Quality of Experience Prediction of HTTP Video Streaming in Mobile Network with Random Forest

Yue Yu(✉), Yu Liu, and Yumei Wang

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
{yuy, liuy, ymwang}@bupt.edu.cn

Abstract. As video is witnessing a rapid growth in mobile networks, it is crucial for network service operators to understand if and how Quality of Service (QoS) metrics affect user engagement and how to optimize users' Quality of Experience (QoE). Our aim in this paper is to infer the QoE from the observable QoS metrics using machine learning techniques. For this purpose, Random Forest is applied to predict three objective QoE metrics, i.e., rebuffering frequency, mean bitrate and bitrate switch frequency, with the initial information of each video session. In our simulation, QoE of four different video streamings are analyzed with eight different system loads. Results show that sufficient prediction accuracy can be achieved for all QoE metrics with the attributes we adopted, especially with low and middle system loads. In terms of type of streamings, the prediction of all metrics for static users performs better than mobile users. Feature selection is also implemented under the highest load to examine the effect of different attributes on each QoE metric and the correlation among attributes.

Keywords: HTTP video streaming · Quality of experience · Random forest Mobile networks

1 Introduction

Video streaming is becoming more and more important in recent years. According to Cisco's forecast [1], video traffic will account for 78% of Internet traffic by 2021. HTTP video streaming is widely used in delivering on-demand multimedia content, with retransmission applied to guarantee data correctness. At the server side, single or several encoded versions are stored, where video files are divided into several chunks (segments). After being downloaded, the chunk is stored in the player's buffer for playback. Before the buffer becomes empty, users can proceed on video playing; otherwise, the video will suffer a rebuffering event.

The QoE concept has emerged mainly with the basic motivation that QoS is not powerful enough to fully express everything nowadays involved in a communication service, which is a multi-dimensional construct and consists of subjective and objective parameters [2]. When it comes to the QoE of HTTP streaming users, according to [3], it highly depends on two crucial factors: (1) the visual quality and its variation and (2) the

X. Liu et al. (Eds.): ChinaCom 2018, LNICST 262, pp. 82–91, 2019.
https://doi.org/10.1007/978-3-030-06161-6_8

frequency and duration of rebuffering events. Different from the Peak Signal to Noise Ratio (PSNR), rebuffering events cannot be directly measured but only predicted from classic QoS metrics [4]. This allows to infer QoE metrics by still relying on QoS monitoring systems. Nevertheless, it is highly complex to map between QoS and QoE metrics, as they often lay in high dimensional spaces and are subject to noise. As a consequence, it is not practical to get a closed form modeling and its experimental validation. Therefore, machine learning techniques are applied to derive the complex relationships between QoS and QoE metrics. In the context of mobile networks, it is challenging for operators to correlate the cell-related parameters like channel state information (CSI) and existing users number to QoE metrics of video consumers, due to the system complexity and difficulty in obtaining the cross-layer information. To overcome this difficulty, we have established a cross-layer simulation program that simulates the behaviors of HTTP video streamings in mobile networks as well as buffer information in user side. Thus we can access all cross-layer information for correlating the QoS parameters in data link or physical layer and the user QoE.

When video streaming service is offered over wireless networks, there are two variability time scales in QoE metrics: flow level (tens of seconds) driven by the departures/arrivals of calls, and wireless channel variability time scale (milliseconds) driven by the fast fading [5]. The analytical results in [5] demonstrate that the flow dynamics have dominant influence on QoE metrics compared to the jittering in the throughput due to the fast fading. Therefore, we model the radio access network in flow level and focus on the video flow behaviors such as arrival, departure, mobilty and rebuffering while reducing the complexity involved by packet-level protocols [6]. In this paper, a flow refers to a video streaming session.

The rest of paper is organized as follows. Section 2 discusses relevant related work. In Sect. 3, we introduce the mobile network and QoE metrics. Prediction performance of four different types of video streaming is shown in Sect. 4. Section 5 concludes the paper and discusses the future works.

