Research On Fault Prediction Model Based On 5G Data Center

Min Tan[1,2], Guo-hui Li[1], Han Yu[2], Ming Wei[2], Ji-wei Hu[2]

[1] School of Computer Science and Technology, Huazhong University of Science and Technology, China
[2] Wuhan Fiberhome Technical Services Co., Ltd., China
mmpphehe@126.com, 1999010279@hust.edu.cn, yuhan@fiberhome.com, mwei@fiberhome.com, hujiwei@fiberhome.com

Abstract

Intelligent operation and maintenance is leading the innovation in the communications industry. With the opening of the 5G commercial era, the existing management model cannot meet the needs of network operation and maintenance for comprehensive cloud and intelligent environment. This article aims at discovering some important problems of the intelligent deployment in big data centers, building a fault prediction model based on key performance indicators, conducting a deep research in conjunction with scenarios of big data centers and providing a complete and effective solution. The key technology studied in this paper will cover 80% of the total number of data center failures, and the prediction success rate 3 days in advance will reach 95%.

Keywords: Data center, Fault prediction model, Machine learning

1 Introduction

On June 6, 2019, China officially issued 5G commercial licenses. Driven by policy support, technological progress and market demand, 5G industry develops rapidly in China. Various industries are actively exploring 5G application scenarios. Big data centers are becoming more and important as infrastructure.

At the same time, the current epidemic situation has also made an effect on the information transformation. The government and industries have taken lots of online measures to develop 5G applications, and achieved good results [1]. After the epidemic, a large number of enterprises will embark on the road of information technology, which will drive a big increase for the demand of big data centers [2].

According to the general idea of server management, this paper proposes specific solutions, including server information collection, optimization of CPU power consumption of mainstream servers, optimization of power supply efficiency of data centers, design of intelligent power supply models, and intelligent predictive analysis of disk faults to improve server performance and Reduce energy consumption [3].

The big data center is not only a traditional storing room for computers and network equipment, but also a part of public computing facilities that embody new development concepts such as innovation and green [4]. It is a carrier to promote the development of new generation information technology such as 5G, artificial intelligence, industrial Internet, cloud computing, etc [5].

The existing management model of the big data center cannot meet the needs of network operation and maintenance for comprehensive cloud and intelligent environment [6]. New technologies such as AI and cloud computing are urgently needed to promote the upgrading of the Intelligent operation and maintenance industry [7].

Aiming at the positioning difficulties encountered in the data centers, this paper builds a fault prediction model based on key performance indicators. At the same time, it studies the machine learning model to find a better fault prediction solution and provides an efficient solution [8].

2 Related Works

The existing operation and maintenance management system of the big data center often takes the video surveillance system as the core, and integrates a variety of sensors to obtain the operating status of different devices to achieve centralized operation and maintenance management of the data center [9]. This type of system collects real-time information such as temperature, humidity, and cabinet indicator status in the equipment room to provide early warnings of abnormal data, but the accuracy rate is low, and manual inspection is required to correct the error warnings. How to correctly and efficiently obtain the operating status of power supply system, air conditioning system, security

*Corresponding Author: Guo-hui Li; E-mail: 1999010279@hust.edu.cn
DOI: 10.3966/160792642021012201005

system, server and other equipment has become one of the core problems faced by the operation and maintenance management system [10].

The current operation and maintenance mode of the data center is time-consuming, labor-intensive, and inefficient. It can no longer meet the needs of business development [11]. The main problems are reflected in the following aspects:

The traditional operation and maintenance mode has high security risks, poor real-time fault handling, and lost or tampered operation records; in addition, manual mistaken operations may cause a shutdown or crash to servers.

Operation and maintenance cannot be effectively supervised, status of failure cannot be checked in time, and resources cannot be reasonably allocated.

Information cannot be shared. Each maintenance operation is an isolated incident. Valuable information in the process is difficult to share and learn from later work, and the statistical report work is repetitive and burdensome.

In recent years, some technology companies are trying to apply the abnormal detection technology to daily operation and maintenance [12].The technical framework is shown in Figure 1.

