

An Objective Assessment Method of VoIP Video Quality Based on Network Parameters

Xiaohan Zhao
Beijing Institute of Technology
Beijing,100081
China
869081077@qq.com

Chengcai Li
Beijing Institute of Technology
Beijing,100081
China
1226619245@qq.com

Fei Wang
Beijing Institute of Technology
Beijing,100081
China
396399295@qq.com

Jing Wang
Beijing Institute of Technology
Beijing,100081
China
wangjing@bit.edu.cn

Zesong Fei
Beijing Institute of Technology
Beijing,100081
China
feizesong@bit.edu.cn

ABSTRACT

With the rapid development of Internet technology and popularization of the fourth generation of mobile communication system, VoIP audio & video communication is becoming more and more prevalent. On account of that all the data are transmitted through the IP network, some uncertain factors in the network such as packet loss, delay and jitter would have a strong impact on the quality of communication. Considering that circumstance, this paper proposes a kind of objective video quality assessment method based on network parameters and builds an evaluation model that could predict an objective value of the quality of video call on the basis of those network parameters. The proposed method mainly takes advantage of the Principal Component Analysis (PCA) algorithm and Neural Network (NN) technology which analyzes the parameters collected in network during VoIP video call and obtains the principle components to reduce the dimension of the data that is divided into training data and testing data further, so that it can predict the subjective score with the principle network parameters. According to the experimental results, the objective values forecasted by the proposed objective model have a strong correlation with the subjective scores of video quality gained by the video call.

CCS CONCEPTS

- Networks → Network performance evaluation → Network performance modeling.
- Networks → Network performance evaluation → Network performance analysis.

KEYWORDS

VoIP, Video Quality Assessment, PCA, Network Parameter, Neural Network

1 INTRODUCTION

Along with the great advance in network technology and the popularization of mobile intelligent devices, the intelligent audio & video communication application software whose transmission is based on the network becomes more and more popular. For example, the call services and experiences that VoIP video APP provides is becoming more convenient and more efficient for the users. Network is everywhere and thus the demand of communication is everywhere. Due to its transmission through the network, VoIP video calls are bounded to be affected by the network situation. Some uncertain factors such as network packet loss, network delay and jitter will severely impact the whole process of VoIP video calls, and further affect the performance of Quality of Experience (QoE). So in such situation, in order to improve the experience and QoE score of VoIP video call, designing a precise and effective VoIP objective video quality assessment method is very necessary. The purpose of this paper is to do such a thing, which puts forward a kind of video quality evaluation model based on network parameters such as packet loss, delay and jitter.

It is known that video quality assessment methods can be roughly divided into two categories: subjective assessment and objective assessment. Among them, subjective evaluation is recognized as the most reliable method of video quality evaluation internationally which is based on people's subjective feelings. Evaluators should be divided into several groups when estimate the same video material in specific environment and obtain the Mean Opinion Score (MOS) [1]. Subjective assessment method has some limits and specifications with reference to ITU-T BT.500 [2] and ITU-T BT.1788 [3].

The subjective video quality assessment method evaluates results quickly, but it is hard to meet the actual demand. So raising an objective assessment method is very necessary. The objective evaluation also has three categories: Full Reference quality assessment, Reduced Reference quality assessment and No

Reference quality assessment. The proposed quality evaluation method in this paper belongs to no reference quality assessment and there are also some other methods: The Fuzzy Mathematics based evaluating [4], The MintMOS based evaluating [5], The Cellular Networks based evaluating [6].

In this paper, Section 2 introduces the evaluation algorithm and model in detail. Experimental assessment results and fitting results are showed in Section 3. Section 4 in this paper gives the conclusion.

2 MODEL AND ALGORITHM

2.1 Specification and Preparation of Model

As we all know, the factors in the actual network that impact VoIP video communication are various and complicated. Therefore, we select some important network parameters such as packet loss, delay, jitter and frame rate. Although it is not comprehensive to evaluate VoIP video quality considering those factors only, this method demonstrates that the predicted values have a high correlation with the subjective scores according to experimental results. In this paper, a VoIP video communication application and controllable network environment are prepared, so the next work can be done successfully.

