2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)

Data-Driven QoE Analysis on Video Streaming in Mobile Networks

Qingyong Wang*, Hong-Ning Dai*, Hao Wang†, and Di Wu‡§
*Macau University of Science and Technology, Macau SAR, China
†Norwegian University of Science and Technology, Norway
‡School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China
§Guangdong Province Key Laboratory of Big Data Analysis and Processing, Guangzhou 510006, China
*wang@cnu.ac.cn, hndai@ieee.org; †hawa@ntnu.no; ‡wudi27@mail.sysu.edu.cn

***Abstract*—There is a substantial body of literature on analyzing Quality of Experience (QoE) of Video Streaming while there are few studies on standardizing QoE assessments. One of recent proposals on standardizing QoE of video streaming is video Mean Opinion Score (vMOS), which can model QoE of video streaming in 5 discrete grades. However, there are few studies on quantifying vMOS and investigating the relation between vMOS and other Quality of Service (QoS) parameters. In this paper, we address this concern by proposing a data-driven QoE analysis framework on video streaming QoE data. Moreover, we conduct extensive experiments on realistic dataset and verify the effectiveness of our proposed model. Our results show that vMOS is essentially affected by many QoS parameters such as initial buffering latency, stalling ratio and stalling times. Interestingly, we have found that a small set of QoS parameters play an important role in determining vMOS.**

***Index Terms*—Quality of Service (QoS), Quality of Experience (QoE), Video Streaming, Mobile Networks**

I. Introduction

It is predicted in [1] that the traffic caused by video streaming will occupy more than 77% of all consumer Internet traffic by 2021, among which mobile video traffic will be more than 55% of all video traffic. The growing demands on video streaming over mobile networks inevitably lead to the challenges in optimizing network resource in order to improve the user *perceptual* experience. Many previous works mainly focused on improving quality of service (QoS) of video streaming over mobile networks. Typical QoS measures include throughput, bandwidth, outage, jitter and delay [2]. However, most of these QoS metrics failed to characterize user perceptual experience, which is also called quality of experience (QoE). In fact, it is more crucial to conduct video quality assessment from QoE than that from QoS [3], [4]. This is because (i) enhancing QoS does not directly improve QoE [5]; (ii) only improving QoS sometimes significantly increases operating expenditure, consequently decreasing the profit of service providers [6].

Therefore, QoE improvement of video streaming over mobile networks has received extensive attention recently. In particular, the work of [7] investigated QoE-driven cross-layer optimization for video transmission in wireless networks. Ramamurthi et al. [8] proposed a resource management scheme at network core in wireless networks to improve video QoE. The work of [9] presented a large-scale measurement-based study on the effects of Internet path selection in video QoE and offers several QoE enhancement schemes.

However, the prerequisite of QoE improvement of video streaming is to quantify QoE appropriately. Video QoE assessment schemes can be generally categorized into *subjective tests*, *objective assessments* and *data-driven analysis* [3]. Compared with subjective tests and objective assessments, data-driven analysis is more promising due to the availability of massive data sets and the accuracy of characterizing user perception while overcoming the drawbacks of subjective tests and objective assessments (such as high cost and insufficient human visual system knowledge). In particular, the work of [6] proposed a data-driven model to quantify the metrics affecting video QoE. Jiang et al. [10] improved video QoE by exploiting data-driven QoE prediction. The work of [11] improved the video bitrate adaptation based on data-driven QoE prediction.

In addition to the above efforts, there are also other solutions on standardizing QoE. One of recent video QoE measurement standards is U-vMOS (User/Unified/Ubiquitous video Mean Opinion Score)[1], which was proposed by Huawei in 2016 [12]. The score of vMOS is essentially established according to Mean Opinion Score (MOS) standardized by International Telecommunication Union (ITU) [13], in which discrete grades from 1 to 5 represent bad, poor, fair, good and excellent, respectively. It is shown in [12] that vMOS at video playback startup is mainly determined by three key factors: *video quality*, *initial buffering delay* and *video freezing duration*, each of which is also affected by multiple QoS variables. Recently, Pan et al. [14] investigated machine learning based bitrate estimation on YouTube video streaming based on Huawei's vMOS assessment model. However, they just gave a mathematical expression of vMOS based on their subjective estimations. *To the best of our knowledge, there is no data-driven QoE analysis on vMOS.*

Therefore, this paper aims to conduct data-driven QoE analysis on vMOS. In particular, we obtain a realistic dataset on video QoE based on SpeedVideo Global Operating Platform (SVGOP) established by Huawei. This dataset contains 88,526 samples and 15 features. This dataset has the following

[1]For simplicity, we use vMOS to represent U-vMOS throughout this paper.