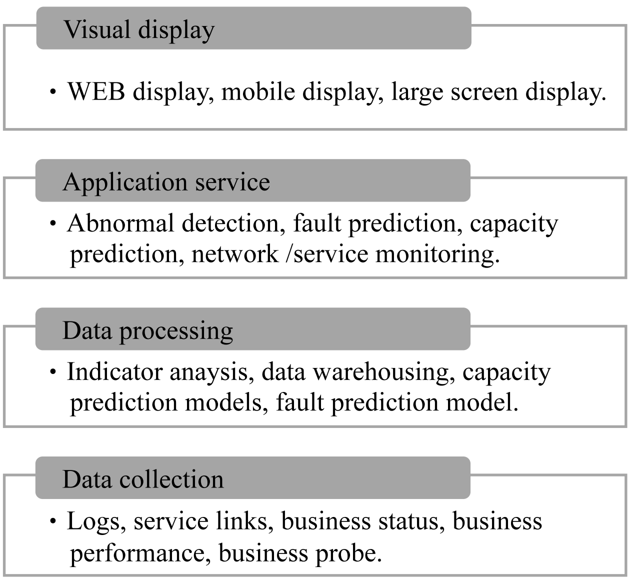

Figure 1. Framework of Abnormal Detection

Abnormal detection technology automatically discovers abnormal fluctuations in the time series data of key indicators through algorithms, and provides decision-making basis for subsequent alarms, automatic stop loss, and root cause analysis [13]. Due to the defects of the existing technology, the accuracy rate and recall rate of anomaly detection algorithms are low, and there are a large number of false positives and false negatives [14]. The main problems include the following:

(a) The frequency of abnormalities is very low. In actual operation and maintenance scenarios, abnormalities rarely occur in business systems, so there are few abnormal records available for analysis.

(b) Anomalous diversity. Because the actual business system is very complex and will be constantly updated and upgraded, resulting in the diversity of abnormal types.

As a commonly used tree model, the GBDT model can naturally perform feature division, feature combination and feature selection on original features, and obtain high-order feature attributes and nonlinear mappings [15]. The logistic regression model has low computational complexity and easy parallel processing, and easy to obtain discrete target values [16]. They have their own characteristics and have been well applied in the prediction of system fault [17].

3 Proposed System

3.1 Principle

With the emergence of 5G networks, big data and cloud computing, network system is faced with massive amounts of data throughput and various requests and operations from a large number of users. For example, massive amounts of data are exchanging in network devices constantly, and applications are receiving requests constantly from a large number of users for posting information, comments, forwarding, and so on.

Compared with traditional data centers, 5G data centers have the characteristics of high density, modularity, flexible networking, and green energy saving. This article focuses on the high-density and modular characteristics of 5G data centers and focuses on failure prediction.

At this time, only the operation and maintenance personnel manually monitor the operation of the entire system or application, and it will be difficult to find the abnormality of the system or application in a timely and comprehensive manner. In order to solve this problem, intelligent operation and maintenance are fully introduced, and machine learning is used to automatically detect abnormal key performance indicators, thereby locating problems that occur in the operation of the system. In this way, the operation and maintenance personnel can quickly solve problems.

Based on data statistics, maintenance personnel need to spend more than 40% of their time on fault locating. It can be said that the speed of network fault locating directly affects the repair efficiency of the entire fault. This project intends to extract key performance indicators that can reflect the operation of the system from the operation and maintenance logs, then build a timing chart of the key performance indicators, and finally use machine learning to automatically detect abnormal key performance indicators, in which way to locate the problems that occur in the operation.

