Computers and Electrical Engineering 000 (2016) 1–11

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Symmetric Matrix-based Predictive Classifier for Big Data computation and information sharing in Cloud ☆

V. Vennila*, A. Rajiv Kannan

Department of Computer Science and Engineering, K.S.R. College of Engineering, Tiruchengode, Tamilnadu, India

ARTICLE INFO

Article history:
Received 17 September 2015
Revised 19 May 2016
Accepted 20 May 2016
Available online xxx

Keywords:
Big Data
Cloud computing
MapReduce function
Cross-Validation
Predictive Bayes Classifier
Symmetric Matrix

ABSTRACT

Big Data requires real-time data-intensive processing that runs on high-performance clusters. In Big Data applications, data collection has grown exponentially. It is highly complex to extract, identify and transmit information using existing software tools. Big Data applications increase the gaps in performance between legitimate classifiers. In this paper, a Parallel Symmetric Matrix-based Predictive Bayes Classifier (PSM-PBC) model is developed for efficient Big Data computation and information sharing in Cloud environment. Initially, the Tridiagonal Symmetric Matrix is constructed on distributed Big Data in parallel. This approach enables an increase in the rate of data computation using a Householder transformation. A Cross-Validated Bayes Classifier then evaluates real-value diagonal search data to improve the prediction rate. Finally, the MapReduce function on Bayes Classes provides efficient predictive analytics regarding Big Data. The experimental evaluations are conducted with Amazon EC2 Cloud Big Data sets and exhibit improvement of the prediction rate by 10.55% along with a reduction in computation time by 40.93% compared to state-of-the-art methods.

1. Introduction

In the use of Big Data, a technical barrier is encountered because moving data across different locations is highly expensive, even if a large main memory is used to hold all data for computing. The size or complexity of the Big Data includes transaction and interaction data sets that exceed regular technical capability in capturing, managing and processing data at a reasonable cost. The main goal in Big Data application is to explore the large volumes of data and extract useful information or knowledge for future actions.

Cloud computing is a parallel distributed computing system that has become frequently used for Big Data analytics. Data Mining with Big Data (DM-BD) was designed in [1] using the Heterogeneous Autonomous Complex Evolving theorem to improve the security and privacy in Cloud environment. Another prototype method, called FlexAnalytics [2], was designed to increase data transmission bandwidth. However, neither method addressed the issues related to space and time complexity.

In this paper, an efficient PSM-PBC model is presented for Big Data computation and information sharing in Cloud environment. The flow of the PSM-PBC model includes three processes. First, the Tridiagonal Symmetric Matrix is constructed on distributed Big Data in parallel. This approach enables faster computation for data extraction and information sharing across the Cloud paradigm using a Householder transformation. The Cross-Validated Bayes Classifier model is used in the

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. R. Tolosana-Calasanz.

* Corresponding author.
E-mail address: v.vennila@ksrce.ac.in (V. Vennila).

http://dx.doi.org/10.1016/j.compeleceng.2016.05.018
0045-7906/

second step of the PSM-PBC model to evaluate real-value diagonal search data for the corresponding query results obtained from each user request, which helps to increase the prediction rate. Finally, the MapReduce function is improved with Bayes Classes to provide predictive analytics about Big Data, for better computation and information sharing.

The rest of the paper is organized as follows: Section 2 discusses related works. In Section 3, the Parallel Symmetric Matrix-based Predictive Bayes Classifier model is described with the help of diagrams. Section 4 presents the results of the evaluation of the proposed PSM-PBC model. Finally, Section 5 provides the study's conclusions.

2. Related works

There are many research challenges in Big Data management for analytics in Cloud environment. Performance analysis for Big Data in Cloud was presented in [3] to reduce the gap between qualitative and quantitative representations. A review of machine learning methods to improve the classification accuracy was presented in [4].