0-7695-6329-5/17/31.00 ©2017 IEEE
DOI 10.1109/ISPA/IUCC.2017.00168

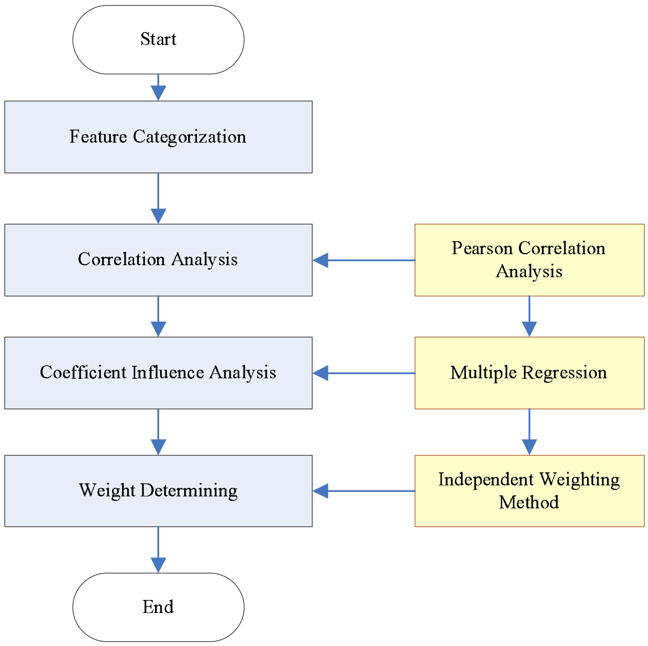

Fig. 1. Our method used in this paper

unique characteristics: 1) heterogeneous data types, 2) positive/negative correlations and 3) dependence of features; these characteristics result in the difficulties in analyzing video QoE data.

To address the above concerns, we propose a QoE data-driven analysis framework. In particular, our analysis framework consists of four components: feature categorization, correlation analysis, coefficient influence analysis and weight determining. We then conduct extensive experiments on the dataset by using our proposed framework and obtain the exact expression of vMOS, which heavily depends on multiple QoS parameters. Our results also offer many useful insights in improving video QoE.

The remainder of this paper is organized as follows. We present the overview of our method in Section II. Section III presents the experimental results. Finally, we conclude this paper in Section IV.

II. QoE Analysis Framework

Fig. 1 shows the flow chart of our proposed method. In particular, we first categorize various QoS parameters into three types. We then apply Pearson correlation analysis to determine the positive and negative correlations of these parameters. We next employ multiple regression to determine the influence of coefficients and obtain the weights of these coefficients by using independent weighting method.

The categorization of the QoS parameters will be given in Section III since it is highly related to the dataset. In this section, we mainly describe correlation analysis in Section II-A, coefficient influence analysis in Section II-B and weight analysis in Section II-C.

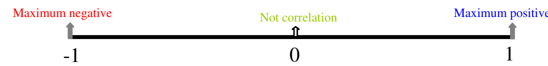

Fig. 2. Correlation standard

A. Correlation Analysis

Correlation analysis is used to analyze the linear correlation between two variables. In this paper, we use Pearson correlation coefficient as the measure of the correlation between two variables. In particular, we define a Pearson correlation coefficient $(r_{xy})_i$ between two variables x_i and y_j as follows,

$$(r_{xy})_{ij} = \frac{\sum_{t=1}^{T}(x_{ti} - \bar{x}_i) \cdot (y_{tj} - \bar{y}_j)}{\sqrt{\sum_{t=1}^{T}(x_{ti} - \bar{x}_i)^2} \cdot \sqrt{\sum_{t=1}^{T}(y_{tj} - \bar{y}_j)^2}}, \quad (1)$$

where T is the sample size, $x = \{x_{t1}, x_{t2}, ..., x_{tm}\}$ and $y = \{y_{t1}, y_{t2}, ..., y_{tn}\}$ represent two feature sets with the same sample size, $\bar{x}_i = \sum_{t=1}^{T} x_{ti}/T$ and $\bar{y}_j = \sum_{t=1}^{T} y_{tj}/T$.