2.2 Model of VoIP Video Material and Network Parameter Collection

In this paper, all the video materials and network parameters are collected in the real network environment, so those records should be authentic, and thus the predicted values from the evaluation model should be convincing. During the video call, the video materials are recorded automatically, the packet loss, delay, jitter and video frame rate are collected per second in a TXT file. The following picture named Figure 1 is the VoIP video call system that consists of WANem and two Android mobile terminals (MT). The WANem is network simulator (NS) that can control network bandwidth, packet loss, delay and jitter parameters.

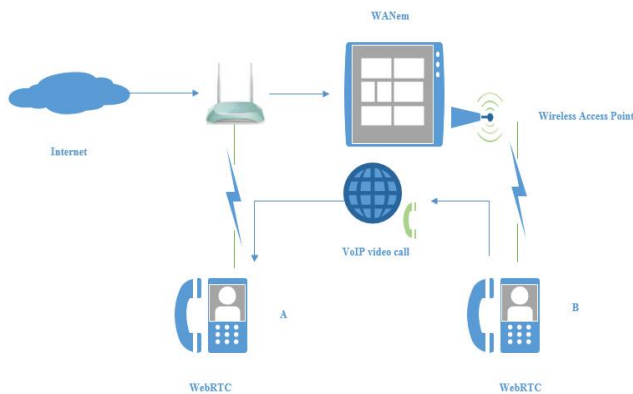


Figure 1: System of VoIP video call and data collection

In the system, the two MTs are installed with VoIP video call application based on WebRTC, one of which is connected to the Router by wireless access point through WANem, and the other one is connected to the Router directly. And the WANem can control

the network path through which B sending data packet to A, so it can affect the packet loss, delay, jitter of B.

Once VoIP video call between A and B is established, A would record the video materials and collect the parameters such as packet loss, delay, jitter of B. Every video call lasts about 30 seconds. Besides, WANem would set different values of bandwidth of B, so that the video quality of the materials recorded in A should have different levels of subjective score that varies from 1 to 5 representing terrible to perfect, which is shown in Table 1.

Table 1: Score and level of video quality

Score	Level of video quality
1.0	Terrible
2.0	Poor
3.0	General
4.0	Good
5.0	Perfect

These scores of the video materials recorded in A are graded according to the Tab 1, but in fact, score 4.0 and score 5.0 are seldom gained. In this paper, all the network parameters collected and the scores of subjective video quality gained are divided into two parts, one named training data and the other named testing data.

2.3 PCA Algorithm

When the multivariate issue is studied, the more variables, the larger amount of calculation and higher computational complexity are observed. As an evaluation method, PCA, abbreviation of Principal Components Analysis, can lower the dimension of high variant space while ensuring minimal data loss.

We assume that there are n indexes, namely $x_1, x_2, x_3, \dots, x_n$, which indicates their respective features of the objects. PCA is applied to these n indexes and m new indexes $z_1, z_2, z_3, \dots, z_m$ ($m < n$) will be obtained which is linear combination of $x_1, x_2, x_3, \dots, x_n$. The variance of Z increases gradually as the following:

$$V(z_1) \geq V(z_2) \geq V(z_3) \geq \dots \geq V(z_n)$$

Therefore, the degree of Z 's variation and the original data it carries both increase [7].

The following is the procedure of PCA:

- (1) Standardize the data by Z-score.

$$y_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad i = 1, 2, \dots, I, \quad j = 1, 2, \dots, J \quad (1)$$

Where x_{ij} is the value of the j th criteria belonging to the i th sample. \bar{x}_j is the sample mean of the j th criteria. σ_j is the sample standard deviation of the j th criteria. In order to establish a suitable model for fitting, we extract network parameters of videos for 30 times in 30 seconds and then determine more than 21 indicators

such as packet loss rate, delay, jitter, variance of frame rates, maximum, minimum, average, median and mode of frames.

Initially, calculating the matrix of correlation coefficients. Then, removing those variables whose correlation coefficients are greater than 0.8. Afterwards, determining the packet loss rate range, median of packet loss rate, delay range, average delay, maximum and minimum jitter buffer, average jitter buffer, maximum frame rate, minimum frame rate, variance of frame rate, average frame rate as the initial indicators and finally standardize the data.

(2) Calculate the covariance matrix of data sheet.

The quotation of covariance matrix between variables X and Y can be written as the following.

$$Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \quad (2)$$

(3) Compute eigenvectors and eigenvalues of covariance matrix.

