ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Hospital Queueing Recommendation System Using Spark Streaming For Big Data

Sneha Girolla, Sunil Rathod

PG Student, Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Pune, Maharashtra, India

Professor, Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Pune, Maharashtra, India

ABSTRACT: Effective patient queue management to attend patient wait delays and patient overcrowding is one amongst the key challenges sweet-faced by hospitals. Gratuitous and annoying waits for long periods end in substantial human resource and time wastage and increase the frustration endured by patients. For every patient within the queue, the overall treatment time of all the patients before him is that the time that he should wait. It'd be convenient and desirable if the patients may receive the foremost economical treatment arrange and recognize the anticipated waiting time through a mobile application that updates in real time. Therefore, we have a tendency to propose a Patient Treatment Time Prediction (PTTP) algorithm program to predict the waiting time for every treatment task for a patient. We have a tendency to use realistic patient knowledge from varied hospitals to get a patient treatment time model for every task. Supported this large-scale, realistic dataset, the treatment time for every patient within the current queue of every task is expected. Supported the anticipated waiting time, a Hospital Queuing-Recommendation (HQR) system is developed. HQR calculates Associate in Nursing predicts an economical and convenient treatment arrange counseled for the patient. Owing to the large-scale, realistic dataset and also the demand for period of time response, the PTTP algorithmic program and HQR system mandate potency and low-latency response. We have a tendency to use Associate in Nursing Apache Spark-based cloud implementation at the National Supercomputing Center in Changsha to attain the for mentioned goals. Intensive experimentation and simulation results demonstrate the effectiveness and pertinence of our planned model to advocate an efficient treatment arrange for patients to attenuate their wait times in hospitals.

KEYWORDS: Apache spark, big data, cloud computing, hospital queuing recommendation, patient treatment time prediction.

I. INTRODUCTION

Most hospitals area unit are overcrowded and lack effective patient queue management. Patient queue management and wait time prediction type a difficult and sophisticated job as a result of every patient may need totally different phases/ operations, like a medical checkup, varied tests, e.g., a sugar level or biopsy, X-rays or a CT scan, minor surgeries, throughout treatment. We tend to decision every of those phases /operations as treatment tasks or tasks during this paper. Every treatment task will have varied time needs for every patient that makes time prediction and recommendation extremely difficult. A patient is sometimes needed to endure examinations, inspections or tests (refereed as tasks) consistent with his condition. In such a case, quite one task could be needed for every patient. a number of the tasks area unit freelance, whereas others may need to attend for the completion of dependent tasks. Most patients should await unpredictable however long periods in queues, anticipating their communicate accomplish every treatment task. During this paper, a Patient Treatment Time Prediction (PTTP) model is trained supported hospitals' historical knowledge. The waiting time of every treatment task is foreseen by PTTP that is that the total of all patients' waiting times within the current queue. Then, consistent with every patient's requested treatment tasks, a Hospital Queuing-Recommendation (HQR) system recommends associate degree economical and convenient treatment set up with the smallest amount waiting time for the patient.

The patient treatment time consumption of every patient within the waiting queue is calculable by the trained PTTP model. We tend to specialize in serving to patients complete their treatment tasks in a very foreseeable time and serving

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

to hospitals schedule every treatment task queue and avoid overcrowded and ineffective queues. We use huge realistic knowledge from varied hospitals to develop a patient treatment time consumption model.Rrealistic patient knowledge area unit analyzed rigorously and strictly supported necessary parameters, like patient treatment begin time, end time, patient age, and detail treatment content for every totally different task. We tend to determine and calculate totally different waiting times for various patients supported their conditions and operations performed throughout treatment.

The accuracy of the information analysis with continuous options, varied improvement strategies of classification and regression algorithms square measure planned. A self-adaptive induction algorithmic rule for the progressive construction of binary regression trees was bestowed in. Tyree et al. introduced a parallel boosted regression tree algorithmic rule for internet search ranking. In, a multi-branch call tree algorithmic rule was planned supported a correlation-splitting criterion. Different improved classification and regression tree strategies. The random forest algorithmic rule is associate ensemble classifier algorithmic rule supported a choice tree, that may be an appropriate information-mining algorithmic rule for giant data. The random forest algorithmic rule is wide employed in several fields like quick action detection via discriminative random forest ballot and Top-K sub volume search strong and correct form model matching exploitation random forest regression ballot, and a giant information analytic framework for peer-to-peer botnet detection exploitation random forests. The experimental leads to these papers demonstrate the effectiveness and relevance of the random forest algorithmic rule. Bernard planned a dynamic coaching technique to enhance the accuracy of the random forest algorithmic rule. In a very random forest technique supported weighted trees was planned to classify high-dimensional screak information. However, the initial random forest algorithmic rule uses a standard direct ballot technique within the ballot method. In such a case, the random forest containing screaky call trees would doubtless cause associate incorrect expected worth for the testing dataset. Varied recommendation algorithms are bestowed and applied in connected fields. Meng et al. planned a keyword-aware service recommendation technique on MapReduce for giant information applications. A travel recommendation algorithmic rule that mines people's attributes and travel-group sorts.

