

Video Quality Prediction over Wireless 4G

Chun Pong Lau, Xiangliang Zhang and Basem Shihada

CEMSE, King Abdullah University of Science and Technology, Saudi Arabia
{lau.pong, xiangliang.zhang, basem.shihada}@kaust.edu.sa

Abstract. In this paper, we study the problem of video quality prediction over the wireless 4G network. Video transmission data is collected from a real 4G SCM testbed for investigating factors that affect video quality. After feature transformation and selection on video and network parameters, video quality is predicted by solving as regression problem. Experimental results show that the dominated factor on video quality is the channel attenuation and video quality can be well estimated by our models with small errors.

Keywords: Video Quality Prediction, Wireless 4G, Superposition Coded Multicasting

1 Introduction

Fourth generation (4G) of mobile communication standards, such as Long Term Evolution Advanced (LTE-Advanced) and Worldwide Interoperability for Microwave Access Release 2 (WiMAX2), provide high speed and large range of wireless connectivity. Because of the large coverage of a base station (BS), users within the coverage area result in different channel quality causing the multi-users diversity problem. In order to provide a reliable video multicast/broadcast service by high speed wireless channels, She *et al.* proposed Superposition Coded Multicasting (SCM) method in [1]. In this scheme, a scalable video bitstream consists of two different quality layers that are modulated into two signals by two different modulations. These two signals are then superimposed into one broadcasting signal for all receivers under the BS coverage. Since the transmission of superimposed signal can be affected by wireless channel conditions, the quality of received video is key concern for both the providers and receivers.

From the video providers' point of view, the Quality of Service (QoS) is determined by the quality of video that customers will receive. If the quality of video can be predicted according to network conditions and video parameters, providers can guarantee certain level of service by taking effective actions. On the one hand, before broadcasting a video, providers can set the input features/parameters to achieve the targeted video quality. On the other hand, during video transmission, operators can adjust the video and network parameters according to this prediction model for various conditions.

There are three main challenges facing the accurate prediction of video quality over wireless channels. First, the behavior of wireless signal is difficult to predict and measure, since it fluctuates and can be significantly affected by complex

and mixed factors, such as attenuation, fading, and interference [2]. Attenuation happens due to the distance between transmitter and receiver. Fading may vary over time, position or radio frequency due to multi-path propagation. Second, multi-user channel diversity challenge where users at different geographic locations within the same BS coverage area have different channel quality [3]. Third, collecting data for analysis and prediction is expensive. In order to accurately predict the video quality, all possible conditions that may affect video quality must be considered, and all corresponding data should be collected. Constructing an emulation testbed with implementation of SCM is often expensive. To the best of our knowledge, the testbed used in this paper is the first built emulation testbed on this research area. This work assists research work without real testbed by building a practical prediction model. It helps to reduce the dependence on the expensive testbed system through estimating transmission results, instead of conducting real experiments.

Video quality prediction can be formalized as a regression problem. Video transmission is described by a set of features, including video and network parameters. Video quality, measured by a common quality metric called Peak Signal-to-Noise Ratio (PSNR), is the target for prediction [4]. Given a raw data, feature selection is first used for selecting the useful features from the data set. Three popular regression approaches, k -nearest neighbor algorithm (k -NN), Support Vector Machines (SVM) and Neuron Network (NN) are employed for prediction. Nominal type feature (*modulation*) and a ratio type feature (*power boosting*) are transformed together into one ratio type feature. The improved results from advance approaches including the transformed feature demonstrate the main factors affecting video quality and should gain more attention in video transmission.

The rest of this paper is organized as follows. Section 2 introduces the background and related work of video quality prediction problem. Section 3 describes the methodology for prediction, which consists of data collection on real testbed, feature transformation and selection, and employed prediction techniques. Section 4 reports experimental results. Section 5 finally concludes and gives perspectives.

2 Background

Researchers face many challenges on video quality prediction. One of the difficulties is the videos are affected by numerous parameters. Through collecting data from simulations of 3G and WLAN network, Khan *et al.* predicted the video quality with both application and network level parameters [5]. However, this was limited to four parameters, which are frame rate, send bitrate, packet error rate, and link bandwidth. These limited number of features cannot represent the network condition well especially when network condition becomes complex in the wireless 4G.