For matrix A, if the equation $AX=\lambda X$ has non-zero solution, λ is the eigenvalue of A and the non-zero solution X is its eigenvectors. And the eigenvectors can be attained from the formula below.

$$|A - \lambda I| = 0 \quad (3)$$

(4) Select eigenvectors $u_1, u_2, u_3, \dots, u_j$ which corresponds to relatively big eigenvalues to gain a new data set.

Principle components extraction needs to ensure the information of original data as much as possible, which can be obtained by calculating the cumulative contribution rate. The follow E is used to express cumulative contribution rate:

$$E = \frac{\sum_{k=1}^m \lambda_k}{\sum_{j=1}^J \lambda_j} \quad (4)$$

In this paper, we set $E > 85\%$ (85% is enough to express the original data) and get six principal components in the following Table 2 [8].

Table 2:Extraction sums of squared loadings

Component	Eigen value	Variance %	Cumulative %
F1	4.182	32.169	32.169
F2	3.838	29.524	61.693
F3	1.213	9.331	71.024
F4	0.905	6.965	77.989
F5	0.706	5.433	83.422
F6	0.600	4.619	88.040

(5) Calculate principal components through the following formula. The transition matrix is showed in Table 3 which can transform a 13-dimensions matrix to a 6-dimensions matrix.

$$z_k = \sum_{j=1}^J u_j x_j \quad (5)$$

Table 3: Transition matrix

	F1	F2	F3	F4	F5	F6
X1	0.381	-0.139	-0.006	0.375	-0.059	-0.118
X2	0.226	-0.178	0.211	0.412	0.771	0.132
X3	0.325	-0.027	0.219	0.484	-0.524	0.085
X4	0.244	0.359	0.245	-0.195	0.084	0.015
X5	-0.386	0.014	0.358	0.159	-0.070	-0.265
X6	0.039	0.421	0.216	-0.170	0.238	-0.234
X7	0.271	0.304	0.246	-0.140	-0.201	0.374
X8	-0.329	0.005	0.483	0.175	-0.077	-0.382
X9	0.146	0.417	0.231	0.015	0.059	0.046
X10	-0.147	0.400	-0.332	0.303	0.005	0.043
X11	-0.385	0.040	0.196	0.250	-0.024	0.534
X12	0.076	0.366	-0.365	0.337	-0.003	-0.407
X13	-0.327	0.279	-0.202	0.208	0.080	0.307

2.4 Neutral Network Model

Neutral Network is in essence a kind of computing model made up of large numbers of neutrons (nodes) and connections between them. In Neutral Network, each neutron represents an output function while the connection between every two nodes indicates a weight. Output of Neutral Network will vary according to network connection methods, weights and output functions. Based error back propagation algorithms, BP (Back Propagation) is multi-tier back propagation Neutral Network [9] whose relevancy of non-linear parameters is relatively high. So BP is an ideal prediction method which consists of an input layer, one or more hidden layer and an output layer.

The learning process of networks is made up of forward propagation of data and back propagation of error. During the process of training, the information goes through the input layer and the hidden layer and finally is transferred to the output layer. Afterwards, a comparison will be made between the output value and the actual value. If there is an error, the error will be propagated backward to modify layer by layer node weights of each layer to reduce the error until the output results meet the accuracy requirements [10].

The neutron i of BP's hidden layer and output layer is determined by the following formula:

$$O_i = f_i(\sum w_j x_j - \theta_i) \quad (6)$$

In the above formula, f_i represents activation function corresponding to neutrons. The activation function of the hidden

layer neurons adopted in this paper is Sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ [11] while that of the output layer neurons is linear function $f(x) = x$. Additionally, θ_i is the threshold of i , x_j represents the input of every neuron and w_j corresponds to the weight of the connection between the input and the neuron.

Specific procedures are taken as the following:

- (1) Initialize the data, determine the number of nodes in each layer, set each weight and threshold to a relatively small number.
- (2) Input and output the samples, and study each sample.
- (3) Calculate the output of the output layer neurons according to the activation function of the input samples and each node.
- (4) Compare calculation results with the actual output, and then calculate the error of each layer.
- (5) Update the threshold of each node of input layer, hidden layer and output layer and the weight between the connections according to the calculation results.
- (6) If the error is less than the figure of the learning accuracy, the process of study is finished. Otherwise, restudy from step (2).