2 Related Work

QoE has recently gained momentum as a way to assess the perceived quality of users during videos watching. Authors of [7] studied the QoE with TCP information. Authors of [8,9] utilized flow-level model to investigate the video performance metrics, where the correlation between video rebuffering and the proposed performance metrics is not clear. Machine learning has been widely used to study both subjective and objective QoE to deal with the complexity of finding correlation between the parameters. Authors of [10] used machine learning to study the correlation between users' engagement and application metrics, such as buffer times. In [11], the cell-related parameters were first used as the research focus, but they just researched whether rebuffering occurred. Studies like the one presented in [12] proposed a QoE predicting module for adaptive HTTP streaming, without taking the traditional bitrate-constant streaming into account. Although many services have already made the migration towards adaptive streaming, their platforms continue to maintain backward compatibility with traditional bitrate-constant streaming. The investigation performed in [13] predicted QoE factors focusing on the hidden and context information, while consideration of up to 50 associated variables may increase the complexity of attribute extraction and the construction of the predictive model.

The authors of [11] used cell-related parameters (e.g., physical throughput and number of active flows) with Support Vector Machine (SVM) to predict whether a flow will encounter a rebuffering event. We consider this work as a starting point for our research and present two further contributions: (1) Instead of merely focusing on rebuffering/non-rebuffering, we bring insight into the relationship between cell-related QoS metrics and three main QoE metrics, namely rebuffering frequency, the video quality, and its variation. (2) In terms of machine learning tools, the Random Forest algorithm is adopted, which outperformes SVM in multiple classification problems and supports feature (attribute) selection analysis.

3 System Description

In this section, model of radio access network based on the flow-level concept are presented firstly. Then we show four types of HTTP streamings in our simulation. At last, we introduce the recorded attributes and QoE metrics.

3.1 Radio Access Network

Based on the concept of flow-level model in paper [8], a cell is modeled by a set of K capacity regions denoted as $R = \{R_1, \ldots, R_K\}$. In each region, physical throughputs are supposed to be homogeneous and thus, on the downlink, users are served with the same physical throughputs. Users in a cellular network are classified into static users and mobile users. The physical throughput of static users is assumed to be constant, and that of mobile users may randomly vary with time when a mobility envent occurs.

As for traffic characteristics, we follow the classical assumption that streaming flows with beginning physical throughput R_k arrive as a Poisson process with rate $\lambda_k = p_k\lambda$, where λ is the overall flow arrival rate in the cell and p_k stands for the traffic proportion with physical throughput R_k, where $\sum_k p_k = 1$. With the stability condition in paper [8], the maximum flow arrival rate, λ_{max}, guaranteeing the system stability, can be obtained. In our simulation, eight flow arrival rates normalized by the maximum value λ_{max} were demonstrated, since traffic arrival rate, λ, varies along hours in the real network. For each λ, simulator generates $m = 10^6$ streaming arrivals for the training of the Random Forest.

3.2 HTTP Video Streaming

Generally speaking, the video streaming can be categorized into two types

- Fixed bitrate streaming (also called progressive download). This is the original implementation of the HTTP video streaming and maintains a fixed bitrate for each chunk during the whole video downloading process.
- Adaptive streaming. Adaptive video streaming can switch among several optional bitrates according to the measured throughput, γ. Given the preset discrete set

$V = \{v_1, \ldots, v_M\}$, where $v_M > \ldots > v_1$, users select a video bitrate, v, for the next chunk as below, where $i = 1, \cdots, M-1$.

$$v = \begin{cases} v_M, & \gamma \geq v_M \\ v_i, & v_i \leq \gamma < v_{i+1} \end{cases} \tag{1}$$

In order to provide a solution which will be compatible with current and previous video streaming technologies, four types of streamings are simulated. Table 1 lists the four types of streamings in our simulation.

Table 1. Types of streamings.

