ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

HEALTH CARE ANALYTICS WITH HADOOP BIG DATA PROCESSING

V.Kavitha, (PG Scholar)
Department of Computer Science and Engineering
Kalasalingam Institute of Technology
Krishnankoil, Virudhunagar dist,Tamil Nadu
Kavitharadha24@gmail.com

S.Kannudurai, (Assistant professor)
Department of Information Technology
Kalasalingam Institute of Technology
Krishnankoil, Virudhunagar dist,Tamil Nadu
kannuduraivlp@gmail.com

Abstract- Maintaining large amount of data at one system takes long time to access to reduce this problem use replicas of that data at large number of disks. It reduces our time to access that data. The data in healthcare is increasing rapidly and is probable to increase significantly in coming years. This has different kinds of health data like EHR, genomic, behavioral and public health which can be processed with big data processing. Although many technologies are used for big processing with health care records, which uses predictive analysis however which is not sufficient for all kind of health records. In this paper Fuzzy C-means Clustering Algorithm which creates centroid-based clustering from a fixed set of examples. That helps to maintain all kind of health data with low costs, and Provides right intervention to the right patient at the right time. It's potentially beneficial for all the components of a healthcare system like provider, payer, patient, and management. This includes that the health care data should be properly analyzed so that which group or gender are attack by the diseases most.

Keywords: Healthcare, Hadoop, big data, clustering

I. INTRODUCTION

Big data is an emerging technology which is mostly used in all kind of data processing methods. Research labs, hospitals and other medical institutions are trying to make maximum advantage of big data analytics to reduce healthcare costs by changing the models of treatment delivery also used for curing diseases, improving profits, predicting epidemics and increasing the quality of human life by preventing deaths. Hadoop is the only technology that allows health care to store data in its native form.

Hadoop is a fashionable computing tool for big data processing. Which uses map reduce as a default algorithm to work with large quantity of data sets. Map reduce divides its work into two kind of phases i.e.) mapper phase and reducer phase. Each phase has key/value pairs as an input and output. That key/value pairs are chosen by the user of the program.

Map phase is just a data training phase. Map reduce runs the various tasks in parallel, managing all communications and data transfers between the different parts of the system and providing for redundancy and tolerance. Map () simply performs filtering, sorting and the reduce () performs the summary operation.

II. RELATED WORKS

1). Healthcare Data Analysis using Dynamic Slot Allocation in Hadoop International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-3 Issue-5, November 2014 (IJRTE) , Aditi Bansal, Ankita Deshpande, Priyanka Ghare, Seema Dhikale, Balaji Bodkhe

In this new era of big data even health care needs to be modernized, this includes that the health care data should be properly analyzed so that we can deduce that in which group or gender, diseases attack the most. This gigantic size of analytics will need large computation which can be done with help of distributed processing, Hadoop. Map Reduce, a popular computing paradigm for large-scale data processing in cloud computing. However, the slot-base Map Reduce system (e.g., Hadoop MRv1) due to its unoptimized resource allocation, can suffer from poor performance. To address it, the framework in this paper optimizes the resource allocation. Due to the static pre-configuration of distinct map slots and reduce slots which are not fungible, many a times slots can be severely under-utilized.

2). A Survey on Applications of Big Data Analytics in Healthcare International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-5, November 2015, Shubham Borikar, Mohan Bhagchandani, Raunak Kochar, Ketansing Pardeshi, Manisha Gahirwal

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

The data in healthcare is increasing rapidly and is expected to increase significantly in coming years. Healthcare services although armed with modern technologies for curing the diseases grapples when it comes to preventing the diseases beforehand. Adoption of Big Data solutions will play an important role in transforming the outcomes of the healthcare industry by promoting evidence based reasoning and providing patient centric treatment. In this age of Big Data we can provide solutions to identify individuals who are prone to certain lifestyle diseases.

3). A Survey: Big Data Analytics on Healthcare System, Contemporary Engineering Sciences, Vol. 8, 2015, no. 3, 121 - 125 HIKARI Ltd, www.mhikari.comhttp://dx.doi.org/10.12988/ces.2015.412255, R. Sathiyavathi.

