Computer Networks 91 (2015) 101–116

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Understanding viewer engagement of video service in Wi-Fi network

Yanjiao Chen[a,*], Qihong Chen[a], Fan Zhang[a], Qian Zhang[a], Kaishun Wu[b], Ruochen Huang[c], Liang Zhou[c]

[a] *Hong Kong University of Science and Technology, Hong Kong*
[b] *Shenzhen University, China*
[c] *Nanjing University of Posts and Telecommunications, China*

ARTICLE INFO

Article history:
Received 15 March 2015
Revised 20 August 2015
Accepted 22 August 2015
Available online 31 August 2015

Keywords:
Viewer engagement
Quality metrics
Video content
Viewer demography
Personalized prediction model

ABSTRACT

With the dramatic growth of online video services and video traffic, video service providers and network operators have keen interest in improving viewer engagement. Viewer engagement is mainly influenced by four aspects: service quality metrics (e.g., rebuffer time), network quality metrics (e.g., physical-layer data rate), video content (e.g., video length) and viewer demography. Previous works only partially consider some of these factors due to limitation of the dataset. In this paper, we develop an experimental platform with more than 50 self-deployed routers in our university campus, collecting information regarding all four aspects of engagement-related factors. Correlation and information gain analysis show that different viewer groups and video content types have different engagement patterns. Furthermore, we analyze each factor's significance in determining viewer engagement. Finally, we propose to build personalized models to better predict viewer engagement, with bootstrapping customized models for new viewers.

1. Introduction

It is estimated that the sum of all forms of videos (TV, Video-on-Demand (VoD), Internet, and P2P) will be around $80\sim90\%$ of global consumer traffic by 2017 [1]. Video streaming over the Internet, especially through mobile network, is becoming more and more popular. Throughout the world, Internet video traffic will be 69% of all consumer Internet traffic by 2017 [1], and mobile video traffic will be over one third of mobile data traffic by the end of 2018 [2]. The majority of video traffic to mobile is still Wi-Fi, not 3G or 4G [3].

Video service is user-centric, therefore, it is better to evaluate video quality from the viewers' perspective, known as Quality-of-Experience (QoE). Traditionally, QoE is measured by subjective test, which directly solicits viewers' evaluation scores under the laboratory environment. However, due to its high cost, subjective tests are often conducted with limited viewers, video types and test conditions. Furthermore, subjective test cannot be used for real-time monitoring.

With the availability of large-scale online video data, data-driven approaches emerge for viewer satisfaction analysis. Instead of focusing on viewer experience, researchers resort to viewer engagement, which can be measured by quantifiable metrics (e.g., viewing time ratio, the number of watched videos and the probability of return), and conforms with the business models of subscription-based or advertisement-based video services. If a viewer is more engaged with a video, it is more likely that he will subscribe to the video service, and more ads can be displayed to

* Corresponding author. Tel.: +1 (416) 526-6035.
E-mail addresses: chenyj.thu@gmail.com (Y. Chen), qchenag@ust.hk (Q. Chen), fzhangaf@ust.hk (F. Zhang), qianzh@ust.hk (Q. Zhang), wu@szu.edu.cn (K. Wu), ruochen90@qq.com (R. Huang), 13913936864@163.com (L. Zhou).

http://dx.doi.org/10.1016/j.comnet.2015.08.006
1389-1286/

him. In [4], the authors collect the data from the content providers' websites to study the influence of service quality metrics (e.g., rebuffer time) on viewer engagement, and predict viewer engagement by linear regression. As an extension, in [5], the authors further study the impact of video content on viewer engagement; and propose decision tree-based prediction model to characterize non-linear and non-monotonic relationship between viewer engagement and other factors. In [6], the authors use the dataset from the network operator to study how network quality metrics affect viewer engagement, and predict viewer engagement with regression tree models.

Different from the existing works, we develop a third-party experimental platform consisting of more than 50 self-deployed routers in the university campus. Having collected the HTTP packets using TCPDUMP, we perform deep packet inspection to extract all four aspects of engagement-related factors[1]: service quality metrics, network quality metrics, video content and viewer demography. We first use correlation and information gain to analyze the influence of a certain factor on viewer engagement. We have the following observations.

- *Viewer group differentiation*. Different viewer groups have different sensitivities towards the same factor. For example, video popularity has a more significant influence on female viewers than male viewers.
- *Importance of data rate variance*. According to the analysis, average data rate has little influence on the viewer engagement, while data rate variance has relatively high correlation with viewer engagement. The role of data rate variance has never been mentioned in previous works.
- *Poor viewer engagement on mobile devices*. The viewing time ratio on mobile devices is much lower than that on personal computers. Moreover, we find that, for mobile devices, average transmission data rate is negatively correlated with viewer engagement. One possible reason is that the rate adaptation of the mobile devices is too aggressive, resulting in high retransmission and low throughput, which leads to poor viewer engagement.

