International Journal of Future Generation Communication and Networking
Vol. 9, No. 12 (2016), pp. 129-138
http://dx.doi.org/10.14257/ijfgcn.2016.9.12.12

Network Traffic Identification Algorithm Based on Neural Network

Fei Wu, Yong Ye, Hongfa Li, Shilong Ni and Jiangwen Su

Fujian Yirong Information Technology Co. Ltd
1526332417@qq.com

Abstract

In this paper, a network traffic identification model is established using a multilayer excitation function quantum neural network which is suitable for data classification. Because the conventional quantum neural network has different target function in the training of the weights of the network and the sigmoid function of the neurons in the hidden layer, the coupling effect of the two parameters is not processed. This will result in the middle and later stage of the training iteration process, and it may be possible to reduce the objective function value of a kind of parameter, and make the objective function value of another kind of parameter increase. In order to avoid this situation, using LM algorithm to optimize, using the same objective function not only as the target function of the network weight, but also the function of translational spacing of sigmoid function of neurons in the hidden layer, and the training objective is to minimize the sum of squared error of the neural network output and the desired value. Finally, the recognition performance of the proposed algorithm is compared with that of the conventional quantum neural network and LM-BP neural network. The results show that the convergence rate of the proposed algorithm is the fastest and the convergence accuracy is the highest.

Keywords*: network traffic identification, quantum neural network, Levenberg-Marquardt algorithm, session flow characteristics*

1. Introduction

With the continuous development of Internet technology, a variety of network services and application types are emerging, which put forward higher requirements for Internet governance. Meanwhile, network security is increasingly verifying the great significance of real-time and effective detection for network traffic [1].

(1) Network traffic detection system assists in achieving the goals of billing and controlling bandwidth. Network traffic detection system adopts traffic detection technology to differentiate network traffic and then realizes differentiated control and billing. The adoption of this technology, on the one hand, improves network efficiency and user experience; on the other hand, it increases benefits of basic network operators. Previously, the "pipeline" operation of operators, without differentiating pipe traffic, was a single business control model and billing model. With the dramatic increase in network size and number of users, business models became diversified and there were flaws in operators' control, billing and security management. New application traffic which adopts Peer to Peer network (P2P), Voice over Internet Protocol (VoIP) and Streaming Media Technology took up 60% -80% of total bandwidth. However, its failure in differentiated billing resulted in the blocking of business and uneven distribution of network resources. After the adoption of detection system, Quality of Service (QoS) is guaranteed, and the best-effort service type transforms into the differentiated service model, which ensures the normal and orderly operation of critical businesses.

ISSN: 2233-7857 IJFGCN

(2) Network traffic detection system provides protection for internal network. Network traffic detection system not only can protect internal network from the harm of external network, but can also regulate the behavior of internal network staff accessing to the network to prevent internal leakage. Trojans, viruses, worms, Denial of Service attacks (DoS) pose threats to internal network security. Detection system can carry on real-time monitor to network and run the solution quickly according to the changes in network parameters. Before causing damage, it can prevent network attacks by combining with IPS, IDS and Against Virus (AV). If the invasion had happened, it can reduce the loss to the maximum extent so as to achieve visible, controllable and traceable effects. Security and defense concept is often confined to conventional gateway levels and network boundaries, such as firewalls, vulnerability scanning, security audit, anti-virus, IDS, *etc*. Statistics of the Chinese Ministry of Public Security show that 80% of network security incidents are due to internal network factors and only the remaining 20% of security incidents come from external network factors. Internet behavior management can view traffic visits of overall or a specific IP according to network traffic statistics logs, standardize network applications that can be available to in working hours and shield other access, which improves work efficiency and ensures the security of confidential information within the network.

