ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

A Hospital Queueing Recommendation System Using Patient Treatment Time Prediction Algorithm

M.NANDHINI [1], V.MATHESWARAN [2]

1. P.G. Student, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

2. Assistant Professor, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

Abstract: Effective patient queue management to minimize patient wait delays and patient overcrowding is one of the major challenges faced by hospitals. Unnecessary and annoying waits for long periods result in substantial human resource and time wastage and increase the frustration endured by patients. For each patient in the queue, the total treatment time of all the patients before him is the time that he must wait. It would be convenient and preferable if the patients could receive the most efficient treatment plan and know the predicted waiting time through a mobile application that updates in real time. Therefore, we propose a Patient Treatment Time Prediction (PTTP) algorithm to predict the waiting time for each treatment task for a patient. We use realistic patient data from various hospitals to obtain a patient treatment time model for each task. Based on this large-scale, realistic dataset, the treatment time for each patient in the current queue of each task is predicted. Based on the predicted waiting time, a Hospital Queuing-Recommendation (HQR) system is developed. HQR calculates and predicts an efficient and convenient treatment plan recommended for the patient. Because of the large-scale, realistic dataset and the requirement for real-time response, the PTTP algorithm and HQR system mandate efficiency and low-latency response.

1. INTRODUCTION

Currently, most hospitals are overcrowded and they are not efficient in providing proper 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 time 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 compare to unstructured data e.g. Text files, sensor data, log data, web data, social networking data or different

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

varieties of data. For Big Data management Hadoop is used.

2. LITERATURE SURVEY

1. Qiang Li, Qi Han, Limin Sun, "Collaborative Recognition of Queuing Behavior on Mobile Phones"2015 IEEE Transactions on Mobile Computing

Most of the smartphones have built in sensors that can measure motion, location, orientation and various other environmental conditions. These sensors can provide raw data with high precision and accuracy for inferring and recognizing queuing behavior. Furthermore we observed that people frequently carry smartphones when they are not at home. We can use a collaborative approach for queuing recognition based on smartphones by following queuing rules(First Come First Server). We can use a prototype of QueueSense with clients on smartphones which uses Android platforms and a server in cloud. Smartphones make use of widely available sensors such as accelerometer, Bluetooth and compass to sense individual activities. Queuing features are calculated based on queuing properties in terms of individual activities and support vector machine(SVM) is used to automatically detect whether the people are queuing or not on smartphones. The cloud backend process multi-lines scenarios and provide estimation of queue length and waiting time. Agglomerative hierarchical clustering is used on server side to divide queuers into different lines based on changing rate of relative position of queuers.

2. Hong-lei-Zhang, Peng-fei Li, Xin-hang Li, Sheng-qiang Chi, Jing-song Li*,"Design and Implementation of Clinical Data Center Based on Hadoop",2015 7th International Conference on Information Technology in Medicine and Education

Hospital Information generally contains an information systems such as Electronic Medical Records(EMR) and Picture archiving And Communication System(PACS).Hospital data is center generally stores the Structured And Unstructured Data. Most data used in the EMR is Structured Data which includes information of a patient, information of an treatment, diagnostic information and the reports. This above data is stored in the hadoop cluster with the help of an JDBC/ODBC interface and then the data is stored in an HDFS with the help of an MAPREDUCE and the HIVE .The Structured data in the HDFS is written using an HIVE and its SQL like Query language HQL. Data is Processed in the system we need to give a connection to the database then we need to check an existence of a table i.e if a Table exists already then we need to add a Partition otherwise we need to create a Table and then we need to add a Partition. Then we need to check an availability of an data if an data is available then just update the data otherwise write the data. and then we need to disconnect with the database.

3. Kapil Pandey Anand Gadwal Prashant Lakkadwala ,"Hadoop Multi Node Cluster Resource Analysis ",2016 Symposium on Colossal Data Analysis and Networking (CDAN)

Hadoop is a framework which provides distributed processing of large data sets across cluster using a simple programming model. Mainly Apache

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

Hadoop Framework consists MapReduce and Hadoop distributed file system. Hadoop distributed file system, as Map reduce provide a simple programming model well as other related projects e.g. Apache Hive, Apache HBase etc. The basically three important parameters of hadoop cluster; they are CPU, MEMORY and DISK. All the nodes are dedicated to work for the hadoop tasks only. This can be useful for efficient management of cluster to provide capacity, scalability and performance of cluster such that provisioning of resources will be efficient.