Type	Description
Type I	Static and adaptive streaming
Type II	Static and fixed bitrate streaming
Type III	Mobile and adaptive streaming
Type IV	Mobile and fixed bitrate streaming

3.3 Recorded Attributes

We aim to take a step closer to exploring the correlation of each user's initial QoS metrics and user's QoE by recording complete buffer statistics. Therefore, we develop an simulator that simulates the actual behavior (e.g., playback, rebuffering, and mobility) and buffer state of each user in a radio access network, driven by some flow-events. In Fig. 1, we present an illustration of a video session life time in the event-driven simulatorm, where the buffer state will switch as the corresponding flow-event occurs and the chunk events mean downloading of a new chunk. Fine-grained information about the video session in our simulation program is recorded, including the bitrate of each video segment, the bitrate switching between adjacent video segments, and the number of rebuffering events during video downloading.

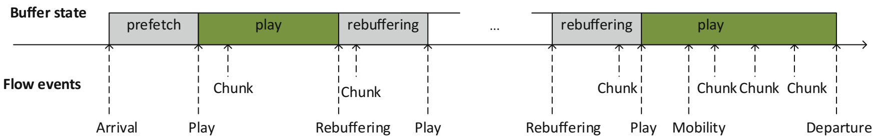

Fig. 1. An illustration of a video session life time in the event-driven simulator.

Table 2 presents all the data output by our simulator for j-th user, which can be summarized into two sets: (1) attributes set: the initial attributes recorded when user j arrives and (2) targets set: the total number of rebuffering events and the set of selected bitrates recorded during departure of user j.

Table 2. Parameters generated for j-th user in our simulations.

Set	Symbol	Description	Unit
Attributes	R_j	Physical throughput recorded at arrival	Mbps
	T_j	Video duration	S
	F_j	Numbers of flows in cell of each region	Vector
	$\lvert F_j \rvert$	Total number of flows in cell	Null
	F_j^r	Numbers of flows in rebuffering of each region	Vector
	$\lvert F_j^r \rvert$	Total number of flows in rebuffering	Null
Targets	N_j	Number of rebuffering events encountered	Null
	S_j	Set of bitrates selected	Vector

3.4 The QoE Metrics

In this subsection, we present three main QoE metrics reflecting the perceived video quality of users and the discretization for classification

- Rebuffering frequency (RF): The ratio of the number of rebuffering events to the duration of the session.
- Mean bitrate (MB, only for adaptive streaming) : The average of the bitrates weighted by the duration each bitrate is played for.
- Bitrate switch frequency (SF, only for adaptive streaming): The ratio of the number of bitrate switches to the duration of the session.

These metrics are difficult to be predicted in its raw continuous form. To simplify the classification and create a predictive model, we have further processed the metrics by labeling the data as shown in Eqs. (2)–(4).

$$RF_{label} = \begin{cases} "no\ rebuffering", & RF = 0 \\ "mild\ rebuffering", & 0 < RF < L_{rf} \\ "severe\ rebuffering", & L_{rf} \leq RF \end{cases} \tag{2}$$

where we adopt $L_{rf} = 0.1$, since [14] showed that with rebuffering ratio over 0.1, most of users abandon the video because of the quality degradation.

$$MB_{label} = \begin{cases} "low\ bitrate", & v_1 \leq MB < L_{mb1} \\ "middle\ bitrate", & L_{mb1} \leq MB < L_{mb2} \\ "high\ bitrate", & L_{mb2} \leq MB \leq v_M \end{cases} \tag{3}$$

where v_1 and v_M are the minimum and maximum values of the optional bitrates and we set L_{mb1} as 1.5, L_{mb2} as 2, the medians of the optional bitrates.

$$SF_{label} = \begin{cases} "no\ switch", & SF = 0 \\ "mild\ switch", & 0 < SF < L_{sf} \\ "severe\ switch", & L_{sf} \leq SF \end{cases} \tag{4}$$

where L_{sf} is set to 0.3, which distinguishes mild and severe switch in this paper.

4 Simulation Analysis

In this section, we analyze the prediction performance of machine learning among different types of HTTP streaming with recorded attributes. We adopt the simulation configuration in [11] and set the optional bitrates as 1, 1.5, 2, 2.5 Mbps.