(3) Network traffic detection system brings new growth points of interest for operators. Operators’ traditional profitable business are gradually replaced by OTT (Over the top) businesses. Afterwards, the issue of how to charge for data traffic will be put on the agenda. The management of oriented traffic in Internet companies, differentiated billing and other critical businesses cannot do without traffic detection technology. SMS and voice calls are the main profitable businesses for operators, accounting for about 70% of total revenue at the beginning of 2013. With the development of OTT (Over the top), operators’ main profitable businesses are replaced by chat software, microblog, *etc*. Internet companies skip over operators and conducts user-oriented development directly. Thus operators are at risk of becoming "pipeline". In the first three quarters of 2013, operator's profit growth rate declined. If appropriate measures were not taken, this phenomenon would be more serious. The use of traffic detection system not only provides basic conditions for operators’ traffic control, but also provides effective ways for sustained profitability.

Traditional ways for network traffic classification are: (1) port recognition based network traffic classification method; (2) data packet based network traffic classification method. Although the traditional network traffic classification methods have simple algorithm and high efficiency, they no longer apply to today’s complex and diverse types of Internet services and applications due to their own limitations. Now the widely applied network traffic classification methods are: (1) statistical characteristic based network traffic classification method (2) supervised machine learning based network traffic classification method; (3) unsupervised machine learning based network traffic classification method. The supervised machine learning based algorithms are divided into Bayesian based algorithm, decision tree based algorithm, support vector machine based algorithm and neural network based algorithm while the unsupervised machine learning based algorithms are divided into model-based method, density-based method and division-based method *etc*. [2-3].

2. Traffic Identification Model Introduction

The network traffic identification model in this paper is established using a multilayer excitation function quantum neural network which is suitable for data classification and its process is shown in Figure 1.

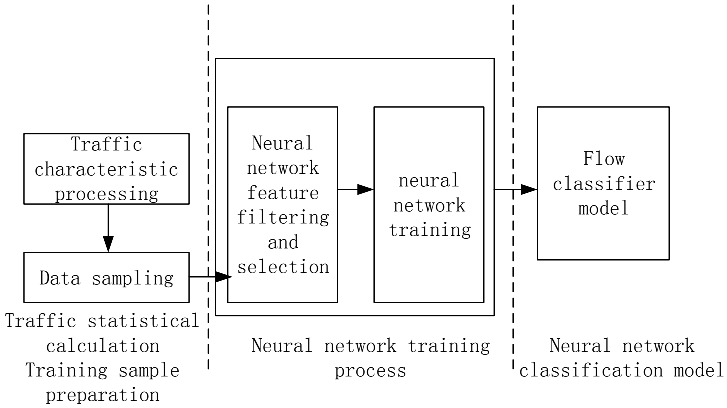

Figure 1. The Process of Establishing Traffic Identification Model based on Neural Network Algorithm

There are 12 types of network traffic which are identified and studied in this paper as shown in Table 1.

Table 1. Network Traffic Classification

NO	Category	NO	Category
A	WWW	G	MULTIMEDIA
B	P2P FILE SHARING	H	INTERACTIVE
C	P2P AUDIO VIDEO	I	DATABASE
D	P2P INSTANT MASSANGING	G	BULK
E	ATTACK	K	SERVICES
F	GAMES	L	MAIL

The extraction of session flow characteristics for network traffic identification is essential for the establishment of recognition model. There are more than 200 kinds of session flow characteristics for establishing recognition model, but most apply only to TCP protocol traffic. Since the current UDP protocol has been widely applied, especially its application in the P2P technology, the UDP protocol cannot be ignored. So the 40 kinds of session flow characteristics capable of identifying UDP protocol and TCP protocol used in this article are: mean and variance of data packets distribution of session flow sending and receiving data, the mean and variance of interval of session flow sending and receiving data, session flow duration, the number of packets of session flow sending and receiving data, the ratio of numbers of bytes of session flow sending and receiving data, the ratio of numbers of session flow sending and receiving data packets and the number of bytes of session flow sending and receiving data, *etc*. [4-5].

