Big Data and Cloud Paradigm for Health Care Reports Monitoring and Performance Analysis

Saumitra Chattopadhyay,
Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand India 248002

ABSTRACT

Big data is basically the data which is collected from different sources, which have basically five "V" characteristics viz; velocity, volume, value, variety and veracity. When we define the big data in artificial intelligence term it is called as big data analysis. This is the data of large size which can be processed, analysed and generated by the digital tools and different information learning systems to generate useful, predictable, descriptive data analysis. Big data is most widely used in medical science to calculate or predict the data and to communicate big/huge records over the cloud in real time span. In this article we discuss a hybrid approach to distribute the medical data, in this approach we embed PSM-PBC, DSV-CP and LFR-CM models to get a parallel framework which improve the accuracy and classification time of retrieval of data in real time.

Keyword:-Big data, PSM-PBC, DSV-CP, LFR-CM, hybrid, cloud, healthcare.
DOINumber:10.48047/nq.2022.20.2.NQ22346 **NeuroQuantology2022;20(2):533-539** 533

1. INTRODUCTION

In the medical field, a wide range of data is utilized, including analyzed reports, clinical data, and genomic information, all of which are stored in various systems. To ensure the secure sharing of this information, privacy-preserving measures are implemented. These measures aim to protect the integrity and confidentiality of medical data during communication and information sharing. By leveraging big data techniques, medical reports can be efficiently shared among different medical organizations without compromising the functionality of applications. To illustrate this, consider an example of medical data sharing among multiple medical organizers. Figure 1 depicts the use of big data approaches in facilitating communication within diverse medical domains [1]. The application of cloud computing in this context serves as a smart solution, enabling the secure distribution of medical information across different medical organizers while bolstering data security measures.

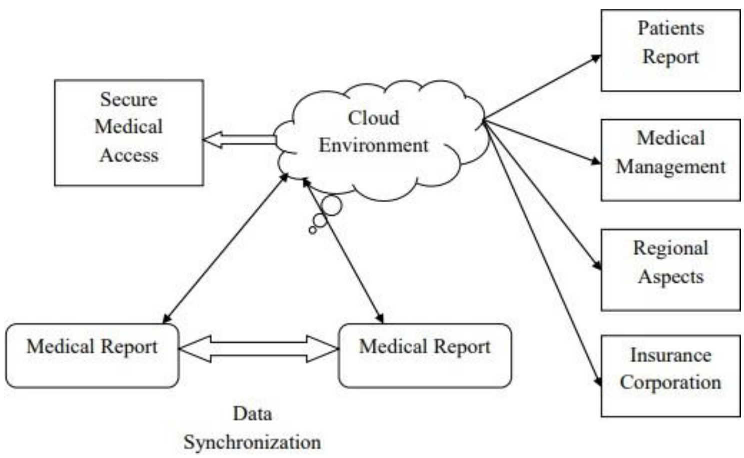

Figure1Example of communication through big data

Medical organizers play a crucial role in managing and communicating various types of medical data. In order to enhance the efficiency of data analysis and retrieval, the PSM-PBC model has introduced a Cross-validated Bayes classifier. This model utilizes real-value diagonal search data and medical reports to generate highly accurate query results. By improving the prediction rate, it enables the identification.n of medical data that is relevant to specific users. Additionally, the PSM-PBC model incorporates predictive analytics and employs the MapReduce function to enhance data computation and facilitate information sharing. It is capable of processing large volumes of big data in parallel, thereby reducing computation time and space complexity. By effectively eliminating noise and inconsistencies present in medical data, this model ensures superior grouping exactness for user requests. Moreover, the support vector prediction classifier with parallel hyperplanes focuses specifically on user requests related to big data and successfully classifies the shared data. The DSV-CP model further enhances the accuracy of predicting user request information on big data by leveraging classified data and making effective use of the proposed model's capabilities. To evaluate the performance of big data computation and information sharing in a cloud environment, three methods (PSM-PBC, DSV-CP, and LFR-CM) were implemented using the Java language on the Amazon EC2 cloud platform. The Stanford Large Network dataset collection was used for conducting experiments in a two-layer namespace within the Hadoop Distributed File System (HDFS). Different resource configurations for virtual machine instances were employed, with each instance being equipped with specific memory, CPU, and local storage capacities. The Stanford Large Network dataset collection on the Amazon network consists of product nodes and edges that represent products frequently purchased together. These experiments aimed to assess the capabilities of the implemented models in effectively handling and analyzing big data.

