Understanding the Relationships between Performance Metrics and QoE for Over-The-Top Video

Weiwei Li *, Petros Spachos §, Mark Chignell*, Alberto Leon-Garcia*, Leon Zucherman‡ and Jie Jiang‡
*Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada
§School of Engineering, University of Guelph, Guelph, ON, Canada
‡Technology Strategy and Operations, TELUS Communications Company, Toronto, Canada

***Abstract*—In this paper, we study the relationships between Quality of Service (QoS) and Quality of Experience (QoE) in a session-based Over-The-Top (OTT) video service. A number of Performance Metrics (PMs) with and without the existence of failures during a video are examined. As QoE factors, Technical Quality (TQ) and Acceptability are used. We analyze the correlation between QoS performance metrics and QoE factors, and find new PMs should be employed because failures are included in QoE evaluation. We also summarize the relationships between QoS metrics and QoE factors through machine learning approaches. Using decision tree, we have a general idea about the relationships between PMs and QoE factors. We also understand the impact caused by failures and the value of rating scales.**

I. Introduction

Telecommunications should provide services which satisfy consumers, while the dramatic growth on data traffic is stressing network. Video service has occupied an importance place in network services and video traffic has taken a huge amount of traffic on the Internet. The crucial requirement is to support video services meet a customer's expectation in terms of Quality of Experience (QoE). QoE is a subjective measure of user's perception.

However, current QoE study needs to be investigated and developed systematically. People are still at the stop of determining the methodology of QoE assessment, exploring the connections between Quality of Service (QoS) and QoE, and deriving potential models for QoE estimation. One thing we notice that traditional QoS has focused on the video itself, while ignored that video is only a part of the whole service session for an Over-The-Top (OTT) video streaming. Proposing QoE for next generation network, we realize that we should study QoE based on a life cycle of a video session [1]. It is reasonable since OTT video becomes popular recently.

In this paper, we studied the relationship between PMs and QoE factors collected by subjective experiments. PMs are objective measurable metrics to represent the quality of a video. Two QoE factors are involved in this paper: Technical Quality (TQ) and Acceptability. TQ collects user's opinions from available options to understand QoE, while Acceptability is a binary measure to locate whether user accept the video quality or not. We also discuss the impact of failures happened during a video display. Our previous research has showed that failures should be considered into QoE assessment [2] [3]. As a follow-up study, we investigate whether the addition of failure requires new performance metrics, and how the failures impact user's assessment on QoE.

The remainder of the paper is structured as follows: In Section II, the related work on QoE and PMs is reviewed. In Section III, the Session- Bases QoE and QoS is described. And the data collection process is presented in Section IV. Section V is the analysis overviews followed by Section VI, with the relationships between PMs and QoE factors. We conclude in Section VII.

II. Related Work

Current QoE research has initiated effort to reveal the relationship between QoS and QoE. A. Khan et. al. have studied the impact of QoS parameters on QoE and proposed a QoE adaptation scheme for video applications [4]. R. Imran et. al. have utilized statistical techniques to evaluate QoE performance based on QoS parameters [5]. Alreshoodi and Woods have summarized recent studies on QoE/ QoS correlation modes in [6]. They have summarized three possible approaches for mapping the QoE/ QoS relationship: use QoS to map QoE, use QoE to map QoS and use some QoS and QoE to estimate other QoS and QoE. They admitted that the problem is which approach is efficient enough.

Machine learning is introduced into the QoE/QoE model study. Balachandran et. al. proposed a data-driven machine learning approach to tackle the complex relationship between the quality metrics and QoE measurement [7]. Mushtaq et. al. have discussed different machine learning approaches to assess the correlations between QoS and QoE [8]. Six classifiers are tested based on their data. Chen et. al have discussed QoS parameters impacting users' satisfaction and proposed a video rate adaptation scheme to improve viewer QoE [9], [10].

QoE assessment based on 2-point likert scale ('Yes'/'No') is another way to understand user's perception. Menkovskis et. al. have implemented a QoE model to predict whether the quality of network service is acceptable ('Yes') or unacceptable ('No') [11]. Their mode is based on decision tree, and they declared that the accuracy is over 90%. Other work on acceptability QoE model is proposed by Song et. al. [12] They have generated a logistic regression model to map QoS parameters to acceptability.