WEKA [15], one of the most popular open-source machine learning libraries, is adopted to implement the Random Forest algorithm and to investigate the prediction performance. In classification for each QoE metric, the datasets consist of instance-label pairs (X_j, Y_j), where $j = 1, \cdots, m$. X_j consists of all attributes of user j, and Y_j corresponds to each category label. For example, the prediction of the rebuffering frequency can be expressed as a three-class classification problem with instance-label pairs $(X_j, RF_{label,j})$. With the feature selection algorithms, Random Forest evaluates the predictive power of each attribute and its redundancy with each other, and tends to select attributes that have a high correlation with the target but have a low correlation with each other. Effective feature selection can significantly reduce the difficulty of attribute extraction and the complexity of the predictive model. In addition, the Random Forest algorithm can evaluate the information gain which represents the worth of each attribute in the construction of the predictive model.

In our simulation, eight flow arrival rates normalized by the maximum value λ_{max} are demonstrated to show the performance at each load. Under each load, we present the respective prediction performance for four different HTTP video streamings, as shown in Fig. 2, 3, and 4. In general, when load increases, prediction performance decreases due to the increase of uncertainty.

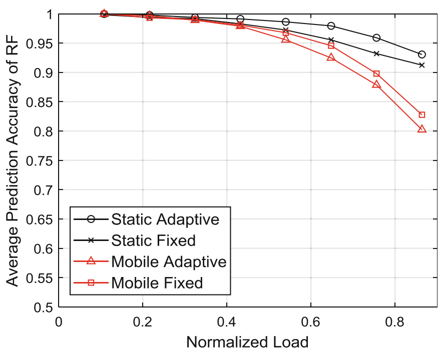

Fig. 2. The average prediction accuracy of the rebuffering frequency.

Figure 2 shows the average prediction accuracy of the rebuffering frequency. In general, sufficient accuracy can be achieved especially when the load is low. With respect to mobility of streamings, the simulation results show that static users can achieve more than 90% of accuracy even in large load, which is a significant improvement over previous approaches [16] where the achieved accuracy was approximately 84% for a binary classification and the severity of rebuffering was

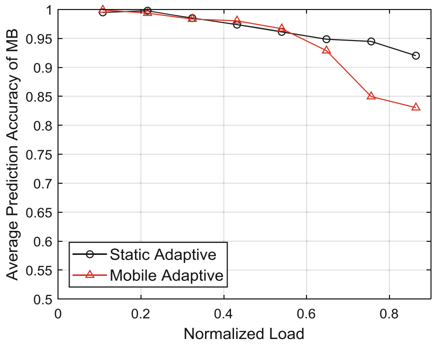

Fig. 3. The average prediction accuracy of the mean bitrate.

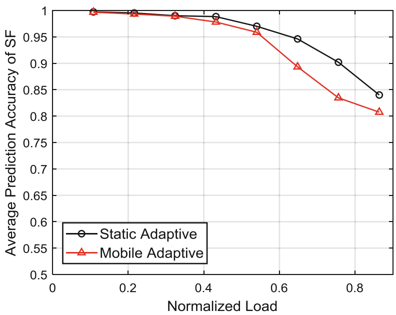

Fig. 4. The average prediction accuracy of the bitrate switch frequency.

unclear. However, rebuffering frequency of mobile users is much more difficult to be predicted when load is large. In terms of fixed or adaptive property, there is no general rule saying that fixed bitrate is easier to be predicted than adaptive streaming, where mobility plays a more important role. We list the results of feature selection for each type of streamings under the highest load in Table 3. As presented in Table 2, $F_{j,\ k}$ means the number of flows of region k in the cell, where $k = 1, \ldots, K$.

Firstly, for mobile users, the physical throughput R_j is not selected, which means that the initial physical throughput can not provide enough information to predict the rebuffering. Secondly, the abandon of $\left|F_j^r\right|$ suggests a high redundancy between $\left|F_j\right|$ and $\left|F_j^r\right|$, which may be good news for operators that they do not need to know more application information from users' side. Further, experiments show that, using the remaining attributes can achieve almost the same accuracy as overall attributes, but with reduced feature extraction overhead.

