ISSN : 0974 - 7435 Volume 10 Issue 24

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(24), 2014 [15422-15429]

Quality analysis algorithms research of 4G mobile network

Huang Jing
Department of information science and Engineering, Hunan City University,Hunan Yiyang, 413000, (CHINA)

ABSTRACT

In terms of network operating QoS analysis, we propose a cascading network QoS analyzing algorithm based on K-means and C4.5 algorithm. The algorithm was testified to be suitable for multiple analyze requirements. Also based on the monitoring data captured from the real Internet, the algorithm was proved to be effective and efficient The algorithm processes the data captured from the project User-oriented Active 4G Network Measurement System. The system is deployed at multiple wireless network accessing points, conducting 4G network QoS monitoring 24 hours. The paper covers experiments to find a proper K and C4.5 discrete value, and use the KKZ algorithm to initialize the cluster center values. Based on the six traces of monitoring data, we compared the performances of cascading network QoS analyzing algorithm, K-mean algorithm and C4.5 algorithm. As a result, the cascading algorithm was highly efficient and reduces the noise of single algorithm; also it proved to be suitable to several types of monitoring data..

KEYWORDS

4G mobile network; Quality analysis; K-means.

INTRODUCTION

With the development of 4g communication technology, Internet gradually becomes a network of transfer music, pictures, video and other multimedia content, especially hosting services become more and more rich by the mobile Internet. This requires the Internet provide a more reliable network operators[1]. Due to the different kinds of application of network quality of service requirements vary, some strong network real-time application requirements, some applications require strong network stability. It is an important subject to provide network service quality index, according to different kinds of applications, secondly according to network service quality monitoring data, the service level of network running the discriminate is another important issue. In this paper, based on the active network measurement system of running in the existing domestic province mobile Internet main nodes, the network service quality measurement data obtained. Secondly analyzes the running quality indicators, and find out the key factors affecting the quality of 4G network operation. Finally based on the several key factors, this paper designed a simple and applicable to the current network running quality of the 4G network analysis algorithm, can carry on the analysis to the existing network running quality, and to the existing 4G network operation quality of anomaly detection value judgment. Based on the network to collect the 4G network running quality monitoring data of a series of experiments, the reliability and validity of this algorithm is verified.

The traditional network quality analysis is usually refers to the network traffic anomaly judgment and early warning. General packet, or flow records such as network data was analyzed, combining with data mining, correlation analysis and other methods to potential threats in the network and operation of abnormal analysis and alarm.

Some scholars put forward a macro and systematic method to detect abnormalities, the main idea is using the technology of data mining to discover useful to be able to describe the system program and user behavior patterns, and the characteristics of the relevant system to calculate can be used to distinguish all kinds of anomalies and the invasion of the classifier[2]. Some scholars put forward an abnormal value method based on ARIMA (Autoregressive integrated moving average) to detect and remove the QoS[3]. This paper combines ARIMA predict and outliers to predict short-term and long-term monitoring data, the data used is real network end-to-end QoS delay. Clustering method is proposed by KLAUS JULISCH to deal with the intrusion detection alarm information, they found some root causes for more than 90% of the alarm message[5]. This paper mainly tells the story of how to identify the alarm root cause of a lot of, how to remove the root cause, and reduces the load of the alarm. Some scholars put forward the automatic extraction of a communication network, the description, and similarity matching and relevancy computation mode of the data mining technology. Technology based on the extracted from time series and defined patterns, for system monitoring, discover and QoS model under the network environment in correlation analysis[6]. Some scholars proposed a statistical base for the median prediction algorithm, the text of the model is similar to the basic genetic algorithm of some heuristic algorithm[7]. Core idea from historical data extracted a rule, to predict future data. Some scholar's research method is by examining the characteristics of distribution in baton, the abnormal differentiate into different categories. In this paper, the method is suitable for the wan traffic. The concrete method is to use entropy to detection and classification of anomalies. In detecting abnormal, paper introduces the method of multi path space, to show how it used to detect more than the flow characteristics of the road. Some scholars based on the network perspective technology. Put forward a new method for large-scale network active warning, no control of the network can be real-time monitoring and early warning actively. When network abnormal behavior occurs, can accurately detect abnormal sources, narrowing the scope of abnormal location, for large-scale network abnormal behavior detection and prevention has important practical significance. Some scholars use ARIMA model to forecast and analysis on the abnormal values in wireless network.