3. Quantum Neural Network

Generally, quantum neural network consists of three layers, namely, the input layer of n_i input neuron, the hidden layer of n_h neurons and the output layer of n_o neurons. The hidden layer neurons of quantum neural network studied in this paper used multi-sigmoid function, as follows

$$\begin{aligned}\tilde{h}_{j,k} &= \frac{1}{n_s}\sum_{r=1}^{n_s} h_{j,k}^{r} \\ &= \frac{1}{n_s}\sum_{r=1}^{n_s} sgm\left(\beta_h\left(\sum_{r=1}^{n_s} v_{ji}x_{i,k} + b_j^1 - \theta_j^r\right)\right)\end{aligned}, \tag{1}$$

where $\tilde{h}_{j,k}$ is the output of the j^{th} neuron of the hidden layer to the input of x_k; $h_{j,k}^{\gamma}$ is the output of the j^{th} neuron of the hidden layer to the input of the first sigmoid function; And n_s is the layers of the multi-layer activation function, namely the number of sigmoid function in excitation functions; β_k is the slope parameter of the sigmoid function in the hidden layer; θ_j^{γ} is the translation spacing of the first sigmoid function in the j^{th} neuron; v_{ji} is the connection weight of the j^{th} neurons and the i^{th} input neuron; and b_j^1 is the bias of the j^{th} neuron in the hidden layer. The sigmoid function can be denoted by $sgm()$

$$sgm(\tau) = 1/\left(1+\exp(-\tau)\right). \tag{2}$$

Traditional quantum neural network is divided into two training process, the first one is the training for the network weights v and ω, in order to minimize the sum of squared errors between the neural network output value and expectations

$$\{v,\omega\} = \arg\min\left\{E = \sum_{k=1}^{K}\sum_{i=1}^{n_o}\left(y_{i,k} - \hat{y}_{i,k}\right)^2\right\}, \tag{3}$$

where k is the number of training samples; E is the objective function of weights v and ω; $y_{j,k}$ is the expectation the i^{th} neuron output of the input x_k; $\hat{y}_{j,k}$ is the actual value of the i^{th} neuron output of the input x_k;

The second training process is the training for the translational spacing θ of sigmoid function of the neuron in the hidden layer to minimize the output sum variance of hidden layer neuron

$$\begin{aligned}&\{\theta\} = \arg\min\left\{G = \sum_{j=1}^{n_h}\sum_{m=1}^{n_o}\sigma_{j,m}^2\right\} \\ &\arg\min\left\{\sum_{j=1}^{n_h}\sum_{m=1}^{n_o}\sum_{x_k:x_k\in C_m}\left(\left\langle\tilde{h}_{j,Cm}\right\rangle - \tilde{h}_{j,k}\right)^2\right\}\end{aligned}, \tag{4}$$

where G is the objective function of translation spacing of neural sigmoid function in the hidden layer; C_m is an input data set of class m; $\langle\tilde{h}_{j,Cm}\rangle$ is the output mean value of the j^{th} neurons in input data set class m in the hidden layer:

$$\left\langle\tilde{h}_{j,Cm}\right\rangle = \frac{1}{|C_m|}\sum_{x_k:x_k\in C_m}\tilde{h}_{j,k}\ , \tag{5}$$

where $|C_m|$ is the total number of data in set C_m.

In the above two training process, using gradient descent method for the objective function, the weights of the network v and ω as well as the translational spacing θ are iterated in neurons in the hidden layer. In the iterative process, these parameters can be adjusted. Since different objective function are used in in the training for weights v, ω and translational spacing θ, the coupling parameters of the two types of processing training iteration are not disposed. This will result in the middle and later stage of the training iteration process, and it may be possible to reduce the objective function value of a kind of parameter, and make the objective function value of another kind of parameter increase.

Setting the objective function of translational spacing θ of sigmoid function of neurons in the hidden layer is to classify the data, whose processing unit is the same type of data. So you need to re-set the objective function [6-7].