Figure 3 shows the average prediction accuracy of the mean bitrate. As mentioned earlier, the mean bitrate is only meaningful for adaptive video streaming. In terms of adaptive streaming alone, overall, over almost 85% accuracy is achieved even at high loads. Similarly, the prediction accuracy of static users can still reach more than 90%

Table 3. Attributes selected and respective information gain for RF.

Static adaptive		Static fixed		Mobile adaptive		Mobile fixed	
Attribute	Gain	Attribute	Gain	Attribute	Gain	Attribute	Gain
R_j	0.382	R_j	0.344	T_j	0.141	T_j	0.086
T_j	0.023	T_j	0.03	$\lvert F_j \rvert$	0.301	$\lvert F_j \rvert$	0.444
$\lvert F_j \rvert$	0.144	$\lvert F_j \rvert$	0.214	$F_{j,1}$	0.082	$F_{j,1}$	0.110
$F^r_{j,3}$	0.086			$F_{j,2}$	0.10	$F_{j,2}$	0.138
				$F_{j,3}$	0.15	$F_{j,3}$	0.219
				$F_{j,4}$	0.164	$F_{j,4}$	0.257
				$F_{j,5}$	0.121	$F_{j,5}$	0.179

even in high load and the impairment of mobility on predictions reduces the prediction performance for mobile users. Table 4 presents the results of feature selection for adaptive users under the highest load.

Table 4. Attributes selected and respective information gain for MB.

Static adaptive		Mobile adaptive	
Attribute	Gain	Attribute	Gain
R_j	0.254	T_j	0.044
T_j	0.002	$\lvert F_j \rvert$	0.57
$\lvert F_j \rvert$	0.308	$F_{j,3}$	0.294
$F_{j,3}$	0.195	$F_{j,4}$	0.324
$F_{j,4}$	0.219	$\lvert F^r_j \rvert$	0.461

Table 5 presents the confusion matrix for MB for static adaptive users under the highest load. The confusion matrix provides specific prediction accuracy of each class. We can see that the classification errors occur between instances *"Low"* and those with *"Middle"*, also between *"Middle"* and *"High"*, however, significantly fewer misclassifications between *"Low"* and *"High"*. Possible reasons include the classifier's inability to correctly identify marginal cases which are close to the MB thresholds, and the subtle differences between instances of different classes.

Figure 4 shows the average prediction accuracy of the bitrate switch frequency. In general, the accuracy of predicting for all loads exceeding 80% can be achieved, and when the load is not so high, the accuracy is above 90%. In addition, higher prediction accuracy for static users can be achieved.

Table 6 presents the results of feature selection for adaptive users under the highest load. The information gain of T_j shows the importance of T_j for predicting SF.

Table 5. Confusion matrix for MB of static adaptive users.

	Predicted label		
Actual label	*"Low"*	*"Middle"*	*"High"*
"Low"	97.2%	2.2%	0.6%
"Middle"	15.7%	73%	11.3%
"High"	2.8%	9.2%	88%

Table 6. Attributes selected and respective information gain for SF.

Static Adaptive		Mobile adaptive	
Attribute	Gain	Attribute	Gain
R_j	0.050	T_j	0.381
T_j	0.126	$\lvert F_j \rvert$	0.083
$\lvert F_j \rvert$	0.066		

5 Conclusions and Feature Works

In this paper, we aim to infer the QoE metrics from the observable QoS metrics with machine learning techniques. Based on the concept of flow-level dynamics, we develop an event-driven simulator to generate datasets, by which we correlate the cell-parameters and users' QoE. We examined the prediction performance of three QoE metrics for different HTTP video streamings along different loads. Then the machine learning technique, i.e., Random Forest, is used to obtain our predictive model along the system loads. Simulation results show that, with the initial information of each video session such as number of flows and radio conditions, sufficient accuracy can be achieved. In terms of type of streamings, the prediction of all metrics for static users performs better than mobile users, due to the increase of uncertainty from mobility, which calls for more information for prediction. We also perform feature selection with the highest load as an example to examine the effect of different attributes on each QoE metric and the correlation among attributes.

