

# Patient Treatment Time Prediction using Interdomain Communication in Big Data and Cloud Computing

Priyanka C M<sup>1</sup>, Prof. Chetana Srinivas<sup>2</sup>M.Tech [SCS,2<sup>nd</sup> Year], Department of Computer Science & Engineering, EWIT, Bengaluru, India<sup>1</sup>Associate Professor, Department of Computer Science & Engineering, EWIT, Bengaluru, India<sup>2</sup>

**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. Therefore, we propose a Patient Treatment Time Prediction (PTTP) algorithm to predict the waiting time for each treatment task for a patient. HQR calculates and predicts an efficient and convenient treatment plan recommended for the patient. Ciphertext-policy attribute-based encryption (CP-ABE) has been a preferred encryption technology to solve the challenging problem of secure data sharing in cloud computing. An efficient file hierarchy attribute-based encryption scheme is proposed in cloud computing. The layered access structures are integrated into a single access structure, and then, the hierarchical files are encrypted with the integrated access structure. The ciphertext components related to attributes could be shared by the files. Therefore, both ciphertext storage and time cost of encryption are saved. The proposed scheme is highly efficient in terms of encryption and decryption.

**KEYWORDS:** Big Data, Cloud Computing, Patient Treatment Time Prediction, Hospital queuing Recommendation

## I. INTRODUCTION

Currently, most hospitals are overcrowded and there is no efficient 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 waiting time of each task is predicted in real time. PTTP algorithm is proposed as learning algorithm for calculating the waiting time. 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”. Now in present days very less amount of data is generated in structured form as compared to unstructured data e.g. Text files, sensor data, log data, web data, social networking data or different varieties of data. With the burgeoning of network technology and mobile terminal, online data sharing has become a new “pet”, such as Facebook, MySpace, and Badoo. Meanwhile, cloud computing [17]–[21] is one of the most promising application platforms to solve the explosive expanding of data sharing. In cloud computing, to protect data from leaking, users need to encrypt their data before being shared. Access control [22], [23] is paramount as it is the first line of defence that prevents unauthorized access to the shared data. Recently, attribute-based encryption(ABE)

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

[24]–[26] has been attracted much more attentions since it can keep data privacy and realize fine-grained, one-to-many, and non interactive access control. Ciphertext-policy attribute based encryption (CP-ABE) is one of feasible schemes which has much more flexibility and is more suitable for general applications.

## II. RELATED WORK

To improve the accuracy of the data analysis with continuous features, various optimization methods of classification and regression algorithms are proposed. A self-adaptive induction algorithm for the incremental construction of binary regression trees was presented in [1]. Tyree *et al.* [2] introduced a parallel boosted regression tree algorithm for web search ranking. In [3], a multi-branch decision tree algorithm was proposed based on a correlation-splitting criterion. Other improved classification and regression tree methods were proposed in [4]. The random forest algorithm is an ensemble classifier algorithm based on a decision tree, which is a suitable data-mining algorithm for big data. The random forest algorithm is widely used in many fields such as fast action detection via discriminative random forest voting and Top-K sub volume search, robust and accurate shape model matching using random forest regression voting, and a big data analytic framework for peer-to-peer botnet detection using random forests [10]. The experimental results in these papers demonstrate the effectiveness and applicability of the random forest algorithm.

Bernard proposed a dynamic training method to improve the accuracy of the random forest algorithm. In, a random forest method based on weighted trees was proposed to classify high-dimensional noisy data. However, the original random forest algorithm uses a traditional direct voting method in the voting process. In such a case, the random forest containing noisy decision trees would likely lead to an incorrect predicted value for the testing dataset. Various recommendation algorithms have been presented and applied in related fields. Meng *et al.* proposed a keyword-aware service recommendation method on MapReduce for big data applications. A travel recommendation algorithm that mines people's attributes and travel-group types was 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 the social network to his direct and indirect friends. To predict the waiting time for each treatment task, we use the random forest algorithm to train the patient treatment time consumption based on both patient and time characteristics and then build the PTPP model. Because patient treatment time consumption is a continuous variable, a Classification And Regression Tree (CART) model is used as a meta-classifier in the RF algorithm. Because of the shortcomings of the original RF algorithm and the characteristics of the patient data, in this paper, the RF algorithm is improved in 4 aspects to obtain an effective result from large-scale, high dimensional, continuous, and noisy patient data. Compared with the original RF algorithm, our PTPP algorithm based on an improved RF algorithm has significant advantages in terms of accuracy and performance. Moreover, there is no existing research on hospital queuing management and recommendations. Therefore, we propose an HQR system based on the PTPP model. To the best of our knowledge, this paper is the first attempt to solve the problem of patient waiting time for hospital queuing service computing. A treatment queuing recommendation with an efficient and convenient treatment plan and the least waiting time is recommended for each patient. Sahai and Waters [24] proposed fuzzy Identity-Based Encryption (IBE) in 2005, which was the prototype of ABE. Latterly, a variant of ABE named CP-ABE [27], [28], was proposed.