Big data and the related technologies have improved health care enormously, from understanding the origins of diseases, better diagnoses, helping patients to monitor their own conditions. By digitizing, combining effectively using big data, healthcare organizations can improve their quality of service by analyzing the effectiveness of a treatment and also the efficiency of the healthcare delivery process by detecting fraud, waste and drug abuse more quickly and efficiently. General goals to use analytics are, we can predict readmission risks, increase the efficiency of clinical care, and finding opportunities for cost savings.

4). A Survey On Big Data Analytics In Health Care, Priyanka Ketal, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5865-5868, Priyanka K *B.V.B.C.ET Hubli,* Prof Nagarathna Kulennavar *B.V.B C.E.T Hubli.*

Health care centers using a new information management approach called as big data analytics. Including big data analytics in health sector provides stakeholders with new insights that have the potential to advance personalized care, improve patient outcomes and avoid unnecessary costs. This paper defines big data analytics and its characteristics, comments on its advantages and challenges in health care.

III. PROPOSED SYSTEM

The data in the health care sector is growing rapidly and is coming from various internal as well as external sources like mobile devices, wearable sensor devices, clinical notes, social media etc. The effective analysis of the present health data can help in providing newer solutions to the present diseases. Providing health care suggestions to the users by generating a report to provide personalized health tips to the users of the system. The health care system is overwhelming not only because of massive volumes but also diversity of data types and the speed at which it managed.

In normal analysis of big data in health care using map reduce that accordingly doing the filtering, sorting and summary operations. To improve the performance of the map reduce, it works faster than before. So here the algorithm of Fuzzy c-means clustering used.

Big data analytics and applications in healthcare are at a nascent stage of development, but advances in platforms and tools can accelerate their maturing process to a greater extent. It's useful to Identify patients at high-risk and ensures that they get the right treatment as they need.

IV.TOOLS AND TECHNOLOGIES

Hadoop is a tool that is perfectly handles big data which is used to work with large amount of data's that has different kind of dataset. For installing hadoop first choose the open source operating. Hadoop works with the implementation of two main factors such that HDFS and Map reduce.

V.SYSTEM ARCHITECTURE

The architecture includes patients, big data analysts and specialized doctors. In the first step the individual inputs of a set of data sets. The individual's diseases from various health records like Electronic Health Records (HER), genomic record, behavioral record and public health records are then compared to all other patients available in the existing database and an initial filtering is done. With this filtering only those patients with whom individual has some disease similarly are kept. On this filtered dataset collaborative filtering is performed. The data analyst's outcome is a list of diseases which the patients have.

Then the doctor's data like specialization, doctor id, location and other details about the doctors are also collected by the analysts. The both patient's and doctor's data are kept in a same big data storage area as a dataset. Which is then processed by the doctors as well as those who are

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

handling the patient's when they are in critical stage of sick.