The value of Pearson correlation coefficient $(r_{xy})_{ij}$ is in a range of $[-1, 1]$ as shown in Fig. 2. In particular, a value of 1 means the maximum positive (i.e., the total positive linear correlation between two variables), -1 means the maximum negative (i.e., total negative linear correlation) and 0 means no linear correlation. Note that when $(r_{xy})_{ij}$ is quite close to 0, it may probably lead to the confused positive and negative relevance as shown in [15]. In this situation, we need some interventions on choosing indicators.

After determining the positive or negative correlation coefficients of features, we need to normalize the features since they are in different units. In particular, we make a conversion from the absolute value to the relative value. In particular, we choose the MAX-MIN scaling method to normalize the positive and negative values. More specifically, we have

- Positive values:

$$u_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}, \quad (2)$$

- Negative values:

$$u_{ij} = \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})}, \quad (3)$$

where x_{ij} represents the original value, u_{ij} represents the value after normalization, $\min(\cdot)$ is the minimum value and $\max(\cdot)$ is the maximum value.

B. Coefficient Influence Analysis

We then conduct multiple regression to analyze the relationship between two or more independent variables and one dependent variable. In particular, we conduct the multiple regression analysis on the QoS features and the scoring factors (i.e., SQuality, SLoading and SStalling). This analysis is implied by that fact that vMOS is a function of the above 3 key factors, i.e., $\text{vMOS} = f(\text{SQuality}, \text{SLoading}, \text{SStalling})$

TABLE I
CATEGORIZATION OF FEATURES

Types	Features	Variables
Video Quality	Average rate of playing phase (kbps)	x_1
	Video total download (DL) rate (kbps)	x_2
	Video bitrate (kbps)	x_3
Initial Loading	Initial max DL rate (kbps)	x_4
	End-to-End (E2E) round-trip time (RTT) (ms)	x_5
	Initial buffering latency (ms)	x_6
	Video Initial buffer download (byte)	x_7
Stalling	Playing time(ms)	x_8
	Playing total duration	x_9
	Stalling times	x_{10}
	Stalling ratio	x_{11}

TABLE II
CORRELATION COEFFICIENTS

Types	Features	Values	Pos/Neg	Chosen
Video Quality	Average Rate of playing phase	-0.0104	-	**+**
	Video total DL rate	-0.028	-	**+**
	Video bitrate	0.8709	+	+
Initial Loading	Initial Max DL rate	0.6164	+	+
	E2E RTT	-0.5550	-	-
	Initial buffering latency	-0.8866	-	-
	Video initial buffer download	0.2138	+	+
Stalling	Playing time	0.9066	+	+
	Playing total duration	0.0371	+	+
	Stalling times	-0.9334	-	-
	Stalling ratio	-0.3371	-	-

[12]. We then give a general expression of the multiple regression equation as follows,

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m, \tag{4}$$

which y_i refers to one of the scoring factors such asvMOS, $\mathrm{SQuality}$, $\mathrm{SLoading}$ and $\mathrm{SStalling}$, x_i represents each of QoS features, β_0 is the offset, β_i is the corresponding regression coefficient.

In order to assess the accuracy (or fitting) of regression, we use the root mean square error (R^2), which is defined as follows,

$$R^2 = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y_j})^2}, \tag{5}$$

where y_j is the dependent variable and $\hat{y_i}$ is the predicated value. In practice, the regression fits successfully only when the level of significance is less than 0.005 [15].

C. Weight Determining

After obtaining coefficients of QoS parameters, we need to evaluate the importance of each QoS parameter. In particular, we use independent weighting method to determine the weights of coefficients [16].

