Big Data Research ••• (••••) •••–•••

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Tensor Decomposition Based Approach for Training Extreme Learning Machines

Nikhitha K. Nair, Asharaf S.

Indian Institute of Information Technology and Management-Kerala (IIITM-K), Thiruvananthapuram, India

ARTICLE INFO

Article history:
Received 30 November 2016
Received in revised form 4 June 2017
Accepted 4 July 2017
Available online xxxx

Keywords:
ELM
TENSOR
PARAFAC
TUCKER
ALS
HOSVD

ABSTRACT

Conventional Extreme Learning Machines utilize Moore–Penrose generalized pseudo-inverse to solve hidden layer activation matrix and perform analytical determination of output weights. Scalability is the major concern to be addressed in Extreme Learning Machines while dealing with large dataset. Motivated by these scalability concerns, this paper proposes a novel tensor decomposition based Extreme Learning Machine which utilize PARAFAC and TUCKER decomposition based techniques in a SPARK platform. This proposed Extreme Learning Machine achieve reduced training time and better accuracy when compared with a conventional Extreme Learning Machine.

1. Introduction

Extreme Learning Machine has become a significant research topic in the field of machine intelligence and big data analytics due to its vibrant characteristics such as extremely fast training rate and universal regression/classification and approximation capability. ELM tend to be an effective solution for the drawbacks of single hidden layer feed-forward neural networks. It has wide applications in domains such as face classification [1,2], image segmentation [1,3] and human action recognition [1,4].

Traditional feed-forward neural networks exhibit potentials such as (1) approximation of complex nonlinear mappings directly from input samples; (2) inefficient handling of large classes of phenomena using different classical parametric techniques [5]. These neural networks have much slower learning algorithms than desired.

Extreme Learning Machines exhibit important characteristics such as 1) introducing minimum error from training process. 2) requirement of smallest normalised weights in the entire process and providing perfect generalisation performance [5,7,10]. These machines reveal uniqueness property with minimum normalised least-squares solution of $H\beta = T$, which is $\hat{\beta} = H^{+}T$ [5,7,10,11]. The most expensive computation in Extreme Learning Machines exist in the process of computing Moore–Penrose generalised inverse of hidden layer matrix [12]. Some of the most popular methods for computing this matrix include singular value decomposition method, iterative method, orthogonalization method and orthogonal method [12–15]. When nonsingular matrices are to be handled, both iterative and orthogonalization method show poor performance [12,13]. The singular value decomposition method tend to be an accurate method. This method prone to be time-consuming when massive data are handled [12,13,16].

The CANDECOMP and TUCKER decomposition techniques are two major tensor decompositions used in the field of data mining, classification and big data analytics. In case of tensors, there does not exist a single generalized Singular Value Decomposition. Both these techniques are considered as extensions of Singular Value Decomposition to adapt higher dimensions [25].

This paper contribute to develop a Tensor decomposition based Extreme Learning Machine on a Apache Spark Platform, which overcome the shortcomings of conventional Extreme Learning Machines implemented on Apache Spark Platform, whose leaning ability is pathetic to huge dataset. PARAFAC decomposition based tensor (ELM-PARAFAC) and TUCKER decomposition based tensor (ELM-TUCKER) are the two proposed scalable tensor decomposition based Extreme Learning Machines. Utilizing these tensor techniques, factor matrices are generated from the training data tensor, which is then trained using Extreme Learning Machines. Therefore, training massive data are made effective with these tensor decomposition techniques.

The rest of this paper is formulated as follows. Section 2 discuss some of the work related to the proposed approach. Section 3 contribute a brief overview of standard Extreme Learning Machines

E-mail addresses: nikhitha.mphilcs3@iiitmk.ac.in (N.K. Nair), asharaf.s@iiitmk.ac.in (Asharaf S.).

http://dx.doi.org/10.1016/j.bdr.2017.07.002
2214-5796/

Please cite this article in press as: N.K. Nair and Asharaf S., Tensor Decomposition Based Approach for Training Extreme Learning Machines, Big Data Res. (2017), http://dx.doi.org/10.1016/j.bdr.2017.07.002

and its algorithm, which is commonly used in classification and regression tasks. Section 4 discuss a brief overview of tensors and its basic preliminaries. Section 5 reflect the core part of this work, which discuss the real problem faced by standard Extreme Learning Machines, introduce the methodology adopted for solving the threat and implementation of proposed algorithm. Section 6 compare the performance of standard Extreme Learning Machines with that of proposed Tensorized Extreme Learning Machines. Conclusions and future scope of this work is drawn in Section 7.

2. Related work

Till now, Support Vector Machines are extensively being used for classification. These machines exhibit learning features such as: a) Mapping training data to a higher dimensional feature space with the help of a nonlinear feature mapping function b) Solution for maximum margin separation of two different classes determined using standard optimizing method. There is intensive computational complexity in the process of training support vector machines since quadratic programming is utilized [8].