The limitations of the current network quality analysis technology is, the existing research only focus on the network of all kinds of abnormal invasion monitoring, lack of quality of service for network link analysis research[8]. Especially, because the lack of research and analysis the quality of network application. And limited literature is based on the research of the historical data, for future long-term a prediction or abnormal network behavior decision, urgently need a can carry on the analysis to the network quality real-time algorithm. This paper proposes a algorithm was able to analyze quality of network early warning. Using data mining technology, the active network measurement system in this paper is measured by network quality monitoring data modeling, using historical data model for the new data for real-time analysis.

THE ALGORITHM DESIGN OF NETWORK RUNNING QUALITY ANALYSIS

4G network quality analysis algorithm is proposed in this paper, which is geared to the needs of the user's active network measurement system of a module, function is abnormal monitoring data of the system were analyzed, and judgment. In the design of the practicality and efficiency of the algorithm is fully considered, and have the corresponding mature software running in the center of the test network servers. Algorithm is the basic idea, using historical data to build training model, the new measurement data and the model matching, tandem type multilevel discriminates algorithm is used to determine whether new measurement data is abnormal data.4G network running quality analysis algorithm includes two main modules, discriminates training module and data analysis module.

Below is the overall design of the algorithm in the system architecture diagram.

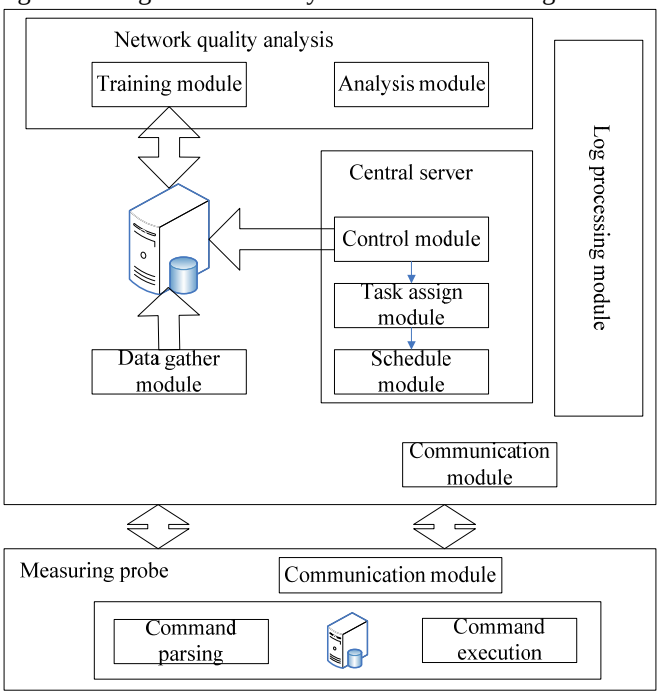

Figure 1 : The overall design architecture diagram of the algorithm

As shown in the above, the algorithm designed to run in the center of the user-oriented active network measurement system on the server. Central server is the core component of the whole system, the basic function is to manage the operation of the whole system and the distributed monitoring task. Central server consists of control module, task allocation module, scheduling module and other modules. Central server is doing the communication through communication module and measuring probe. The management of the monitoring task is mainly monitoring the running state of the monitoring probe, and. Monitoring probe is a basic measurement unit of network measurement, carrying specific measurement task, it includes the command parsing module, command execution module and communication module, etc. All monitoring data through the communication module is sent back to a central server. Center server data acquisition module will be monitoring data are collected, unified stored in the database. All of the tasks and monitoring data is stored in the database.

The algorithm at runtime, you first need to read from the database monitoring tasks, training module for statistical modeling, data analysis module using the modeling data to analyze the measured data of the new, and make records. Figure 2 is the quality analysis algorithm flow chart of the whole 4G network running.