Data mining software Weka is adopted in this training where three layers BP neural network model is established as Figure 2, and its input are 6 principal components extracting by PCA's dimensionality reduction.

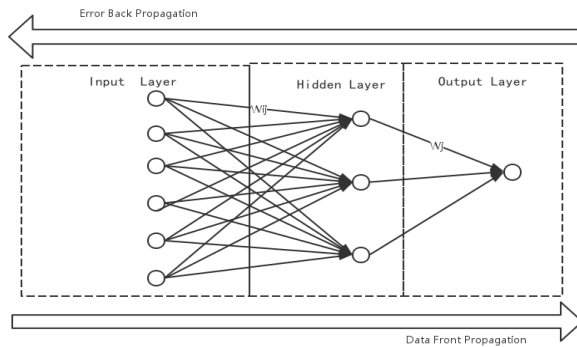


Figure 2: Structure of Neutral Network

In our Neutral Network, the weights and thresholds can be seen in the Table 4.

Table 4.:Weight and threshold of neural network

Layer	Node	Input Node	Weights	Threshold
Hidden Layer	Hidden Node 1	F1	-0.2471	-1.1849
		F2	-0.4119	
		F3	-0.8831	
		F4	-0.7656	
		F5	0.7063	
		F6	2.0534	
	Hidden Node 2	F1	5.6969	0.8660
		F2	1.1353	
		F3	1.6974	
		F4	2.8256	
		F5	-0.7166	
		F6	-2.1470	
	Hidden Node 3	F1	-0.6417	-1.6036
		F2	2.4842	
		F3	-0.4190	
		F4	1.2415	
		F5	1.3374	
		F6	1.5301	
Output Layer	Output Node	HiddenNode 1	-1.2432	0.9242
		HiddenNode2	-1.8667	
		HiddenNode3	1.5100	

3 RESULTS AND DISCUSSION

3.1 Evaluation Indexes

In order to evaluate the merits and demerits of the fitting model, we choose the following indicators to compare the predicting results and the actual result.

- (1) Correlation coefficient (CC)

The closer CC correlation coefficient to 1, the higher the linear correlation degree of between variables. The relevancy between predicted results and the real value [12] is determined by the following formula:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

x_i and y_i represent real score and predicting score.

- (2) Mean absolute error (MAE)

The MAE is the mean of the absolute value of the difference between actual and estimated values of all samples.

$$MAE = |ActualMOS - PredictedMOS| \quad (8)$$

(3) Root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{(ActualMOS - PredictedMOS)^2}{n}} \quad (9)$$

These three indicators can effectively describe the correlation between the predictive results and the real values and the accuracy of the algorithm.

3.2 Results and Analysis

In order to effectively establish the prediction model, all the 60 samples that include video call materials and corresponding network parameters are divided into testing set and training set at the ratio of 1:4. In this paper, a total of 60 samples, including 45 training sets and 15 test sets.

The actual values and predicted values of testing set are shown in following Table 5:

Table 5: MOS values of testing set

Actual MOS	Predicted MOS	Error
3	2.427	-0.573
2.5	2.978	0.478
2.5	1.526	-0.974
3	3.345	0.345
1	1.08	0.08
3	3.198	0.198
1.5	1.102	-0.398
3	3.098	0.098
3.5	3.383	-0.117
1	1.046	0.046
3	3.39	0.39
3.5	3.603	0.103
3	3.237	0.237
1	1.393	0.393
1	0.806	-0.194

And the performance is shown in Table 6:

Table 6: Performance of testing set

Performance	CC	MAE	RMSE
Test set	0.9267	0.3085	0.3888

It can be seen from Table 5 that the predicted MOS of testing data are very close to actual MOS statistically. And the performance of testing set in Table 6 shows that the predicted MOS values have a high correlation coefficient and low error.

Figure 3 and Figure 4 are the scatter diagrams of all predicted values of testing set and training set in which the horizontal axis represents prediction results and actual results as ordinate. The line of y=x is in the middle and the others are the 0.5 MOS points error lines.

