ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Efficient E-health system using Deep Sparse Mining in Big data

Gousia Ahmed*, Prof. Deepa Amne**

*M.Tech Student, Dept. of CSE, BIT Ballarpur, Gondwana University, Maharashtra, India

** Assistant Professor, Dept. of CSE, BIT Ballarpur, Gondwana University, Maharashtra, India

ABSTRACT: Big Data is transforming healthcare, business, as e-Health disease becomes one of key driving factors during the innovation process. Look into BDeHS (Big Data e-Health Service) to fulfil the Big Data applications in the e-Health service domain. Existing Data Mining technologies such cannot be simply applied to e-Health services directly.Whilst manydimensions of big data still present issues in its use and adoption, such as managing the volume, variety, velocity, veracity, and value, the accuracy, integrity, and semantic interpretation are of greater concern in clinicalapplicationIn this paper we explain why the existing Big Data technologies such as Hadoop,MapReduce cannot be simply applied to e-Health services directly. Our design of the BDeHS for heart disease that supplies data operation management capabilities and e-Health meaningful usages.

KEYWORDS: Big Data Technologies; e-Health Solutions; Big Data as a Service; Data mining; Hadoop map reduce.

I. INTRODUCTION

The delivery of professional health care services required a patient to be physically collocated with a medical provider in earlier days. Patients living in regions with inadequate health services had to either travel long distances for care or accept substandard medical services. Those seeking access to medical literature and educational health resources were relegated to visiting specialized medical libraries, if they could access these resources at all. Patient data stored in the files of a primary care physician were not readily accessible by specialists, pharmacies, insurance companies, hospitals or labs. Each healthcare provider housed individual data and imaging snapshots of patients. There were physical, economic, and knowledge barriers to receiving optimal healthcare services. Big data that concerns a large volume complex and growing data sets with several autonomous sources. Big data is rapidly growing in all science and engineering .Big data is characterized as (vvvvc) (ie) volume, velocity, variable, variety and complexity. In the recent years big data is growing in all science and engineering here we going to implement BDeHS for heart disease.

The electronic medical record (EMR) initiative has resulted data streams from all types of patients at the hospital, insurance companies and doctor's office. A single patient stay generates thousands of data elements, including diagnoses, procedures, medications, medical supplies, digital image, lab results and billing. These need to be validated, processed and integrated into a large data pools to enable meaningful analysis. Multiplying this by all the patient stays across the health processing systems and combining it with the large number of points where data is generated and stored and the scope of the big data challenge begins to emerge. Hadoop is another framework for dealing with big data. It provides a set of general primitives for doing batch processing. It supports the running of applications on large clusters of commodity hardware. The Hadoop framework transparently provides both reliability and data motion to applications.

II. RELATED WORK

In the last few years, technological improvement opens newpossibilities to healthcare and medicine practice, but carriers someinherit risks and leaves decision makers with numerousunanswered questions about quality, security and other importantmatters. Some surveys and approaches have showed theimportance of data quality of end-users, particularly in healthcare domain.E-Health monitoring applications have some particularitiesconcerning the importance on data quality. On the one hand, successful healthcare delivery and planning strongly rely on data(e.g. sensed data, diagnosis, administration information); thehigher quality of the data, the better will be the patient assistance.On the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

other hand, these applications are also particularly exposedto a contextual environment (i.e. patients' mobility,communication technologies performance, informationheterogeneity…) that has an important impact on informationmanagement and application achievement. Motivated by theseobservations, we study the related data quality issues over the specificities of e-Health monitoring applications.

A. *Data Collection and medical information:*
As retrieved data grows, users and providers are more and more concerned about data quality. Data quality remains an important aspect of information management and becomes a research domain increasingly active. Specific research approaches and well established quality managements programs as Six Sigma and Total quality Management have been adapted to data quality assessment. Data quality often takes several perspectives (i.e. user view, product view …), in the literature there is no single definition of data quality accepted by researchers or specialists. Recently, data quality was better defined as "contextual". This means that the user (i.e. quality analyst) defines their own perspective of quality for each proposed use of data and within its particular context of use according to the application domain and goal. Thus, according to high-data quality appears when data fits its intended use in operations, decision-making and planning.