Future works will consider more attributes to improve the prediction accuracy in high loads, especially for mobile users. More QoE metrics like start-up delay will be researched to completely study the perceived quality by HTTP video streaming. The application of other machine learning models such as Neural Networks may improve the prediction accuracy.

Acknowledgements. This work has been sponsored by Huawei Research Fund (grant No. YBN2016110032) and National Science Foundation of China (No. 61201149). The authors would also like to thank the reviewers for their constructive comments.

References

1. Cisco visual networking index: Global mobile data traffic forecast update 2016–2021 white paper. https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/vni-infographic.html. Accessed 06 June 2018
2. Patrick Le Callet, S.M., Perkis, A.: Qualinet White Paper on Definitions of Quality of Experience (2012). http://www.qualinet.eu/index.php. Accessed 06 June 2018
3. Yin, X., Jindal, A., Sekar, V., Sinopoli, B.: A control-theoretic approach for dynamic adaptive video streaming over HTTP. In: Proceedings of the 2015 ACM Conference on Special Interest Group on Data, pp. 325–338, London (2015)
4. Dimopoulos, G., Leontiadis, I., Barlet-Ros, P., Papagiannaki, K.: Measuring video QoE from encrypted traffic. In: Proceedings of the 2016 Internet Measurement Conference, pp. 513–526, Santa Monica (2016)
5. Xu, Y., Elayoubi, S., Altman, E., El-Azouzi, R.: Impact of flowlevel dynamics on qoe of video streaming in wireless networks. In: 2013 Proceedings IEEE INFOCOM, pp. 2715–2723, Turin (2013)
6. Bonald, T., Proutiere, A.: A queueing analysis of data networks. Queueing Netw. **154**, 729–765 (2011)
7. Singh, K.D., Aoul, Y.H., Rubino, G.: Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC. In: 2012 IEEE Consumer Communications and Networking Conference (CCNC), pp. 127–131, Las Vegas (2012)
8. Bonald, T., Elayoubi, S., Lin, Y.-T.: A flow-level performance model for mobile networks carrying adaptive streaming traffic. In: IEEE Globecom, San Diego (2015)
9. Lin, Y.-T., Bonald, T., Elayoubi, S.: Impact of chunk duration on adaptive streaming performance in mobile networks. In: 2016 IEEE Wireless Communications and Networking Conference, Doha (2016)
10. Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, L., Zhang, H.: Developing a predictive model of quality of experience for internet video. In: Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM, pp. 339–350, New York (2013)
11. Lin, Y.-T., Oliveira, E.M.R., Jemaa, S.B., Elayoubi, S.E.: Machine learning for predicting QoE of video streaming in mobile networks. In: 2017 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, Paris (2017)
12. Chen, Z., Liao, N., Gu, X., Wu, F., Shi, G.: Hybrid distortion ranking tuned bitstream-layer video quality assessment. In: IEEE Trans. Circuits Syst. Video Technol. **26**(6), 1029–1043 (2016)
13. Vasilev, V., Leguay, J., Paris, S., Maggi, L., Debbah, M.: Predicting QoE factors with machine learning. In: IEEE International Conference on Communications (ICC) 2018, Kansas City (2018)
14. Krishnan, S., et al.: Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. IEEE/ACM Trans. Netw. **21**(6), 2001–2014 (2013)
15. WEKA: Data Mining Software in Java. https://www.cs.waikato.ac.nz/ml/weka. Accessed 26 May 2018
16. Aggarwal, V., et al.: Prometheus: toward quality-of-experience estimation for mobile apps from passive network measurements. In: Proceedings of the 15th Workshop on Mobile Computing Systems and Applications, p. 18. ACM, Santa Barbara (2014)