4. Aditya Bhardwaj, Vanraj,Ankit Kumar, Yogendra Narayan, Pawan Kumar, "Big Data Emerging Technologies: A CaseStudy with Analyzing Twitter Data using Apache Hive", Proceedings of 2015 RAECS UIET Panjab University Chandigarh 21-22nd December 2015

Large amount of data is created by users in daily life. which requires huge amount of storage and various techniques to discover knowledge from data. Hadoop architecture is of two main components HDFS(Hadoop Distributed File Sysytem) and Mapreduce for Big Data Analytics. There are various technologies belongs to Hadoop Hbase for storing large dataset, Apache pig is scripting language for processing of large data set, Hive is designed for OLAP is fast and Scalable, Scoop is used for import and export data from RDBMS to Hadoop, Zookeeper is used for distributed application and flume is for moving large amount of data to centralized data.

[5] Jiawei Han, Yanheng Liu, Xin Sun," A Scalable Random Forest Algorithm Based on MapReduce",2013 IEEE

Random forest is most popular data classification and regression algorithm for machine learning. This system introduced a Scalable Random Forest Algorithm which is based on MapReduce Technique .The algorithm is divided into three stages: initializing, generating and voting. SMRF algorithm has main objective of improving the traditional random forest algorithm based on MapReduce model. SMRF algorithm provide scalable performance, and it can negotiate with the distributed computing environments to decide its trees scale.

2.1 EXISTING SYSTEM

To improve the accuracy of the data analysis with continuous features, various optimization methods of classification and regression algorithms are proposed. 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. Yang et. al. 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. Adomavicius and Kwon introduced new recommendation techniques for multi-criteria rating systems. Adomavicius and Tuzhilin introduced an overview of the current generation of recommendation methods, such as content-based, collaborative, and hybrid recommendation approaches. However, there is no effective prediction algorithm for patient treatment time consumption in the existing studies.

Patient queue management and wait time prediction form a challenging and complex job because each patient might require different operations, such as a checkup, various tests, e.g., a sugar level or blood test, X-rays or a CT scan during treatment. So there are five major methodologies used in this system Big Data management with Historical Dataset, Preprocessing of data, Use Learning Algorithm PTTP(patient Treatment Time prediction)with base of RF (Random Forest) Algorithm Calculate the Waiting Time in Hospital Queue Recommendation. A random forest optimization algorithm is performed for the PTTP model. The queue waiting time of each treatment task is predicted using the trained PTTP model. A parallel HQR system is introduced, and an efficient and convenient treatment plan is recommended for each patient. The patient may undergo various treatment operations such as CT scan, MR scan and a payment task. These set of treatment operations are submitted to decision maker and recommendation module via mobile interface. The predicted waiting Time of all of the treatment tasks is calculated by PTTP model. After this a treatment recommendation with least waiting time is advised

2.1.1 Disadvantages

- There is no effective prediction algorithm for patient treatment time consumption.
- There is no existing research on hospital queuing management and recommendations.

2.2. PROPOSED SYSTEM

It proposes a PTTP algorithm and an HQR system. Considering the real-time requirements, enormous data, and complexity of the system, it employs big data and cloud computing models for efficiency and scalability. The PTTP algorithm is trained based on an improved Random Forest (RF) algorithm for each treatment task, and the waiting time of each task is predicted based on the trained PTTP model. Then, HQR recommends an efficient and convenient treatment plan for each patient. Patients can see the recommended plan and predicted waiting time in real-time using a mobile application. The RF algorithm is improved in 4 aspects to obtain an effective result from large-scale, high dimensional, continuous, and noisy patient data. Therefore, we propose an HQR system based on the PTTP model. This is the first attempt to solve the problem of patient waiting time for hospital queuing service computing.

Random forest is most popular data classification and regression algorithm for machine learning. This system introduced a Scalable Random Forest Algorithm which is based on MapReduce Technique .The algorithm is divided into three stages: initializing, generating and voting. SMRF algorithm has main objective of improving the traditional random forest algorithm based on MapReduce model. SMRF algorithm provide scalable performance, and it can negotiate with the distributed computing environments to decide its trees scale.

2.2.1 Advantage of Proposed System:

- Extensive experimentation and application results show that the PTTP algorithm achieves high precision and performance.
- Compared with the original RF algorithm, PTTP algorithm based on an improved RF algorithm has

significant advantages in terms of accuracy and performance.

- A treatment queuing recommendation with an efficient and convenient treatment plan and the least waiting time is recommended for each patient.

3 SYSTEM MODEL

Prediction based on analysis and processing of massive noisy patient data from various hospitals is a challenging task. Some of the major challenges are the following:

(1) 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. Moreover, because of the manual operation and various unexpected events during treatments, a large amount of incomplete or inconsistent data appears, such as a lack of patient gender and age data, time inconsistencies caused by the time zone settings of medical machines from different manufacturers, and treatment records with only a start time but no end time.

(2) 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 and various circumstances, different periods, and different conditions of patients. For example, in the case of a CT scan task, the time required for an old man is generally longer than that required for a young man.

(3) There are strict time requirements for hospital queuing management and recommendation. The speed of executing the PTTP model and HQR scheme is also critical.