## III. ALGORITHM

To build the PTPP model based on both, patient and time characteristics, a PTPP algorithm is proposed. The PTPP model is based on an improved RF algorithm and is trained from the massive, complex, and noisy hospital treatment data. Most of the data in hospitals are massive, unstructured, and high dimensional. Hospitals produce a huge amount of business data every day that contain a great deal of information, such as patient information, medical activity information, time, treatment department, and detailed information of the treatment task. The time consumption of the treatment tasks in each department might not lie in the same range, which can vary according to the content of tasks

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

and various circumstances, different periods, and different conditions of patients. There are strict time requirements for hospital queuing management and recommendation.

**Algorithm:** Process of the Random Forest-Based PTTP Algorithm

**Input:**

$S_{Train}$ : the training datasets;  
 $k$ : the number of CART trees in the RF model.

**Output:**

$PTTP_{RF}$ : the PTTP model based on the RF algorithm.

```

1: for  $i = 1$  to  $k$  do
2:   create training subset  $s_{traini}$   $\leftarrow sampling(S_{Train})$ ;
3:   create OOB subset  $s_{OOBi}$   $\leftarrow (S_{Train} - s_{traini})$ ;
4:   create an empty CART tree  $h_i$ ;
5:   for each independent variable  $y_j$  in  $s_{traini}$  do
6:     calculate candidate split points  $v_s \leftarrow y_j$ ;
7:     for each  $v_p$  in  $v_s$  do
8:       calculate the best split point  $(y_j, v_p) \leftarrow$ 
          $arg \min[\sum_{x \in R_L} (y_i - c_L)^2 + \sum_{x \in R_R} (y_i - c_R)^2]$ ;
9:     end for
10:    append node  $Node_{(y_j, v_p)}$  to  $h_i$ ;
11:    split data for left branch  $R_{L(y_j, v_p)} \leftarrow \{x/y_j \leq v_p\}$ ;
12:    split data for right branch  $R_{R(y_j, v_p)} \leftarrow \{x/y_j > v_p\}$ ;
13:    for each data  $R$  in  $\{R_{L(y_j, v_p)}, R_{R(y_j, v_p)}\}$  do
14:      calculate  $\emptyset(v_{pl}/y_j) \leftarrow \max_i \emptyset(v_i/y_j)$ ;
15:      if  $(\emptyset(v_{p(LR)}/y_j) \geq \emptyset(v_p/y_j))$  then
16:        append subnode  $Node_{(y_j, v_{p(LR)})}$  to  $Node_{(y_j, v_p)}$  as
          multi-branch;
17:        split data to two forks  $R_{L(y_j, v_{pL})}$  and  $R_{R(y_j, v_{pR})}$ ;
18:      else
19:        collect cleaned data for leaf node  $D_{leaf} \leftarrow$ 
           $(IL \leq y_j \leq OL)$ ;
20:        calculate mean value of leaf node  $c \leftarrow$ 
           $1/k \sum D_{leaf}$ ;
21:      end if
22:    end for
23:    remove  $y_j$  from  $s_{traini}$ ;
24:  end for
25:  calculate accuracy
    $CA_i \leftarrow (I(h_i(x)=y))/(I(h_i(x)=y) + \sum I(h_i(x)=z))$  for  $h_i$ 
   by testing  $s_{OOBi}$ ;
26: end for
27:  $PTTP_{RF} \leftarrow H(X, \Theta_j) \leftarrow 1/k \sum_{i=1}^k [CA_i \times h_i]$ ;
28: return  $PTTP_{RF}$ .
```

## IV. SYSTEM MODEL

- A PTTP algorithm is proposed based on an improved Random Forest (RF) algorithm. The predicted waiting time of each treatment task is obtained by the PTTP model, which is the sum of all patients' probable treatment times in the current queue.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

