International Journal of Scientific Research and Engineering Development-– Volume X Issue X, Year

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

REVIEW AND ANAYLYSIS ON PATIENT WAITITNG TIME IN HOSPITAL QUEUING SYSTEM IN A BIG DATA ATMOSPHERE

Miss.Komal Mahadev Parit.
Lcube Creation, Pune
Software Development Department, India
komalparit295@gmail.com

Mr. Prashant Bajirao Patil.
Source One Management Services Pvt. Ltd
Software Development Department, India
prashant.patil033@gmail.com

----------**********----------

Abstract:

Due to the enormous growth in population hospitals are congested because of this it becomes difficult for hospital management system to control or to reduce the patient waiting time while receiving treatment in hospital. While enchanting treatment doctors declared number of different test to identify the disease to give appropriate treatment. Thus while calculating this entire test we have to interval in a line. A patient has to wait till the whole patient before him or she gets preserved. Unnecessary waiting time not only discarded patient time but also provide prevention during waiting in line. It would be additional suitable if patient could get the predicted waiting time and handling strategy on mobile application which displays the management strategy and predicted waiting time on real time. Therefore to rise the effectiveness and to chance the patent requirement we come up with a new method called PTTP with HQR i.e. Hospital Queuing-Recommendation System is established. In this technique PTTP algorithm predict the management time on the basis of hospital records. On the basis of this waiting time HQR recommend the treatment strategy for the patient. As the patient records is of in huge quantity we have use Hadoop to achieve the goal.

Keywords — **Big data, Pig, Data Mining Technique, Hospital Historical dataset, PTTP algorithm, Random forest.**

----------**********----------

I.INTRODUCTION

Suitable to the enormous increase in people hospitals are congested because of this it materialize to complex hospital administration organization to decrease patient waiting time while delivery treatment in hospital. Organization of patient queue and deduction of waiting time is tremendously tough job. As the waiting time approximation need the earlier material of the time taken by dissimilar tasks to comprehensive such as time taken by Vaccination [1], Medicine, X-ray, CT-scan, blood test, sugar test etc.

typically a patient has to go from side to side quantity of test during treatment point out by doctor which is depending on his or her inspection situation. Accordingly a patient has to stay for the time till all the patient antiquity to him gets treated [17, 13]. Thus avoidable waiting for such a extended time misuse the patient time and also gives trouble.

In this paper we review a approach to reduction the waiting time of patient [19] and to assist hospital connotation deal to skillfully state the patient row arrangement. To realize this job we have regard as patient evidence from hospitals such as patient age, gender, disease type, doctor name, task name, treatment start time and treatment end interval. On the commencement of this information set we have premeditated predicted time for our arrangement means exercise phase completed using pig script to implementation of random forest in PTTP prototypical [7, 13]. The approximation for predict time is depend of number of landscapes such as genders , age group, treatment start time and end time etc. that talk about PTTP model with hospitals important information. The normal waiting time is designed by this algorithm. And depending on this waiting time Hospital Queuing preparation predict the treatment planning for patient to reduction the waiting time through the number of test development. Because of the big info convenience from the patient description we have believe massive information for flexibility and facility. To calculate the waiting time an superior report of Random Forest Algorithm is used to make the planned PTTP structure [22] [7, 11]. Our main objective is to contribution the most outstanding quality treatment, inside time. We would comparable to contribution you to strongly and successfully take distribution of your hospital system by PTTP perfect using random forest using Pig Script [31,30] in big environment.

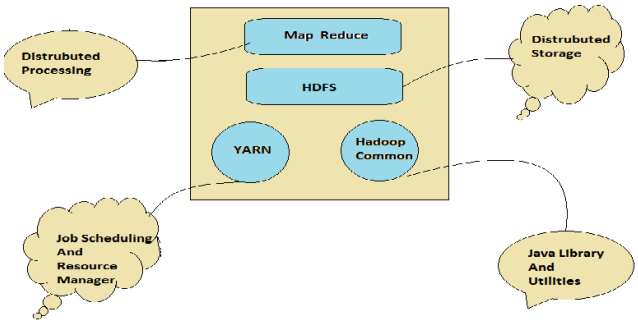
II.ASSOCIATED WORK

A) BACKGROUND

1995, RF technique is projected initially by Tin Kam Ho [12, 8]. In 2001, Leo Breiman establishes RF does not more robust because of Law of huge records [12, 4]. In 2006, Lin and Jeon recognized a relationship between RF and adaptive nearest neighbor technique [12, 9]. In 2008, G. Biau provides theorems that launch the worldwide regularity of averaging system [12, 10], which improved the essential hypothesis of RF. Random Forest includes below characteristics:

i) Its correctness is as brilliant.
ii) Its relatively strong to noise.
iii) It's quicker than bagging or enhancing.
iv) It's gives cooperative local calculation of fault, might, association and erratic consequence.
v) Its honest and basically parallelized.