The data sources and key performance indicators are shown in the following Table 1:

Table 1. Key performance indicators

Classification	Module	Indicator
Business	Global statistics	Business request times, Call success times, Failure times
		No abnormal time for business call
		Business statistics
	State statistics	Failure index, including request number statistics, success / failure statistics
		Intranet quality, export delay, DNS response, business resource utilization rate
	Business event	Failed business and call details
Alarm	Business system alarm statistics	Number of Business system alarm
	Monitoring system alarm statistics	Number of Monitoring system alarm
Capacity	Resource pool capacity	CPU utilization
		Total memory utilization
		Total storage utilization
	Resource usage trends	CPU usage trend-statistical minimum and maximum, can increase the growth rate calculation
		Memory usage trend-statistical minimum and maximum, can increase the growth rate calculation
		Storage usage trend-statistical minimum and maximum, can increase the growth rate calculation
Network	Network quality	Network quality status
	Application link	Call between applications
	Intranet matrix	Network quality status between modules

3.2 Implementation

In this paper, the fault location system is divided into two parts: offline model training and online real-time detection.

Offline model training system is shown in Figure 2.

Offline model training system includes following steps:

- Data preprocessing: Including data missing value filling, data normalization and other processing. At the same time, abnormal data oversampling is adopted for data imbalance to achieve data balance.
- Feature Engineering: Extracting the original features, statistical features and wavelet features of the data, for example: The mean, variance and slope of the data in a sliding window. Then integrating these features together to form a feature set.
- Model training: Due to the diversity of key performance indicators, it is difficult for a single model to adapt to all KPIs, so the logistic regression model and GBDT model are used for training, and then the results of each model are weighed to obtain the final result.
- Model evaluation: The model evaluation results consider the data imbalance, so F1-Socre and ROC / AUC / PRC are used to evaluate the performance of the model.

In this paper, by analyzing the historical data of failures, PTN, OTN, OLT, ONU, SDH servers and other theme libraries are constructed according to the type of equipment. Each category of theme library corresponds to a different set of weight coefficients. Through online detection, the accuracy of fault location

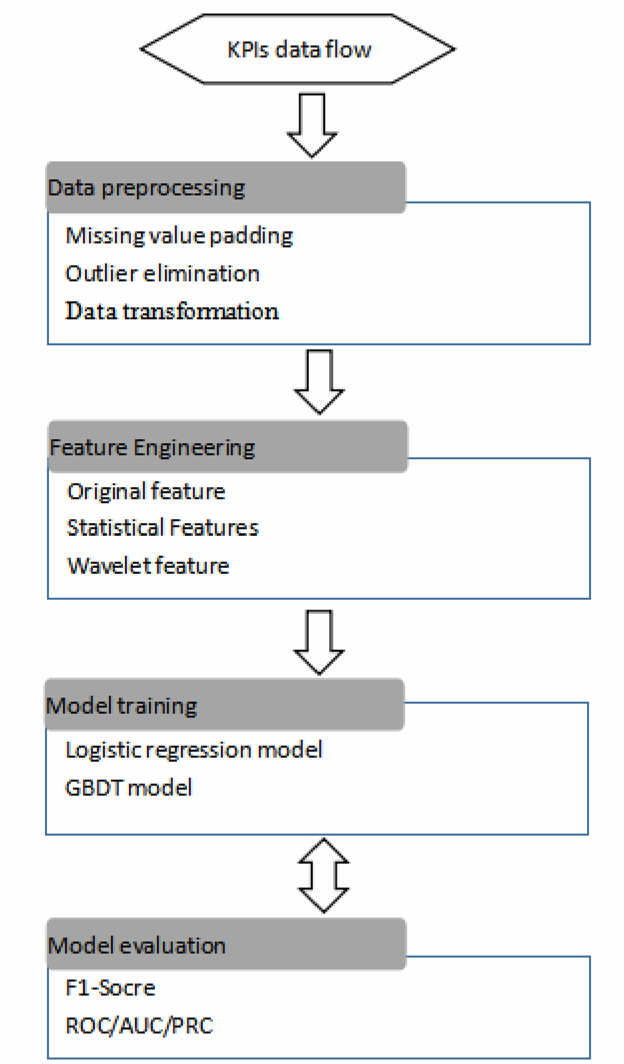

Figure 2. Offline model training system

is continuously improved. The online real-time detection system is shown in Figure 3.