In summary, in order to avoid reducing the objective function value of a kind of parameter, and making the objective function value of another kind of parameter increase in the middle and later stage of the training iteration process, we set E to be not only the objective function of network weights $\mathbf{v}$ and ω but also the objective function of translational spacing θ of sigmoid function of neurons in the hidden layer. The training objective is to minimize the sum of squared error of the neural network output and the desired value

$$\{v,\omega,\theta\} = \arg\min\left\{E = \sum_{k=1}^{K}\sum_{i=1}^{n_o}\left(y_{i,k} - \hat{y}_{i,k}\right)^2\right\}. \tag{6}$$

We set the objective function is denoted by the square sum of error of k set of samples

$$\begin{aligned} E &= \sum_{k=1}^{K}\sum_{i=1}^{n_o}\left(y_{i,k} - \hat{y}_{i,k}\right)^2 = \sum_{k=1}^{K}\sum_{i=1}^{n_o}\left(c_{i,k}\right)^2 \\ &= \sum_{i=1}^{N}\left(v_i\right)^2 = v^T v \end{aligned}, \tag{7}$$

where $v^T = \left[v_1\, v_2 \cdots v_N\right] = \left[c_{1,1}\, c_{2,1} \cdots c_{n_o,1} c_{1,2} \cdots c_{n_o,K}\right]$ and $N = K \times n_o$.

The network weights $\mathbf{v}$, ω ,and the translational spacing θ of sigmoid function of neurons in the hidden layer is consolidated as a parameter vector

$$\begin{aligned} x^T &= \left[x_1\, x_2 \cdots x_N\right] = \left[v_{1,1}\, v_{1,2} \cdots v_{n_h,n_i} b_1^1 \cdots b_{n_h}^1 \right. \\ &\quad \left. \omega_{1,1} \cdots \omega_{n_o,n_h}\; b_1^2 \cdots b_{n_o}^2\; \theta_1^1 \cdots \theta_{n_h}^{n_o}\right] \end{aligned}, \tag{8}$$

where $n = n_h\left(n_i + 1\right) + n_o\left(n_h + 1\right) + n_h n_s$.

For x, using Newton's Method to update we obtained

$$x_{k+1} = x_k - A_k^{-1} g_k, \tag{9}$$

where $A_k = \nabla^2 E|_{x=x_k}, g_k = \nabla E|_{x=x_k}$.

The gradient of the objective function for x is expressed as

$$\nabla E = 2J^T\left(x\right)v\left(x\right) \tag{10}$$

where $J(\mathrm{x})$ is the Jacobian matrix of the objective function.

Hesse matrix of the objective function is expressed as

$$\nabla^2 E = 2J^T\left(x\right)J\left(x\right) + 2S\left(x\right), \tag{11}$$

where S(x) is usually small so it is negligible.

We rearrange the equation and obtain the updated Gauss - Newton law expressed as

$$x_{k+1} = x_k - \left[J^T\left(x_k\right)J\left(x_k\right)\right]^{-1} J^T\left(x_k\right)v\left(x_k\right), \tag{12}$$

The updating method for LM algorithm is

$$x_{k+1} = x_k - \left[J^T\left(x_k\right)J\left(x_k\right) + \mu_k I\right]^{-1} J^T\left(x_k\right)v\left(x_k\right) \tag{13}$$

where $\mu_k I$ is the correction factor [8].

In summary, through the improved training methods, the network training iterative processes of quantum neural are as follows: Step 1: To input a training sample for quantum neural network, and get the output error $\mathrm{c} = \mathrm{y} - \hat{\mathrm{y}}$ between the actual output and expected output of the neural network, and then solve the square sum of error of each sample E and Jacobi matrix J.

Step 2: Solve Δx_k by the updated formula of LM algorithm.

Step 3: Solve the square sum E' of error by using $\Delta x_k + x_k$. If the new square sum of error is smaller than the former one, then divide the correction factor μ by translational spacing θ of the sigmoid function. Then set $x_{k+1} = \Delta x_k + x_k$ and go back to Step 2.

If the calculated value of gradient of the objective function for x is smaller than the set value, or the number of training iterations reaches the set maximum value, or the sum of squared error decreases to the set target value, then it is determined that the quantum mortality rate has reached convergence and iterative training is terminated [9-10].