To further analyze the significance of each factor in determining viewer engagement, we leverage Bayesian theory to compute the posterior probability of viewer engagement conditioned on a certain viewing condition. We find that, for a particular factor, its influence on viewer engagement is significant within a certain range. For example, when the rebuffer time is $30 \sim 50$ s, its influence on viewer engagement is most significant; when the rebuffer time is fewer than 30 s or greater than 50 s, its influence on viewer engagement can be ignored. Therefore, the network operator who wants to increase viewer engagement can focus on decreasing rebuffer time while it is in the range $30 \sim 50$ s, but investing the resources elsewhere (e.g. decreasing the startup delay) if the rebuffer time is fewer than 30 s.

[1] Our deep packet inspection-based method can be applied to the unencrypted HTTP packets used by most of China's video websites, but not to encrypted HTTPS and HTTP/2 packets used by video websites such as YouTube. Existing methods that extract information from the client- or server-side logs are able to work on HTTPS and HTTP/2 packets after decryption [4,5]. Methods that do not rely on application-layer information can also deal with HTTPS and HTTP/2 packets [6].

To predict viewer engagement, we first build a general model for the entire dataset and major viewer groups. Then, we personalize the system, training the model only on a single viewer's data, achieving better prediction accuracy. However, personalized prediction models demand the viewers' historical data for training, which is unavailable for a new viewer. Therefore, we build customized models for bootstrapping the training process for a new viewer, based on historical data from viewers who share similar attributes with the new viewer.

Our major contributions are as follows.

- We build a third-party experimental platform to collect a dataset with all four engagement-related factors: service quality metrics, network quality metrics, video content and viewer demography. We also carry out a survey to solicit viewers' subjective opinions on their usual viewing habits. (Section 3)
- We analyze the influence of different factors on viewer engagement through correlation and information gain analysis. We further analyze engagement patterns for different viewer groups and content types. (Section 4)
- We analyze the significance of different factors in determining viewer engagement. (Section 5)
- We propose personalized models for predicting individual viewer's engagement, and bootstrap the training process of new viewers with customized models. (Section 6)

2. Related work

Large-scale measurement studies have been carried out for various video services, including online VoD service [7,8], live VoD service [9], the YouTube traffic [10–12] and mobile video service [13]. Different from these measurement studies which present general viewer behaviors, we focus on an in-depth understanding of the influence of different factors on viewer engagement.

Recently, data-driven analysis on viewer engagement has drawn much attention. In [4], the authors study the influence of video service quality metrics on viewer engagement for different video types, and quantify such influence by linear regression. However, the linear regression cannot capture the non-linear and non-monotonic relationship between viewer engagement and other factors. To address this problem, a decision-tree based prediction model is developed in [5,14], but it can only predict viewer engagement in a coarse-grained level. In [6], the authors study the influence of cellular network conditions on viewer engagement, and build fine-grained prediction model based on regression tree algorithm. In [15], the authors verify the causal relationship between video service quality metrics and viewer engagement, but it does not work on predicting viewer engagement. All of the previous works collect data from either content providers' websites or cellular network operators.

Compared with existing data-driven analysis on viewer engagement, our work is different in three ways. First, our dataset includes both service quality metrics and network quality metrics, as well as information on video content and viewer demography. Second, we have a more in-depth

Fig. 1. A typical video session.

analysis on the significance of different factors in shaping viewer engagement. Third, we propose to establish personalized models instead of general models for better predicting viewer engagement. Nevertheless, we extract video-related information from the packet header, which can only deal with unencrypted HTTP packets, but not encrypted HTTPS and HTTP/2 packets. Existing works that collect and analyze data from the client- or server-side logs do not have this problem as the packets are decrypted by the browser [4,5]; the methods that do not use application-layer information also averts this difficulty [6].

3. Preliminaries and dataset

In this section, we first introduce the background of video service and viewer engagement. Then, we describe the data collection process. Finally, we present general statistics of our dataset.

3.1. Background

Fig. 1 shows a typical video session. After the viewer initiates a video request, a certain amount of data has to be downloaded in the buffer before the video starts playing (*startup state*). During playing, the video player fetches the data in the buffer to play to the viewer; and meanwhile, more data are downloaded from the server (*playing state*). If the rate of data consumption exceeds the rate of downloading (e.g., due to poor network capacity), the buffer will be exhausted. In this case, the video player has to pause to fill its buffer to a certain level before start playing again (*rebuffer state*). The viewer may choose to quit before the video completes, or watch to the end of the video.