B. *MapReduce of Collected Data:*
Mapreduce is a parallel programming model that is used to retrieve the data from the Hadoop cluster. In this model, the library handles lot of messy details that programmers doesn't need to worry about. For example, the library takes care of parallelization, fault tolerance, data distribution, load balancing, etc. This splits the tasks and executes on the various nodes parallel, thus speeding up the computation and retrieving required data from a huge dataset in a fast manner. This provides a clear abstraction for programmers. They have to just implement (or use) two functions: map and reduce. The data are fed into the map function as key value pairs to produce intermediate key value pairs.

III. PROPOSED ALGORITHM

In the spare uncertain and incomplete data are describing the features for Big Data applications. In every human being sparse, number of data points is very less for drawing reliable conclusions and this is normally a complex of the data dimensionality issues where data in a high-dimensional space do not show clear trends or distributions. Such uncertain data are a special type of data reality where each data field is no longer deterministic but is subject to some random error distributions. Data's like incomplete data's are refer to the missing of data field values for some samples the missing values can be caused by different realities such as the malfunction of a sensor node or some systematic policies to intentionally skip some values.

Description of the Proposed Algorithm:
The data collected consists of several diseases and their related symptoms and the remedies required to cure that disease. Let this document be D1.The queries consist of the symptoms that are suffered by the patient. Let the document consisting of these symptoms be D2. The words present in D1 and D2 are matched by the process called semantic matching. This process is performed with the help of Jaccard distance formula.

$$J.D.\ (D1,D2) = \frac{\sum All\ Common\ words\ between\ D1\ and\ D2}{Total\ words\ in\ D1\ and\ D2}$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

IV. PSEUDO CODE

Step 1: Let the words showing symptoms present in queries be Qi and the words showing symptoms present in data Sets i.e,Sk.
Step 2: Semantic matching is performed between Qi and Sk with the help of Jaccard distance formula i.e, J.D.(Qi,Sk)
Step 3: According to the value of Jaccard distances, a score is calculated $Score_k$.

Step 4: Based on the scores calculated, a mean score is calculated Sd
Step 5: If the mean value Sd is greater than the Score of disease present in the data sets, then that disease is short listed.

If Sd> Symptoms
Shortlist the disease
Else
Remove the disease

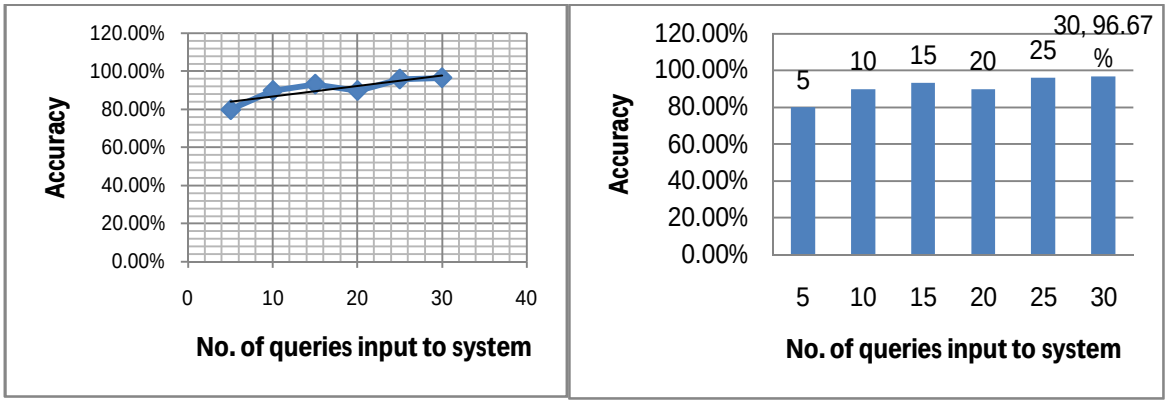
Step 6: Collect all the remedies of shortlisted disease and display the disease with highest score.

V. RESULTS

As per the applied MapReduce functions and the algorithm used for mining, the result analysis can be given by the following table. Here the accuracy of the system is evaluated by the number of relevant results and the number of queries input to the system.