We denote the weight of a coefficient β_i by w_i, which can be calculated by the following equation,

$$w_i = \frac{\zeta_i}{\sum_{i=1}^{m}\zeta_i}, \tag{6}$$

where ζ_i can be calculated by

$$\zeta_i = \sqrt{\frac{\beta_i}{\sum_{i=1}^{m}\beta_i}}. \tag{7}$$

III. EXPERIMENT RESULTS

In this section, we conduct the analysis on the sample dataset SpeedVideo Global Operating Platform (SVGOP) established by Huawei[2]; SVGOP is a specific application of vMOS in mobile networks throughout the world. In particular, the dataset contains 88,526 samples with 11 features (i.e., QoS parameters) and 4 scoring factors. We first categorize the features in Section III-A, and then conduct correlation analysis in Section III-B. Next, Section III-C presents the multiple regression results. Finally, we determine the weights based on independent weighting method in Section III-D.

A. Categorization of features

It is shown in [12] that vMOS is a function of $\mathrm{SQuality}$, $\mathrm{SLoading}$ and $\mathrm{SStalling}$; this implies that we should categorize the features into three types: 1) Video Quality related features, 2) Initial Loading related features, 3) Stalling related features. Therefore, we categorize 11 features into 3 types as shown in Table I. We denote each of these 11 features by variable x_i (i = 1 to 11). In addition, we also denote the scoring factors vMOS, $\mathrm{SQuality}$, $\mathrm{SLoading}$ and $\mathrm{SStalling}$ by y_1, y_2, y_3 and y_4, respectively.

B. Correlation Analysis

We then conduct Pearson correlation analysis on these features. In particular, we calculate the correlation coefficients according to Eq. (1) and obtain the values as listed in Table II.

Let us first analyze the first type of features. As shown in Table II, we can see that the coefficient of Video bitrate is 0.8709, which is relatively close to 1 (i.e., the maximum positive), implying that Video bitrate is quite correlated with $\mathrm{SQuality}$ (i.e., y_2). However, Table II also show that Average rate of playing phase and Video total DL rate have the values of -0.0104 and -0.028, respectively. These two values are quite close to 0. In this case, we need to take some interventions on the indicators of the coefficients. According to the guidance given in [15], we choose positive indicators for both Average rate of playing phase and Video total DL rate (see the plus symbols in bold fonts in Table II).

Regarding to the second type (i.e., Initial Loading), it is shown in Table II that both Initial Max DL rate and Video initial buffer download are positively correlated to $\mathrm{SLoading}$ (i.e., y_3) while both E2E RTT and Initial buffering latency are negatively correlated to $\mathrm{SLoading}$. Among them, the coefficient of Initial buffering latency is -0.8866 (i.e., the maximum absolute value), implying that it strongly affects QoE in terms of Initial Loading.

We next analyze the third type of features. Table II shows that Stalling times is strongly negatively correlated with $\mathrm{SStalling}$ (i.e., -0.9334), implying that it has a strong influence

[2]http://speedvideo.huawei.com/

TABLE III
CORRELATION COEFFICIENT WITH vMOS

	SQuality	SLoading	SStalling
vMOS	0.0328	0.8131	0.9320

on QoE in terms of stalling. Besides, we can see from Table II that Playing total duration has a less influence on QoE (i.e., 0.0371). This result implies that users may be more concerned with Stalling times than Playing total duration.

Finally, we conduct correlation analysis on vMOS and the three scoring factors SQuality, SLoading and SStalling. Table III shows the results. It is shown in Table III that all the three factors are positively correlated to vMOS. In contrast to SQuality and SLoading, SStalling is more dominant in users QoE since it has the maximum positive value 0.9320 (i.e., quite close to 1).

C. Multiple Regression Analysis

We next conduct multiple regression analysis. In particular, we give the regression equation on 11 features as follows,

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11}, \quad (8)$$

where x_i ($i = 1$ to 11) corresponds to each of 11 features as given in Table I and β_i is the regression coefficient.

