Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 6, June -2017

Queue Recommendation System for Hospital Application using Parallel Patient Treatment Time Prediction Algorithm in Big Data

Mrs. Sneha Girolla1, Mr. Sunil Rathod

Department of Computer Engineering,Dr. D.Y. Patil School of Engineering,
Department of Computer Engineering,Dr. D.Y. Patil School of Engineering,

Abstract -*The major problems faced by hospital is patients wait delay and patient overcrowding. various examinations, inspection or tests must be done by patient usually according to his medical conditions. Similarly, there are various reasons for a patient to wind up his visit in hospital as soon as possible. For this an effective queue management must be maintained which gives an ease to fast track treatment process. But, patient queue management and wait time prediction brings challenges and complications because each patient requires different phases of treatmentand operations such as check-up. Therefore, a Random Forest Algorithm(RFA) is used to categorize the patients on big data platform. Furthermore, this implementation is applied to Time Prediction for each patient. This is where technology comes into scenario developing system to overcome the queue management and providing effective patient waiting time for each treatment using Apache Spark for real time data analysis using Spark Streaming parallel to RFA with integration of Scala. Hospital Queueing Recommendation System (HQR) is developed for Patient Treatment Time Prediction (PTTP).*

Key Words*: Big data, Apache Spark, Hospital Queue Recommendation (HQR), Random Forest Algorithm, Patient Treatment Time Prediction (PTTP).*

I. INTRODUCTION

Standing in queue to avail a service can cause extreme boredom, annoyance and even rage to patients. Patientsare often forced to wait in line whenever the service facility is busy. Most hospitals are overcrowded and are not efficient to provide queue management due to multiple treatment executing simultaneously. Providing such queue management and waiting time prediction is most challenging and tedious job as each patient go through different operations such as check-up, different tests like X-ray, CT scan, blood tests, sugar level etc. Each taskvaries in inferentiality of treatment with types as independent where tasks are independent results whereas dependent task wait for other to complete the dependent tasks. Mostly patients must wait in different queues for different treatments apart from task variation. An approach to resolve waiting time delay and predict the waiting time for each independent and dependent task Parallel Patient Treatment Time Prediction Algorithm (PTTP)is suggested where it will calculate each treatment task wait time depending upon queue of patients already in waiting state for the same treatment. Which results in proving predicted time required per treatment in a shortest duration of waiting time in real time. PTTP algorithm is proposed as learning algorithm for calculating the patients waiting time to avail any service.PTTP uses Random Forest(RF) algorithm for its implementation which include classification and regression for big data store and retrieval. For efficient queue management recommendation Hospital Queue Recommendation(HQR) system is implemented. In the Computer System, there are mainly three types of Resources Software, Hardware and Data. Data is the most important resource of computer system, because computing is purely based on data. A massive unstructured data is called Big Data. Basically, the term big data not only refers to as large volume of data but also other features such as difference in storage formats, graphical representation etc; that differentiate it from the concepts of ‘‘massive data or large volume of data”.

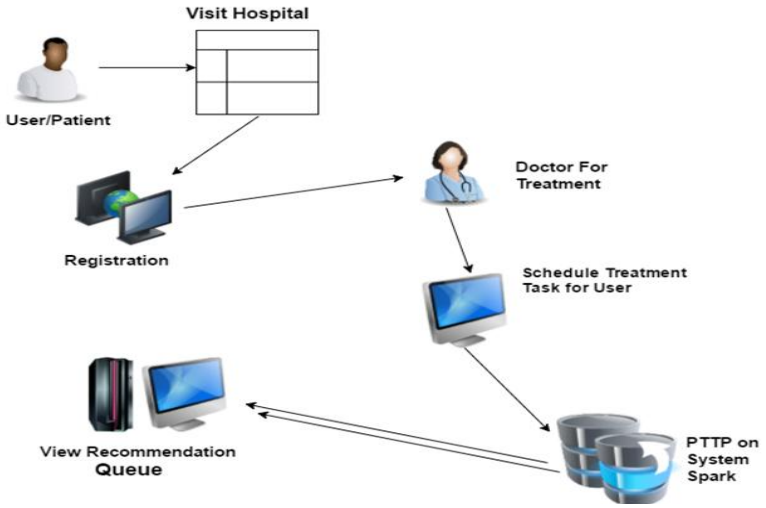

Fig 1: Basic Flow Diagram.

The PTT consumption of every patient within the waiting queue is calculated by trained PTTP model. We tend to specialize in serving patients to complete their treatment tasks in a very foreseeable time with effective assumptions and serving hospitals to schedule every treatment task queue according to each patient and avoid overcrowding and long waiting queues. We use huge realistic knowledge called as Big Data from various hospitals which contains information in depth regarding patients, doctors, hospitals and approximate treatment time which helps in patient treatment time consumption model.