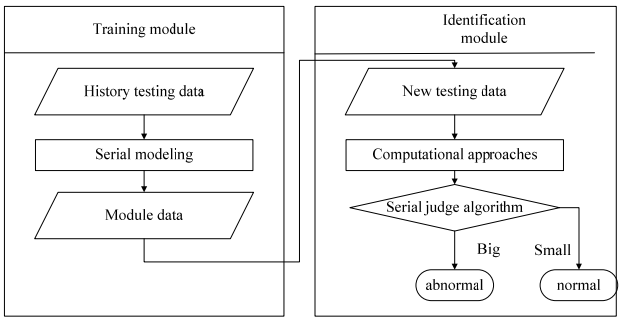

Figure 2 : The flow chart of the whole algorithm

ALGORITHM PROCESS OF NETWORK QUALITY ANALYSIS

The data processing of the algorithm is a basic network performance measurements and web browsing performance measurement data. Both have multiple dimension measurement data, due to the different dimension has certain relevance, will increase the complexity of processing, and analysis all the dimension is not necessarily useful for results[9]. So this article selects the principal component analysis (PCA), converts multidimensional measurement data to low dimensional index, the

network quality key factor, and through the factor analysis got the score of each factor. First of all, it is necessary to introduce the basic concept of principal component analysis and factor analysis and related principles.

A random variable X1, X2,..., Xp, sample standard deviation for S1, S2, and Sp. As a standardized transformation:

$$C_j = a_{j1}x_1 + a_{j2}x_2 + ... + a_{jp}x_p,\ j = 1,2,...,p \quad \textbf{(1)}$$

Are defined as follows:

(1) $C_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p$, and make the variance biggest, C_1 has described as the first principal component.

(2) If $C_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p$, then $C_2 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p$ perpendicular to the ($a_{11}, a_{12}, ..., a_{1n}$), and C_2 has the largest variance, then C_2 is said as the second principal component.

(3) Similarly, there is a third, forth, fifth. principal components, and most can choose P principal component.

The purpose of factor analysis is to represent most of the original variable information with fewer factors. The ideas of factor analysis can be represented by the mathematical model available. Suppose there are P variables and all (or) treated by standardized mean to 0, the standard deviation is 1. Each variable of the original can be a linear combination of k (k < p) factors $f_1, f_2, f_3, ..., f_k$, as shown in the following type:

$$\begin{cases} x_1 = a_{11}f_1 + a_{12}f_2 + a_{13}f_3 + ... + a_{1k}f_k + \varepsilon_1 \\ x_2 = a_{21}f_1 + a_{22}f_2 + a_{23}f_3 + ... + a_{2k}f_k + \varepsilon_2 \\ x_3 = a_{31}f_1 + a_{32}f_2 + a_{33}f_3 + ... + a_{3k}f_k + \varepsilon_3 \\ ..... \\ x_p = a_{p1}f_1 + a_{p2}f_2 + a_{p3}f_3 + ... + a_{pk}f_k + \varepsilon_p \end{cases} \quad \textbf{(2)}$$

Equation (2) is the mathematical expression of factor analysis, also can use matrix is expressed as the form of $X = AF + s$. Where F is called factor, due to the factor in each of the original variables can be express linear, so also known as the common factor. Factor can be understood as a high-dimensional space perpendicular k axis; A known as the factor loading matrix, called factor loading, is the first I the original variable in the first j A load factor. S called special factors, said parts of the original variables cannot be explained by factors. The mean is zero.

Are selected in this paper, the nine indicators as shown in TABLE 1, respectively for X_1: DNS delay, X_2: connection time delay, the X_3: response delay, X_4: web delay, X_5: page speed, the X_6: embedded documents delay, X_7 : embedded file speed, by X_8: total delay, X_9: general speed. This a few indicators is to evaluate the quality of the mobile network users to access web services monitoring indicators. Raw data samples such as shown in TABLE 1 below.

