Predicting the waiting time of patients in hospital by using RF algorithm and Design of HQR system.

Chetan Shelake, Soham Patwardhan, Sourabh Teli, Krantikumar Mhetar

Student,Dept of Computer Engineering, JSCOE PUNE,MAHARSHTRA, INDIA

---***---

Abstract - *Patients wait delay and patient overcrowding is one of the major problems faced by hospital. A patient is usually required to undergo various examinations, inspection or tests according to his conditions. This waiting time increases the frustration on patients. Patient Queue Management and wait time prediction form challenging and complicated job because each patient might require different phases and operations such as check-up and various tests. Random Forest Algorithm(RFA) is used for Data mining of big data Furthermore, this implementation can also be applied to Time Prediction. Use of HBase will give historical data of patients. HQR and RF are parallelized on Hadoop Platform. Android Platform is used for providing Graphical User Interface.*

Key Words: *Big data,Cloud computing,Hospital Queue Recommendation(HQR),Hadoop,Random Forest algorithm.*

1.INTRODUCTION

Currently, most hospitals are overcrowded and they are not efficient in providing proper queue management. Providing Patient queue management and waiting time prediction is challenging and tedious job as each patient vary in different operations such as checkup, different tests like X-ray, CT scan, blood tests, sugar level. Some of the tasks are independent whereas some tasks are waiting to complete other dependent tasks. Most patients must have to wait in different queues for different treatments. In order to complete required treatment in a shortest duration of time waiting time of each task is predicted in real time. PTTP algorithm is proposed as learning algorithm for calculating the waiting time[1].Patient Treatment Time Prediction(PTTP) uses RF algorithm for its implementation. Based on this Hospital Queue Recommendation(HQR) system is diagnosed. In the Computer System we have mainly three types of

Resources, they are Software, Hardware and Data. Data is the most important resource of computer system, because whatever computing we are doing is just because of data.

 The massive unstructured data is called Big Data. Basically, The term big data not only means large volume of data but also other features that differentiate it from the concepts of "massive data or large volume of data". [3] Now in present days very less amount of data is generated in structured form as compare to unstructured data e.g. Text files, sensor data, log data, web data, social networking data or different varieties of data. For Big Data management Hadoop is used.

Hadoop is a framework that provides distributed processing of large data sets across cluster using a simple programming model. It is an open source data management which uses distributed processing.[3]

2. LITERATURE SURVEY

Patient queue management and wait time prediction form a challenging and complex job because each patient might require different operations, such as a checkup, various tests, e.g., a sugar level or blood test, X-rays or a CT scan during treatment. So there are five major methodologies used in this system Big Data management with Historical Dataset, Preprocessing of data, Use Learning Algorithm PTTP(patient Treatment Time prediction)with base of RF (Random Forest) Algorithm Calculate the Waiting Time in Hospital Queue Recommendation. A random forest optimization algorithm is performed for the PTTP model. The queue waiting time of each treatment task is predicted using the trained PTTP model. A parallel HQR system is introduced, and an efficient and convenient treatment plan is recommended for each patient. The patient may undergo various treatment operations such as CT scan, MR scan and a payment task. These set of treatment operations are

submitted to decision maker and recommendation module via mobile interface. The predicted waiting Time of all of the treatment tasks is calculated by PTTP model. After this a treatment recommendation with least waiting time is advised [1]

Most of the smartphones have built in sensors that can measure motion, location, orientation and various other environmental conditions. These sensors can provide raw data with high precision and accuracy for inferring and recognizing queuing behavior. Furthermore we observed that people frequently carry smartphones when they are not at home. We can use a collaborative approach for queuing recognition based on smartphones by following queuing rules(First Come First Server). We can use a prototype of QueueSense with clients on smartphones which uses Android platforms and a server in cloud. Smartphones make use of widely available sensors such as accelerometer, Bluetooth and compass to sense individual activities. Queuing features are calculated based on queuing properties in terms of individual activities and support vector machine(SVM) is used to automatically detect whether the people are queuing or not on smartphones. The cloud backend process multi-lines scenarios and provide estimation of queue length and waiting time. Agglomerative hierarchical clustering is used on server side to divide queuers into different lines based on changing rate of relative position of queuers[2].

Hospital Information generally contains an information systems such as Electronic Medical Records(EMR) and Picture archiving And Communication System(PACS).Hospital data is center generally stores the Structured And Unstructured Data. Most data used in the EMR is Structured Data which includes information of a patient, information of an treatment, diagnostic information and the reports. This above data is stored in the hadoop cluster with the help of an JDBC/ODBC interface and then the data is stored in an HDFS with the help of an MAPREDUCE and the HIVE .The Structured data in the HDFS is written using an HIVE and its SQL like Query language HQL.