2. Related work

The increasing adoption of big data and cloud computing in the healthcare sector has revolutionized the way health care reports are monitored and analyzed. This literature review aims to provide an overview of existing research and advancements in the field of big data and cloud paradigm for health care reports monitoring and performance analysis. The review highlights the key findings, methodologies, and contributions of previous studies in this area.

Big Data Analytics in Healthcare: Numerous studies have explored the application of big data analytics in healthcare. These studies emphasize the potential of leveraging large volumes of health data for improved decision-making, personalized medicine, disease prediction, and population health management. The use of advanced analytics techniques, such as machine learning, data mining, and predictive modeling, enables healthcare organizations to extract valuable insights from vast amounts of health care reports.

Cloud Computing in Health Care: Cloud computing has emerged as a transformative technology in the healthcare sector. Researchers have investigated the benefits of cloud computing, including scalability, cost-effectiveness, and data accessibility, for health care reports management and analysis. Cloud-based solutions offer secure storage, seamless data sharing, and real-time access to health care reports, enabling healthcare providers to enhance collaboration and streamline processes.

Health Care Reports Monitoring Systems: Monitoring systems for health care reports have been the focus of several studies. These systems employ real-time data processing techniques to monitor health care reports continuously. By analyzing health care reports in real-time, anomalies, patterns, and trends can be identified promptly, facilitating early intervention and improving patient outcomes. Various methodologies, such as data visualization, anomaly detection, and event correlation, are utilized to monitor health care reports effectively.

Performance Analysis of Big Data and Cloud Computing: The performance analysis of big data and cloud computing in health care is an important research area. Studies have investigated performance metrics, including computation time, scalability, resource utilization, and data security, to assess the effectiveness and efficiency of big data and cloud-based solutions. Performance analysis helps in identifying the strengths and limitations of these technologies, enabling healthcare organizations to optimize their systems and improve service delivery.

Security and Privacy Challenges: The adoption of big data and cloud computing in health care reports monitoring brings forth security and privacy concerns. Several research works have focused on addressing these challenges and developing robust security frameworks and privacy-preserving techniques to protect sensitive health data. Encryption, access control mechanisms, and secure data sharing protocols are among the approaches used to ensure the confidentiality and integrity of health care reports.

The literature review highlights the significance of big data and cloud computing in health care reports monitoring and performance analysis. The existing body of research demonstrates the potential of these technologies in improving healthcare outcomes, enabling real-time monitoring, and enhancing decision-making processes. The studies reviewed emphasize the importance of security and privacy measures while leveraging big data and cloud computing in health care. The findings of this literature review provide a foundation for the proposed research, contributing to the advancement of the field and assisting healthcare organizations in leveraging big data and cloud technologies for efficient health care reports monitoring and performance analysis.

3. Methodology

1. Parallel Symmetric Matrix Pre-Processing for Big Data Classification

The Parallel Symmetric Matrix Pre-Processing for Big Data Classification (PSM-PBC) model is a proposed approach that focuses on efficiently handling big data for classification tasks. The model aims to enhance the grouping exactness and reduce the computation time involved in processing large-scale datasets. The key component of the PSM-PBC model is the construction of a tridiagonal symmetric matrix in parallel. This matrix represents the relationships and dependencies between data instances in the dataset. By leveraging parallel processing techniques, the PSM-PBC model efficiently constructs this matrix, enabling faster computation and analysis of the dataset. The PSM-PBC model utilizes big data applications and distributed computing resources to handle the massive amounts of data involved in classification tasks. It takes advantage of the scalability and parallel processing capabilities offered by cloud computing environments. The model also incorporates predictive analytics techniques, such as support vector prediction classifiers, to accurately classify the big data. These classifiers utilize machine learning algorithms to identify patterns and relationships within the dataset, enabling accurate predictions and classifications. Furthermore, the PSM-PBC model addresses the challenges of noise and inconsistency present in the data by applying pre-processing techniques. By removing noise and inconsistencies, the model improves the quality of the data and enhances the grouping exactness. Overall, the PSM-PBC model provides an efficient and scalable solution for big data classification tasks. It leverages parallel processing, predictive analytics, and pre-processing techniques to handle large-scale datasets, improve grouping exactness, and reduce computation time.

