RESEARCH ARTICLE   OPEN ACCESS

SURVEY ON PATIENT WAITING TIME IN HOSPITAL QUEUING SYSTEM IN A BIG DATA

Mr. Prashant Patil.
Department of Computer Engg, LKCT, Indore
RGPV University, India

Dr.Sanjay Thakur.
Department of Computer Engg, LKCT, Indore
RGPV University, India

ABSTRACT

Due to the huge increase in population hospitals are overcrowded because of this it becomes difficult for a hospital management system to control or to minimize patient waiting time while getting treatment in hospital. While taking treatment doctors mentioned several different tests to diagnose the disease to give proper treatment. Thus while evaluating this entire test we have to wait in a queue. A patient has to wait until the entire patient before him or she gets treated. Unnecessary waiting time not only waste patient time but also give frustration during waiting in the queue. It would be more convenient if a patient could get the predicted waiting time and treatment plan on an application that shows the treatment plan and predicted waiting time in real time. Therefore to increase the efficiency and to meet the patent requirement we come up with a new technique called PTTP with HQR i.e. Hospital Queuing-Recommendation System is developed. In this method, the PTTP algorithm predicts the treatment time based on hospital data. Based on this waiting time HQR recommend the treatment plan for the patient. As the patient data is off in a huge amount we have use Apache Hadoop to achieve the goal.

Index Terms - Big data, Pig, Data Mining Technique, Hospital Historical dataset, PTTP algorithm, Random forest.

I.INTRODUCTION

Appropriate to the enormous rise in residents hospitals are congested because of this it happens to complex hospital management organizations to reduce patient waiting time while receiving treatment in hospital. The organization of patient queue and guess of waiting time is an extremely harsh job. As the waiting time estimation require the earlier information of the time taken by different tasks to complete such as time taken by Injection [1], Medicine, X-ray, CT-scan, blood test, sugar test, etc. usually a patient has to go from side to side number of test during treatment point out by doctor which is depending on his or her checkup situation. Consequently, a patient has to stay for the time until all the patient history to him gets treated [17, 13]. Thus avoidable waiting for such a long time misuse the patient time and also gives disturbance.

In this dissertation, we present a way to decrease the waiting time of patient [19] and to assist hospital association deal to expertly state the patient row system. To achieve this job we have regarded as patient information from hospitals such as patient age, gender, disease type, doctor name, task name, treatment start time and treatment end time. At the beginning of this information set, we have designed predicted time for our system means the training phase completed using pig script to implementation of random forest in the PTTP model [7, 13]. The estimation for predict time depends on several features such as gender, age group, treatment start time and end time, etc. that talk about the PTTP model with hospitals significant information. The average waiting time is calculated by this algorithm. And depending on this waiting time Hospital Queuing arrangement predicts the treatment arrangement for the patient to decrease the waiting time throughout the number of test progression. Because of the big information accessibility from the patient narration we have to believe huge information for flexibility and facility. To calculate the waiting time a superior report of the Random Forest Algorithm is used to prepare the planned PTTP structure [22] [7, 11]. Our major objective is to grant the most excellent quality treatment, within time. We would similar to assist you to securely and effectively take delivery of your hospital system by the PTTP model using a random forest using Pig Script [31,30] in a big environment.

II.RELATED 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 the 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 the below characteristics:

i) Its correctness is excellent.
ii) It's comparatively strong to noise.
iii) It's faster than bagging or boosting.
iv) It gives a helpful domestic approximation of fault, might, association and erratic significance.
v) It's straightforward and simply 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 big data; it is a collection of huge datasets that cannot be processed using conventional computing methods. Big data is not simply a data; rather it has become an entire theme, which occupies a variety of tools, methods, and frameworks. Hadoop File System was developed by a 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 an extremely big amount of data and provides easier access. To store such enormous data, the records are stored diagonally various 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], the 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

To improve the accuracy of the data analysis with continuous features, different optimization methods of classification algorithms are proposed. There are special kinds of studies for data mining techniques in medical databases. We identify the following categories:

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 SUMMARIZE REVIEWS AND CHALLENGES IN MINING MEDICAL/HOSPITAL DATA IN GENERAL.

F. Hosseinkhah et al. [32] represent challenges in data mining on medical databases. The 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. STUDIES OF CLASSIFICATION TECHNIQUES IN HADOOP USED FOR PREDICTION.

The **Self-Adaptive Induction** algorithm for the incremental construction of binary regression trees was presented by R. Fidalgo-Merino and M. Nunez. [2].

Tyree, K. Q. Weinberger, K. Agrawal, and J. Paykin [3] introduced a parallel boosted **Regression Tree** algorithm for web search ranking.

In G. Chrysos, P. Dagritzikos, I. Papaefstathiou, and A. Dollas [5] proposed **Classification And Regression Tree (CART)** algorithm.

The **Random Forest Algorithm** [7] is an ensemble classifier algorithm based on a decision tree, which is a suitable data-mining algorithm for big data.

Keyword-Aware Service proposition technique on MapReduce for big information applications. A travel recommendation algorithm that mines people's attributes were proposed in [15].

Yang et. al. [16] introduced a **Bayesian-Inference-Based Recommendation** system for online social networks, in which

a user propagates a content rating query along with the social network to his direct and indirect friends.

3. STUDIES THAT PRESENT NEW TECHNOLOGIES AND ALGORITHMS

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 the best framework that is used to preprocessing the huge amount of data and get the right prediction from historical data in little time using their scripting languages like Pig. Kafka is also a framework that is used for streaming data.

IV.CONCLUSION

In this paper, [1] a random forest in the PTTP model implement based on big data. The queue waiting time of every treatment task is a guess based on the trained PTTP model dataset. General experimentation and purpose results show that our PTTP algorithm implementation achieves high accuracy and presentation. Hospitals' information quantity is rising every day. As a result, [12, 7, and 15] an incremental PTTP algorithm based on torrent information and a more convenient prediction. The prediction time finds out which is used for testing purposes. To decrease patient waiting time using the PTTP model in a 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] ``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.

[3] ``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.

[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] ``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.

[6] ``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.

[7] ``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.

[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] ``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.

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

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

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

[33] 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