Introduced a Bayesian-inference-based recommendation system for on-line social networks, during which a user propagates a content rating question on the social network to his direct and indirect friends. Adomavicius and Kwon introduced new recommendation techniques for multi-criteria rating systems.

II. LITRATURE SURVEY

1. Self-Adaptive Induction of Regression Trees

Authors: Rau´ l Fidalgo-Merino and Marlon Nu´n˜ez

Description: A progressive regression tree learning methodology has been conferred that is in a position to induce functions that modification at totally different rates and speeds in information streams that embody presence of noise and virtual drift in examples for issues with unknown conditions. As way as we all know, there's no different algorithmic program ready to construct and adapt regression trees to information streams whose underlying operate changes over time likewise as having different dynamics gift (changes in noise, virtual drift, totally different speeds of changes in operate, And partial or complete changes) while not an a priori parameterization.

2. Parallel Boosted Regression Trees for Web Search Ranking

Authors: Stephen Tyree, Kilian Q. Weinberger, Kunal Agrawal

Description: We have bestowed a parallel rule for coaching gradient boosted regression trees. To our data, this can be the primary work that expressly parallelizes the development of regression trees for the aim of gradient boosting. Our approach utilizes the facts that gradient boosting is thought to be strong to the classification accuracy of the weak learners which regression trees square measure of strictly restricted depth. We've shown that our approach provides spectacular (almost linear) speedups on many large-scale web-search information sets with none important sacrifice in accuracy. Our methodology applies to multicore shared-memory systems likewise on distributed setups in clusters and clouds (e.g. Amazon EC2). The distributed setup makes our methodology significantly enticing to real-world settings with terribly giant information sets.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

3. Correlation based splitting criterionin multi branch decision tree
Authors: NimaSalehi-Moghaddami, HadiSadoghiYazdi,HaniehPoostchi
Description: In this paper, a replacement DT algorithmic program is projected, the Correlation primarily based Multi Branch Support Vector (C-MBSV) algorithmic program. Within the projected DT algorithmic program, the correlation idea is employed to search out the most effective ripping variable, and so it's relevant ripping thresholds is such as by employing a SVM. Since the algorithmic program will come multiple thresholds, the tree is split to at least one or additional branches at every node. If the instances during a sub-node belong to at least one category, then the sub-node is considered a leaf node, else we have a tendency to still split the sub-node till all leaf nodes area unit generated. Some leaves that don't result in any choices were conjointly generated, that we have a tendency to refer to as Unclassified Regions.

4. A New Framework for Distributed Boosting Algorithm
Authors: Nguyen Thi Van Uyen
Description: The main contribution of our algorithmic program is that we tend to provide a framework for building a boosting classifier in distributed knowledgebases the image of the one that's engineered on the entire data while not transmittal data between sites. Our technique is acceptable and economical for distributed databases that can't be incorporate at a single\location. Within the future, we are going to do experiments on additional datasets to judge our algorithmic program's stability and on distributed atmosphere to assess the potency of our planned algorithm in long dimension.

5. Fast Action Detection via Discriminative Random Forest Voting and Top-K Sub volume Search
Authors: Gang Yu, Junsong Yuan
Description:We have developed a brand new system for the spatio-temporal localization of human actions in video sequences. The system improves upon the state of the art in 2 aspects. First, we tend to planned a random forest-based pick technique to figure the numerous the interest points, that achieves a multiple orders-of-magnitude speed-up compared with the nearest-neighbour-based marking theme.

III.PROPOSED SYSTEM

In this paper, we have a tendency to propose a PTTP algorithm and HQR system. Considering the time period necessities, enormous data and quality of the system, we have a tendency to use huge information and cloud computing models for potency and measurability. The PTTP algorithmic rule for trained data supports improved Random Forest (RF) algorithmic rule for every treatment task and the waiting time of every task is foretold by using trained PTTP model. Then, HQR recommends an economical and convenient treatment arrangement for every patient. Patients will see the counseled arrangement and foretold waiting time in time period employing a mobile or web application. Intensive experimentation and application results show that the PTTP algorithmic rule achieves high preciseness and performance. The anticipated waiting time of every treatment task is obtained by the PTTP model that is that the total of all patients' probable treatment times within the current queue. An HQR system is planned for anticipated waiting time. A treatment recommendation with an economical and convenient treatment is arranged and also the least waiting time is suggested for every patient. The PTTP algorithm and HQR system squares the measures parallelized on the Apache Spark cloud platform at the National Supercomputing Center in Changsha (NSCC) to realize the said goals. Intensive hospital information square measure holds on within the Apache HBase, and a parallel solution is utilized with the MapReduce and Resilient Distributed Datasets (RDD) programming model. The rest of the paper is organized as follows. Section one reviews connected work. Section three details a PTTP algorithmic rule And an HQR system. The parallel implementation of the PTTP algorithmic rule and HQR system on the Apache Spark cloud surroundings is elaborate in Section four. Experimental results and evaluations square measure given in Section five with relevancy the advice accuracy and performance.