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

Mok et. al. and Pessemier et. al. have studied the impacts of impairments on QoE directly instead of correlating network performance to QoE [13], [14]. We focus on finding the impact of impairment and failures, which has a close explanation about the QoS side compared to their work.

Dorian et. al. have proposed the impact of video quality on QoE factor, however, they did not discuss the possible failures although they mention the concept of a video session life, which included "stopped/ exit" [15].

III. Session-based QoE and QoS

A. Session QoE: impairments and failures

With the popularity of OTT video streaming, research proposed that user perception of a video service should be studied during the life-cycle of a video session [15].

The concept of session includes:

- request a video by user,
- wait for the video to start, watched the video along with some possible impairments, and
- quit the video service normally or abnormally.

Previous QoE assessment focused on Integrity impairments. Integrity concerns the degree to which the video session unfolds without impairment [16]. However, it ignores whether a video can successfully start and/ or normally end. This is why we proposed two more components, Accessibility and Retainability, to understand QoE in the whole session [1]. Accessibility concerns whether a session can start, and Retainability describes the capability to continue a session (with/ without impairments) till a normal end (video completion or user ends the display).

B. QoE Factors

In our previous research for impairments and failures, we discussed TQ, CQ (Content Quality), OX (Overall eXperience), and Acceptability. [17], [18] We found that TQ and Acceptability reflects user's perception to TQ, while CQ represents the content of videos. OX is mainly determined by TQ, not CQ. Therefore, we focus on two factors to represent user's perception:

Technical Quality (TQ): The level of TQ is determined by the rating scale as mentioned before. We employed three different scales:

- Scale A: Excellent, Good, Fair, Poor, Bad.
- Scale B: Excellent, Good, Fair, Poor, Bad, Terrible.
- Scale C: Excellent, Good, Fair, Poor, Bad, Terrible, Worst Possible.

1) *Scale A* follows MOS (Mean Opinion Score), which recommends by the ITU standard [16].
2) *Scale B* extends the option of Scale A on the negative side by adding one more choice as the bottom. We are interested in whether user's evaluation about impairments and failures tends to go to negative side, when more failures are shown.
3) *Scale C* also extends on the basis of MOS. However, we provide two more negative choices. The design for Scale B and Scale C is to tell whether the opinion score are stable even worse opinion scores are provided.

Acceptability: Two levels are included in Acceptability, 'Yes' and 'No'. The employment of Acceptability is widely discussed recently in [12]. Some online video websites start to use a similar binary rating instead the previous 5-points rating, such as 'like/dislike' in YouTube and Facebook [19], [20].

C. QoS Performance Metrics

For the objective metrics, we will discuss the following four video quality metrics:

i Rebuffering Number (RN): The number of rebuffering times during a video session.
ii Rebuffering Ratio (RR): The ratio of total rebuffering time versus the total display time of a video (rebuffering time + content viewing time).
iii Non-interruption Content Viewing Ratio (VR_c): The content display time after the last impairment versus the total content viewing time.
iv Non-interruption Viewing Ratio (VR_s): The content display time after the last impairment versus the total video length.

VR_c and VR_s are new performance metrics proposed with the addition of failure types, since the definition of Integrity assumes that user can watch the whole video although encountered impairments. We found that the length of content viewing time impact the evaluation of Retainability failures [2]. That is why we listed VR_c and VR_s as performance metrics. We will further discuss the association between VR_c/VR_s and QoE evaluation in Section VI.

IV. Data collection and dataset

A. Experiment Procedure

In our experiment, each subject watched 31 videos in about 90 minutes. The average length of videos is $96.7s$. Each video has one type of impairment/ failure. Each participant encountered the same amount of impairments/ failures, in a randomized order. At the end of each video, there were four questions about TQ, CQ, OX and Acceptability. The experiment procedure for each subject is the same as the experiment we described in [17], although we change the types of impairments and failures, and introduce different scales for the test.

B. Impairment and failure types

In this experiment, we have videos without any impairment/ failure (I0), videos encountered Accessibility failure (A), three types of Integrity impairments (I1-I3), and three types of Retainability failures (R0-R2). Performance metrics of corresponding types are listed in Table I.