In this paper, we quantify viewer engagement by *viewing time ratio*, that is, the actual time that a viewer watches a video clip divided by the total video length [5]. The more time a viewer spends watching a video, the more *engaged* he is. When studying potential factors that may affect viewer engagement, we consider service quality metrics, network quality metrics, video content, and viewer demography.

Service quality metrics

- *Startup delay*. As shown in Fig. 1, startup delay is the time between the viewer requests the video and the video actually begins playing, advertisement time included.
- *Rebuffer time*. As shown in Fig. 1, when the buffer is depleted, the player pauses to rebuffer. We use the total rebuffer time during a video session to quantify the rebuffer events. We also consider rebuffer time ratio, that is, the total rebuffer time divided by video length.

Network quality metrics

- *PHY data rate*. IEEE 802.11 standard designated a series of available physical-layer data rate for the Wi-Fi network[2]. For example, 802.11b supports {1, 2, 5.5, 11} Mbit/s and 802.11g supports {6, 9, 12, 18, 24, 36, 48, 54} Mbit/s. We use the mean and the variance to quantify the PHY data rate.
  - *Average data rate*. Average data rate, measured in megabit per second (Mb/s), is the time average of the PHY data rate during a video session.
  - *Data rate variance*. We use data rate variance to represent how stable the data rate is during the video session. Low data rate variance indicates relatively stable wireless environment, while high data rate variance indicates disturbing wireless environment. Let $r_1, r_2, \dots, r_n$ denote the data rate samples, and μ_r denote the average data rate. We adopt the following three metrics to quantify data rate variance.
    - Standard deviation $V_1 = \sqrt{\sum_i (r_i - \mu_r)/n}$
    - Absolute data change $V_2 = \sum_i |r_i - r_{i-1}|$
    - p-norm $(p = 3)$ $V_3 = (\sum_i |r_i - \mu_r|^p)^{1/p}$
- *Signal strength*. The signal strength is measured by the antenna in dBm.

Video content

- *Video length*, the total time duration of the video.
- *Video popularity*. We use the number of (previous) views of video to represent its popularity.
- *Device types*, classified as personal computers (PC) and mobile devices.

Viewer demography

- *Gender*. Female or male.
- *Grade (of learning)*. In our dataset, the grade of viewers include 4th year undergraduate (UG4), 1st year postgraduate (PG1) and 2nd year postgraduate (PG2).

3.2. Experimental platform and dataset

The dataset in [4,5] is collected from the video player instrumentation, which includes only service quality metrics but no network quality metrics. The dataset in [6] is collected from the network operator, which includes only network quality metrics but no service quality metrics. With the help of our experimental platform, we are able to collect both

[2] Note that the PHY data rate is not video encoding bitrate. We cannot identify video bitrates from the HTTP packet header. In the following of the contexts, we will use data rate to refer to PHY data rate without confusion.

Table 1
Indicators in HTTP packet header.

Event	Indicator
Initiate request	http contains "click"
Start playing	http contains "/vc?"
Video length	http contains "/player/addPlayerDurationReport?"
Actual watching time	"http contains "/tslog?"
Pause	http contains "e=pause"
Resume	http contains "e=play"

Fig. 2. Video length.

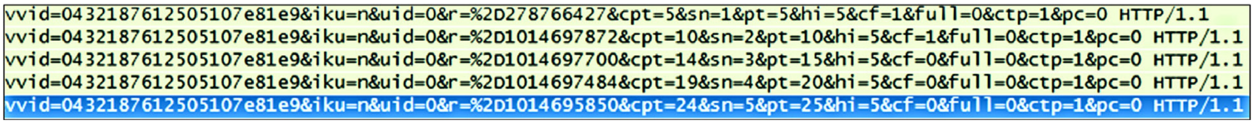


Fig. 3. Video length.

network quality metrics (from the routers) and the service quality metrics (derived from the HTTP packets). Moreover, we indicate the basic viewer demography information from the location of the routers, which is never considered in existing works.

We deployed 50 routers in the offices and student dormitories across the university campus, providing free Wi-Fi access to the users. All the users who access the routers are notified of the experiment and are asked to sign an e-agreement for data collection. For data processing, we only parse the unencrypted HTTP packet header to get service quality metrics and network quality metrics, without intruding the user privacy. The data collection process lasted for three months from April to June in 2014. The final dataset consists of 11996 unique video sessions and 469 unique viewers. Our dataset is collected independently from the content providers and the network operators. Though it is smaller in scale compared with previous works, it contains more detailed information on each video session, allowing us to gain more in-depth insight on the influence of each factor on viewer engagement.