3.1PTTP model basedon improved algorithm RF algorithm

To predict the waiting time for every patient treatment task, the patient treatment time consumption supported completely different patient characteristics and time characteristics should initial be calculated. The time consumption

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

of every treatment task might not be same vary, that varies consistent with the content of tasks and varied circumstances, completely different periods, and completely different conditions of patients. Therefore, we tend to use the RF rule to coach patient treatment time consumption.

Based on each patient and time characteristics so build the PTTP model. As a result of the restrictions of the initial RF rule and also the characteristics of hospital treatment information, the RF rule is improved in four aspects to get a good result from large-scale, high dimensional, continuous, and rackety hospital treatment information. All of the chosen options of the info are utilized in the coaching method, rather than m options designated indiscriminately, as is completed within the original RF rule, as a result of the options of the information are restricted and also the data are already clean of unessential options like patient name, address, and number. As a result of the target variable of the treatment information is patient treatment time consumption that may 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.

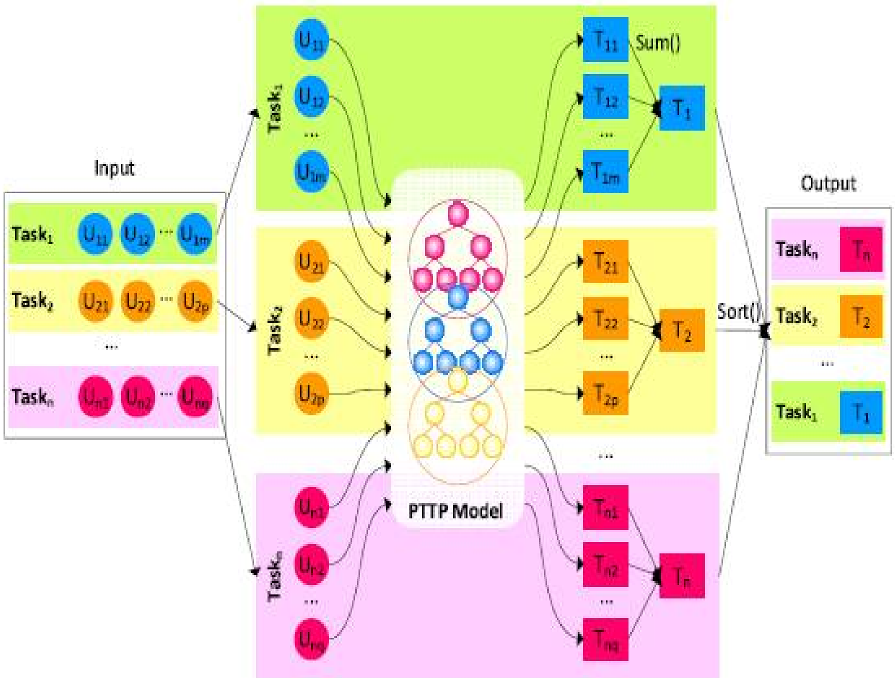

Fig: System Architecture

3.2 Construct Multi-branch for cart model

I have different values, such as the time range (0 - 23) and day of week (Monday - Sunday). Therefore, to construct the regression tree model felicitously, a multi-branch regression tree model instead of two-fork tree model is used constructing the CART, which is the second optimization aspect of the RF algorithm. After the tree node split into two forks by variable *yj*and value *vp*in step (2), the same variable *yj*continues to be selected to calculate the best split point *vpL*for the data in the left branch and *vpR*for the data in the right branch taking.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Website: www.ijircce.com
Vol. 5, Issue 5, May 2017

IV. SIMULATION AND RESULTS

Performance Measures Used:

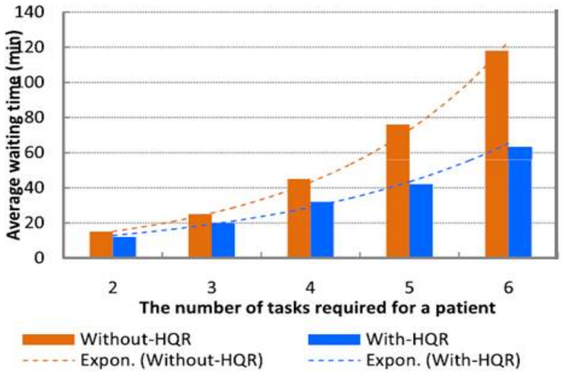

Fig. Average waiting time for patients.