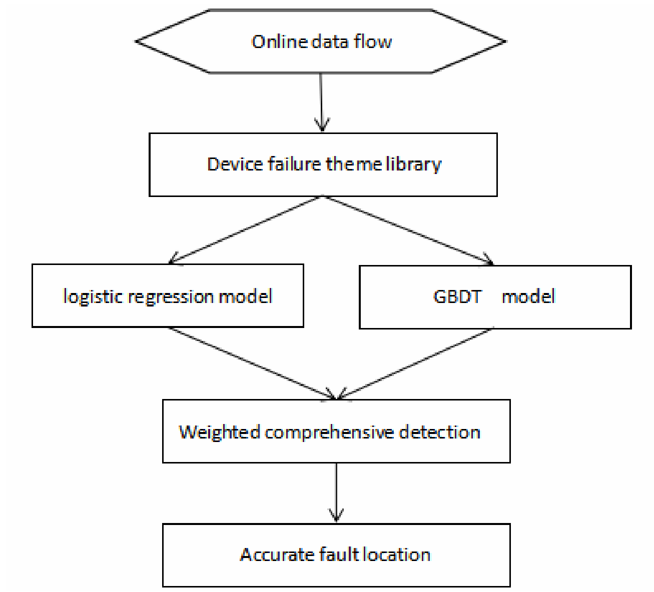

Figure 3. Online real-time detection system

4 Experiments and Results

Due to the diversity of key performance indicators, this paper chooses logistic regression(LR) model and GBDT model to form a weighted model, combined with key performance indicators of operation and maintenance equipment. (for example, CPU usage, memory usage, disk IO, network card throughput, network equipment Port threshold, etc.)

In this paper, the isolated basic data is effectively fused and associated with the combination of the optimal weight coefficient method, and comprehensively utilizes the useful information of the single-item fault location model to build a fault location model for big data centers. Training provides versatile and efficient information support for fault location determination and maintenance decisions.

In this paper, different data structure types of Monte Carlo method generates simulation data, and then passes the 10-fold cross-validation to generate training set and test set, calculating the training set and test set respectively Accuracy.

4.1 LR fault Location Model

Logistic regression model is a generalized linear regression analysis model. In this paper, through the analysis of key performance indicators, the probability of various equipment failures in the data room is predicted.

The logistic regression model in this paper uses the maximum likelihood method for prediction, and its function is expressed as:

$$l(\beta)=\sum_{i=1}^{m}\{y_i(\beta_0+X_i\beta)\}-\ln[1+\exp(\beta_0+X_i\beta)]\}$$

β_0 is the coefficient of the constant term.

$\beta_j(j=1,2,...,m)$ is the coefficient corresponding to explanatory variable.

β is the coefficient vector $\beta=(\beta_1,\beta_2,...,\beta_m)^{\mathrm{T}}$ consisting of these coefficients.

In this paper, the key performance indicators explain variables, and the predicted results response variables. Here, 1000 sets of key performance indicators are taken as observation samples and combined with the theme library of the device to make predictions. It should be noted that the selected key performance indicators should be real-time acquired, reducing the use of default values to improve the accuracy of prediction.

The results are shown in Figure 4.

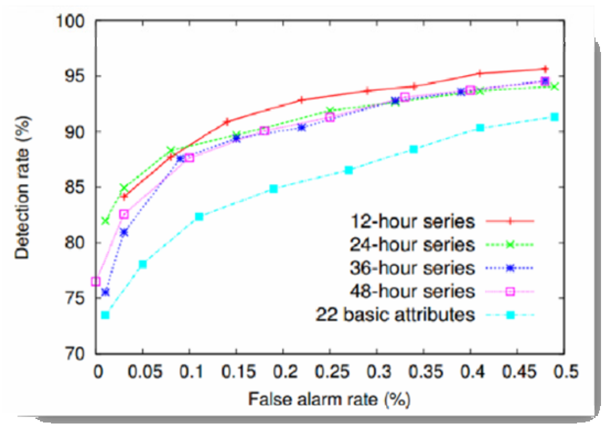

Figure 4. LR model prediction result

4.2 GBDT Fault Location Model

The GBDT model mainly includes the following steps:

- Preprocess the historical fault data as the training set of the algorithm, and use the key performance indicators to construct the data block. The more effective information the data block contains, the less the default value, and the more accurate the prediction will be.
- Set the number of samples m to be replaced, and sample m times for the training set D. Each sample constitutes a new sub-training set D_i (i = 1, 2, 3 ... m), and the number of samples is n (n < N).
- Using the generated sub-training data set Di, the Bagging algorithm is used to generate m weak classifiers composed of the GBDT algorithm in parallel, thereby forming the GBDT algorithm.
- The previous weak classifier will generate m prediction results, and use the voting method to determine the final strong classifier prediction results.

The results are shown in Figure 5.

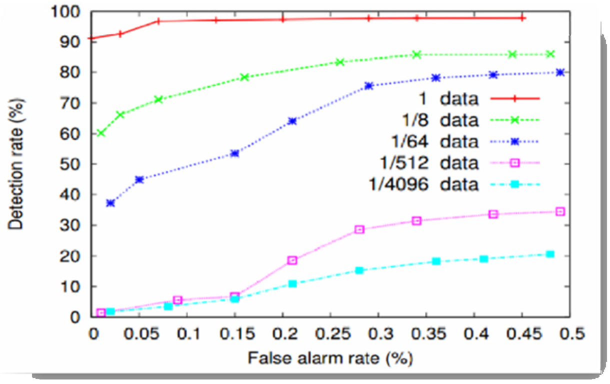

Figure 5. GBDT model prediction result

In Figure 4 and Figure 5, the ordinate represents the accuracy of the failure prediction, and the abscissa represents the probability of false alarm reporting. The LR mode in Figure 4 uses fixed time series operation and maintenance data as input, The shorter the time series, the more accurate the prediction. While the GBDT in Figure 5 uses fixed The operation and maintenance data block as input, the more complete the data block information, the more accurate the prediction.

4.3 Optimal Weight Coefficient Model

Data preprocessing is to process the eigenvalues, positive and negative infinity values in the value of variables. This paper is carried out through a combination of expert marking and machine learning.

Data set division is to reduce the risk of overfitting, the data in the training set is divided into two parts. One part of the data is used to train the GBDT model, and the other part of the data is used to train the LR model by obtaining new features through the trained GBDT model.

LR is a linear model with limited learning ability. Feature engineering is especially importan. Existing feature engineering experiments are mainly focused on finding distinguishable features and combinations, but the effect is limited. The characteristics of the GBDT algorithm can be used to discover distinguishable features and reduce labor costs in feature engineering. According to the practice of this paper, the Optimal weight coefficient model has a obvious effect.

The single fault location model can be expressed as:

$$f_i(x), i = 1,2,\ldots, M$$

The combined fault location model can be expressed as:

$$f(x) = \sum_{i=1}^{M} w_i f_i(x)$$

Where w is the weight coefficient of the single prediction model, it can be expressed as

$$\sum_{i=1}^{M} w_i = 1, 0 \le w_i \le 1$$

The calculation methods according to the weight coefficient can be divided into two categories: the optimal weight coefficient method and the non-optimal weight coefficient method. In this paper, the maximum or minimum value is obtained through certain limited conditions, so as to obtain the weight coefficient range, and then according to the scene, the weight coefficient is set according to the arithmetic average method, the reciprocal variance method, and the recursive weighting method, it can be expressed as:

$$\min(\max)\theta = \theta(w_1, w_2, \ldots, w_M)$$

$$\sum_{i=1}^{M} w_i = 1$$

$$0 \le w_i \le 1, i = 1,2,\ldots, M$$

It needs to be clarified that min is used for error failure alarm rate, and max is used for failure prediction accuracy rate.