Table IV lists the regression results, where "**Coeff**" stands for coefficient. Note that we assess the accuracy of the multiple regression model by the coefficient of determination (R^2). In practice, we require $R^2 > 0.8$ so that the regression fits successfully. Our multiple regression analysis is also conducted according to the three types of features (see our categorizations in Section III-A).

It is shown in Table IV that the first group of multiple regression results has the accuracy with $R^2 = 0.871$, implying that the regression model fits well. Besides, Table IV also shows that the coefficient of determination for the second group of regression results is $R^2 = 0.895$ (i.e., the regression model fits). Among 4 features in this group, the coefficient of Initial buffering latency is -2.263 (i.e., the maximum absolute value); it indicates that Initial buffering latency plays an important role in QoE. Similarly, the regression in the third group also fits successfully with the coefficient of determination $R^2 = 0.906$. One thing to note that the coefficient of Play time is 0. This is because the significant level is 0.014 (i.e., > 0.005) during the regression. Thus, we ignore it in the regression according to [15]. Moreover, the coefficient of Stalling times is -1.234, implying Stalling times plays an important role in QoE; it confirms with our aforementioned observation in Section III-B.

Finally, we conduct the regression on vMOS and SQuality, SLoading and SStalling together and obtain the coefficients as shown in Table V.

As shown in Table V, the value $R^2 = 0.977$ implies that the regression fits well. Besides, Stalling has the largest coefficient (i.e., 0.710) compared with Video Quality and Initial Loading; this indicates that Stalling plays an important role in vMOS.

TABLE IV
REGRESSION COEFFICIENTS AND MODEL PRECISION

Type	**Feature**	**Coeff**	**Value**	R^2
Video Quality	Average Rate of playing phase	β_1	0.943	0.871
	Video total DL rate	β_2	1.113	
	Video bitrate	β_3	1.003	
Initial Loading	Initial Max DL rate	β_4	0.294	0.895
	E2E RTT	β_5	-0.217	
	Initial buffering latency	β_6	-2.263	
	Video initial buffer download	β_7	0.214	
Stalling	Playing time	β_8	0.001	0.906
	Playing total duration	β_9	0.006	
	Stalling times	β_{10}	-1.234	
	Stalling ratio	β_{11}	0.906	

TABLE V
COEFFICIENTS AND MODEL PRECISION

	Video Quality	Initial Loading	Stalling	R^2
Value	0.002	0.471	0.710	0.977

Integrating the two stages of multiple regressions together, we finally obtain the regression equation as follows,

$$y_1 = 0.8127 + 0.0019 \cdot x_1 + 0.0022 \cdot x_2 + 0.0020 \cdot x_3 + 0.1385 \cdot x_4 - 0.1022 \cdot x_5 - 1.0659 \cdot x_6 + 0.1215 \cdot x_7 + 0.0043 \cdot x_9 - 0.8761 \cdot x_{10} + 0.6433 \cdot x_{11}. \quad (9)$$

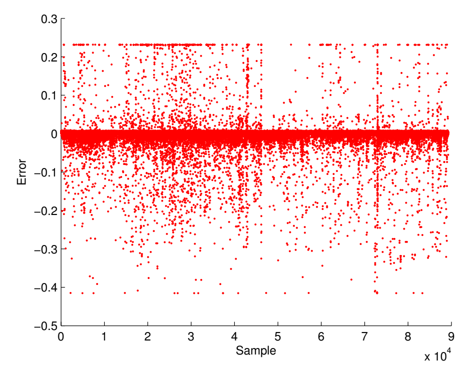

Fig. 3. Residual Error Plot

We use the residual plots to evaluate the accuracy of the regression model. Fig. 3 shows the residual plots. We can see from Fig. 3 that the fitting rate is 98.18% under the confidence interval of 99%, implying that the regression fits.

Furthermore, we use t-statistics to assess the accuracy of our regression. Table VI presents the t-statistic and standard error values of regression coefficients, where the t-statistic indicates the statistical significance of the relationship between dependent and independent parameters. We can see from Table VI that our regression fits well.