C. Experiment description

The number of subjects in this experiment is 108. We divided subjects into four groups and each group finished their experiment under different test conditions. $G1$ followed the ITU standard model, i.e. the rating scale is a 5-point

TABLE I
PERFORMANCE METRICS FOR INTEGRITY IMPAIRMENTS AND FAILURES

Type	RN	RR	VR_c	VR_s
I0	0	0	1	1
I1	1	$\frac{t_d}{t_d+t_v}$	$\frac{t_v-t_{f1}}{t_v}$	$\frac{t_v-t_{f1}}{t_v}$
I2	2	$\frac{2t_d}{2d_f+t_v}$	$\frac{t_v-t_{f2}}{t_v}$	$\frac{t_v-t_{f2}}{t_v}$
I3	3	$\frac{3t_d}{3t_d+t_v}$	$\frac{t_v-t_{f3}}{t_v}$	$\frac{t_v-t_{f3}}{t_v}$
R0	0	0	1	$\frac{t_{R0}}{t_v}$
R1	1	$\frac{t_d}{t_d+t_{R1}}$	$\frac{t_{R1}-t_{f1}}{t_{R1}}$	$\frac{t_{R1}-t_{f1}}{t_v}$
R2	2	$\frac{2t_d}{2t_d+t_{R2}}$	$\frac{t_{R2}-t_{f2}}{t_{R2}}$	$\frac{t_{R2}-t_{f2}}{t_v}$
AF	1	1	0	0

$*t_d$: The duration of each rebuffering.
$*t_v$: The content time of each video.
$*t_{Ri}$: The content viewing time for Ri, i=0, 1, 2
$*t_{fi}$: The time point of the i-th rebuffering happened at the content time, i=1, 2, 3.

TABLE II
GROUPS ARRANGEMENT: RATING SCALES AND IMPAIRMENT/ FAILURES TYPES

Group	Subject No.	Rating scale	Impairment/failure	Symbols
$G1$	36	A	I0-I3	A_I
$G2$	24	A	I0-I3, R0-R2, AF	A_IF
$G3$	24	B	I0-I3, R0-R2, AF	B_IF
$G4$	24	C	I0-I3, R0-R2, AF	C_IF

scale (Scale A), and they only evaluated Integrity impairments during the experiment. $G2$, $G3$, and $G4$ evaluated both Integrity impairments and failures. At the same time, to further explore the impact of Retainability and Accessibility failures, we employed various rating scales on $G2$, $G3$, and $G4$. The purpose of various scales is to extend negative choices for rating, because our experiment introduces more negative scenarios (failure types) than usual. Table II shows that details about rating scales and group arrangement.

In our data analysis, we will use 5 to -1 to represent these ratings, i.e. Excellent = 5, Good = 4, Fair= 3, Poor= 2, Bad= 1, Terrible = 0, and Worst Possible = -1.

V. ANALYSIS OVERVIEW

In this paper, we are interested in the following questions:

- Whether the addition of failures requires new PMs or current PMs are enough to sketch the impact of failures?
- Whether the understanding about QoE based on Integrity impairment keeps the same after the addition of failures?
- Is an extended scale necessary for failures?
- Is the impact of PMs on TQ and Acceptability the same or not?

A. Correlation

Correlation is a statistical measure of association between two variables. A correlation coefficient is a direct approach to reflect relationships between a pair of variables. However, the disadvantages is that correlation cannot reveal the interactions if there are more than two variables.

Correlation can help us to decide whether a specific PM related to QoE factor or not. In Section VI, we first use the Kendall correlation to measure the relevance between multiple PMs and QoE factors. As recommended in [15], the Kendall correlation is a rank correlation which does not have any assumption on the distribution or the joint distribution of variables; while Pearson correlation assumes a linear correlation between variables. Considering that VR_s and VR_c are related to failures which are rarely discussed before, we are interested to figure out whether they should be used as a performance metric.

B. Machine Learning Classifier Approaches

Correlation is used to explore the relationships between two variables, and we use machine learning classification to analyze the complex relationships among selected PMs and QoE factors, and to compare the impact under variable test conditions (the addition of failures and the change of scales). Machine learning classifier is a black box approach to analyze the association between PMs and QoE factors. The advantage is that it provides a clear output (QoE factors) by the input (PMs). Although at the same time, it hides details from data. That is why, as stated in [7], the accuracy of prediction decreases when the requirement of granularity becomes higher.