TABLE 1 : Performance measurement monitoring data of the sample of web browsing

DNS delay	Connection time delay	Response time delay	Web delay	Page speed	Embedded file time delay	Inline text speed	Total time delay	Total rate
250	114	410	8885	3429	10303	10988	19553	7412
8410	140	444	7945	1923	9870	11469	26365	5497
6530	104	400	10190	1885	10421	10997	27246	5371
8070	109	825	0684	1681	9198	10177	28063	4466
275	109	360	8680	3499	5729	8225	24794	6497
235	120	395	8940	3412	9146	9946	18441	6653
9930	130	509	8519	1707	10464	8693	29044	4224
7110	114	485	10834	1756	13433	8427	31493	4602
255	114	380	9625	3173	11706	9670	21701	6678

First, principal component analysis is conducted to get common degrees between the indexes, such as shown in TABLE 2. The table is the initial solution of factor analysis. The first column in the table is a common degree, and the factor analysis of the initial variable when it said if using principal component analysis (PCA) to extract feature take all nine characteristics of the original, so the original data of all variance can be explained.

TABLE 2 : The common degrees of indexes

communality		
	The initial	**Extract**
DNS delay	1.000	.657
The connection	1.000	.645
The response time delay	1.000	.467
Web delay	1.000	.859
Page speed	1.000	.571
The built-in time delay	1.000	. nV
Embedded speed	1.000	750
The total time delay	1.000	.962
Total rate	1.000	.880

Second, get the characteristic values of each principal component factor contribution rate and cumulative contribution rate, are shown in TABLE 3 below.

TABLE 3 : Factor statistics of each principal component

ingredients	**The initial eigenvalue**		
	Factor eigenvalues	**The variance contribution rate**	**The cumulative contribution rate**
1	2.370	26.332	26.332
2	2.001	22.238	48.570
3	1.228	13.64	62.211
4	1.	73.510	11.299
5	.792	8.805	2.316
6	.10	8.40	90.729
7	.676	7.506	98.235
8	.148	1.644	99.879
9	.011	.121	100.000

Form factor is the first column number, meaning respectively characteristic and value, variance contribution rate and cumulative variance contribution rate. Statistics shall be carried out in accordance with the characteristic value from big to small order arrangement; the table also includes the contribution rate of each principal component and the cumulative contribution rate. The eigenvalue of the first principal component is 2.37, it represents the nine variables of the total variance of 26.332. So on, if you can see, the original data is extracted total variance is more than 80% for the standard, for web browsing performance measurement of measurement data, keep five principal component is more appropriate. At this point the total information loss of original variables is less, the factor analysis results are ideal. So for this kind of monitoring data, we can take five data instead of the original nine indicators.

Further, this paper made a factor analysis; it is according to the characteristics of scattered with the size of the main points. The abscissa is a factor in the graph, the number of ordinate is characteristic root. We can see from the chart, starting from the fifth characteristic root, and other features of the main component of decreases. Further evidence that the first five feature extracting is the right choice.

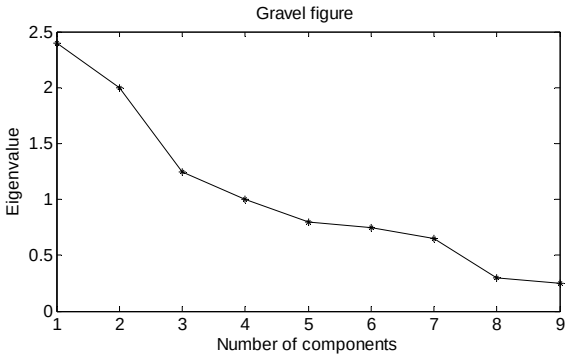

Figure 3 : The gravel of factor analysis

EXPERIMENTAL RESULTS AND VERIFICATION

Experimental data declaration

The System is deployed in several cities of the southern province. Each city has deployed more monitoring probe in different position. Monitoring probe used in the form of a card 2 g and 3G mobile access network, and network quality test on the specified website. System deployment diagram as shown in Figure 6 to 8. In the provincial capital deployed a central server, and the measured results sent by the real-time of the monitoring of the whole province deployment to the central server for storage and management. In charge of server configuration testing tasks, the management of the network probe, storage management and generate the statistics of the measurement result.