Garcia *et al.* proposed a framework to find a parametric content description of videos [6]. Assuming that video content influences the video quality, they

proposed to decompose the input video to spatial and temporal features which highlight a strong adequation with the perceived video quality. However, prediction of video quality was not quantified in their approach.

Dalal *et al.* investigated the feature selection for predicting video quality after network transmission [7]. Network statistic such as packet lost rate and error rate were considered as features. Two methods correlation ranking and principal component analysis (PCA) were investigated. However, video they studied were collected from a wired network.

Liu *et al.* proposed a framework for video prediction after encoding process, which can reduce time cost for choosing correct parameters for video encoding process [8]. Their approach was simple and efficient. However, they did not consider the parameters of network transmission for the prediction.

All of the above reported studies omitted the wireless network characteristics and assumed a simulation based data. Therefore, we attempt to collect real-life data and perform more accurate prediction considering the wireless network characteristics. In the next section, we introduce our methodology.

3 Methodology

This section introduces our proposed approach of video quality prediction. We first introduce the testbed for data collection. Then, as pre-processing is a crucial part of prediction, we introduce feature transformation, normalization and ranking mechanisms. Finally, k -NN, SVM and NN are introduced to build the regression models for video quality prediction.

3.1 Testbed Description and Data Collection

Video data in this paper is collected by a real-time SCM testbed, which consists of two network emulators, two personal computers (PC) and one variable attenuator. In the transmitter side, one PC is connected to one network emulator acted as the video server and BS. The other PC is connected to the second emulator acted as the mobile receiver and mobile device. Videos are transmitted over the emulators connected by coaxial radio frequency cables through the variable attenuator acting as the wireless channel with different values of channel attenuation. The emulators are implemented by National Instruments NI PXIe-1062Q chassis with embedded controller PXIe-8130, IF transceiver PXIe-5641R, RF Up converter PXIe-5610 and RF Down converter PXIe-5600. Figure 1 shows the system architecture of SCM with transmitter on the top. The video data passes through all layers and become superposition coded signal broadcast over the emulated wireless channel. Receiver receives the broadcasting signals that go through all the layers to reconstruct the video data.

Video streams are encoded as H.264-SVC [9] bitstream, and are stored in the video server. Each video sequence consists of two layers, base layer and enhancement layer. Base layer contains most of the important information of the video and should be correctly received and decoded in order to reconstruct

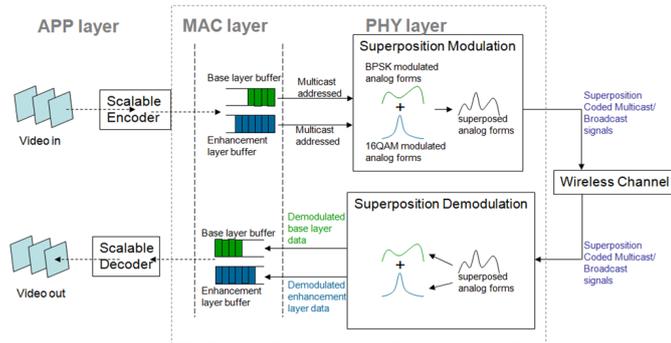


Fig. 1: SCM System Architecture

the video after transmission. Enhancement layer is additional information of the base layer to improve the video quality when successfully decoded. It is expected that enhancement layer bitrate is higher than base layer bitrate since the encoding quantization parameters (QP) of enhancement layer is lower than the base layer QP. Different number of group of pictures (GOP) and intra period are considered as the video features/parameters in the system.

From the network prospective, transmission power, and reception power sensitivity are controlled by the transmitter and receiver emulators in physical layer. Channel attenuation is controlled manually by the variable attenuator. Assume that during each of the transmission, the attenuation value is constant to remain a static transmission environment to reduce complexity. Carrier frequency, modulation, power boosting, cyclic prefix and bandwidth are selected from the MAC layer control software.

A video transmission, when completed, is described by a set of video and network features. The first three columns of Table 1 show the features considered in this work. The last row is the prediction target PSNR, which is commonly used to measure the quality of the reconstructed video after network transmission. It is calculated by comparing the video frame by frame and pixel by pixel between the received video and the original video before encoding and transmission. A higher PSNR value introduces a higher video quality received. Given a video with its known feature values, our target is to predict the PSNR after it is received.