B) BIG DATA



Big Data [29] establish with large-volume [28] [27], various, independent sources with spread and decentralized organize, and inquire about to discover difficult and developing associations between data. Big data means really a big data; it is a collection of huge datasets that cannot be processed using conventional computing method. Big data is not simply a data; rather it has become a entire theme, which occupy a variety of tools, methods and frameworks. Hadoop File System was developed by distributed file system (DFS) design. It is run on good hardware. Unlike other distributed

systems, [28] [27] HDFS is highly fault tolerant and considered using low-cost hardware. HDFS holds extremely big amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. HDFS also construct applications accessible to parallel processing.

The user of the **MapReduce** as two functions: *Map* and *Reduce*. *Map [28]*, written by the user, takes an input pair and produces a set of *intermediate* key/value pairs and passes them to the *Reduce* function. Apache Pig is a concept over MapReduce. It is a tool or platform which is used to examine better sets of data representing them as data flows. Pig is usually used through **Hadoop**. Using **Pig Latin [28]**, developer can do MapReduce tasks easily without having to type complex codes in Java using pig script. To examine data using **Apache Pig [27]**, developers need to write scripts using Pig Latin language. All these scripts are internally transformed to Map and Reduce tasks. [27][28] Apache Pig has a component known as **Pig Engine** that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs.

III. LITERATURE SURVEY

Title	Objective	Contribution	Result
Data Mining with Big Data	To inject randomness into the data to ensure a number of privacy goals.	To fill most frequently observed values or to build learning models to predict possible values for each data field, based on the observed values of a given instance.	An effective data analysis and prediction platform to achieve fast response and real-time classification for such Big Data.
A Hybrid Chemical Reaction Optimization Scheme for Task Scheduling on Heterogeneous Computing Systems	making decisions about the execution order of tasks and task-to-processor mapping.	(1)A Gaussian random walk approach is proposed to search for optimal local candidate solutions. (2) A left or right rotating shift method based on the theory of maximum Hamming distance is used to guarantee that our HCRO algorithm can escape from local optima.	DAG tasks much better than the existing algorithms in terms of make span and speed of convergence.
Fast Action Detection via Discriminative Random Forest Voting and Top-K	To achieve efficient and robust action detection	Characterize a video as a collection of spatio-temporal interest points, and locate actions via finding spatiotemporal video sub	Detection speed is several orders of magnitude faster than existing methods.

Subvolume Search		volumes of the highest mutual information score towards each action class.	
Scheduling Precedence Constrained Stochastic Tasks on Heterogeneous Cluster Systems	To map tasks of a parallel application onto processors of a cluster system and order their executions, so that task precedence constraints are satisfied and the minimum make span is achieved	Invent the concepts of stochastic bottom level and stochastic dynamic level (SDL), and to assign a task to a processor such that the SDL of the task on the processor is optimal according to a unique definition	Clearly demonstrate that the proposed stochastic task scheduling algorithm significantly outperforms existing algorithms
Robust and Accurate Shape Model Matching Using Random Forest Regression-Voting	To minimize the uncertainty of the predictions locally on the node level of every tree independently but	Investigate the impact of the choice of parameters	A fully automatic shape model matching system where a sequence of RFRV-CLMs was initialized using our own
	across all trees in the forest, as well as Neighborhood Approximation Forests which apply RFs to efficiently retrieve nearest-neighbor images.		implementation of Hough Forests
Travel Recommendation by Mining People Attributes and Travel Group Types From Community-Contributed Photos	To predict the locations (, with probabilities) that the user (with certain attributes) might like to visit from location.	Background destination popularity and the people's attribute distributions in each city.	To exploit social contexts in travel photo streams predicting their travel group types.
Self-Adaptive Induction of Regression Trees	To predict the departure delay of a flight depending on its characterization.	Treatment of a leaf or a no coherent node, and updating statistics of visited nodes.	False alarms are only detected when conditions of massive noise are present

Bayesian-Inference-Based Recommendation in Online Social Networks	To allow the root (querying user) to calculate the conditional probabilities	To measure the rating similarity between friends in an online social network by a set of rating conditional probabilities.	Achieve same MAE as best KNN MAE with coverage of 73 percent when setting probability threshold to 0.37.
KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications	A personalized service recommendation list and recommending the most appropriate services to the users effectively	(1) A keyword-aware service recommendation method, named KASR, is proposed in this paper, which is based on a user-based Collaborative Filtering algorithm.	KASR significantly improves the accuracy and scalability of service recommender systems over existing approaches.
An Advanced MapReduce: Cloud MapReduce, Enhancements and Applications	Bridging the gap between heavy-weight HOP (Hadoop Online Prototype) and light-weight	1. Developing C-CMR which has great functional and performance edge over conventional CMR and 2. Consolidating	CMR is more efficient and runs faster than other implementations of the MR framework.
		CMR, S-CMR and CCMR into a unified framework for processing huge amounts of data in the cloud efficiently.	