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. 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.

To predict the waiting time for each patient treatment task, the patient treatment time consumption based on different patient characteristics and time characteristics must first be calculated. The time consumption of each treatment task might not lie in same range, which varies according to the content of tasks and various circumstances, different periods, and different conditions of patients. Therefore, the system uses the RF algorithm to train patient treatment time consumption based on both patient and time characteristics and then build the PTTP model. Because of the limitations of the original RF algorithm and the characteristics of hospital treatment data, the RF algorithm is improved in 4 aspects to obtain an effective result from large-scale, high dimensional, continuous, and noisy hospital treatment data.

All of the selected (cleaned) features of the data are used in the training process, instead of m features selected randomly, as is done in the original RF algorithm, because the features of the data are limited and the data are already cleaned of unnecessary features such as patient name, address, and telephone number.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

Because the target variable of the treatment data is patient treatment time consumption, which is a continuous variable, a CART model is used as a meta-classifier in the improved RF algorithm. At the same time, some independent variables of the data are nominal data, which have different values such as time range (0 - 23) and day of week (Monday - Sunday). In such a case, the two-fork tree model of the traditional CART cannot fully reflect the analysis results. Therefore, to construct the regression tree model felicitously, a multi-branch model is proposed for the construction process instead of the two-fork model of the traditional CART algorithm. [7] proposed a secure hash message authentication code. A secure hash message authentication code to avoid certificate revocation list checking is proposed for vehicular ad hoc networks (VANETs). The group signature scheme is widely used in VANETs for secure communication, the existing systems based on group signature scheme provides verification delay in certificate revocation list checking. In order to overcome this delay this paper uses a Hash message authentication code (HMAC).

Although we have removed part of the error in the preprocessing, other types of noisy data might also exist. In some treatment tasks, the time consumption is the time interval between one patient and the next in the same treatment. For example, in a payment task, assume that the operation time point of the last patient in the morning is ‘‘12:00:00’’ and the operation time point of the first patient in the afternoon is ‘‘14:00:00’’. The time consumption of the former is ‘‘7200 (s)’’ and is considered as incorrect data because it is larger than the normal value of ‘‘100 (s)’’. However, the value ‘‘7200 (s)’’ of time consumption has not always been incorrect data, such as in a blood examination task. Therefore, we cannot simply designate one value of time consumption as noisy data; each must be classified according to treatment data features. Then, it must be identified and remove the noisy data. In calculating the average value of the data in each leaf node of the regression tree, noisy data are removed to reduce their influence on accuracy. The original RF algorithm uses a traditional direct voting method in the prediction process. In such a case, a RF containing noisy decision trees would likely lead to an incorrect predicted value for the testing dataset. Therefore, in this paper, a weighted voting method is employed in the prediction process of the RF model. Each tree classifier corresponds to a specified reasonable weight for voting the testing data. A tree classifier that has high accuracy in the training process will have a high voting weight in the prediction process. Hence, the classifier improves the overall classification accuracy of the RF algorithm, and reduces the generalization error. Compared with the original RF algorithm, our PTTP algorithm based on the improved RF algorithm, has significant advantages in terms of accuracy and performance.

Process of the Hospital Queuing Recommendation

Input: X: the treatment data of the current patient;

Output: Ts(X): the recommended tasks with predicted waiting time.

1: create map Ts(X) ← HashMap < string, double>;

2: for each Taski in X do

3: create array Ui[] ← patients-in-waiting of Taski ;

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

4: for each patient Uik in Ui do

5: predict time consumption Tik ← PTTPRF ;

6: end for

7: calculate predicted waiting time Ti ← 1 Wi Pm k=1 Tik ;

8: append waiting time Ts(X) ←< Taski, Ti >;

9: end for

10: sort map Ts(X) in an ascending order;

11: for each < Taski, Ti > in Ts(X) do

12: if (Taski has dependent tasks) then

13: put records of the dependent tasks before Taski ;

14: end if 15: end for

16: return Ts(X).

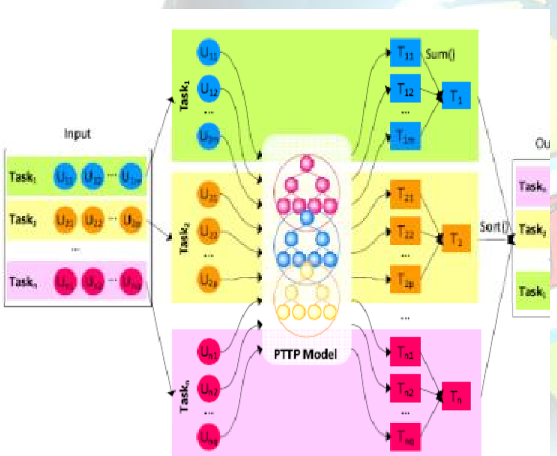

Fig.1 Process of the HQR system based on the PTTP model.