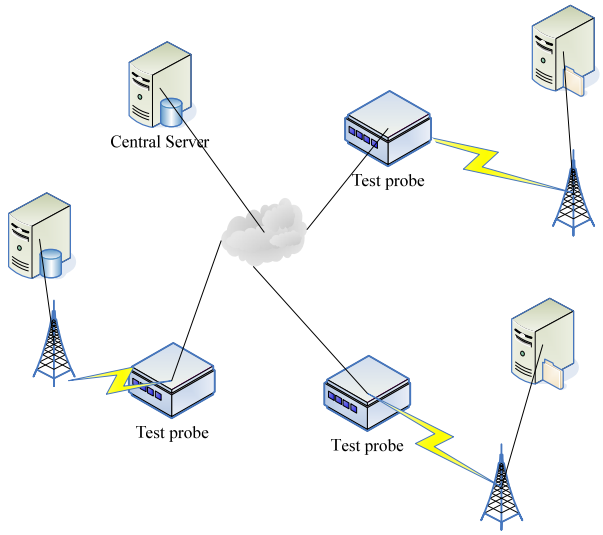

Figure 4 : System deployment diagram

After the system deployed in the province, the network monitoring is done from September 2013 to June 2014 in continuous. Measurement of target selection is a bigger 250 sites, domestic traffic basic network performance monitoring and specific network service quality monitoring. Test goal of comprehensive and long test time span to ensure the reliability of the experimental data. The following table is used in this paper, the experiment data, the six covered the 3G and 4G two types of access[10].

TABLE 4 : The experimental data

Data number	Monitoring type	Network type	Acquisition time	Normal data	Abnormal data	total
P1	Basic performance measurement	3G	In November 2013	686	40	726
P2	Basic performance measurement	4G	In March 2014	704	46	750
P3	Basic performance measurement	3G	In May 2014	703	37	740
W1	Web browsing performance	2G	In December	1300	142	1442

	measurement		2013			
W2	Web browsing performance measurement	3G	In April 2014	1270	130	1400
W3	Web browsing performance measurement	4G	In June 2014	1303	120	1423

Clustering number selected in the experiment

As stated earlier, k-means algorithm if improper selection of cluster number K will get poor clustering results. For in this paper, the algorithm will directly affect the monitoring data analysis results. So you need to analyze the experimental data, the selection of a suitable for data clustering number K in this paper.

There are many kinds of algorithms to the k-means algorithm to optimize the number of K. This article uses the more widespread use of minimizing the SSE as K value optimization measures. The calculating formula of SSE is as formula (2). The basic idea of this algorithm is k-means algorithm of K value optimization algorithm into local minimization problem of SSE, all possible K values listed in the first place, according to the distribution of the SSE and combined with the distribution of the clustering result data to choose K values.

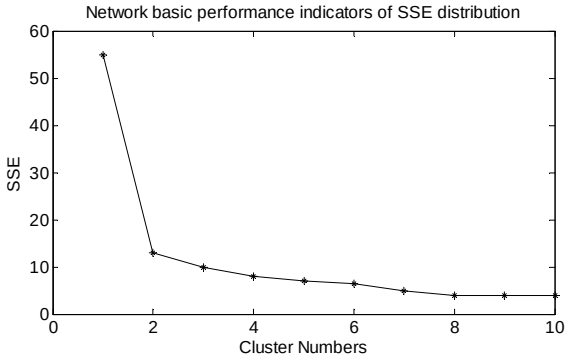

Figure 5 : Network basic performance indicators of SSE distribution

In this paper, first, the basic network performance measurement data of P1 as sample for analysis. Figure 5 is the index of the SSE and clustering number distribution, the abscissa represents clustering number, for basic network performance measurement data, this paper choose the clustering number from 1 to 10. Ordinate represent different clustering number corresponding SSE value.

Experimental results and analysis

This section is integrated to the k-means with C4.5 decision tree algorithm is analyzed and the result of the instructions, and the experimental results with single k-means algorithm comparing with C4.5 decision tree algorithm. Experiments using the 3.1 section describes the data sample. In this paper, we use the five indicators to compare the results of different algorithms.

1. True Positive Rate and TPR, refers to the ratio of abnormal monitoring data is right.

2. False Positive Rate, FPR, refers to the normal monitoring data is wrong judgment as the ratio of abnormal.

3. Precision, refers to the correct judgment of abnormal monitoring data of all data in the given as the ratio of abnormal monitoring data.