To improve the accuracy of the data investigation with nonstop features, different optimization approaches of classification algorithms are projected. There are different categories of trainings for data mining methods in medical databases. We identify the following groupings:

1. Studies that summarize reviews and challenges in mining medical/hospital data in general.
2. Studies of classification techniques in Hadoop used for prediction.
3. Studies that present new technologies and algorithms.
4. Studies that present new frameworks, tool and applications in medicine and healthcare system to prevent patient time

1. STUDIES THAT ENCAPSULATE REVIEWS AND EXPERIMENTS IN MINING HEALTH/HOSPITAL RECORDS IN GENERAL.

F.Hosseinkhah et al. [32] represent challenges in data mining on medical databases .They spotlight on the uniqueness of medical data and provide some investigational outcome on a gastritis medical database using data mining techniques of Naïve Bayes, Neural Network, and Association rule. S. K. Wasan, V. Bhatnagar and H. Kaur [33] observe the collision of data mining techniques, as well as artificial neural networks, on medical diagnostics.

2. LESSONS THAT EXTANT NOVEL TECHNOLOGIES AND PROCEDURES

To predict the waiting time for every treatment task, we utilize the random forest algorithm to train the patient treatment time utilization based on both patient and time characteristics and then build the PTTP model. Because patient treatment time utilization is a permanent variable, a Classification and Regression Tree (CART) model is used as a meta-classifier in the RF algorithm.

4. STUDIES THAT PRESENT NEW FRAMEWORKS, TOOL AND APPLICATIONS IN MEDICINE AND HEALTHCARE SYSTEM TO PREVENT PATIENT TIME

Hadoop is best framework which is used to pre-processing the huge amount of data and gets right prediction from historical data in little time using their scripting languages like Pig. Kafka is also framework which is used to streaming data.

IV.CONCLUSION

In this paper, [1] a random forest in PTTP typical implement constructed on big data and apache spark related is proposed. The line waiting time of each management job is guess based on the trained PTTP model dataset. General investigation and determination results show that our PTTP algorithm implementation accomplishes high correctness and performance. Hospitals' information quantity is rising each day. As a result, [12, 7, and 15] an incremental PTTP algorithm constructed on flow information and a supplementary suitable prediction .The prediction time invention out which is used to testing purpose. To reduction patient waiting time using PTTP model in big environment.

V.REFERENCES

[1]" A Parallel Patient Treatment Time Prediction Algorithm and Its Applications in Hospital Queuing-Recommendation in a Big Data Environment" , 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).

[2] ``Trees weighting random forest method for classifying high-dimensional noisy data," in *Proc. IEEE 7th Int. Conf. e-Business Eng. (ICEBE)*, H. B. Li, W. Wang, H. W. Ding, and J. Dong, Nov. 2010, pp. 160_163.

[3] ``Analysis of a random forests model," *J. Mach. Learn. Res.*,G. Biau, vol. 13, no. 1, pp. 1063_1095, Apr. 2012.

[4] ``Correlation based splitting criterionin multi branch decision tree," *Central Eur. J. Comput. Sci.*, N. Salehi-Moghaddami, H. S. Yazdi, and H. Poostchi, vol. 1, no. 2, pp. 205_220, Jun. 2011.

[5] ``https://www.javatpoint.com/hdfs" By javatpoint.

[6]F.Hosseinkhah, H.Ashktorab, R.Veen, M. M. Owrang O., Challenges in Data Mining on Medical Databases IGI Global (2009) pp 502-511

[7] S.K. Wasan, V. Bhatnagar, H.Kaur, The Impact Of Data Mining Techniques On Medical Diagnostics, Data Science Journal, Volume 5, (2006) pp. 119-126

[8] " *Mach. Learn.*, vol. 45, no. 1, pp. 5_32, L. Breiman, ``Random forests, Oct. 2001.

[9] ``Fast action detection via discriminative random forest voting and top-K sub volume

search," *IEEE Trans. Multimedia*, G. Yu, N. A. Goussies, J. Yuan, and Z. Liu, vol. 13, no. 3, pp. 507_517, Jun. 2011.

[10] ``Robust and accurate shape model matching using random forest regression-voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, vol. 37, no. 9, pp. 1862_1874, Sep. 2015.

[11] ``Big data analytics framework for peer-to-peer botnet detection using random forests," *Inf. Sci.*, K. Singh, S. C. Guntuku, A. Thakur, and C. Hota, vol. 278, pp. 488_497, Sep. 2014.