No. of queries input to system	No. of relevant results	Accuracy
5	4	80.00%
10	9	90.00%
15	14	93.33%
20	18	90.00%
25	24	96.00%
30	29	96.67%

As per the table showing accuracy of the results, the graph can be plotted between the number of queries input to the system and the accuracy. Fig below shows the graph where the accuracy is increasing as per the increase in number of queries input. The following fig shows the graph for the accuracy of the system according to the queries input to the system and their relevant results.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

VI. CONCLUSION AND FUTURE WORK

This paper discusses about the big data and its characteristics, methods and challenges and suggests how to overcome the underlying problems being faced by the health care industry. Also presents the big ideas to fix the healthcare system in India. The implementation part of this paper can be done using HDFS (Hadoop File System) for the huge data storage and Hadoop Map Reduce with Deep Sparse mining algorithm. The use of big data analytics across the healthcare organization and healthcare industry will mine the doctor's lab transcript's using text mining and correlation to patient outcomes and location aware application analytics for enhancing customer experience. Achieving better outcomes at lower costs has become very important for health care which can be achieved through the implementation of this paper using Hadoop HDFS and MapReduce to uncover the information lying in big health data sets. Current master's thesis proposed and designed a system which is considered as the first part of today's modern E-health system and the rest two parts have to be developed as a future work of this project. More in detail is modern RPMS with all three parts where 'Data Mining' and 'Critical Patient Monitoring' parts dedicated as a future work. The health care for different kinds of patients has to be provided accordingly; hence intelligent analysis of measurement results is highly important for clinicians as well as patients

REFERENCES

1. K.Abinaya: Data Mining with Big Data e-Health Service Using Map Reduce: International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 2, February 2015
2. Hoffman, S., &Podgurski, A.: Big bad data: Law, public health, and biomedical databases. The Journalof Law, Medicine & Ethics, 41(Journal Article), 56-60. doi: 10.1111/jlme.12040, 2013
3. Ola, O., &Sedig, K. The challenge of big data in public health: an opportunity for visual analytics.Online J Public Health Inform. 5;5(3):223. doi: 10.5210/ojphi.v5i3.4933, 2014
4. Raghupathi W: Data Mining in Health Care. In Healthcare Informatics: Improving Efficiency and Productivity:211–223:2010
5. Y.-C. Chen, W.-C. Peng, and S.-Y. Lee, "Efficient Algorithms for Influence Maximization in Social Networks," Knowledge and Information Systems, vol. 33, no. 3, pp. 577-601
6. P. Domingos and G. Hulten, "Mining High-Speed Data Streams," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '00), pp. 71-80, 2000.
7. R. Ahmed and G. Karypis, "Algorithms for Mining the Evolution of Conserved Relational States in Dynamic Networks," Knowledge and Information Systems, vol. 33, no. 3, pp. 603-630
8. Chawla, N. V., & Davis, D. A. Bringing big data to personalized healthcare: a patient-centered framework. Journal of General Internal Medicine, 28 Suppl 3(3), S660-665. doi: 10.1007/s11606-013-2455-8,2013.
9. Feldman, B., Martin, E., &Skotnes, T. Big data in healthcare: Hype and hope,2012
10. Wasan, SiriKrishan, VasudhaBhatnagar, and HarleenKaur. "The impact of data mining techniques on medical diagnostics." Data Science Journal 5.19: 119-124,2006
11. J. S. Kahn, V. Aulakh and A. Bosworth, "What It Takes: Characteristics Of The Ideal Personal Health Record", Health Affairs, 28, pp. 369-376,2009
12. Adler-Milstein, J., &Jha, A. K. Healthcare's "Big Data" challenge American Journal of Managed Care, 19(7), 537-538, 2013.

BIOGRAPHY

Miss Gousia Ahmed, M.Tech Student, Dept. of CSE, BIT Ballarpur, Gondwana University, Maharashtra, India. Completed B.E. in Computer Technology from Nagpur University (2012).

Mrs Deepa A. Amne, Assistant Professor, Dept. of CSE, BIT Ballarpur, Gondwana University, Maharashtra, India. Completed M.Tech.in Computer Science &Engg from JNTU (2013), B.E.in Computer Technology from Nagpur University (2007).