4.4 Model Test Result

The positioning accuracy of the three models actually tested is shown in the Table 2:

Table 2. Model test result

Model	Sample	Precison	Recall	Fl-score	Support
LR	“0”	0.92	0.4	0.56	9001
	“1”	0.62	0.96	0.74	9005
	total	0.77	0.68	0.65	18006
GBDT	“0”	0.92	0.8	0.74	9052
	“1”	0.82	0.92	0.86	9013
	total	0.87	0.86	0.8	18065
Optimal Weight Coefficient	“0”	0.98	0.95	0.92	9021
	“1”	0.93	0.93	0.94	9031
	total	0.95	0.94	0.93	18052

“0” indicates the prediction accuracy of the sample with value 0, while “1” indicates the prediction accuracy of the sample with value 1, and “total” means the overall prediction accuracy of the sample.

In the test, the optimal weight coefficient model has a better fault location effect, and the accuracy rate is increased to more than 95%.

Overall, the prediction accuracy of the combined prediction model is significantly improved compared to the single prediction model. When the correlation of the explanatory variables is stronger, the larger the sample size, the more the advantages of the combined forecasting model will become.

At present, the combination method has received more and more attention in various fields, various forms of model combination have been continuously studied, and the weight selection method has also been continuously developed. In future research, multiple weight selection methods can be used Compare the effects.

5 Conclusion

Through the technical research of 5G data center failure prediction in this paper, the scope and accuracy of failure coverage have been greatly improved. The key technology studied will cover 80% of the total number of data center failures, and the prediction success rate 3 days in advance will reach 95%.We will study the intelligent monitoring and abnormality based on spatiotemporal convolutional neural network Detection Technology. Using time and space collaborative information to intelligently monitor the status of data center equipments.

Under the trend of 5G network transformation, intelligent operation and maintenance technology has become a core innovation capability. The use of various new technologies makes operation and maintenance services more convenient. Colleagues also have a huge impact on the traditional operation and maintenance industry. With the gradual advancement of commercial deployment, intelligent operation and maintenance is also constantly undergoing technological innovation and evolution, and gradually formed a set of practical and efficient solutions, and ultimately achieve the purpose of improving operation and maintenance efficiency and reducing labor costs.

With the development of 5G commercialization, the demand for intelligent operation and maintenance is also increasing, and the environment facing it is becoming more and more complicated. Information construction does not happen overnight. The company will purchase different hardware devices in different periods and face different needs, resulting in many different types of network equipment, massive different types of servers, various virtualization solutions, different operating systems, and diverse Application software and database. Faced with a large-scale and complex-structured operation and maintenance environment, how to effectively monitor the operation of the entire system and discover abnormalities in the system in time becomes particularly important.

Aiming at discovering some important problems of the intelligent deployment in big data centers, building a fault prediction model based on key performance indicators, conducting a deep research in conjunction with scenarios of big data centers and providing a complete and effective solution.

References

[1] P. Guo, K. Wang, A.-L. Luo, M.-Z. Xue, Computational Intelligence for Big Data Analysis: Current Status and Future Prospect, *Journal of Software*, Vol. 26, No. 11, pp. 3010-3025, November, 2015.

[2] B. Kang, G. Jo, H. Choo, A History-based Predictive Handover Scheme for Wireless Networks, *Journal of Internet Technology*, Vol. 19, No. 5, pp. 1487-1497, September, 2018.

[3] X.-H. Zhao, Y. Shi, S.-Z. Chen, TS-SMOSA: A Multi-Objective Optimization Method for Task Scheduling in Mobile Edge Computing, *Journal of Internet Technology*, Vol. 20, No. 4, pp. 1057-1068, July, 2019.

[4] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, *The Annals of Statistics*, Vol. 28, No. 2, pp. 337-374, April, 2000.

[5] S. Mukherjee, N. Sharma, Intrusion Detection using Naive Bayes Classifier with Feature Reduction, *Procedia Technology*, Vol. 4, pp. 119-128, 2012.

[6] A. Fernández, S. Río, V. López, A. Bawakid, M. Jesus, J-M. Benítez, F. Herrera, Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks, *WIREs Data Mining and Knowledge Discovery*, Vol. 4, No. 5, pp. 380-409, September/October, 2014.

