Xu, 2015) studied the theoretical feasibility of ELM and proved that via suitable activation functions, such as polynomials, NadarayaWatson and sigmoid functions, the ELMs can attain the theoretical generalization bound of the neural networks with all connections adjusted. Before ELM, the similar idea has been adopted earlier in Jaeger (2001); Jaeger and Haas (2004) as the echo state network method (ESN), in Maass, Natschläger, and Markram (2002) as the liquid state machine (LSM), in Lowe (1989) as RBF network with random centers, in Schmidt, Kraaijveld, and Duin (1992) as the feedforward neural networks with random weights and in Pao (1989) as the random vector functional-link network (RVFL). The relationships between ELMs and these earlier work please refer to the literatures (Huang, 2015; Huang, Huang, Song, & You, 2015).

The output function of ELM with L hidden nodes for generalized SLFNs is:

$$f_L(\mathbf{x}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i h_i(\mathbf{x}) = \mathbf{h}(\mathbf{x})\boldsymbol{\beta} \tag{1}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_L]$ is the output weights between the hidden nodes to output nodes. $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \dots, h_L(\mathbf{x})]$ is the output vector of the hidden layer with respect to the input $\mathbf{x}$; $h_i(\mathbf{x}) = g(\mathbf{w}_i, b_i, \mathbf{x})$, $\mathbf{w}_i \in \mathbf{R}^d$, $b_i \in R$ is the output of the ith *random hidden node* where the hidden node parameters (e.g. $\mathbf{w}$ and b) need not be tuned, in particular, can be *randomly* generated according to any continuous sampling distribution probability. $g(\mathbf{w}_i, b_i, \mathbf{x})$ is a nonlinear piecewise continuous function satisfying ELM universal approximation capability theorems (Huang, Chen et al., 2006). For example, such nonlinear piecewise continuous function can be but are not limited to:

(1) Sigmoid function

$$g(\mathbf{w}, b, \mathbf{x}) = \frac{1}{1 + \exp(-\lambda(\mathbf{w} \cdot \mathbf{x} + b))}. \tag{2}$$

(2) Fourier function

$$g(\mathbf{w}, b, \mathbf{x}) = \sin(\mathbf{w} \cdot \mathbf{x} + b). \tag{3}$$

(3) Hardlimit function

$$g(\mathbf{w}, b, \mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} - b \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(4) Gaussian function

$$g(\mathbf{w}, b, \mathbf{x}) = \exp(-b\|\mathbf{x} - \mathbf{w}\|^2). \tag{5}$$