In our experiments throughout the paper, the video transmitted is a 40 seconds length movie trailer, which is a standard video provided by Durian Open Movie project named as "Sintel" [10]. Each test of experimental operation takes around 2 minutes. In addition, calculating the PSNR value between received video with original video requires 1 to 2 minutes. 985 testes were conducted in total. Ten random chosen examples are shown in Table 2.

3.2 Feature Transformation

Two network features called *Base Layer Modulation* and *Enhancement Layer Modulation* are of our interest while predicting video quality. The possible values of these two features are BPSK, QPSK, 16QAM and 64QAM, which are difficult to be directly used for prediction. Therefore, *Base Layer Modulation* and *Base*

Table 1: Video and network features, with statistic, Pearson correlation coefficient and P-value w.r.t. target PSNR

	Features	Type	Min	Max	Range	Count	Pearson Corr. Coef.	P-value
Video	Base bitrate	Ratio	399.5	739	339.5	7	0.116650	2.4360e-4
	Enh. bitrate	Ratio	777.88	3074.25	2296.4	7	0.101631	1.4037e-3
	Overall bitrate	Ratio	1189.3	3473.7	2284.3	6	0.109720	5.6144e-4
	Base Encoding QP	Interval	20	28	8	3	-0.107587	7.1926e-4
	Enh. Encoding QP	Interval	35	40	5	2	0.028953	3.6404e-1
	Group of Pictures	Interval	2	8	6	3	-0.074944	1.8652e-2
	Intra Period	Interval	2	32	30	5	-0.114500	3.1721e-4
Network	Transmission Power	Ratio	0	0	0	1	NaN	NaN
	Reception Power Sen.	Ratio	-10	0	10	2	-0.028953	3.6404e-01
	Attenuation	Ratio	2	21	19	18	-0.502685	3.3726e-64
	Carrier Frequency	Ratio	2.51	2.51	0	1	0.000000	1
	Base Modulation	Nominal	BPSK	BPSK	N/A	1	Transformed	Transformed
	Base Power Boosting	Ratio	-3	2	5	4	Transformed	Transformed
	Base Prob. of Tx Sym. Err	Ratio	2.328e-4	0.8858	0.0884	9	-0.190526	1.6672e-9
	Enh. Modulation	Nominal	QPSK	16QAM	N/A	2	Transformed	Transformed
	Enh. Power Boosting	Ratio	-11	0	11	7	Transformed	Transformed
	Enh. Prob. of Tx Sym. Err	Ratio	0.0789	0.5253	0.4465	9	0.151654	1.7421e-6
	Cyclic Prefix	Interval	0.25	0.25	0	1	NaN	NaN
	Bandwidth	Interval	5	10	5	2	0.145872	4.2871e-6
	Target	PSNR	Ratio	25.26	66.91	41.64	461	

Table 2: Examples of video records

Base Bitrate	Enh. Bitrate	Overall Bitrate	B. QP	E. QP	GOP	I. P.	Tx P.	Rx P.	At.	Freq.	B. Mod	B. PB	E. Mod	E. PB	CP	BW	PSNR
520.63	1686.87	2207	25	35	2	4	0	0	2	2.51	BPSK	-1	QPSK	-7	1/4	10	58.5195
520.63	1686.87	2207	25	35	2	4	0	0	9	2.51	BPSK	-1	QPSK	-11	1/4	10	55.2172
739.	1634.75	2373	25	35	2	2	0	0	7	2.51	BPSK	2	16QAM	0	1/4	10	58.9206
442.	1201.25	1643	25	35	4	8	0	0	7	2.51	BPSK	2	16QAM	0	1/4	10	58.9995
415.13	991.62	1406.75	25	35	8	32	0	0	4	2.51	BPSK	2	16QAM	0	1/4	10	56.7126
411.5	777.88	1189	28	35	4	16	0	0	2	2.51	BPSK	-1	QPSK	-7	1/4	5	56.3021
410.63	1221.5	1189	25	35	4	16	0	0	2	2.51	BPSK	-1	QPSK	-7	1/4	10	55.9479
411.5	777.88	1189	28	35	4	16	0	0	11	2.51	BPSK	-1	QPSK	-11	1/4	5	47.0903
411.5	777.88	1189	28	35	4	16	0	0	5	2.51	BPSK	-1	QPSK	-5	1/4	5	55.5979
399.5	3074.25	3473	20	40	4	4	0	-10	10	2.51	BPSK	-1	16QAM	-5	1/4	10	52.9213