After training the PTTP model for each treatment task using historical hospital treatment data, a PTTP-based hospital queue recommendation system is developed. An efficient and convenient treatment plan is created and recommended to each patient to achieve intelligent triage. Assume that there are various treatment tasks for each patient according to the patient's condition, such as examinations and inspections. Let Tasks= {Task1, Task2, . . . , Taskn} be a set of treatment tasks that the current patient must complete, and let Ui = {Ui1, Ui2, . . . , Uim} be a set of patients in waiting the queue for Taski .

3.1 MODULE DESCRIPTION

DATA PREPROCESSING

Prediction based on analysis and processing of massive noisy patient data from various hospitals is a challenging task. Some of the major challenges are the following: 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. Moreover, because of the manual operation and various unexpected events during treatments, a large amount of incomplete or inconsistent data appears, such as a lack of patient gender and age data, time inconsistencies caused by the time zone settings of medical machines from different manufacturers, and treatment records with only a start time but no end time.

In the pre-processing phase, hospital treatment data from different treatment tasks are gathered. Substantial numbers of patients visit each hospital every day. Let S be a set of patients in a hospital, and a patient who has been registered and his information is represented by si.

CHOOSE THE SAME DIMENSIONS OF THE DATA

The hospital treatment data generated from different treatment tasks have different contents and formats as well as varying dimensions. To train the patient time consumption model for each treatment task, we choose the same features of these data, such as the patient information, the treatment task information, and the time information (start time and end time). Other feature subspaces of the treatment data are not chosen because they are not useful for the PTTP algorithm, such as patient name, telephone number, and address.

PTTP MODEL BASED ON THE IMPROVED RF ALGORITHM

Although we have removed part of the error data in the pre-processing, other types of noisy data mentioned above might exist. Therefore, the third optimization aspect of the RF algorithm is to reduce the impudence that the noisy data have on the algorithm accuracy. A box-plot-based noise removal method is performed in the value calculation of each CART leaf node.

Accuracy and Robustness Analysis

To evaluate the accuracy and robustness of our improved-RF- based PTTP algorithm, it implemented the PTTP algorithm based on the original random forest. The accuracies of the PTTP algorithm and PTTP-ORF algorithm are analysed under different ratios of noisy data. A noise removal method is introduced in the training process of the regression tree model. The effect of noise removal is validated and analysed. Six groups of leaf node data in the regression tree models are discussed in experiments, the specific conditions of the six groups of leaf nodes.

PERFORMANCE EVOLUTION

To evaluate the performance of the PTTP algorithm, four groups of historical hospital treatment data are trained at different scales of the Spark cluster. The sizes of these datasets are 50GB, 100GB, 300GB, and 200GB. The scale of slave nodes of the Spark cluster in each case increases from 5 to 80. By observing the average execution time of the PTTP algorithm in each case, different performances across various cases are compared and analysed. The results are presented. The advantage of the parallel algorithm in cases of large-scale data is greater than in cases of small-scale data. The benefit is more obvious when the number of slave nodes increases. As the number of cluster nodes increases from 5 to 80, the average execution time of the PTTP model decreases from 879 to 285 s for 300GB of data, and decreases from 328 to 81 s for 50GB of data.

REPORT GENERATION

A 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. The queue waiting time of each treatment task I 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. Hospitals' data

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

volumes are increasing every day. The workload of training the historical data in each set of hospital guide recommendations is expected to be very high, but it need not be.

4. CONCLUSION

In this system it ha been reviewed on the technologies which are being used for hospital queuing behavior. Further it introduced different techniques used to implement them. Present methods include hadoop, sql, Hbase, RF algorithm. Later on it compared the different techniques used by researchers in their systems, such as extension in RF algorithm, storing of structure data into the database etc. This comparison will help us in building our system more convenient and useful. From the research we have proposed the system which will predict time required for particular task. For training purpose the system uses RF algorithm and it develops HQR system for patients. Patients can check their waiting time by using mobile phones and thus can reduce their frustration.

FUTURE SCOPE

A 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. The queue 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 PTTP algorithm and HQR system achieve high precision and performance. Hospitals' data volumes are increasing every day. The workload of training the historical data in each set of hospital guide recommendations is expected to be very high, but it need not be. Consequently, an incremental PTTP algorithm based on streaming data and a more convenient recommendation with minimized path-awareness are suggested for future work.

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 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 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.

[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.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 17, March 2017

[7] Christo Ananth, M.Danya Priyadharshini, "A Secure Hash Message Authentication Code to avoid Certificate Revocation list Checking in Vehicular Adhoc networks", International Journal of Applied Engineering Research (IJAER), Volume 10, Special Issue 2, 2015,(1250-1254)

[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.