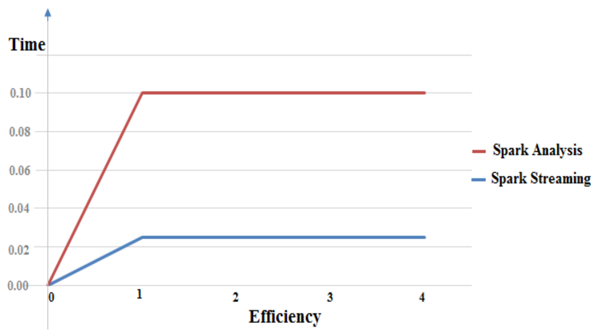

Fig: Time & Efficiency chart

RESULT ANALYSIS:

Input:

Here, Whole System taken many more attribute for the input purpose but here author mainly focuses on the Time and performance of system. Based some few attributes we will getting following analytical result for our proposed system.

EXPECTED RESULT:

No.	Feature Name	Value Range of Each Feature
y1	Patient Gender	Male, Female.
y2	Patient Age	Age
y3	Department	Dept. in Hospital
y4	Dr. Name	All Dr. in Hospital
y5	Task Name	Treatment Task of Patients
y6	Start Time	Start time of Treatment task
y7	End Time	End time of Treatment task
y8	Time Range	Time range of Treatment time in a day.
y9	Time Consumption	End Time – Start Time.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Parameter	Existing	Proposed
A	10	4
B	10	5
C	8	8
D	10	3
E	8	2

Figs: Result Table

Where,

A = Computation Cost.
B = Time Consumption.
C = Scalable.
D = Waiting Time.
E = User Friendly.

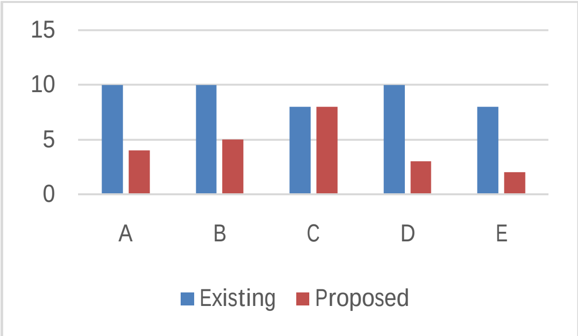

Fig: Time line chart of Result Analysis

V. CONCLUSION

In this paper, a PTTP algorithmic rule supported huge knowledge and therefore the Apache Spark cloud setting is projected. A random forest improvement algorithmic rule is performed for the PTTP model. The queue waiting time of every treatment task is expected supported the trained PTTP model. A parallel HQR system is developed, and an economical and convenient treatment set up is usually recommended for every patient. Intensive experiments and application results show that our PTTP algorithmic rule and HQR system win high preciseness and performance. Hospitals' knowledge volumes are increasing daily. The work of coaching the historical knowledge in every set of hospital guide recommendations is predicted to be terribly high, however it needn't be.

ACKNOWLEDGMENT

We might want to thank the analysts and also distributers for making their assets accessible. We additionally appreciative to commentator for their significant recommendations furthermore thank the school powers for giving the obliged base and backing.

REFERENCES

[1] R. Fidalgo-Merino and M. Nunez, ``Self-adaptive induction of regression trees,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1659_1672, Aug. 2011.
[2] S. Tyree, K. Q. Weinberger, K. Agrawal, and J. Paykin, ``Parallel boosted regression trees for Web search ranking,'' in *Proc. 20th Int. Conf. WorldWide Web (WWW)*, 2012, pp. 387_396.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

[3] N. Salehi-Moghaddami, H. S. Yazdi, and H. Poostchi, ``Correlation based splitting criterionin multi branch decision tree,'' *Central Eur. J. Comput.Sci.*, vol. 1, no. 2, pp. 205_220, Jun. 2011.
[4] G. Chrysos, P. Dagritzikos, I. Papaefstathiou, and A. Dollas, ``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.*, vol. 9, no. 4, pp. 47:1_47:25, Jan. 2013.
[5] 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.
[6] 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.
[7] L. Breiman, ``Random forests,'' *Mach. Learn.*, vol. 45, no. 1, pp. 5_32, Oct. 2001.
[8] G. Yu, N. A. Goussies, J. Yuan, and Z. Liu, ``Fast action detection via discriminative random forest voting and top-K subvolume search,'' *IEEETrans. Multimedia*, vol. 13, no. 3, pp. 507_517, Jun. 2011.
[9] 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.
[10] 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. 488_497, Sep. 2014.