4. Total accuracy, refers to all the normal and abnormal data by identifying the correct ratio.

5. F Measure refers to the accuracy and recall rate of the weighted harmonic mean, computation formula is as follows:

$$F = \frac{2PR}{P+R} \quad \textbf{(3)}$$

On the above equation, P represents precision rate and R represents recall rate. Recall rate refers to the correct judgment of abnormal data of actual all the percentage of abnormal data..

This section to use k-means and cascade of C4.5 decision tree algorithm to determine the results were analyzed, and the analysis of the first is the basic performance measurement data analysis of the results.

Diagram are separate k-means algorithm, C4.5 algorithm separately, the discriminant efficiency of k-means and C4.5 algorithm of cascade, one thousand one hundred percent is used for the test data in the experiments of cross validation method (the original data into 10 portions, each time 9 data do the training sample, a test sample from cycle average) ten times.

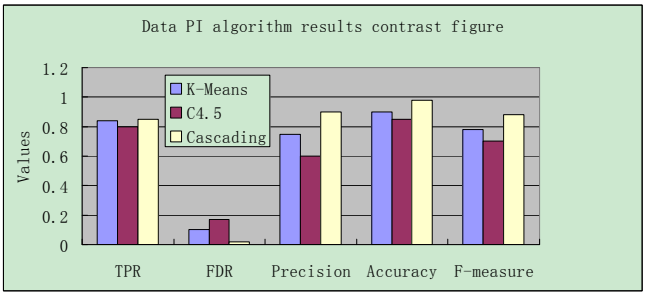

Figure 6 : Data PI algorithm results contrast figure

Figure 6 is the result comparison chart of the separate k-means algorithm, C4.5 decision tree algorithm and two kinds of algorithm of serial algorithm (Cascading). The different classification criterion of the algorithm is used in the Pl data. The effectiveness of the proposed algorithm is presented in this paper.

CONCLUTION

Aiming at the need of network quality of service analysis, this paper proposes a new 4G network quality analysis algorithm. The algorithm is based on cascade algorithm of k-means algorithm and C4.5 algorithm, the algorithm can be divided into discriminant two training module and analysis module, training module of historical data modeling, identification module using the modeling data for integrated decision of the new measurement data. This article compares the cascade algorithm, k-means algorithm separately, separate performance of C4.5 algorithm, compared with accuracy, misjudgment rate index, etc. Results show that: in the monitoring data, the cascade performance of the algorithm is more than the performance of the single k-means and C4.5 algorithm.

REFERENCES

[1] I.Miloucheva, U.Hofimnn, P.A.Gutierrez; Spatio-tenporal QoS pattern analysis in large scale internet environment//Interactive Multimedia on Next Generation Networks, Springer, 282-293 **(2003)**.

[2] Z.U.M.Abusina, S.M.S.Zabir, A.Ashir et al.; An engii^ering approach to dynamic prediction of network performance from application fogs, International Journal of Network Management, **15(3)**, 151-162 **(2005)**.

[3] A.Lakhina, M.Crovella, C.Diot; Mining anomalies using traffic feature distributions: ACM SIGCOMM Conputer Communication Revfew, ACM, **35**, 217-228 **(2005)**.

[4] C.A.AHernandez Suarez, O.J.Salcedo Parra, A.Escobar Diaz; An ARIMA model for forecasting Wi-Fi data network traffic values, Ingenieria e Investigation, **29(2)**, 65-69 **(2009)**.

[5] P.Ferguson, G.Huston; Quality of service: delivering QoS on the Internet and in corporate networks, Wiley New York, **(1998)**.

[6] I.T.Jo Diffe; Princ5>al component anafysis, Springer-Verlag New York, **(1986)**.

[7] S.T.K.Koutroumbas, Pattern recognition, Fourth Edition, 408-445 **(2008)**.

[8] Jianming liu, Feng Zhao, yue-xia zhang; 3G and 4G wireless communication technology in the application of ICT network model [J], Power System Communications, **30(201)**, 1-4 **(2009)**.

[9] Guohua Hu, Shu-jie Yuan, Tan Min, 4G mobile 4G communication technology and safety defects analysis [J], 4G communication technology, **(7)**, 155-157 **(2008)**.

[10] R.Denis, S.Philippe; Convergence and competition on the way towards 4G[C], Ottawa: Radio and Wireless Symposium, 277-280 **(2007)**