Layer Power Boosting are transformed to *Base Layer Probability of transmitted symbol error*. Similarly for Enhancement layer, *Enhancement Layer Modulation* and *Enhancement Layer Power Boosting* are transformed to *Enhancement Layer Probability of transmitted symbol error*. The idea of the transformation is considering the transmitted symbol energy and the receiver symbol decision area. *Probability of symbol error* means a transmitted symbol falls into other symbols decision area with the amount of Gaussian noise added onto it. The probability of symbol error is calculated as follows according to [11]. Assuming that base layer modulation is always BPSK, the probability of correctly detecting the abscissa of each SPC symbol in the i^{th} region is:

$$P_{i,1} = Q\left(\frac{-\mu_{y_i}}{\sigma_{y_i}}\right) \quad (1)$$

where μ_{y_i} is the mean for received symbol abscissa coordination and σ_{y_i} is the variance. *Base layer probability of transmitted symbol error* is:

$$P_b = 1 - \frac{2\sqrt{m_2}}{M} \sum_{i=0}^{\sqrt{m_2}-1} P_{i,1} \quad (2)$$

where $M = m_1 \times m_2$, m_1 is the base layer modulation number and m_2 is the enhancement layer modulation number.

After the symbol is decoded on base layer by successive interference cancellation (SIC), the remained symbol will be decoded by enhancement layer demodulator. The standard symbol error equation for m_2 -QAM demodulation is expressed as follow:

$$P_{m_2QAM} = 2 \left[2 \left(1 - \frac{1}{\sqrt{m_2}} \right) Q \left(\sqrt{\frac{3}{m_2 - 1} \frac{E_2}{N_0}} \right) \right] - \left[2 \left(1 - \frac{1}{\sqrt{m_2}} \right) Q \left(\sqrt{\frac{3}{m_2 - 1} \frac{E_2}{N_0}} \right) \right]^2 \quad (3)$$

where $Q(\cdot)$ is known as Q-function, E_2 is the remained energy after SIC of symbol for enhancement layer decoding process, N_0 is the Gaussian noise power.

Finally the *probability of enhancement symbol error* is calculated as follow:

$$P_e = 1 - (1 - P_{m_2QAM})(1 - P_b) \quad (4)$$

Transformed feature P_b and P_e range from 0.00024 to 0.08858 for *base layer* and 0.0789 to 0.5253 for *enhancement layer*.

3.3 Feature Normalization

Since values of raw video features are on different scale, normalization is performed to linearly scale them to unit range. An original value x of a feature is normalized to be x^* :

$$x_i^* = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (5)$$

where x_{min} is the smallest value and x_{max} is the largest value of the feature. To further investigate the improvement that can be made by normalization, original and normalized data are compared in the experimental results.

3.4 Feature Ranking

In order to eliminate irrelevant and redundant features in the raw data, supervised feature selection is performed in data preprocessing phase. Features are ranked by their Pearson correlation coefficients R_f with respect to the target PSNR in training data. As a measure of the linear dependence of a feature and the target, $R_f = +1$ indicates a strong linear relationship between them, $R_f = -1$ represents a negative linear relationship and $R_f = 0$ shows the independence of them. For a feature X_f , its coefficient w.r.t. target Y is defined as:

$$R_f = \frac{\text{cov}(X_f, Y)}{\sqrt{\text{var}(X_f)\text{var}(Y)}} \quad (6)$$

Features with R_f close to 0 will be eliminated when learning a prediction model.