[12] ``Dynamic random forests," *Pattern Recognit. Lett.*, S. Bernard, S. Adam, and L. Heutte, vol. 33, no. 12, pp. 1580_1586, Sep. 2012.

[13] ``Self-adaptive induction of regression trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, R. Fidalgo-Merino and M. Nunez, vol. 33, no. 8,pp. 1659_1672, Aug. 2011.

[14] ``Parallel boosted regression trees for Web search ranking," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, S. Tyree, K. Q. Weinberger, K. Agrawal, and J. Paykin, 2012, pp. 387_396.

[15] ``KASR: A keyword-aware service recommendation method on MapReduce for big data applications," *IEEE Trans. Parallel Distrib. Syst.*, S. Meng, W. Dou, X. Zhang, and J. Chen, vol. 25, no. 12, pp. 3221_3231, Dec. 2014.

[16] ``Travel recommendation by mining people attributes and travel group types from community-contributed photos," *IEEE Trans. Multimedia*, Y.-Y. Chen, A.-J. Cheng, and W. H. Hsu, vol. 15, no. 6, pp. 1283_1295, Oct. 2013.

[17] ``Bayesian-inference-based recommendation in online social networks," *IEEE Trans. Parallel Distrib. Syst.*, X. Yang, Y. Guo, and Y. Liu, vol. 24, no. 4, pp. 642_651, Apr. 2013.

[18] ``New recommendation techniques for multi criteria rating systems," *IEEE Intell. Syst.*, G. Adomavicius and Y. Kwon, vol. 22, no. 3, pp. 48_55, May/Jun. 2007.

[19] ``Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, G. Adomavicius and A. Tuzhilin, vol. 17, no. 6, pp. 734_749, Jun. 2005.

[20] ``Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, vol. 26, no. 1, pp. 97_107, Jan. 2014.

[21] ``MapReduce: Simplified data processing on large clusters," *Commun. ACM*, J. Dean and S. Ghemawat, vol. 51, no. 1, pp. 107_113,Jan. 2008.

[22] ``Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in *Proc. USENIX NSDI*,2012, M. Zaharia *et al.*, pp. 1_14.

[23] ``A hybrid chemical reaction optimization scheme for task scheduling on heterogeneous computing systems," *IEEE Trans. Parallel Distrib. Syst.*, Y. Xu, K. Li, L. He, L. Zhang, and K. Li, vol. 26, no. 12, pp. 3208_3222, Dec. 2015.

[24] ``Scheduling precedence constrained stochastic tasks on heterogeneous cluster systems," *IEEE Trans. Comput.*, K. Li, X. Tang, B. Veeravalli, and K. Li, vol. 64, no. 1, pp. 191_204, Jan. 2015.

[25] ``An advanced MapReduce: Cloud MapReduce, enhancements and applications," *IEEE Trans. Netw. Service Manage.*, D. Dahiphale *et al.*, vol. 11, no. 1, pp. 101_115, Mar. 2014.

[26] ``Fast and interactive analytics over hadoop data with spark," in *Proc. USENIX NSDI*, 2012, M. Zaharia *et al.*, pp. 45_51.

[27] ``Introduction to Big Data," Haifeng Li ADP Innovation Labs 135 W 18th ST New York, NY 10011 haifeng.li@adp.com., October 31, 2015.

[28]``https://www.tutorialspoint.com/hadoop,"Hadoop Tutorials, by tutorials library.

[29] ``MapReduce: Simplified Data Processing on Large Clusters," Jeffrey Dean and Sanjay Ghemawat jeff@google.com, sanjay@google.com.

[30] ``Hadoop: The Definitive Guide," *Tom White*.

[31] ``HC-CART: A parallel system implementation of data mining classification and regression tree (CART) algorithm on a multi-FPGA system," *ACM Trans. Archit. Code Optim.*, G. Chrysos, P. Dagritzikos, I. Papaefstathiou, and A. Dollas, vol. 9, no. 4, pp. 47:1_47:25, Jan. 2013.

[32] ``A new framework for distributed boosting algorithm," in *Proc. Future Generat. Commun. Netw. (FGCN)*, N. T. Van Uyen and T. C. Chung, Dec. 2007, pp. 420_423.

[33] ``A streaming parallel decision tree algorithm," *J. Mach. Learn. Res.*, Y. Ben-Haim and E. Tom-Tov, vol. 11, no. 1, pp. 849_872, Oct. 2010.

[34] "Patient waiting time prediction in hospital queuing system using improved random forest in big data" Prashant Patil; Sanjay Thakur 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) Year:2019 | Volume:1 | Conference Paper | Publisher: IEEE