[7] H.-H. Chen, L.-S. Huang, J.-B. Chen, T.-L. Pao, Fuzzy Decision Load-balancing Algorithm for Cloud-based Terminal Services, *Journal of Internet Technology*, Vol. 19, No. 3, pp. 689-695, May, 2018.

[8] A. Dixit, F. Hao, S. Mukherjee, T.V. Lakshman, R. Kompella, Towards an elastic distributed SDN controller, *ACM SIGCOMM Computer Communication Review*, Vol. 43, No. 4, pp. 7-12, October, 2013.

[9] X.-C. Sun, Z.-J. Wang, H. Chu, Q.-R. Zhang, An Efficient Resource Management Algorithm for Information Centric Networks, *Journal of Internet Technology*, Vol. 17, No. 5, pp. 1007-1015, September, 2016.

[10] S.-C. Li, L.-D. Xu, S.-S. Zhao, 5G Internet of Things: A survey, *Journal of Industrial Information Integration*, Vol. 10, pp. 1-9, June, 2018.

[11] M. Li, H.-X. Yin, F.-Y. Xing, J.-C. Wang, H.-H. Wang, Design of Control Plane Architecture Based on Cloud Platform and Experimental Network Demonstration for Multi-domain SDON, *Journal of Physics: Conference Series*, Vol. 679, 012022, 2016.

[12] I. Abaker, T. Hashem, I. Yaqoob, N.-B. Anuar, S. Mokhtar, A. Gani, S.-U. Khan, The rise of "big data" on cloud computing: Review and open research issues, *Information Systems*, Vol. 47, pp. 98-115, January, 2015.

[13] B. Zeng, S. Yang, X.-W. Yin, Robotic Relocalization Algorithm Assisted by Industrial Internet of Things and Artificial Intelligence, *Journal of Internet Technology*, Vol. 21, No. 5, pp. 1517-1530, September, 2020.

[14] M.-S. Kim, J.-K. Lee, J.-H. Park, J.-H. Kang, Security Challenges in Recent Internet Threats and Enhanced Security Service Model for Future IT Environments, *Journal of Internet Technology*, Vol. 17, No. 5, pp. 947-955, September, 2016.

[15] R. Fei, Y.-B. Zhu, Q.-Z. Yao, Q.-Z. Xu, B. Hu, A Deep Learning Method Based Self-Attention and Bi-directional LSTM in Emotion Classification, *Journal of Internet*

Technology, Vol. 21, No. 5, pp. 1447-1461, September, 2020.
[16]H. F. Rashvand and H. C. Chao, *Dynamic Ad-Hoc Networks*, The Institution of Engineering and Technology, 2013.
[17] Z. H. Zhou, *Machine Learning (Chinese Edition)*, Beijing: Tsinghua University Press, 2016.

Biographies

Min Tan received his M.S. degree in Software Engineering from University of Science and Technology of China, HeFei, China, in 2010. He is now a Ph.D. student of Electronics and Information in Huazhong University of Science and Technology. His research interest is Artificial Intelligence and Information Communication.

Guo-hui Li received his Ph.D. degrees from Huazhong University of Science and Technology in 1999. He is a Professor in Department of Computer Science and Technology, Huazhong University of Science and Technology, China. His research interests include Big data analysis, and open environment machine learning.

Han Yu received his M.S.degrees in project management from Huazhong University of Science and Technology in 2013. He is a senior engineer in Wuhan Fiberhome Technical Services Co., Ltd., China. His research interests include Network management and artificial intelligence.

Ming Wei received his M.S.degrees in Mechanical engineering from Huazhong University of Science and Technology. He is a senior engineer in Wuhan Fiberhome Technical Services Co., Ltd., China. His research interests include Network optimization, deep learning and cloud computing.

Ji-wei Hu received his M.S.degrees in project management from Central China Normal University. He is an engineer in Wuhan Fiberhome Technical Services Co., Ltd., China. His research interests include Network communication and information digitization.