3.5 Regression Models

Three regression algorithms are employed in this paper for predicting video quality, k -Nearest Neighbor, Support Vector Machine and Neural Network.

k -Nearest neighbor algorithm (k -NN) is a lazy learning method. The quality of a testing video is estimated based on its k nearest neighbors in training data, which have similar feature values to the testing video. Assuming that its k most similar videos have PSNR value $T_j, j = 1..k$, the quality of the video i is

$$T_i = \frac{\sum_{j=1}^k w_{i,j} \times T_j}{\sum_{j=1}^k w_{i,j}} \quad (7)$$

where $w_{i,j}$ is the weight function inversely proportional to the distance of the training object j . It is defined as: $w_{i,j} = \frac{1}{c+(d_{i,j})^2}$ where c is the kernel width parameter and $d_{i,j}$ is the distance between video i and j . Considering the feature correlation coefficients, we define distance $d_{i,j}$ by $d_{i,j} = \sqrt{\sum_f R_f \times (x_{i,f} - x_{j,f})^2}$ where $x_{i,f}$ and $x_{j,f}$ is the value of feature f in video i and j .

To acquire the best results, the number of nearest neighbors k is set from 1 to 10, and the parameter c is tested by values in $[10^{-4} - 10^{-3}]$ stepped by 5×10^{-5} .

Support Vector Machine (SVM) is considered as one of the most robust and powerful algorithms for classification and regression. In this work, Epsilon-SVR, the regression model of LIBSVM 3.12 [12], is employed for video quality prediction. After testing of various kernel functions, Radial Basis Function (RBF) kernel is selected for its best performance: $\phi(x, x_i) = e^{-\gamma\|x-x_i\|^2}$. The setting of SVM parameters C (*cost*) and kernel parameter γ (*gamma*) are explored by searching in $[1 - 100]$ for C and in $[10 - 400]$ for γ . Experimental results on different settings of C and γ will be demonstrated in Figure 3 of the next section.

Neural Network (NN) has been widely used in predicting video quality based on application and network parameters [5, 13, 14]. Given enough hidden units, any continuous function can be uniformly approximated to arbitrary accuracy [15]. We applied two to three hidden layers NN with back-propagation learning algorithm. Different numbers of hidden neurons are evaluated ranged from $1 \times F$ to $3 \times F$, where F is the number of features. Results of NN with different topology settings will be reported in next section.

4 Experimental Results

This section presents the experimental results on 985 video testings. 5-fold cross validation is applied for reliable evaluation reason. In order to investigate the

impact of feature normalization and transformation introduced in last section, four different cases summarized in Table 3 are evaluated and compared.

Table 3: Experimental Evaluation Cases

		Modulation and boosting	
		Excluded	Transformed
Normalization	Without	Case 1	Case 3
	With	Case 2	Case 4

Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2) are considered as the criteria to evaluate the performance of the prediction methods. RMSE is used to quantify the difference between the predicted and actual PSNR value. A smaller value indicates a better result with higher accuracy. RMSE is calculated as $\sqrt{\sum_{i=1}^N (T_i - T_{i-act})^2 / N}$, where T_i and T_{i-act} are the predicted and actual PSNR respectively, and N is the number of testing videos.

R^2 is another performance measurement of prediction. It is calculated by $R^2 = 1 - \frac{SSE}{SST}$, where the total sum of square $SST = \sum_{i=1}^N (T_{i-act} - \frac{1}{N} \sum_{i=1}^N T_{i-act})^2$ and sum of squares error $SSE = \sum_{i=1}^N (T_{i-act} - T_i)^2$. R^2 is in $[0, 1]$, where the maximum value 1 indicates the best prediction model.

4.1 Feature Ranking

As introduced in section 3.4, each feature is ranked by its Pearson correlation coefficient R_f . The last two columns of Table 1 show the R_f and corresponding P-values of all features. Those features with $-0.03 < R_f < 0.03$ or NaN^1 are eliminated, such as Enhancement Encoding QP (0.028953), Transmission Power (NaN), Reception Power Sensitivity (-0.028953), Carrier Frequency (0.000000), and Cyclic Prefix (NaN). Their large P-values (> 0.05) also confirm that these feature and the target value have no significant correlation.

4.2 Importance of Transforming Modulation and Power Boosting

To investigate the importance of transforming modulation and power boosting features presented in Section 3.2, we compare results of Case 1 and 2 with that of Case 3 and 4.