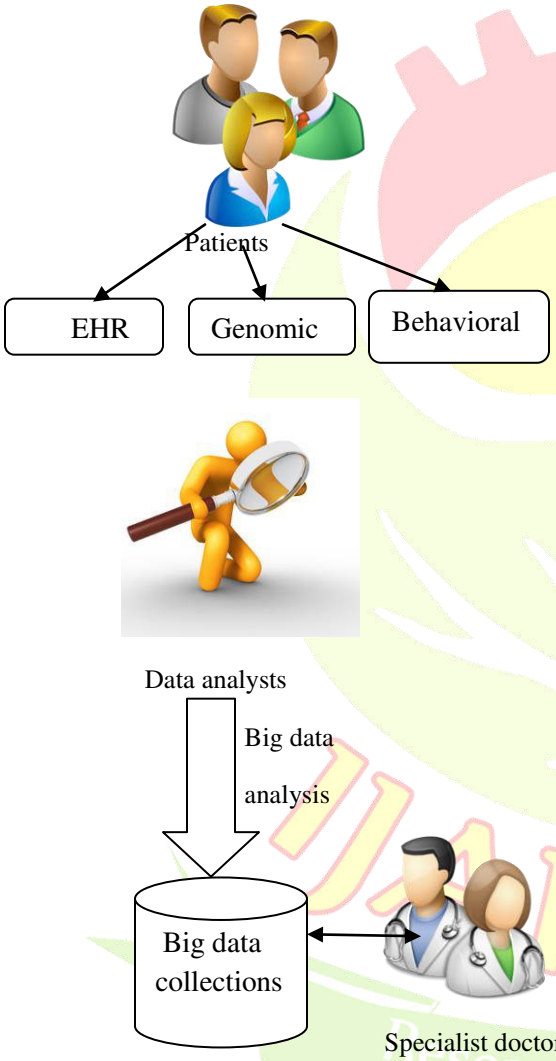

Fig 1. System architecture for proposed system

VI. ALGORITHMS

In this paper the algorithm of Fuzzy c-means clustering algorithm is used to improve the performance of map reduce.

Fuzzy c-means clustering algorithm

Here an algorithm is used to clustering the data's from the dataset which has the similarity of one another. It usually uses predictive based analysis which helps to avoid duplicate data's as enter in the system and repetitions. This algorithm is centroid-based clustering algorithm in which dataset is group into n-clusters.

STEP 1: Let X = {x1, x2, x3 ..., xn} be the set of data points
STEP 2: V = {v1, v2, v3 ..., vc} be the set of centers.

STEP 3: Randomly select 'c' cluster centers.

STEP 4: Calculate the fuzzy membership 'μij' using:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

STEP 5: compute the fuzzy centers 'vj' using:

$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (\mu_{ij})^m), \forall j = 1, 2, ...c$$

STEP 6: Repeat step 2) and 3) until the minimum j value is achieved or $|| U(k+1) - U(k)|| < \beta$.

Where,

k is the iteration.

β is the termination criterion between [0,1].

U = (μij)n*c is the fuzzy membership matrix.

J is the objective function.

VII. CONCLUSION

The big data is a growing technology which maintains all kind of data's in the real world. Though there are several challenges like combining heterogeneous data, infrastructure issues, insufficient real time processing, data quality that must be addressed, Big Data has the potential to transform and revolutionize the way healthcare systems use technologies to gain valuable insight from the data repositories. In the future we are sure to see widespread use of big data analytics across the different areas of healthcare industry. This paper provides the mechanism to improve the quality of big data analytics by improving the performance quality of map reduce, whose proper selection can give

promising results. Big data is helps to maintain over the all type of health care analysis and gives the wonderful results to the user.

REFERENCES

[1] Aditi Bansal, Ankita Deshpande, Priyanka Ghare, Seema Dhikale, Balaji Bodkhe "Healthcare Data Analysis using Dynamic Slot Allocation in Hadoop" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-3 Issue-5, November 2014 (IJRTE)

[2] Shubham Borikar, Mohan Bhagchandani, Raunak Kochar, Ketansing Pardeshi, Manisha Gahirwal, A Survey on Applications of Big Data Analytics in Healthcare International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-5, November 2015.

[3] R. Sathiyavathi A Survey: Big Data Analytics on Healthcare System, Contemporary Engineering Sciences, Vol. 8, 2015, no. 3, 121 - 125 HIKARI Ltd, www.mhikari.comhttp://dx.doi.org/10.12988/ces.2015.412255 .

[4] Priya nka K *B.V.B.C.ET Hubli,* Prof Nagarathna Kulennavar *B.V.B C.E.T Hubli.*A Survey On Big Data Analytics In Health Care, Priyanka Ketal, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5865-5868.

[5] Divyakant Agrawal, UC Santa Barbara, Philip Bernstein, Microsoft Elisa Bertino, Purdue Univ. "Big Data White pdf", from Nov 2011 to Feb-2012.

[6] Durham, E.-E.A.; Rosen, A.; Harrison, R.W., "Optimization of relational database usage involving Big Data a model architecture for Big Data applications," in Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on , vol., no., pp.454-462, 9-12 Dec.2014.