The neural networks are normally trained with the help of a finite training dataset [5,7]. Huang and Babri [5,7,9] show that a single-hidden layer feed-forward neural network is capable to learn 'N' distinct observations with atmost 'N' hidden nodes using any nonlinear activation functions. These feed-forward neural networks were mostly used in combination with back-propagation (BP) algorithms for optimizing different parameters of neural networks. These algorithms belong to the criteria of gradient based methods. The main issues addressed by these methods include low convergence rate and problem of not achieving the proper global minimum [6]. The backpropogation algorithms deal with several issues such as: 1) Learning algorithm achieves low convergence rate, when learning rate is too small. 2) Learning algorithm stops easily at the local minimum, when the learning algorithm is located much far above the specified global minimum. 3) The traditional neural networks become over-trained by using these algorithms and achieve worse generalisation performance [7].

Infinitely differentiable activation functions allow random assigning of hidden layer biases and input weights in single-hidden layer feed-forward neural network. After randomly choosing weights in the input layer and biases in the hidden layer, the entire single layer feed-forward neural networks are treated as a linear system. The analytical determination of output weights of the neural networks are performed using simple inverse operation in the output matrices of the hidden layer [5].

These all concepts are combined together to form a simple generalized learning algorithm named Extreme Learning Machines (ELM). These Extreme Learning Machines have proven to exhibit thousand of times faster training speed when compared with traditional feed-forward network algorithms. It also achieve better performance in generalization [5,7].

The computation of Moore–Penrose generalised inverse of hidden layer matrix tend to be the most expensive and tedious computation while dealing with Extreme Learning Machines. According to Courrieu, fast computation of Moore–Penrose generalised inverse matrices are achieved using a full rank Cholesky factorization method [12,13]. Tensor product matrix Extreme Learning Machine (TPM-ELM) is another approach stated for computing Moore–Penrose pseudo-inverse of a large matrix [12,16,17]. Different variants of Extreme Learning Machines exist to overcome their shortcomings, which includes pruning ELM, two-stage ELM, voting-based ELM, fully complex ELM, incremental ELM, error-minimized ELM and symmetric ELM [39].

Tensor decomposition techniques tend to be among the emerging tools to analyze multi-dimensional aspects of data in many engineering and scientific fields such as machine learning, signal processing and chemometrics [18–22].

For improved data analysis, data with multiple aspects or modalities are collected, correlated and organized as tensors. There is a loss of internal structure information, when matrix factorization techniques are applied to matricized data. Tensor decomposition techniques provide multiple perspective stereoscopic view rather than flatten view provided by matrix factorization methodologies [18].

In matrix, there is a quadratic nature of increase in number of elements as product of rows and columns. In third-order tensors, number of elements in it increases cubically, as product of columns, rows and tubes. Thus for handling large tensors, there is always a need for tensor decomposition techniques [23].

Over these years, tensors are used for various purposes in the field of data mining and analytics [24]. There are different tensor decomposition approaches that are applied to multi-dimensional data. The higher-order generalizations of Singular Value Decomposition include Tucker decomposition, PARAFAC decomposition (parallel factor analysis) or CANDECOMP (canonical decomposition), Higher-Order Singular Value Decomposition (HO-SVD) and Higher-Order Orthogonal Iteration (HOOI) [23].

PARAFAC decomposition is used for decomposing a large tensor into a sum of rank-one tensors for finding the latent factors. TUCKER decomposition is mostly used for tensor compression as well as finding relations that exist between the latent factors [18, 25–29]. Since there exist an interaction among all pairs of factors, TUCKER decomposition tends to be considered as a generalised version of CANDECOMP decomposition [26].

The tensorization refers to the process of utilizing lower-dimensional original data for creating data tensors that have practical existence [30]. In case of data analysis, higher-order tensors are obtained by the process of tensorization of large-scale matrices or vectors. Nowadays, tensor decomposition techniques are applied for low tensor rank approximations. This works well for big data analytics [31].

Dealing with tensors and their decomposition approaches, intermediate data explosion problem is one of the major issues faced when handling large tensors [25–27]. GIGATENSOR, a scalable approach to tensor decomposition is capable of handling tera-scale tensors using MapReduce framework and their open source implementation in Hadoop platform [27].

The Alternating Least Square (ALS) approach is one of the popular methods used in PARAFAC and TUCKER tensor decomposition. While considering three variables and fixing two of them, the pseudo-inverse calculated by this method gives the solution of minimal norm for those variables [27–29].

The big data analytics always tends to adapt novel technologies that has the capability to handle immense datasets with high speed. Multiway analysis through tensor networks and tensor decompositions prone to be such emerging technologies to handle multidimensional big datasets. The idea behind these tensor decompositions are generating factor matrices by decomposing large training data tensors while interconnected lower-order tensors represent higher-order tensors using tensor networks. The challenge behind these tensor decompositions include finding an approach to analyze large-scale multiway data and perform different tasks such as classification, blind-source separation and compression [32].

Tensor decompositions have also shown immense applications in the area of multi-task learning. In this, multiple related tasks are learned simultaneously so that re-use of knowledge segregated from each task takes place. The multi-task ideas which is based on matrix factorization are generalised to tensor factorization, so that knowledge is shared flexibly in connected and convolutional deep neural networks [33].

Please cite this article in press as: N.K. Nair and Asharaf S., Tensor Decomposition Based Approach for Training Extreme Learning Machines, Big Data Res. (2017), http://dx.doi.org/10.1016/j.bdr.2017.07.002