In this paper, our target is to dig out the primary association between PMs and QoE factors. We find that the machine learning classifier is enough. Therefore, we first go through four simple and widely-used classifiers discussed in [6]–[8]. And then we select one classifier which is most stable accompanying high accuracy across all cases. The candidate methods are Naive Bayes, Logistic Regression, k-NN Classification, and Decision Tree.

VI. IDENTIFY RELATIONSHIPS BETWEEN PMS AND QOE FACTORS

In this section, we will analyze the relationship between PMs and QoE factors. Note that we do not consider Accessibility failure, which gets the lowest rating score across three scales.

A. Performance Metrics versus QoE factors

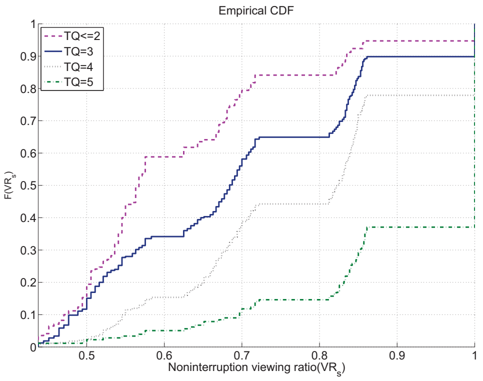
Table III summarizes the Kendall correlation between PMs and TQ under four cases. We find that the correlation between RN and TQ changes a lot when failures are introduced. It means RN cannot reflect the impact caused by rebuffering with the appearance of failures. On the other side, RR measures the impact of rebuffering in a stable manner even with failures.

At the same time, the absolute value of RR is the same as the absolute value of VR_c under same cases, which indicates that RR and VR_c represent a same property from different aspects. Note that the Pearson correlations of (RR, TQ) and that of (VR_c, TQ) are slightly different. It seems that VR_s represents the impact of failures different from RR. Hence, we select RR and VR_s for further analysis.

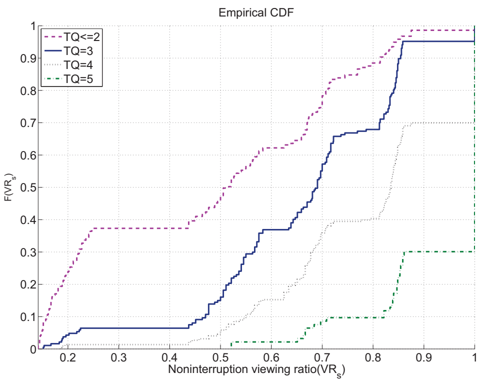
We also plot the empirical CDF of VR_s based on TQ levels in Figure 1. From these plots, it is clear that VR_s represent the characteristic of TQ levels no matter failures are included in the evaluation or not.

TABLE III
CORRELATION COEFFICIENTS BETWEEN PMS AND TQ.

PM \ TQ	A_I	A_IF	B_IF	C_IF
RN	-0.4259	-0.2761	-0.2596	-0.2306
RR	-0.3872	-0.4166	-0.3833	-0.3912
VR_c	0.3872	0.4166	0.3833	0.3912
VR_s	0.3872	0.4944	0.4574	0.4762

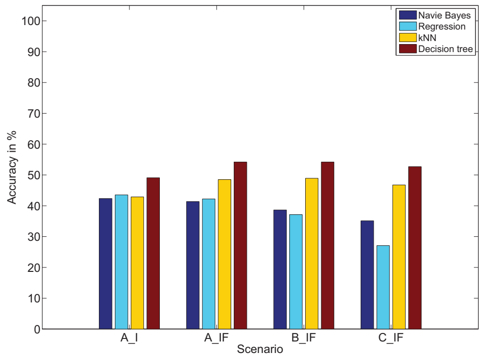


(a) A_I case

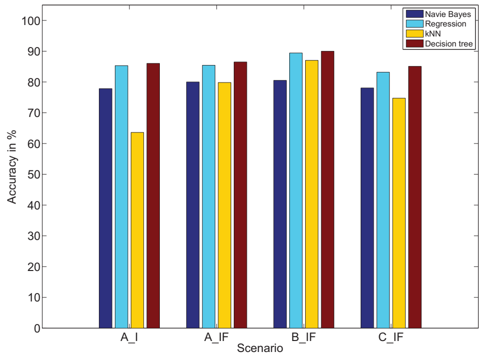


(b) A_IF case

Fig. 1. Empirical cdf for RR_s, A_I



(a) $TQ = f(PMs)$



(b) $Accept = f(PMs)$

Fig. 2. Machine learning classifiers

B. Machine Learning Classifiers Comparison

We use classification method to model the relationship, because TQ and Acceptability in our experiment are categorical variables. Four approaches are compared as mentioned in Section V. K-fold-cross-validation will be employed to find the one provides the highest mean accuracy.