4. Experimental Analysis

4.1. Experimental Data Acquisition and Processing

The data packet acquisition process is shown in Figure 2.

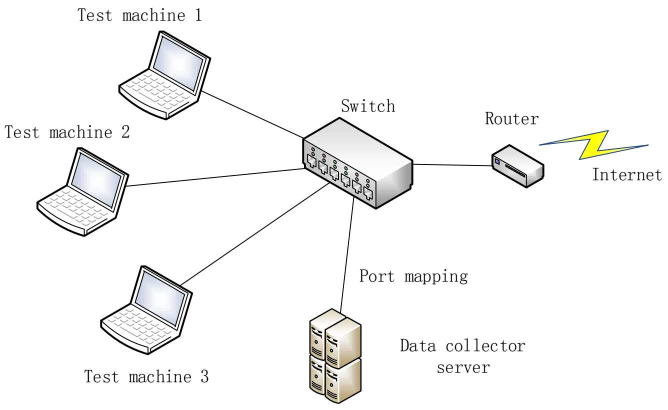

Figure 2. Network Traffic Data Acquisition Process

While collecting the data, running a kind of service or application in table 1 at the same time in N sets of test machines, a total 80000 pieces of session data flow data are mirrored to collection server by port mapping. Since network traffic are divided into 12 class types, therefore, there are overall pieces of session data. From March 2, 2015 to 2015 March 8 and from April 22, 2015 to April 29, 2015, we collected a set of data for each period so we obtained two sets of experimental data for training and performance testing of network traffic classification system. Half of the data was randomly selected for training the identification model established by the neural network, the other half is used for test after the training.

3.2. Detection Model Performance Evaluation Criteria

Precision and recall are used to evaluate the performance of the identification algorithm

$$recall = \frac{TP}{TP+FN} \times 100\% , \tag{14}$$

$$precision = \frac{TP}{TP+FP} \times 100\% , \tag{15}$$

where TP (True Positive) is the number of samples correctly identified as a certain type by the identification model; FN (False Negative) is the number of samples not correctly identified as a certain type by the identification model; and FP (False Positive) is the number of samples wrongly identified as a certain type by the identification model.

4.3. Establish Identification Model

According to traffic type of the session flow characteristics for identification, this paper determines the number of input neurons of quantum neural network algorithm for traffic identification is 40 and the number of output neurons is 12. And the hidden layer uses three-layer multi -sigmoid function as neurons while the output layer uses purelin function. The translational spacing of neurons in the hidden layer is set to [-1.5 0 1.5] and the slope is set to 1.5. The expected error of quantum neural network is 0.001 and the maximum number of training iterations is 10^3.

The network traffic identification model by LM-BP neural network algorithm and conventional quantum neural network algorithm is compared with the LM- quantum neural network algorithm studied in this paper.

Since LM- quantum neural network algorithm of this study needs matrix multiplication, matrix inversion and other operations, so the computational complexity is much more than the other two algorithms and the average time for one iteration is relatively longer. When the number of neurons in the hidden layer is set to 38, the average time for iteration is 3.624s, while under the same conditions, conventional quantum neural network algorithm only needs 1.375s and LM-BP neural network algorithm needs 2.812s. However, the average time for one iteration is fairly long time does not determine that the algorithm convergence is slow.

The number of iterations using three algorithms to obtain the convergence value comparison is shown in Figure 3. The algorithm of this paper can use a minimum of hidden layer neurons to achieve convergence and while at the same number of neurons in the hidden layer, the number of iterations to achieve convergence reaches the least, so it has the fastest convergence rate.