Figure 2a and 2b show the RMSE and R^2 of k -NN on each case. The best setting of kernel parameter c when calculating $w_{i,j}$ was exhaustively explored. Figure 2 presents the results with the best setting. As we can see for all values of k , Case 3 and 4 are significantly better than Case 1 and 2 with lower RMSE values and higher R^2 . We can then conclude that feature transformation plays a very important role on correctly predicting video quality.

Figure 3 presents the prediction performance of SVM. As observed in k -NN, better results are obtained from Case 3 and 4 rather than Case 1 and 2. The

¹ The 0 or NaN of R_f for some features is caused by their single distinct value, as given in the 7th column (Count) of Table 1

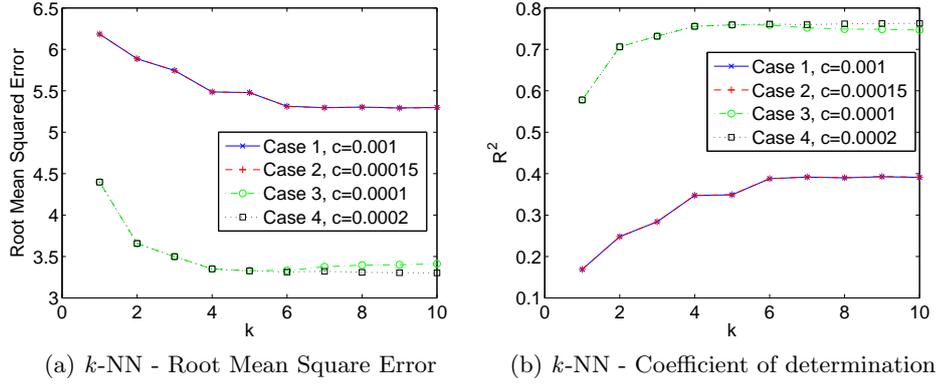


Fig. 2: Prediction performance of k -NN on each case with the best setting of c

setting of SVM parameter C and γ is tested by combinations of values indexed by k on x -axis. The best prediction is on Case 4 when $C = 9$ and $\gamma = 120$ ($k = 332$) resulting RMSE = 3.218 and $R^2 = 0.7669$.

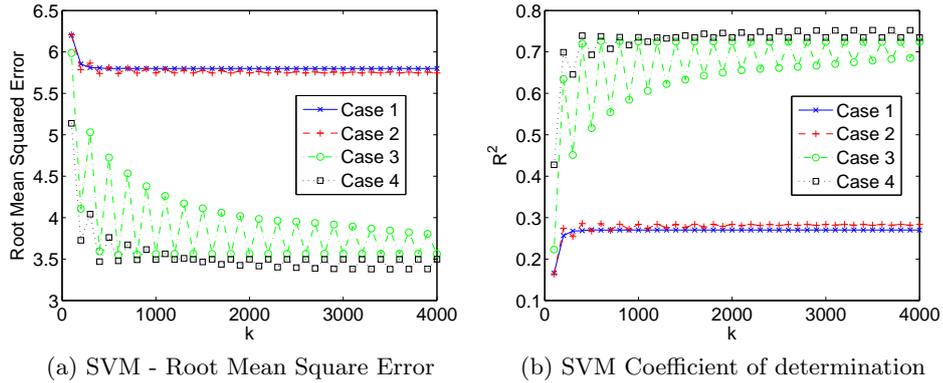


Fig. 3: Prediction performance of SVM on each case (sampled C and γ setting combinations)

Figure 4 shows the results of NN with different network topology setting indexed by k , e.g., $k = 1$ when NN has $F \times F$ neurons (2 hidden layers, each of which has F hidden neurons), $k = 5$ when NN has $F \times F \times F$ neurons (3 hidden layers), where F is the number of features after selection. The same observation can be found that Case 3 and 4 perform better than Case 1 and 2. The best prediction results with lowest RMSE and highest R^2 happened on Case 3 when NN has $3F \times 3F$ neurons ($k = 4$).