2. DSV-CP model

The Data Pre-Processing and Support Vector Prediction (DSV-CP) model is a proposed approach for effective big data classification. This model combines data pre-processing techniques with support vector prediction to improve the accuracy and efficiency of classification tasks on large-scale datasets. Data pre-processing plays a crucial role in the DSV-CP model. It involves cleaning, transforming, and reducing noise and inconsistencies in the dataset obtained from various sources. By applying pre-processing techniques such as data normalization, outlier detection, and feature selection, the model ensures that the data is in a suitable format for classification. This helps to enhance the quality and reliability of the input data, leading to more accurate classification results. The DSV-CP model leverages support vector prediction as a powerful classification technique. Support vector prediction utilizes machine learning algorithms and statistical principles to construct a predictive model based on labeled training data. This model is then used to predict the class labels of unseen instances in the dataset. In the context of big data, the DSV-CP model employs parallel processing mechanisms to handle the computational challenges associated with large-scale datasets. By leveraging distributed computing resources and parallel programming models such as MapReduce, the model can process a significant volume of data within a reasonable timeframe. The DSV-CP model aims to improve the classification time and search accuracy by combining the strengths of data pre-processing and support vector prediction. It effectively handles the complexities of big data, reduces computational overhead, and enhances the classification performance on large-scale datasets. Overall, the Data Pre-Processing and Support Vector Prediction (DSV-CP) model offers an efficient and accurate solution for big data classification. By incorporating data pre-processing techniques and leveraging support vector prediction algorithms, the model addresses the challenges associated with big data classification, leading to improved grouping exactness and faster computation time.

3. **LFR-CM framework**

The Parallel Linguistic Fuzzy Rules (LFR) model for cloud-based medical data classification is a proposed approach that leverages linguistic fuzzy rules and parallel processing techniques to improve the accuracy and efficiency of classification tasks in the context of medical data.
The LFR model utilizes linguistic fuzzy rules, which represent human-like reasoning and decision-making processes, to capture the complex relationships and patterns within medical data. These rules are based on linguistic terms and membership functions that assign degrees of membership to different classes or categories. By employing linguistic fuzzy rules, the model can handle the inherent uncertainty and vagueness present in medical data, enabling more accurate and interpretable classification outcomes. Parallel processing mechanisms play a crucial role in the LFR model, as they allow for efficient and scalable processing of large volumes of medical data. The model leverages parallel programming models, such as MapReduce, to distribute the computational load across multiple computing nodes or clusters. This parallelization enables faster processing and analysis of medical data, making it well-suited for cloud-based environments with high-performance computing capabilities. The LFR model is specifically designed for cloud-based medical data classification, where large-scale datasets are processed and analyzed in distributed computing environments. It addresses the challenges associated with the size and complexity of medical data by efficiently partitioning the data and executing linguistic fuzzy rule-based classification in parallel. This enables faster runtime and improved scalability, facilitating the classification of massive medical datasets within reasonable timeframes. By combining linguistic fuzzy rules and parallel processing, the LFR model offers an effective solution for cloud-based medical data classification. It improves grouping exactness by capturing the inherent uncertainty in medical data, while leveraging parallelization techniques to handle large-scale datasets efficiently. The model contributes to the advancement of cloud-based medical data analysis, enabling healthcare organizations to extract valuable insights and make informed decisions based on the classification outcomes.