- An HQR system is proposed based on the predicted waiting time. A treatment recommendation with an efficient and convenient treatment plan and the least waiting time is recommended for each patient.
- The PTPP algorithm and HQR system are parallelized. The parallel solution is employed with the MapReduce and Resilient Distributed Datasets (RDD) programming model.  
In the RDD programming model, each RDD object supports two types of operations, i.e., transformation and action. Transformation operations include a series of operations on an RDD object, such as *map()*, *filter()*, *flatMap()*, *mapPartitions()*, *union()*, and *join()*. Then, a new RDD object is returned from each transformation operation. Action operations include a series of operations on an RDD object, such as *reduce()*, *collect()*, *count()*, and *countByKey()*, that compute a result and callback to the driver program or save it to an external storage system.
- We propose the layered model of access structure to solve the problem of multiple hierarchical files sharing. The files are encrypted with one integrated access structure.
- Security involves two aspects: file ciphertext confidentiality and content key ciphertext confidentiality.

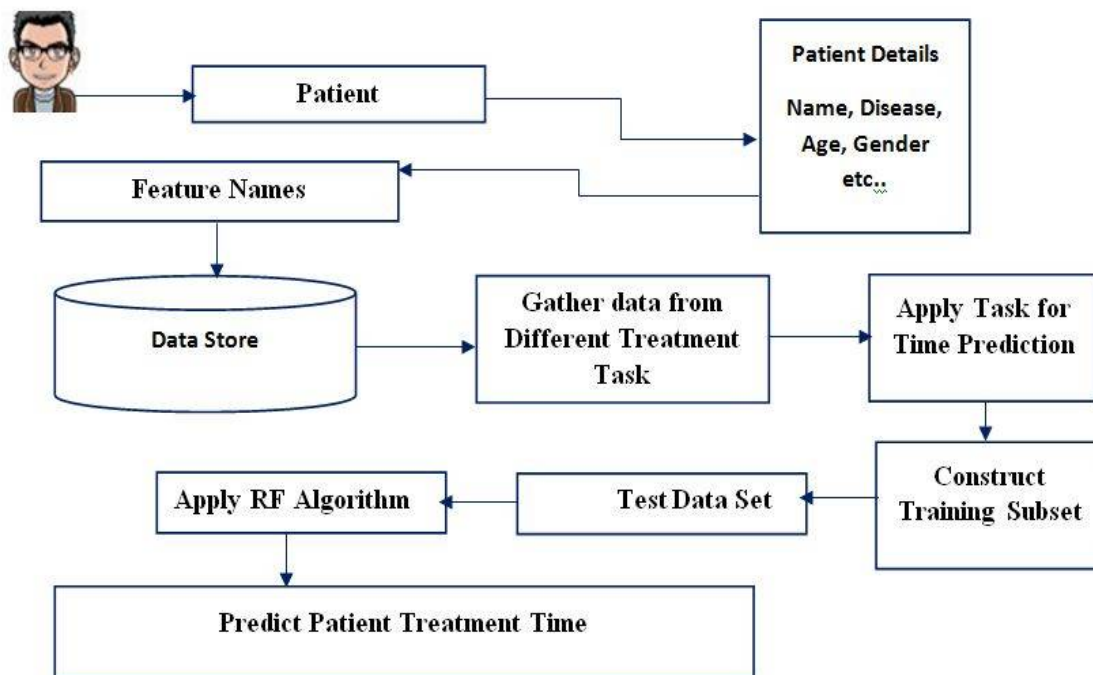


Figure 1: System architecture

**Patient:** It includes patient details such as name, disease, age, gender etc.,

**Data Store:** Here all the patient details are stored in database.

**Time Prediction:** Gather all the data from different treatment task from database and apply the task for time prediction

**Training Subset:** Construct the training subset for each task from time prediction and test dataset.

**Patient treatment time prediction:** Apply Random Forest algorithm for test dataset.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

## V. CONCLUSION

In this paper, a PTPP algorithm based on big data and the cloud environment is proposed. A random forest optimization algorithm is performed for the PTPP model. The queue waiting time of each treatment task is predicted based on the trained PTPP 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 PTPP algorithm and HQR system achieve high precision and performance. Hospitals' data volumes are increasing every day. PTPP algorithm can reduce the influence of noisy data effectively and achieve good robustness.

The hierarchical files are encrypted with an integrated access structure and the ciphertext components related to attributes could be shared by the files. Therefore, both ciphertext storage and time cost of encryption are saved. The proposed scheme has an advantage that users can decrypt all authorization files by computing secret key once.