Results of k -NN, SVM and NN consistently show that prediction excluding modulation and power boosting features (Case 1 and 2) gives worse results than

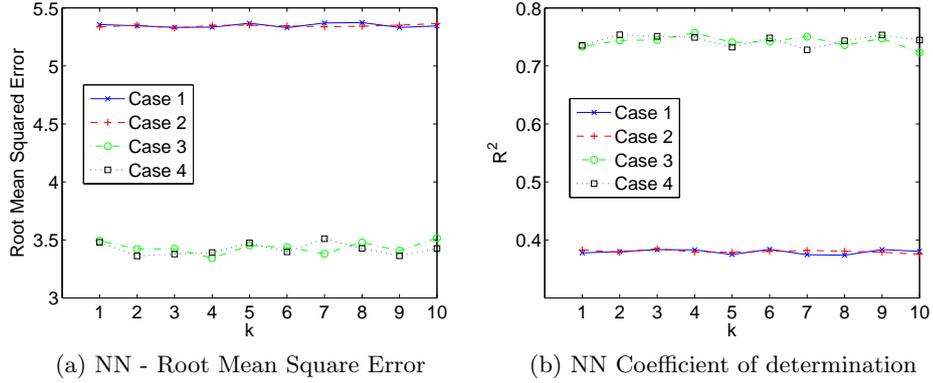


Fig. 4: Prediction performance of NN with different combination of hidden layers and neurons

transforming them into probability of symbol errors (Case 3 and 4). In other words, transformation of modulation and power boosting is essential.

4.3 Effects of Normalization of Data

Since Case 3 and 4 achieve better prediction results, we study the normalization effect on these two cases in this subsection (Case 1 and 2 are exempted). In k -NN, normalization does not make significant effect. In Figure 2, normalized data (Case 4) performs exactly the same as original data (Case 3) when k is less than 4. Normalized data (Case 4) leads to better results than original data (Case 3) when k increases.

Normalization stabilized the performance of SVM when varying parameter setting. Fluctuating curves of Case 3 and smoothing curves of Case 4 in Figure 3 indicate that normalized data can produce better (lower RMSE and higher R^2) and more stable results. However, NN does not significantly benefit from normalization.

In summary, normalization can generally further improve the prediction performance.

4.4 Comparison of Regression Models

Table 4 compares the prediction performance of k -NN, SVM, and NN on Case 3 and Case 4. Case 1 and 2 are excluded because their prediction results are much worse than that of Case 3 and 4. The best RMSE and R^2 value of each model were presented when its model parameters are set appropriately. In general, three models give similar results, and they all produce small RMSE and high R^2 . SVM on Case 4 can lead to the smallest RMSE and the highest R^2 .

Figure 5 shows the predicted and actual PSNR values for the best prediction of all models, k -NN on Case 4 when $k=10, c=0.0002$, SVM on Case 4 when $C=9, \gamma=120$, and NN on Case 3 when topology is $3F \times 3F = 30 \times 30$, where F is the number of features after feature transformation and selection. Videos on x -axis are ordered by their actual PSNR values. We can observe that all prediction

Table 4: The best predictions of k -NN, SVM, and NN on Case 3 and 4

		k -NN	SVM	NN
RMSE	Case 3	3.32864	3.27009	3.34251
	Case 4	3.30163	3.21764	3.36173
R^2	Case 3	0.75914	0.75922	0.75707
	Case 4	0.76283	0.76689	0.75371

methods can give good prediction when actual PSNR is high. Especially, the prediction of SVM is more accurate when actual PSNR is larger than 54dB. The prediction deviates more from the true target when actual PSNR is relatively low. However, the deviation is only around 5% of the target on average. These less accurate predictions are due to the lack of effective learning examples. Videos with low PSNR have worse quality. In a well-developed transmission system, the number of damaged videos is smaller than the number of undamaged videos.

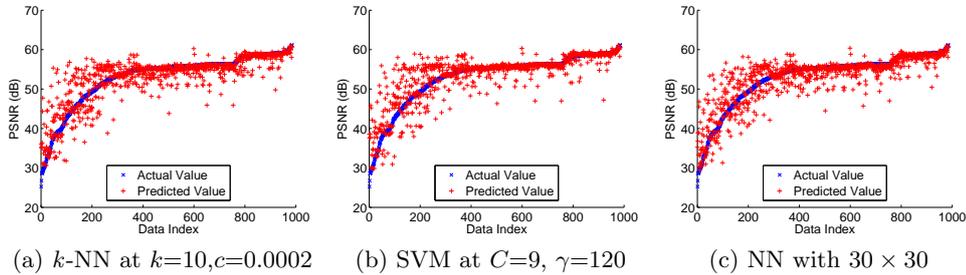


Fig. 5: Comparison of predicted PSNR and actual PSNR

5 Conclusion

This paper studies the problem of video quality prediction in wireless 4G network transmission. A real testbed is built for collecting videos transmitted under various network conditions. The raw transmission data is pre-processed by feature transformation, normalization and ranking. Regression models are learned by three different algorithms and used for predicting the quality of video.