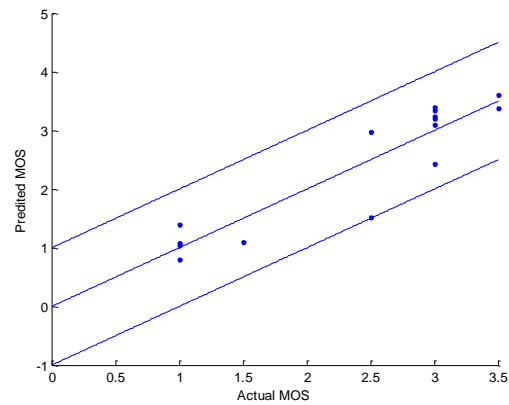


Figure 3: Scatter diagram of testing set

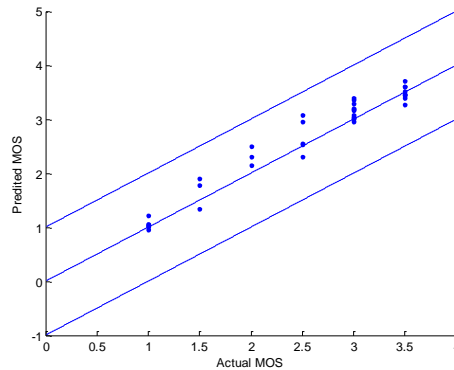


Figure 4: Scatter diagram of training set

From the scatter diagrams in the figures, it is easily seen that the predicted MOS has a good linear relationship with actual MOS of those video materials including testing set and training set. So conclusion can be drawn that the proposed objective VoIP video quality evaluation method according to network parameters is effective and reliable.

4 CONCLUSIONS

In this paper, we propose a kind of objective video quality assessment method based on network parameters and establish a corresponding model that could predict the value of the quality of video call of VoIP. Compared with actual subjective values of video materials collected during VoIP video communication, the predicted results are close to the subjective values and have a high correlation with them. Considering above all, the proposed algorithm and assessment model are proved to be accurate and effective. Moreover, it can be used in the real-time quality monitoring in VoIP video call.

ACKNOWLEDGMENTS

This work was supported by Beijing Natural Science Foundation No.4152047 and 111 Project of China under Grant B14010.

REFERENCES

- [1] ITU-R Recommendation P.800. 1996. *Methods for subjective determination of transmission quality*. International Telecommunication Union, Geneva, Switzerland.
- [2] ITU-R Recommendation BT.500-11. 2002. *Methodology for the subjective assessment of the quality of television pictures[S]*. International Telecommunication Union, Geneva, Switzerland.
- [3] ITU-R Recommendation BT.1788. 2007. *Methodology for the subjective assessment of video quality in multimedia application[S]*. International Telecommunication Union, Geneva, Switzerland.
- [4] L. Li, H. Yan. 2005. *Evaluation the Quality of Steaming Media on Fuzzy Mathematics Theory*. Telecommunications Science, vol. 5.
- [5] M. Venkataraman and M. Chatterjee. 2011. *Inferring Video QoE in Real Time*. IEEE Network.
- [6] G. Gomez. 2005. *End-to-End Quality of Service over Cellular Networks*. Wiley.
- [7] Fu Xiang, Ji Changming. 1999. *A Comprehensive Evaluation of the Regional Water Resource Carrying Capacity—Application of Main Component Analysis Method*. Resources and Environment in the Yangtze Basin, vol.8.
- [8] Li Yan, Zeng Zhenxiang, Zhang Min, Yu Shujiang. 1999. *Application of Primary Component Analysis in the Methods of Comprehensive Evaluation for many Indexes*. Journal of Hebei University of Technology, vol.1.
- [9] Fang Xu. 2011. *The Research and Improvement of the Face Recognition Method Based on the BP Neural Network*. Computer Knowledge and Technology, vol.7.
- [10] Zhou Zheng. 2008. *Surey of Progress in BP Neural Network*. Shanxi Electronic Technology, vol.2.
- [11] Li Xiaofeng, Liu Guangzhong. 2000. *The Improvement of BP Algorithm and Its Application*. Journal of Sichuan University (Engineering Science Edition), vol.32.
- [12] Yan Likun. 2003. *Application of Correlation Coefficient and Biased Correlation Coefficient in Related Analysis*. Journal of Yunnan University of Finance and Economics, vol.19.