## ACKNOWLEDGEMENT

First and foremost, Our sincere thanks to Dr. Arun Biradar, Head, Department of Computer Science and Engineering, for his valuable suggestions and guidance. We express our sincere gratitude to our Principal, Dr. K Channakeshavalu, for his constant support and encouragement in preparation of our work. And we thank all the faculties and East-West Institute for their support.

## 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. World Wide Web (WWW)*, 2012, pp. 387\_396.
- [3] 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.
- [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," *IEEE Trans. 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.
- [11] S. Bernard, S. Adam, and L. Heutte, "Dynamic random forests," *Pattern Recognit. Lett.*, vol. 33, no. 12, pp. 1580\_1586, Sep. 2012.
- [12] H. B. Li, W. Wang, H. W. Ding, and J. Dong, "Trees weighting random forest method for classifying high-dimensional noisy data," in *Proc. IEEE 7th Int. Conf. e-Business Eng. (ICEBE)*, Nov. 2010, pp. 160\_163.
- [13] G. Biau, "Analysis of a random forests model," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 1063\_1095, Apr. 2012.
- [14] 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. 3221\_3231, Dec. 2014.
- [15] Y.-Y. Chen, A.-J. Cheng, and W. H. Hsu, "Travel recommendation by mining people attributes and travel group types from community-contributed photos," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1283\_1295, Oct. 2013.
- [16] X. Yang, Y. Guo, and Y. Liu, "Bayesian-inference-based recommendation in online social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 642\_651, Apr. 2013.
- [17] C.-K. Chu, W.-T. Zhu, J. Han, J.-K. Liu, J. Xu, and J. Zhou, "Security concerns in popular cloud storage services," *IEEE Pervasive Comput.*, vol. 12, no. 4, pp. 50\_57, Oct./Dec. 2013.
- [18] T. Jiang, X. Chen, J. Li, D. S. Wong, J. Ma, and J. Liu, "TIMER: Secure and reliable cloud storage against data re outsourcing," in *Proc. 10th Int. Conf. Inf. Secur. Pract. Exper.*, vol. 8434, May 2014, pp. 346\_358.
- [19] K. Liang, J. K. Liu, D. S. Wong, and W. Susilo, "An efficient cloudbased revocable identity-based proxy re-encryption scheme for public clouds data sharing," in *Proc. 19th Eur. Symp. Res. Comput. Secur.*, vol. 8712, Sep. 2014, pp. 257\_272.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

- [20] T. H. Yuen, Y. Zhang, S. M. Yiu, and J. K. Liu, "Identity-based encryption with post-challenge auxiliary inputs for secure cloud applications and sensor networks," in *Proc. 19th Eur. Symp. Res. Comput. Secur.*, vol. 8712, Sep. 2014, pp. 130–147.
- [21] K. Liang *et al.*, "A DFA-based functional proxy re-encryption scheme for secure public cloud data sharing," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 10, pp. 1667–1680, Oct. 2014.
- [22] T. H. Yuen, J. K. Liu, M. H. Au, X. Huang, W. Susilo, and J. Zhou, "*k*-times attribute-based anonymous access control for cloud computing," *IEEE Trans. Comput.*, vol. 64, no. 9, pp. 2595–2608, Sep. 2015.
- [23] J. K. Liu, M. H. Au, X. Huang, R. Lu, and J. Li, "Fine-grained twofactor access control for Web-based cloud computing services," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 484–497, Mar. 2016.
- [24] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Advances in Cryptology*. Berlin, Germany: Springer, May 2005, pp. 457–473.
- [25] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attributebased encryption for fine-grained access control of encrypted data," in *Proc. 13<sup>th</sup> ACM Conf. Comput. Commun. Secur.*, Oct. 2006, pp. 89–98.
- [26] W. Zhu, J. Yu, T. Wang, P. Zhang, and W. Xie, "Efficient attribute-based encryption from R-LWE," *Chin. J. Electron.*, vol. 23, no. 4, pp. 778–782, Oct. 2014.
- [27] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attributebased encryption," in *Proc. IEEE Symp. Secur. Privacy*, May 2007, pp. 321–334.
- [28] L. Cheung and C. Newport, "Provably secure ciphertext policy ABE," in *Proc. 14th ACM Conf. Comput. Commun. Secur.*, Oct. 2007, pp. 456–465.