The first motivation of this work is to investigate which features are strongly correlated to video quality. From the feature ranking results, we find that attenuation and base layer probability of symbol error are the two important features that highly correlated to the PSNR (target of prediction) with the correlation coefficient $R_f = -0.5027$ and -0.1905 , P-value = $3.372e-64$ and $1.667e-9$, respectively. In addition, according to the evaluation results, we observe that feature transformation of modulation and power boosting for both base layer and enhancement layer significantly affected the prediction accuracy of video quality.

Through pre-processing raw data and exploring the most suitable parameter setting, we built prediction models based on three different regression algorithms. The experimental results demonstrate that these models can accurately predict video quality with small errors and high relevance with target values.

This first attempt of video quality prediction in wireless 4G opens several perspectives for further research. First, the models are useful to reduce the time spent on running real experiments. As we mentioned in the introduction, imitating all of the different channel conditions and conducting video transmitting experiments are very time consuming. Our discovery of key features and prediction model can be used to guide the design of experiments. Second, we will enhance the prediction models to make them perform well on more complex channel conditions.

References

1. She, J., Yu, X., Hou, F., Ho, P.H., Yang, E.H.: A Framework of Cross-Layer Superposition Coded Multicast for Robust IPTV Services over WiMAX. In: IEEE Wireless Communications and Networking Conference WCNC, Ieee (2008) 3139–3144
2. Judd, G., Steenkiste, P.: Characterizing 802.11 wireless link behavior. *Wireless Networks* **16**(1) (June 2008) 167–182
3. Knopp, R., Humblet, P.: Information capacity and power control in single-cell multiuser communications. *IEEE International Conference on Communications ICC '95 Seattle, 'Gateway to Globalization'* **1** (1995) 331 – 335
4. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. *Electronics letters* **44** (2008) 9–10
5. Khan, A., Sun, L., Ifeachor, E.: Content-based video quality prediction for MPEG4 video streaming over wireless networks. *Journal of Multimedia* **4**(4) (2009) 228–239
6. Garcia, M.N., Raake, A.: Towards Content-related Features for Parametric Video Quality Prediction of IPTV Services. In: IEEE International Conference on Acoustics Speech and Signal Processing ICASSP 2008. (2008) 757–760
7. Dalal, A., Olson, J.: Feature Selection for Prediction of User-Perceived Streaming Media Quality. In: *Proceedings of International Symposium on Performance Evaluation of Computer and Telecommunication Systems SPECTS*. (2007) 285–294
8. Liu, Y.x., Kurceren, R., Budhia, U.: Video classification for video quality prediction. *Journal of Zhejiang University SCIENCE A* **7**(5) (May 2006) 919–926
9. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology* **17**(9) (September 2007) 1103–1120
10. Roosendaal, T.: Sintel. In: *ACM SIGGRAPH 2011 Computer Animation Festival. SIGGRAPH '11, New York, NY, USA, ACM* (2011) 71
11. Ho, J.C.C.: Logical superposition coded modulation for wireless video multicasting. Master's thesis, Univ. of Waterloo, Waterloo, Canada (2009)
12. Chang, C., Lin, C.: LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3) (2011) 27
13. Mohamed, S., Rubino, G.: A study of real-time packet video quality using random neural networks. *IEEE Transactions on Circuits and Systems for Video Technology* **12**(12) (2002) 1071–1083
14. Callet, P.L.: A convolutional neural network approach for objective video quality assessment. *IEEE Transactions on Neural Networks* **17**(5) (September 2006) 1316–27
15. Bishop, C.M.: *Pattern recognition and machine learning*. Springer-Verlag New York (2006)