D. Weight Determining

After obtaining the coefficients of the features by using multiple regression method, we then evaluate the importance

TABLE VI
t-STATISTIC AND STANDARD ERROR OF REGRESSION COEFFICIENTS

Feature	t-statistics	Standard Error
Average rate of playing phase	0.270	0.029
Video total DL rate	13.547	0.030
Video bitrate	-5.308	0.001
Initial Max DL rate	-494	0.002
E2E RTT	6.664	0.001
Initial buffering latency	7.249	0.007
Video initial buffer download	16.256	0.013
Playing total duration	0.521	0.002
Stalling times	-243.711	0.005
Stalling ratio	11.922	0.015

of the features. In particular, we use the independent weighting method [16] to calculate the weight of each feature. Table VII lists the weights as well as the rankings of these features.

TABLE VII
WEIGHTS AND RANKING

Features	**Initial Weights**	**Ranking**
Average rate of playing phase	0.0108	10
Video total DL rate	0.0118	8
Video bitrate	0.0112	9
Initial Max DL rate	0.0928	4
E2E RTT	0.0797	6
Initial buffering latency	0.2574	**1**
Video initial buffer download	0.0869	5
Playing total duration	0.0163	7
Stalling times	0.2333	**2**
Stalling ratio	0.2000	**3**

Table VII presents the rankings of all the features. We can see from Table VII that Initial buffering latency, stalling times and stalling ratio rank higher than other features, implying that they play a dominant role in vMOS. In particular, the weights of the three factors occupy 69.01% while other 6 factors occupy another 30.99%.

E. Result analysis

Table VII shows that vMOS is mainly affected by Initial buffering latency, Stalling times and Stalling ratio. We next investigate the influence of them on vMOS based on statistic analysis of the original dataset.

Fig. 4 plots vMOS versus the average of Initial buffering latency. In particular, we first categorize the samples into 5 groups according to different scales of vMOS (i.e., vMOS $= 1, 2, 3, 4, 5$) and then calculate the average value of each group, which is represented by a histogram as shown in Fig. 4. Besides, the bar diagram below shows the proportion of each group to the number of all the samples (in percentage). For example, we can see that the total number of samples with vMOS > 3 counts for nearly 96% over the number of all the samples, implying that most of samples are "good" to users.

Each dot in the red curve in Fig. 4 represents the average value of Initial buffering latency in each group. It is shown in Fig. 4 that there is a significant increment of Initial buffering latency when vMOS is decreased; this implies that vMOS decreases significantly when Initial buffering latency is increased. For example, when Initial buffering latency is

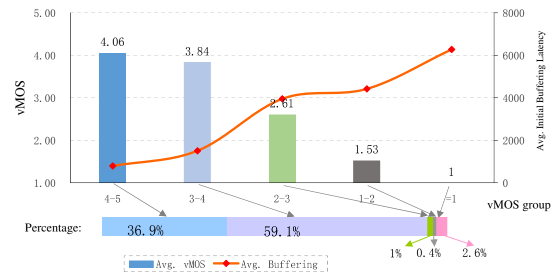

Fig. 4. vMOS versus Initial buffering latency

increased to greater than 3954 ms, vMOS decreases from 3 to 2 (i.e., from "fair" to "poor"); the user QoE decreases significantly.

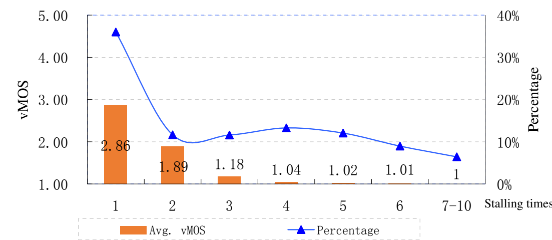

Fig. 5. vMOS versus Stalling times

We next investigate the influence of Stalling times on vMOS. Fig. 5 plots the results of vMOS versus Stalling times. In particular, we categorize the samples into 7 groups according to different values of Stalling times (i.e., 1, 2, 3, 4, 5, 6, 7-10). We then calculate the average value of vMOS of each group, which is shown as the histogram. Besides, Fig. 5 also plots the proportion of each group to the number of all the samples (in percentage), where a blue dot represents the percentage value of each group. We can see from Fig. 5 that the average value of vMOS in the group with Stalling times $= 1$ is 2.86, which is the highest value among all the groups. Moreover, the average value of vMOS drops significantly with the increased Stalling times. This implies that users cannot tolerate too many times of stalling. For example, a small proportion of samples falls into the group with Stalling times 7-10, implying that users felt extremely annoying with the number number of Stalling times. In fact, we have observed from Fig. 5 that there is a significant drop in terms of the proportion when Stalling times = 2 though there is no significant increment or decrease when Stalling times > 2. This may imply that there may exist a threshold on Stalling times. The investigation on the threshold on Stalling times will be left as one of our future works.