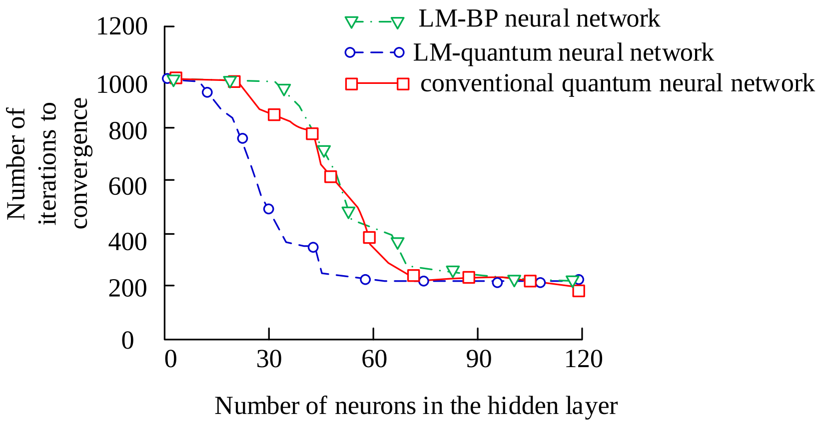

Figure 3. The Number of Iterations Required of Three Kinds of Algorithms

Figure 4 shows, when the number of neurons in the hidden layer is set to 38, from the output of the sum of mean square error of the three algorithms, the use the algorithm of this paper can achieve the fastest decline for sum of mean square error and convergence value is the minimum. Compared to conventional quantum and LM-BP neural network, convergence accuracy is increased by about 10 times.

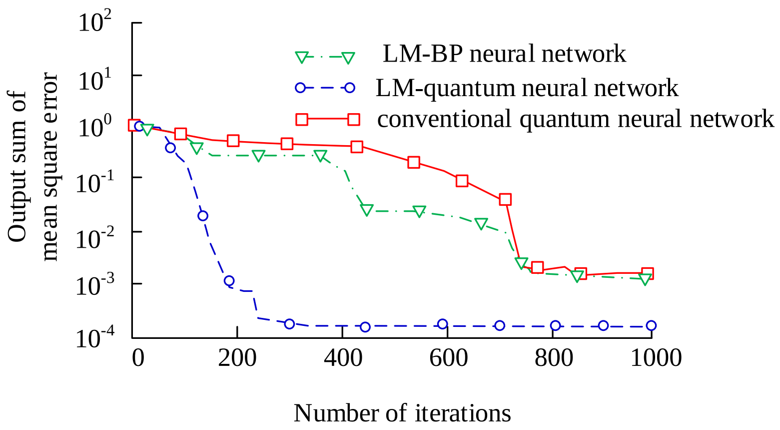

Figure 4. The Output Mean Square Error and Variation of the Three Algorithms

Table 2 lists the comparison of network convergence value and convergence time using three neural network algorithms. The same can be seen that the convergence precision of LM- Quantum Neural Network and the convergence rate in this study are the highest.

Table 2. Convergence and Convergence Time of the Three Algorithms

Algorithm	LM-BP algorithm	Conventional quantum neural network	LM- quantum neural network
Convergence value	8.12×10^{-2}	7.64×10^{-2}	8.76×10^{-3}
Convergence time/s	1120.6	1961.2	768.3

Using the collected test data to test the performance of the network traffic identification model established by the three algorithms, the results are shown in Table 3. The algorithm model in this paper has the best identification accuracy and response rates for 12 types of network applications, respectively reached 91.61% and 93.61%, while identification accuracy and response rates of the other two algorithms are close, but 10% lower than that of the algorithm in this paper.

Table 3. Detection Accuracy and Feedback Rate of Three Detection Models

classification	LM- quantum neural network		LM-BP neural network		Conventional quantum neural network	
	Response rate	Accuracy rate	Response rate	Accuracy rate	Response rate	Accuracy rate
A	94.40	95.50	84.20	87.32	84.30	87.74
B	92.38	94.48	85.21	84.95	85.09	84.49
C	87.87	89.66	79.10	82.24	78.77	82.19
D	92.71	94.69	83.88	86.30	84.08	85.82
E	93.66	96.52	86.70	88.04	87.11	87.83
F	92.16	93.65	81.40	82.05	81.49	82.32
G	90.16	91.71	79.96	81.67	80.31	81.66
H	92.13	92.46	83.04	85.65	83.37	85.18
I	90.27	91.53	81.69	82.99	81.97	83.10
J	90.46	91.83	82.10	83.43	82.55	83.17
K	94.40	97.14	86.37	87.19	86.29	87.47
L	89.01	91.20	78.60	80.72	78.26	80.80