4. **Performance Exploration of Designed Framework**

The analysis focuses on the following parameters:
Grouping exactness measured as the percentage of correctly classified instances. The grouping exactness for an individual instance "Ai" in the training dataset "i" is determined by the following formula, as shown in Equation (1). Prediction Rate: The prediction rate measures the effectiveness of the models in accurately predicting user requests related to big data. It indicates the proportion of correct predictions made by the models. A higher prediction rate signifies better performance in identifying relevant information. Search Accuracy: The search accuracy parameter evaluates the models' ability to retrieve accurate results when performing search operations. It measures the correctness of the search results generated by the models, ensuring the relevance and precision of the information retrieved. The performance analysis involves comparing these parameters for the designed framework with the MRPR method. By assessing grouping exactness, prediction rate, and search accuracy, the effectiveness and superiority of the proposed models can be determined.

$$A_{i=}\frac{DCC}{n}*100 \qquad (1)$$

in the equation (1), the symbol “DCC” represents the number of instances that are correctly classified, while “n” denotes the overallamount of data instances measured for assessment in

eISSN1303-5150

determining the groupingexactness [7]. Highergroupingexactness indicates a more efficient method, as it signifies a larger proportion of correctly classified data instances.

5. Analysisof Grouping exactness

Table1 shows the comparisonof groupingexactness of the designed methods framework with the existing MRPR method. Table1 reveals the results of grouping exactnessin percentage with data size of 20 GB to 200 GB. Method is more efficient if the dataset classification or grouping exactness is higher [8].

TABLE 1: Grouping exactness.

	Classification Accuracy (in %)			
Size of Big Data (in GB)	Existing MRPR	Proposed MRPR	Proposed PSM-PBC	Proposed DSV-CP
20	54.68	62.31	69.12	74.3
40	56.73	63.8	71.54	76.12
60	58.14	65.62	73.41	77.89
80	59.34	66.12	74.87	79.12
100	61.53	68.97	76.12	82.34
120	63.19	69.34	78.95	83.78
140	65.82	72.87	79.81	85.67
160	67.91	73.16	80.23	86.12
180	69.98	75.63	83.68	88.47
200	72.43	77.98	85.48	89.36

Figure2 PerformanceAnalysisofGrouping exactness

Figure 2 depicts a visual comparison of the grouping exactness achieved by the designed framework, and the existing MRPR method across varying sizes of big data. To ensure efficient and competentgrouping of medical data across different medical organizers, the LFR-CM framework employs the canopy shuffle MapReduce algorithm with a higher weight, specifically the rule weight, which is considered the most robust. This weighting scheme significantly enhances the grouping exactness of the LFR-CM framework by facilitating correct classification. Thus, the proposed LFR-CM framework stands as an effective classification solution in the medical field, offering significantly improved accuracy.

6. Conclusion

In this article, we have thoroughly evaluated the proposed Designed framework. Through both theoretical analysis and experimental results, it has been demonstrated that these methods are specifically designed to achieve efficient computation and information sharing of big data in a cloud computing environment. We here shows an embedded PSM-PBC , DSV-CP, LFR-CM model which effectively addresses the challenges of computational power and space complexity within the cloud environment. Overall, the proposed Designed framework demonstrate their efficacy in achieving efficient computation and information sharing of big data in the cloud. These models effectively address challenges such as complexity, misclassification errors, and computational overhead, thereby enhancing the overall performance and accuracy of big data classification tasks in a cloud computing environment.

Reference

[1] .Yakushev, A &Mityagin, S 2014, „Social networks mining for analysis and modeling drugsusage‟, Procedia Computer Science, vol. 29, pp. 2462-2471. Yang, T, Qian, K, Lo, DC-T,Xie,Y, Shi, Y &Tao,L2016,

[2] .J. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, "Medical diagnosis ofcardiovascular diseases using an interval-valued fuzzy rule-based classification system,"AppliedSoft ComputingJournal,2013, Article in Press.