With the surge of social media and the use of data mining techniques, the amount of Big Data has increased rapidly, along with tremendous developments in information technology. The parallel matrix-based method [5] was applied to process large-scale data using MapReduce. MapReduce [6] was applied across distributed data centers to improve reliability. Another dynamic placement strategy [7] applies a heterogeneous Hadoop cluster to reduce the time required for extracting information. Another parallel technique measured Key Performance Indicators [8], resulting in the improvement of the data scan sharing scheme. However, these works did not focus on information sharing in Big Data.

Patient similarity over MapReduce architecture was introduced in [9] to reduce the execution time of retrieving the patient similarity information. Probabilistic generative classifiers were applied [10] to improve the similarity measure obtained for several users using Mahalanobis distance.

Sentiment information from social Big Data was analyzed in [11] for load balancing of memory using a parallel Hadoop Distributed File System (HDFS). With the objective of providing optimization in time [12], a query controller was applied with respect to Big Data analytics. An Extreme Learning Machine (ELM) tree model was designed to minimize the time required to compute heuristics for Big Data classification [13]. However, the prediction rate was not focused in terms of scalable user requests. In PSM-PBC model, the prediction rate is improved by applying the Cross-Validated Bayes Classifier model.

With the increasing amount of data being generated in various fields, complexity issues arise when handling Big Data. A classification algorithm was introduced in [14] to address issues related to execution time. Extraction of information from a large volume of data, the Internet of things [15], was discussed. A review of Big Data in healthcare was presented in [16], focusing on the Bayes Classification model. High dimension [17] and scalable sensor data storage [18] for Big Data were analyzed using an intelligent vehicle model. In [19], a task-oriented resource allocation model was developed in a Cloud where resource allocation tasks are ranked by the pairwise comparison matrix. In [20], an analysis of a huge collection of DMKD literature was performed, to provide a comprehensive picture of current DMKD research and to classify these research activities into high-level categories using a grounded theory approach. In [21], technical issues and surveys of existing works on analytics capabilities for Big Data on the Cloud were highlighted. In [22], the MPI/OpenMP on Beowulf and Apache Spark on Hadoop, which target iterative algorithms, were designed through in-memory computing for Big Data evaluations in Cloud. An effective control with an environmental sensing approach was developed in [23] for Big Data processing in Cloud environment. An efficient peer-to-peer file sharing system was developed in [24] to achieve minimum delay and communication overhead. In [25], Big Data were analyzed by reducing the size of the data and extracting the information in Cloud environment.

Based on the above-mentioned methods, an effective PSM-PBC model is developed to improve Big Data computation in Cloud environment.

3. Parallel Symmetric Matrix-based Predictive Bayes Classifier

A PSM-PBC model is developed for efficient Big Data computation and information sharing in Cloud environment. In the PSM-PBC model, the Tridiagonal Symmetric Matrix is used to enable faster computation for data extraction and information sharing across the Cloud paradigm, which in turn improves the search accuracy. The Cross-Validated Bayes Classifier model in PSM-PBC evaluates real-value diagonal search data for the corresponding user query, aiming to improve the prediction rate. The MapReduce function on Bayes Classes is applied in the PSM-PBC model to reduce space complexity and computational complexity of Big Data. An overview of the PSM-PBC model is shown in Fig. 1.

In Fig. 1, the block diagram of the PSM-PBC model is split into three parts: (i) application of the Tridiagonal Symmetric Matrix to improve the search accuracy on Big Data; (ii) design of the Cross-Validated Bayes Classifier to improve the prediction rate during classification; and (iii) deployment of the MapReduce function with the objective of reducing the space complexity. The detailed description of the PSM-PBC model is presented in the next sections.

3.1. Construction of the Tridiagonal Symmetric Matrix

The first step in the design of the PSM-PBC model is the construction of the Tridiagonal Symmetric Matrix. Raw data collected from users are examined to identify patterns and correlations that may not be obtained during the initial stage

Please cite this article as: V. Vennila, A. Rajiv Kannan, Symmetric Matrix-based predictive classifier for Big Data computation and information sharing in Cloud, Computers and Electrical Engineering (2016), http://dx.doi.org/10.1016/j.compeleceng.2016.05.018