Decision tree provides highest accuracy in almost all scenarios, which agrees with the conclusion in [7]. And it is the most stable classifier across all cases. In the following analysis, decision tree is employed to understand the relationships between PMs and QoE factors. The decision tree and other classifiers we used is implemented by scikit-learn tool [21] and MATLAB.

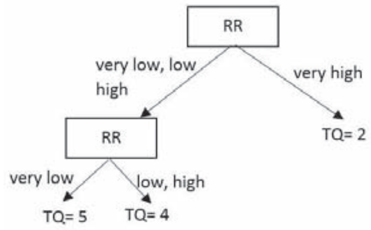
C. The Impact of Failures Appearance

To clearly reveal the effect of PMs under various scenarios by classification tree, we use the compacted decision tree as [7] did. We classify RR and VR_s into four levels: (very low, low, high and very high).

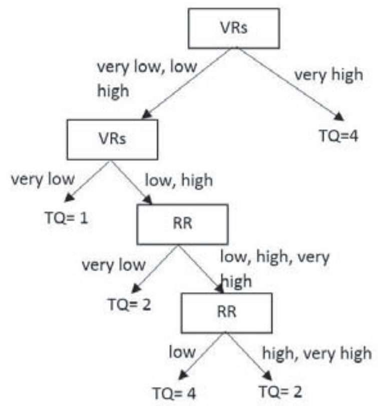
Figure 3 shows the structure of compacted decision tree to interpret $TQ = f(RR, VR_s)$ for A_I and A_IF. We find that if only Integrity impairments appeared, RR is the only consideration for TQ levels in the compacted tree. When failures are taken into account, VR_s plays a more important role to determine TQ levels. It proves that the impact of failures should be considered on QoE of OTT video streaming. And the changed structure indicates that a predictive model for both failures and impairments should consider VR_s first, and the impact of RR should be discussed separated under different VR_s levels.

Also, we notice that the compacted decision tree does not include all possible TQ levels. We can find that $TQ = 1$ are missed in the decision tree based on A_I, while $TQ = 5$ are missed in the case A_IF. It indicates that the failures cause a relative low TQ ratings as we concluded in [2] and [17].Usually, $TQ = 3$ is the most scatter choice for QoE rating [22], that's why $TQ = 3$ is missed in the classifier.

Two reasons for this phenomenon: first, the basis of decisions tree is information gain, which leads the tree biased to samples with larger size under same condition; and second, we compacted the level of PMs instead of using exact values, which provides a general structure for PMs on the price of

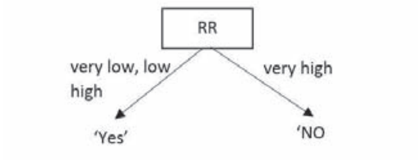


(a) $TQ = f(RR, VR_s)$, A_I

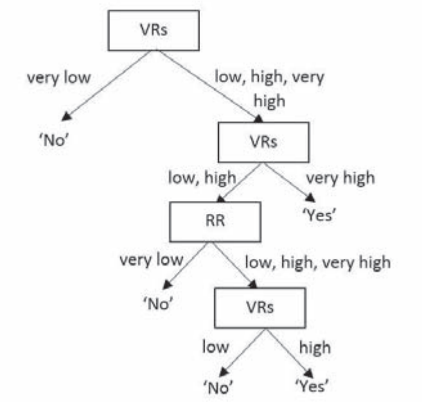


(b) $TQ = f(RR, VR_s)$, A_IF

Fig. 3. Compacted Decision Tree for A_I and A_IF



(a) $Accept = f(RR, VR_s)$, A_I



(b) $Accept = f(RR, VR_s)$, A_IF

Fig. 4. Compacted Decision Tree for A_I and A_IF

TABLE IV
MEAN ACCURACY (%) OF DECISION TREES, A_I AND A_IF

	$TQ = f(RR, VR_s)$	$Accept = f(RR, VR_s)$
A_I	42.64%	80.11%
A_IF	48.19%	81.11%

TABLE V
MEAN ACCURACY (%) OF DECISION TREES, B_IF AND C_IF

	$TQ = f(RR, VR_s)$	$Accept = f(RR, VR_s)$
B_IF	43.05%	82.92%
C_IF	43.75%	82.5%

losing precision of the tree. However, the compacted levels are enough to explore the general relationship between PMs and between PMs and QoE factors.