Data is Processed in the system we need to give a connection to the database then we need to check an existence of a table i.e if a Table exists already then we need to add a Partition otherwise we need to create a Table and then we need to add a Partition. Then we need to check an availability of an data if an data is available then just update

the data otherwise write the data. and then we need to disconnect with the database.[3]

Hadoop is a framework which provides distributed processing of large data sets across cluster using a simple programming model. Mainly Apache Hadoop Framework consists MapReduce and Hadoop distributed file system. Hadoop distributed file system, as Map reduce provide a simple programming model well as other related projects e.g. Apache Hive, Apache HBase etc. The basically three important parameters of hadoop cluster; they are CPU, MEMORY and DISK. All the nodes are dedicated to work for the hadoop tasks only. This can be useful for efficient management of cluster to provide capacity, scalability and performance of cluster such that provisioning of resources will be efficient[4].

Large amount of data is created by users in daily life. which requires huge amount of storage and various techniques to discover knowledge from data. Hadoop architecture is of two main components HDFS(Hadoop Distributed File Sysytem) and Mapreduce for Big Data Analytics. There are various technologies belongs to Hadoop Hbase for storing large dataset, Apache pig is scripting language for processing of large data set, Hive is designed for OLAP is fast and Scalable, Scoop is used for import and export data from RDBMS to Hadoop, Zookeeper is used for distributed application and flume is for moving large amount of data to centralized data.[5]

Random forest is most popular data classification and regression algorithm for machine learning. This system introduced a Scalable Random Forest Algorithm which is based on MapReduce Technique .The algorithm is divided into three stages: initializing, generating and voting. SMRF algorithm has main objective of improving the traditional random forest algorithm based on MapReduce model. SMRF algorithm provide scalable performance, and it can negotiate with the distributed computing environments to decide its trees scale.[6]

3. PROPOSED SYSTEM

Based on the survey we have understood that the system will handle structured data including patient information, treatment information, diagnostic information and reports.

The system stores data in Hadoop cluster by the use of JDBC/ODBC interface in Thrift and after that it stores data in HDFS by Mahout and Hive. For training purpose we use RF algorithm and we develop HQR system for patients. Patients can check their waiting time by using mobile phones and thus can reduce their frustration.
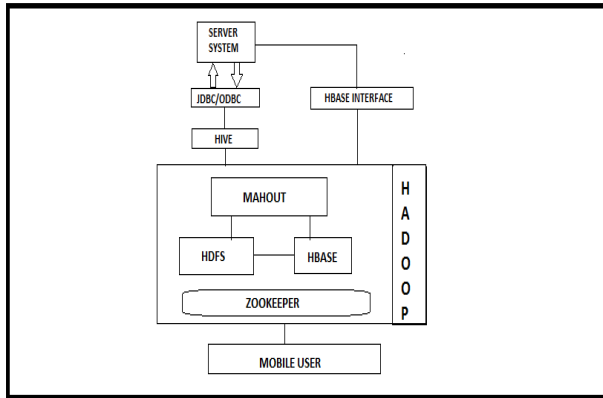

Fig -1: Diagrammatic representation of the proposed system.

4. CONCLUSION

In this paper we have reviewed on the technologies which are being used for hospital queuing behavior. further we have introduced different techniques used to implement them. Present methods include hadoop, sql, Hbase, RF algorithm. Later on we have compared the different techniques used by researchers in their systems, such as extension in RF algorithm, storing of structure data into the database etc. This comparison will help us in building our system more convenient and useful. From the research we have proposed the system which will predict time required for particular task.

ACKNOWLEDGEMENT

REFERENCES

[1] JIANGUO CHEN1,2, (Student Member, IEEE), KENLI LI1,2, (Senior Member, IEEE), ZHUO TANG1,2, (Member, IEEE), KASHIF BILAL3,4, AND KEQIN LI1,2,5, (Fellow, IEEE), "A Parallel Patient Treatment Time Prediction Algorithm and Its Applications in Hospital Queuing-Recommendation in a Big Data Environment", Received March 8, 2016, accepted April 12, 2016, date of publication April 25, 2016, date of current version May 9, 2016

[2] Qiang Li, Qi Han, Limin Sun, "Collaborative Recognition of Queuing Behavior on Mobile Phones"2015 IEEE Transactions on Mobile Computing.

[3] Hong-lei-Zhang, Peng-fei Li, Xin-hang Li, Sheng-qiang Chi, Jing-song Li*,"Design and Implementation of Clinical Data Center Based on Hadoop",2015 7th International Conference on Information Technology in Medicine and Education.

[4] Kapil Pandey Anand Gadwal Prashant Lakkadwala ,"Hadoop Multi Node Cluster Resource Analysis ",2016 Symposium on Colossal Data Analysis and Networking (CDAN)

[5] Aditya Bhardwaj, Vanraj, Ankit Kumar, Yogendra Narayan, Pawan Kumar, "Big Data Emerging Technologies: A CaseStudy with Analyzing Twitter Data using Apache Hive", Proceedings of 2015 RAECS UIET Panjab University Chandigarh 21-22nd December 2015

[6] Jiawei Han, Yanheng Liu, Xin Sun," A Scalable Random Forest Algorithm Based on MapReduce",2013 IEEE

BIOGRAPHIES

Chetan Padmakar Shelake
JSCOE,Computer Department