Realistic patient knowledge which means real time data gathering through real time transactions and registration, is called an area unit which is analysed thoroughly and strictly having support from necessary parameters, like patient treatment start time, end time, patient age and detail treatment content for every task per patient with similar or different task. Hospital data is centre which stores Structured and Unstructured Data. Most data used in the Electronic Medical Records(EMR) is Structured Data which includes information of a patient, information of a treatment, diagnostic information and the reports We tend to determine and calculate approximately accurate different waiting times for different treatment task and for various patients supported medical conditions and operations performed throughout various treatment tasks.

II TAXONOMY CHART

Table 1: Taxonomy chart

services / paper or other sources	Classification	Performance	Random forest	Spark	CART Model	Parallelization	Streaming Data
A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment.	X	✓	X	✓	X	X	X
Robust and accurate shape model matching using random forest regression-voting.	X	X	✓	X	X	X	X
KASR:A keyword-aware service recommendation method on MapReduce for big data applications.	X	✓	✓	X	✓	X	X
Big data analytics framework for peer-to-peer botnet detection using random forests. 2014	✓	X	✓	X	✓	X	X
HC-CART:A parallel system implementation of data mining classification and regression tree (CART) algorithm on a multi-FPGA system .2013	X	✓	✓	X	X	X	X
Proposed system	✓	✓	✓	✓	✓	✓	✓

III. PROPOSED SYSTEM

This paper has developed PTTP algorithm and HQR systemto overcome the problem of wait delays and overcrowding by introducing the concepts on spark computing environment having featuresas maximized response time and minimizing wait delays. To improvise the real-time data steaming for each patient apache spark computing environment is used.

PTTP algorithm based on big data and the Apache Spark cloud environment is proposed. Implementation of random forest optimization algorithm is performed to achieve high degree of dimensionality and classification results.
Because of the restrictions on initial RF rule and characteristics of hospital treatment information, the RF rule is improved in four aspects to get a good result for large-scale, high dimensional, continuous, and rackety hospital treatment information. First, the featured data is used directly without any further classification or sampling. Then, as a result the target variable of the treatment information is patient treatment time consumption which can be a continuous variable, a CART model is employed as a meta-classifier within the improved RF rule. At constant time, some freelance variables of the information are nominal data, that have completely different values like time vary (0 - 23) and day of week (Monday - Sunday). In such a case, the two-fork tree model of the normal CART cannot totally the analysis results. Now for next improvement the noisy data containing inconsistent data cannot be assumed as only value missing form featured data; each must be classified according to the featured data. Then, identify the all variables and parameters used for further calculation in respective algorithms. Therefore, noisy data are removed to reduce their influence on accuracy by calculating the average value of the data in each leaf node of the regression tree.
Waiting time of each treatment task is calculated depending upon the trained data set classification performed on PTTP model which produces the output upon PPT a parallel HQR system is developed which will reproduce an efficient and convenient treatment plan for per patient depending upon various conditions.Conditions vary depending upon time necessities, for example enormous data is generated with high quality of the system each data gathered having different format depending upon treatment task else can be independent task. This data information must be evaluated and use to generate high predictability of results able to justify the recommendations with potency and measurability. For algorithms being used must be efficient to represent an entire range of data without having fault to measurability. Performed task is on trained data which supports which improved Random Forest (RF) algorithmall rules for every treatment task and predicts waiting time of every task. Thus, improvising the processing aspects of random forest described below:

3.1 Data gathered from treatment task:

Data collected from different treatment tasks, for example patient registrations, drug medications, outpatient department(OPD), treatment analysis, medical examination and other related tasks.

3.2Same dimensions are created:

Dimension is a data set composed of singular, non-overlapping data elements. Functionalities provide by dimensions are filtering, grouping and labelling. Table rows of a dimension table are identifieduniquely by a single key field, it is recommended that the key field be a simple integer because a key value used only for joining fields between the fact and dimension tables. The historic data collected from different treatment tasks containing different contents and formats with varying dimensionsshould be unique and consistent. Assessing trained data set to PTTinput model for each patient treatment task, same features are necessary to be chosen, for example aspatient in time information for example while registration it's basic and contact information, the treatment task information for example task id, department name, and the time information inputted valid time during OPD (start time and end time).

3.3 New feature variable calculation:

Feature variable calculation is formulated by analysing the best split point, then the data is divided into two forks for constructing multi-branch for CART model. This data is then used as trained data for PTTP model, which helps in calculations, such as the patient time consumption of each treatment record, overlays of time per patient, and the time range of treatment time.