Figure 4 shows the decision tree of $Accept = f(VR, RR_s)$ for A_I and A_IF. It is obvious that the structures of both trees are close to the corresponding structures shown in Figure 3. Table IV show the mean accuracy based on the above compacted decision tress. The 'Yes'/'No' choice of acceptability leads higher accuracy. It convinces us that the selection of the granularity is important in deriving a predictive model, which stated in [7]. Considering that the available types of PMs might be limited in a OTT video, it is invaluable to discuss whether select TQ or Acceptability as the indicator of QoE.

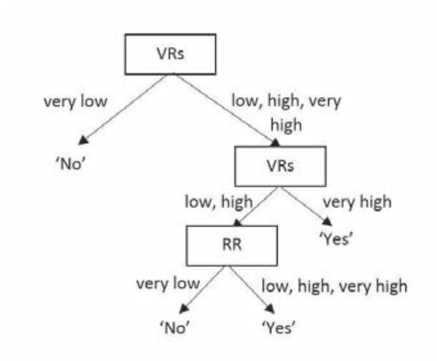
D. Compare Rating scales

The structure of decision trees for $TQ = f(RR, VR_s)$ in both scales are close to the structure shown in Figure 3(b), except that TQ tends to lower levels compared to the tree in A_IF case when RR_s are at the level of 'very low'. It indicates that Scale B and C can help user tell the difference between impairments and failures.

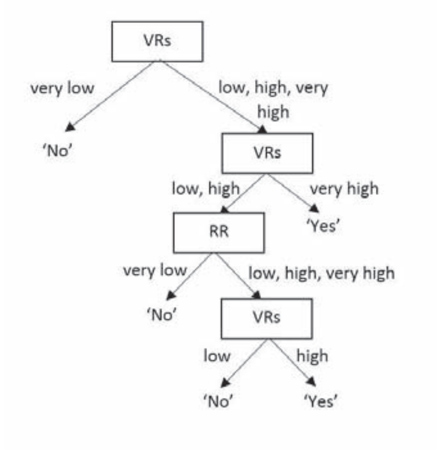
Figure 5 shows the decision tree of $Accept = f(RR, VR_S)$ under B_IF and C_IF. In these scenarios, we not only add failures in the evaluation, but also change the traditional ITU MOS scale to extended new scales. The main concern for this design is to find more details about the impact of Retainability and Accessibility failures. It is obvious that the structures of both trees are close to each other. It indicates that the impact caused by failures is stable across all cases if evaluated by Acceptability. We compare the accuracy of $Accept = f(RR, VR_s)$ and $TQ = f(RR, VR_s)$ shown in Table V. It shows that Acceptability still provides higher accuracy compared to TQ.

VII. CONCLUSIONS

In this paper, we study the relationship between performance metrics and QoE factors through a data-driven machine learning approach. We find that the feature of failures requires new performance metrics to be introduced for QoE evaluations. As a continuing research, we also compare the traditional MOS scale to the binary licker scale. It indicates that multiple levels is not necessary for all OTT video services. Depending on the requirement of accuracy and the purpose of



(a) $Accept = f(RR, VR_s)$, B_IF



(b) $Accept = f(RR, VR_s)$, C_IF

Fig. 5. Compacted decision tree for $Accept = f(RR, VR_s)$,B_IF and C_IF

QoE assessment, acceptability might be a valuable indicator for user's perception.

We also examine whether a extended scale is necessary for the addition of failures. We find that extended scales can help user distinguish different TQ levels.

Generally, depending on the granularity a QoE estimation model wants to achieve and the performance metrics the system can provide, there are different approaches to generate a predictive QoE model.

Acknowledgment

This research was supported by a grant from TELUS and a matching grant from NSERC/CRD.

References

[1] A. Leon-Garcia and L. Zucherman, "Generalizing mos to assess technical quality for end-to-end telecom session," in *Globecom Workshops (GC Wkshps), 2014*, Dec 2014, pp. 681–687.