Fig. 6 plots the results on vMOS against Stalling ratio, in which we categorize samples into 5 groups according to different scales of vMOS (i.e., vMOS $= 1, 2, 3, 4, 5$). It is worth mentioning that we only consider the data samples with Stalling ratio ≥ 1 and ignore those with Stalling ratio $= 0$ (i.e.,

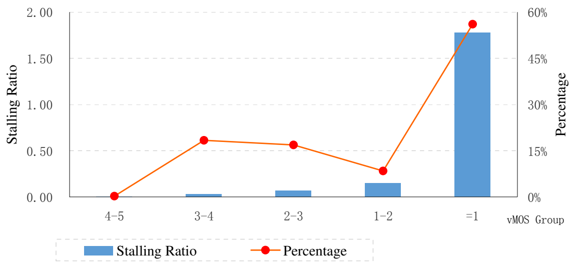

Fig. 6. vMOS versus Stalling ratio

no stalling); this analysis is different from the categorization in Fig. 4 that considers all the samples. The histogram in Fig. 6 represents the average Stalling ratio in different groups and each dot in the red curve represents the percentage of each group. We can see from Fig. 6 that there is a significant increment on Stalling ratio when vMOS is decreased; this implies that Stalling ratio is negatively correlated with vMOS.

Remark 1: In summary, we can see from Fig. 4, Fig. 5 and Fig. 6 that Initial buffering latency, Stalling times and Stalling ratio can significantly affect vMOS. In particular, Initial buffering latency, Stalling times and Stalling ratio are negatively correlated with vMOS.

This result essentially offers us some useful insights in improving video QoE. For example, we may focus on optimizing the network resource to reduce Initial buffering latency, Stalling times and Stalling ratio so that we can significantly improve the video QoE while maintaining relatively low operating expenditure. However, it is not an easy task to achieve this goal because the enhancement of these QoS parameters is also involved with many other technologies, such as cross-layer optimization and distributed resource allocation [10], [17], [18]. For example, we can distribute the most popular videos at the server close to users so that we can significantly reduce the initial buffering latency. However, to determine the popularity of video streaming is challenging since it requires the extensive efforts in analyzing the massive video data [19]. Moreover, it is also difficult to identify the QoS bottlenecks since they are often affected by many factors. For example, video stalling is essentially caused by many factors, such as network congestion, network failure, device mobility and radio spectrum scarcity. There is a challenge in identify the causality of stalling. In the future, we may apply data-driven approach to identify the reason behind video stalling according to different scenarios.

IV. Conclusion

In this paper, we conduct a data-driven analysis on video Mean Opinion Score (vMOS), which is an important measure of user quality of experience of video streaming. In particular, our study is based on a realistic dataset consisting of 88,526 samples and 15 features. This dataset has the characteristics such as heterogeneity of data types, positive and negative correlation of features and dependence of features; they result in the difficulties in analyzing the data. In order to address these challenges, we propose a data-driven analysis framework, which can effectively analyze vMOS and investigate the relation between vMOS and other QoS parameters. In particular, we have the following major findings:

- *vMOS is affected by multiple QoS parameters together*. In particular, we have found that vMOS is essentially affected by 11 QoS parameters.
- *Small set of QoS parameters dominates*. Interestingly, we have found that a small set of QoS parameters has the stronger influence on vMOS than other QoS parameters. For example, the weights of Initial buffering latency, Stalling times and Stalling ratio occupy 69.01% while other 6 parameters only occupy another 30.99%.