3.3Incomplete and inconsistent data is removed:

After feature variable is used for creating CART model then the value is used for mean value calculation of leaf nodes after the noisy data is removed. Calculations for CART model will give negative values of time consumption as inconsistent data which will be removed, for example in time calculations an end time is overlapping start time where end time is less than start time, which can occur when a start time is inputted by from a console and an end time is considered by a machine. Other than feature variable the subspaces which are not in trained data set for PTTP algorithm are considered and added to out-of-box dimension.
Then, HQR system recommends an economical and effective treatment arrangement queue for every patient. Flow of a proposed system architecture considered as patients will be able to see the advised arrangement and foretold waiting time as an output. The recommended waiting queue of every patient's treatment task is calculated by the PTTP model which gives output in time prediction that is the summation of all patient's probable treatment task time within the current waiting queue. HQR system is expected to accept trained and calculated data from PTTP model, and produce treatment task recommendation queue for each patient withinareasonable and suitable treatment task queue which is managedandminimum waiting time is suggested for every patient for every treatment task. Hospital information square measure holds on with the Apache HBase database storage platform and parallel solution is utilized with inherited MapReduce and Resilient Distributed Datasets (RDD) programming model. Apache spark a beginning to Big Data handling: You can run powerful and cost-effective Apache Spark and Apache Hadoop clusters easily on Google Cloud Platform. If you prefer to get more hands-on with your data, use Data process, a managed Spark and Hadoop service that

allows anyone to create and use fast, easy, and cost-effective clusters. As proposed system uses apache Spark, well it is created to handle a wide range of workloads for example iterative algorithms, streaming, batch applications and interactive queries. Separate tools management burden is also reduced by apache spark.
Finally, a hospital queuing recommendation system is generated with the ascending order of treatment tasks with time respective, and performance is increased efficiently due to use of apache spark.

IV. OUTCOMES

PTTP algorithm based on big data and the Apache Spark cloud environment is proposed. A random forest optimization algorithm is performed for the PTTP model. Waiting time of each treatment task is predicted based on the trained PTTP model. A parallel HQR system is developed, and an efficient and convenient treatment plan is recommended for each patient. Extensive experiments and application results show that our PTTP algorithm and HQR system achieve high precision and performance in existing system.

V. CONCLUSION

Hence it is concluded that algorithms being implemented extensively on apache spark computing environment. With effective time prediction model name parallel patient treatment time prediction giving expected results. Extensive experiments and application results show that our PTTP algorithm and HQR system achieve high precision and performance in existing system.From the research, we have proposed the system which will predict time required for singulartreatment task to complete.

REFERENCES

[1] K. Singh, S. C. Guntuku, A. Thakur, and C. Hota, ``Big data analytics framework for peer-to-peer botnet detection using random forests," *Inf. Sci.*, vol. 278, pp. 488497, Sep. 2014.
[2] S. Meng, W. Dou, X. Zhang, and J. Chen, ``KASR: A keyword-aware service recommendation method on MapReduce for big data applications,''IEEE *Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 32213231, Dec. 2014.
[3] S. Tyree, K. Q. Weinberger, K. Agrawal, and J. Paykin, ``Parallel boosted regression trees for Web search ranking," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2012, pp. 387396.
[4] R. Fidalgo-Merino and M. Nunez, ``Self-adaptive induction of regression trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 16591672, Aug. 2011.
[5] G. Yu, N. A. Goussies, J. Yuan, and Z. Liu, ``Fast action detection via discriminative random forest voting and top-K sub volume search," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 507517, Jun. 2011.
[6] N. Salehi-Moghaddami, H. S. Yazdi, and H. Poostchi, ``Correlation based splitting criterion in multi branch decision tree," *Central Eur. J. Comput.Sci.*, vol. 1, no. 2, pp. 205_220, Jun. 2011.
[7] G. Chrysos, P. Dagritzikos, I. Papaefstathiou, and A. Dollas, ``HC-CART: A parallel system implementation of data mining classi_cation and regression tree (CART) algorithm on a multi-FPGA system," *ACM Trans.Archit. Code Optim.*, vol. 9, no. 4, pp. 47:1_47:25, Jan. 2013.
[8] C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, ``Robust and accurate shape model matching using random forest regression-voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1862_1874, Sep. 2015.
[9] N. T. Van Uyen and T. C. Chung, ``A new framework for distributed boosting algorithm," in *Proc. Future Generat. Commun. Netw. (FGCN)*, Dec. 2007, pp. 420_423.
[10] Y. Ben-Haim and E. Tom-Tov, ``A streaming parallel decision tree algorithm," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 849_872, Oct. 2010.
[11] A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment Jianguo Chen, Kenli Li, Senior Member, IEEE, Zhuo Tang, Member, IEEE, Kashif Bilal, Shui Yu, Member, IEEE, Chuliang Weng, Member, IEEE, and Keqin Li, Fellow, IEEE, 1045-9219 (c) 2016 IEEE
[12] https://www.tutorialspoint.com/apache_spark/
[13] https://www.tutorialspoint.com/scala/scala_tutorial.pdf
[14] https://www.cloudera.com/documentation/enterprise/5-6-x/PDF/cloudera-spark.pdf