[2] W. Li, H. Ur-Rehman, M. Chignell, A. Leon-Garcia, L. Zucherman, and J. Jiang, "Impact of retainability failures on video quality of experience," in *Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on*, Nov 2014, pp. 524–531.

[3] W. Li, H.-U. Rehman, D. Kaya, M. Chignell, A. Leon-Garcia, L. Zucherman, and J. Jiang, "Video quality of experience in the presence of accessibility and retainability failures," in *Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine), 2014 10th International Conference on*, Aug 2014, pp. 1–7.

[4] A. Khan, L. Sun, E. Jammeh, and E. Ifeachor, "Quality of experience-driven adaptation scheme for video applications over wireless networks," *IET Communications*, vol. 4, no. 11, pp. 1337–1347, July 2010.

[5] R. Imran, M. Odeh, N. Zorba, and C. Verikoukis, "Spatial opportunistic transmission for quality of experience satisfaction," *J. Vis. Comun. Image Represent.*, vol. 25, no. 3, pp. 578–585, Apr. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jvcir.2013.08.014

[6] M. Alreshoodi and J. Woods, "Survey on qoe\qos correlation models for multimedia services," *CoRR*, vol. abs/1306.0221, 2013. [Online]. Available: http://arxiv.org/abs/1306.0221

[7] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for internet video," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 339–350, Aug. 2013. [Online]. Available: http://doi.acm.org/10.1145/2534169.2486025

[8] M. Mushtaq, B. Augustin, and A. Mellouk, "Empirical study based on machine learning approach to assess the qos/qoe correlation," in *Networks and Optical Communications (NOC), 2012 17th European Conference on*, June 2012, pp. 1–7.

[9] Y. Chen, F. Zhang, F. Zhang, K. Wu, and Z. Q., "Qoe-aware dynamic video rate adaptation," in *IEEE Global Communications Conference (GLOBECOM) 2015*, December 2015.

[10] Y. Chen, Q. Chen, F. Zhang, Q. Zhang, K. Wu, R. Huang, and L. Zhou, "Understanding viewer engagement of video service in wi-fi network," *Comput. Netw.*, vol. 91, no. C, pp. 101–116, Nov. 2015.

[11] V. Menkovski, G. Exarchakos, and A. Liotta, "Machine learning approach for quality of experience aware networks," in *Intelligent Networking and Collaborative Systems (INCOS), 2010 2nd International Conference on*, Nov 2010, pp. 461–466.

[12] W. Song and D. Tjondronegoro, "Acceptability-based qoe models for mobile video," *Multimedia, IEEE Transactions on*, vol. 16, no. 3, pp. 738–750, April 2014.

[13] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching," *Broadcasting, IEEE Transactions on*, vol. 59, no. 1, pp. 47–61, March 2013.

[14] R. Mok, E. Chan, and R. Chang, "Measuring the quality of experience of http video streaming," in *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, May 2011, pp. 485–492.

[15] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," in *Proceedings of the ACM SIGCOMM 2011 Conference*, ser. SIGCOMM '11. New York, NY, USA: ACM, 2011, pp. 362–373. [Online]. Available: http://doi.acm.org/10.1145/2018436.2018478

[16] ITU-T, "Subjective video quality assessment methods for multimedia applications," *Recommendation P.910, Telecommuniation standardization sector of ITU*, Sep, 2009.

[17] P. Spachos, W. Li, M. Chignell, A. Leon-Garcia, L. Zucherman, and J. Jiang, "Acceptability and quality of experience in over the top video," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*, June 2015, pp. 1693–1698.

[18] W. Li, P. Spachos, M. Chignell, A. Leon-Garcia, L. Zucherman, and J. Jiang, in *IEEE Consumer Communications and Networking Conference (CCNC) 2016*.

[19] V. Schöndienst, F. Kulzer, and O. Günther, "Like versus dislike: How facebook's like-button influences people's perception of product and service quality," 2012.

[20] Y. Hou, T. Xiao, S. Zhang, X. Jiang, X. Li, X. Hu, J. Han, L. Guo, L. Miller, R. Neupert, and T. Liu, "Predicting movie trailer viewer's "like/dislike" via learned shot editing patterns," *Affective Computing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[22] T. Hobfeld, R. Schatz, and S. Egger, "Sos: The mos is not enough!" in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, Sept 2011, pp. 131–136.