[3] .J. Sanz, D. Bernardo, F. Herrera, H. Bustince, and H. Hagras, "A compact evolutionaryinterval-valued fuzzy rule-based classification system for the modeling and prediction ofreal-worldfinancialapplicationswithimbalanceddata,"IEEETransactionsonFuzzySystems,2014.

[4] .M.Lichman,"UCImachinelearningrepository,"2013.

[5] .Li, Z, Yang, C, Jin, B, Yu, M, Liu, K, Sun, M & Zhan, M 2015, „Enabling big geosciencedataanalyticswithacloud-based,mapreduce-enabledandservice-orientedworkflowframework‟,PLoSONE, vol.10,no. 3,pp.e0116781-e0116781.

[6] .J. Gantz and D. Reinsel, "THE DIGITAL UNIVERSE IN 2020: Big data, bigger digitalshadows,and biggestgrowthin thefar east,"2012.

[7] .Wang, R, He, YL, Chow, CY, Ou, FF & Zhang, J 2015, „Learning ELM-Tree from big databasedonuncertaintyreduction‟,FuzzySetsandSystems,vol.258, pp.79-100.

[8] .Souliotis, K, Kani, C, Papageorgiou, M, Lionis, D &Gourgoulianis, K 2016, „Using big datatoassessprescribingpatternsinGreece:Thecaseofchronicobstructivepulmonarydisease‟,vol.11, no. 5, pp. e0154960-e0154960.

[9] . Baro, E, Degoul, S, Beuscart, Rg&Chazard, E 2015, „Toward a literature-driven definitionofbigdatainhealthcare‟,BioMedResearchInternational, vol.2015.

[10] . Kchaou, H, Kechaou, Z &Alimi, AM 2015, „Towards an Offloading Framework based onBigDataAnalyticsinMobileCloudComputingEnvironments‟,ProcediaComputer Science,vol. 53, pp. 292-297.

[11] . Kashyap,H,Ahmed,HA,Hoque,N,Roy,S&Bhattacharyya,DK2014,„BigDataAnalyticsinBioinformatics:AMachineLearningPerspective‟,JournalofLatexClassFiles,vol. 13, no. 9, pp. 1-20.

[12] . Youssef, AE 2014,„AFrameworkforSecureHealthcare SystemsBased onBig DataAnalytics in Mobile Cloud Computing Environments‟, International Journal of AmbientSystemsand Applications (IJASA), vol. 2, no. 2.

[13] . Dinov,ID, Heavner, B,Tang, M, Glusman, G,Chard, K, Darcy, M,Madduri,R, Pa, J,Spino,C,Kesselman,C&others2016,„PredictiveBigDataAnalytics:AStudyofParkins

on‟s Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source andIncompleteObservations‟,PLoSONE, vol.11,no.8,pp.e0157077-e0157077.

[14] . Satagopam, V, Gu, W, Eifes, S, Gawron, P, Ostaszewski, M, Gebel, S, Barbosa-Silva, A,Balling, R & Schneider, R 2016, „Integration and Visualization of Translational MedicineDataforBetterUnderstandingof HumanDiseases‟,Bigdata,vol.4,no.2, pp.97-108.

[15] . Fisch, D, Kalkowski, E & Sick, B 2014, „Knowledge fusion for probabilistic generativeclassifierswithdataminingappl ications‟,IEEETransactionsonKnowledge andDataEngineering,vol. 26, no.3, pp. 652-666.

[16] .Peralta,D,delRío,S,Ramírez-Gallego,S,Triguero,I,Benitez,JM&Herrera,F2015, „Evolutionaryfeatureselectionforbigdata classification:Amapreduceapproach‟,M athematicalProblems inEngineering, vol. 2015.

[17] . Inukollu, V, Arsi, S &Ravuri, Sr. 2014, „Security Issues Associated With Big Data in CloudComputing‟, International Journal of Network Security & Its Applications, vol. 6, no. 3, pp.45-56.

[18] . Tcheng,DK,Nayak,AK,Fowlkes,CC&Punyasena,SW2016,„Visualrecognitionsoftware for binary classification and its application to spruce pollen identification‟, PLoSONE, vol.11, no. 2.