Our results have paved the ground for the improvement of video streaming QoE in the future. For example, we may integrate cross-layer optimization and distributed resource allocation schemes together to mitigate the key QoS bottlenecks so as to improve video streaming QoE.

Acknowledgement

The work described in this paper was partially supported by Macao Science and Technology Development Fund under Grant No. 096/2013/A3, the National Key R&D Program of China under Grant 2016YFB0201900, the National Natural Science Foundation of China under Grant 61572538 and Grant 61672170, the Fundamental Research Funds for the Central Universities under Grant 17LGJC23, and the Science and Technology Planning Project of Guangdong Province (No. 2017A050501035). The authors would like to express their appreciation for Gordon K.-T. Hon for his thoughtful discussions.

References

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2016–2021," June 2017.

[2] M. Venkataraman and M. Chatterjee, "Inferring video QoE in real time," *IEEE Network*, vol. 25, no. 1, pp. 4–13, January 2011.

[3] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[4] H. Nam, K.-H. Kim, and H. Schulzrinne, "QoE matters more than QoS: Why people stop watching cat videos," in *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 2016, pp. 1–9.

[5] V. Joseph and G. de Veciana, "NOVA: QoE-driven optimization of DASH-based video delivery in networks," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 82–90.

[6] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a Predictive Model of Quality of Experience for Internet Video," in *Proceedings of the ACM SIGCOMM Conference*, ser. SIGCOMM, 2013.

[7] S. Thakolsri, W. Kellerer, and E. Steinbach, "Qoe-based cross-layer optimization of wireless video with unperceivable temporal video quality fluctuation," in *Proceedings of IEEE International Conference on Communications (ICC)*, 2011.

[8] V. Ramamurthi, O. Oyman, and J. Foerster, "Video-QoE aware resource management at network core," in *Global Communications Conference (GLOBECOM), 2014 IEEE*. IEEE, 2014, pp. 1418–1423.

[9] M. Venkataraman and M. Chatterjee, "Effects of internet path selection on video-QoE: analysis and improvements," *IEEE/ACM Transactions on Networking (TON)*, vol. 22, no. 3, pp. 689–702, 2014.

[10] J. Jiang, V. Sekar, H. Milner, D. Shepherd, I. Stoica, and H. Zhang, "CFA: A Practical Prediction System for Video QoE Optimization," in *NSDI*, 2016, pp. 137–150.

[11] Y. Sun, X. Yin, J. Jiang, V. Sekar, F. Lin, N. Wang, T. Liu, and B. Sinopoli, "CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction," in *Proceedings of ACM SIGCOMM*. ACM, 2016, pp. 272–285.

[12] "Huawei: Technical White Paper on Requirements of Mobile Bearer Network for Mobile Video Service," http://www-file.huawei.com/~/media/CORPORATE/PDF/white%20paper/Technical-White-Paper-on-Mobile-Bearer-Network-Requirements-for-Mobile-Video-Services.pdf, March 2016.

[13] I. T. U. (ITU), "Subjective video quality assessment methods for multimedia applications," p. 910, 2008.

[14] W. Pan, G. Cheng, H. Wu, and Y. Tang, "Towards QoE assessment of encrypted YouTube adaptive video streaming in mobile networks," in *IEEE/ACM 24th International Symposium on Quality of Service (IWQoS)*, 2016.

[15] R. R. Wilcox, *Introduction to robust estimation and hypothesis testing*. Academic press, 2011.

[16] C. R. Genovese, K. Roeder, and L. Wasserman, "False discovery control with p-value weighting," *Biometrika*, vol. 93, no. 3, pp. 509–524, 2006.

[17] Y. Wang and X. Lin, "User-provided networking for QoE provisioning in mobile networks," *IEEE Wireless Communications*, vol. 22, no. 4, pp. 26–33, 2015.

[18] Z. Su, Q. Xu, and Q. Qi, "Big data in mobile social networks: a qoe-oriented framework," *IEEE Network*, vol. 30, no. 1, pp. 52–57, January 2016.

[19] Z. Su, Q. Xu, F. Hou, Q. Yang, and Q. Qi, "Edge caching for layered video contents in mobile social networks," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–1, 2017.