5. Conclusion

In this paper, a network traffic identification model is established using a multilayer excitation function quantum neural network which is suitable for data classification. And the performance of the model is tested experimentally:

(1) In order to avoid reducing the objective function value of a kind of parameter, and making the objective function value of another kind of parameter increase in the middle and later stage of the training iteration process, we use LM algorithm to optimize, using the same objective function not only as the target function of the network weight, but also the function of translational spacing of sigmoid function of neurons in the hidden layer, and the training objective is to minimize the sum of squared error of the neural network output and the desired value.

(2) Since LM- quantum neural network algorithm of this study needs matrix multiplication, matrix inversion and other operations, so the computational complexity is much more than the other two algorithms and the average time for one iteration is relatively longer. However, the average time for one iteration is fairly long time does not determine that the algorithm convergence is slow.

(3) Using the algorithm of this paper can achieve the fastest decline for sum of mean square error and convergence value is the minimum. Compared to LM-BP and conventional quantum neural network, LM- quantum neural network algorism has the best ability to identify network traffic.

References

[1] Y. Shi-wei, Zhu Ke-jun, D. Feng-qin, “A dynamic all parameters adaptive BP neural networks model and its application on oil reservoir prediction”, Applied Mathematics and Computation, vol. 195, no. 1, **(2008)**, pp. 66-75.

[2] A. Este, F. Gringoli and L. Salgarelli, “Support vector machines for TCP traffic classification”, Computer Networks, vol. 53, no. 14, **(2009)**, pp. 2476-2490.

[3] L. Zhu, Y. Rui-xi and G. Xiao-hong, “Accurate classification of the internet traffic based on the SVM Method”, IEEE International Conference on Communications. Glasgow, Scotland: IEEE, **(2007)**, pp. 1373-1378.

[4] B. Raahemi, Z. Wei-cai and L. Jing, “Peer-to-peer traffic identification by mining IP layer data streams using concept-adapting very fast decision tree”, Proceedings of the 20th IEEE International Conference on Tools with Artificial Intelligence, Dayton, OH, USA: IEEE, **(2008)**, pp. 525-532.

[5] A. Callado, J. Kelner and D. Sadok, “Better network traffic identification through the independent combination of techniques”, Journal of Network and Computer Applications, vol. 33, no. 4, **(2010)**, pp. 433-446.

[6] S. Jian, Z. Xiong-wei and S. Xin-jian, “A new training algorithm for quantum neural networks”, Signal Processing, vol. 27, no. 9, **(2011)**, pp. 1306-1312.

[7] M. Indrajit and R. Srikanta, “Comparing the performance of neural networks developed by using Levenberg-Marquardt and Qusai-Newton with the gradient descent algorithm for modeling a multiple response grinding process”, Expert Systems with Applications, vol. 39, no. 3, **(2012)**, pp. 2397-2407.

[8] Z. Hong-yan and G. Zheng, “Novel interpretation for Levenberg-Marquardt algorithm”, Computer Engineering and Applications, vol. 45, no. 19, **(2009)**, pp. 5-8.

[9] B. M. Wilamowski and H. Yu, “Improved computation for Levenberg-Marquardt training”, IEEE Transactions on Neural Networks, vol. 21, no. 6, **(2010)**, pp. 930-937.

[10] W. Jin-ming, W. Geng and Z. Guo-hong, “Speaker recognition method based on quantum neural networks”, Journal of PLA University of Science and Technology (Natural Science Edition), vol. 13, no. 3, **(2012)**, 242-246.

Authors

Wu Fei, he is involving in the project information system for integrated network management system. Now he is working in State Grid Fujian Electric Power Company in NO.257, Wusi Road, Fuzhou, Fujian Province.