The data collection platform consists of two major parts: the client routers and the storage server. We use the off-the-shelf *MERCURY MW4530R 750 Mbps Dual Band Wi-Fi Wireless Gigabit Router* for the client routers. The original operating system is overwritten by OpenWrt [16], an operating system based on Linux kernel. TCPDUMP is used [17] to capture all the uplink and downlink packets between the users and the router. Meanwhile, the routers monitor the wireless network parameters by iw [18], a CLI (command-line interface) configuration utility provided by OpenWrt for wireless devices. The storage space of a router is only 16 MB, too small to keep all the packets, which are usually more than 10 GB per day for a router. In addition, it is burdensome to collect data from all routers distributively. Therefore, we leverage a distant server with 10 TB storage space. Each router copies and sends the collected data packets and the log files of wireless parameters to the server[3], through Network File System (NFS) protocol [19], a distributed file system for file access between different computers. At the same time, we can remotely manage all the routers with the help of the server.

To parse the captured packets, we first use the open-source packet analyzer software Wireshark to identify different events during the viewing session. The indicators in the packet header as used in the Wireshark filters are listed in Table 1.

Viewing time ratio. To get the viewing time ratio, we have to know the total video length and the part of the video that the user has actually watched. The video length information is included in the packet that contains "/player/addPlayerDurationReport?", with the field named "totalsec=" as shown in Fig. 2. Through intensive experiments, we have found that the packet with "/tslog?" will periodically report the actual video watching time by "pt=", as shown in Fig. 3. Note that "pt" denotes the duration of the video but not the total time spent, for example, the user may spend a total of 15 s, but only watches 10 s of the video due to rebuffer time, and the video length is 20 s. In this case, "pt" = 10, and the viewing time ratio is 10/20 = 0.5. We use the "pt" of the last "tslog?" packet to represent the actual video watching time.

Rebuffer time. The rebuffer time is the total time spent minus the actual video watching time and the pausing time[4].

[3] We have verified that the transmission of the copied packets will not congest the network and has little influence on the transmission of regular packets.

[4] The calculation of the rebuffer time is also affected by users' fast-forward or backward operation. This situation is too complicated and we leave it to be solved in our future work.

The total time spent is given by:

$$T_{total} = t_{end}^{/tslog?} - t^{/vc?} \tag{1}$$

in which $t^{/vc?}$ is the time stamp of the packet that indicates the start of the video (Table 1); $t_{end}^{/tslog?}$ is the time stamp of the last "/tslog?" packet. The actual video watching time is the "pt" of the last "/tslog?" packet. We found that every time the video is paused, we receive a packet containing "e = pause", and when the video is resumed, we received a packet containing "e = play". Therefore, the total pause time is:

$$T_{pause} = \sum_{i} (t_i^{e=play} - t_i^{e=pause}) \tag{2}$$

in which $t_i^{e=pause}$ and $t_i^{e=play}$ is the time stamp of the ith pause and resume events respectively. The rebuffer time can be calculated as $T_{rebuffer} = T_{total} - T_{actual} - T_{pause}$.

Startup delay. The startup delay can be easily derived as $T_{startup} = t^{/vc?} - t^{click}$, in which $t^{/vc?}$ is the time stamp of the packet that indicates the start of the video, and t^{click} is the time stamp of the packet that indicates the user request for the video.

We use TSTAT [20], a traffic statistics and analysis tool to process the packets by seeking for indicators in Table 1 and extracting corresponding information. The available version of TSTAT can only analyze the statistics of YouTube traffic but not other video websites. Therefore, we adapt TSTAT to parse HTTP packets from other 11 video websites. We reconstruct the video URL to crawl the video popularity information from the original video website.

The PHY data rate and signal strength are sampled every 1 s by iw in the router. Sampling the data rate with higher frequency will interrupt the normal operation of the routers since each router only has one CPU. One problem with cross-layer analysis is that the time granularity of physical layer metrics (e.g., PHY data rate, signal strength) and application layer metrics (e.g., rebuffer time, startup delay) are intrinsically different. The physical layer metrics have much finer time granularity (e.g., tens of μs) than the application layer metrics (e.g., tens of s). The sampling process resolves this problem by providing physical layer metrics at a coarser time granularity. More ideally, in the future work, we hope to monitor the physical layer metrics continuously and then aggregate them to match the time granularity of application level metrics.

3.3. Subjective survey

To have a better understanding of the viewers' usual viewing behavior, we carry out an anonymous survey among residents of the student dormitories in which we have deployed our routers. We have collected a total number of 165 questionnaires, which help us interpret some of the observations from the video dataset. Interestingly, there are some contradictions between the subjective survey and the video dataset. The questions in the subjective survey include the following parts.

- Basic information: e.g., gender and grade.
- General viewing behavior: e.g., the number of days to access video websites each week, average duration of each access.
- Engagement evaluation: e.g., average viewing time ratio, startup delay tolerance and rebuffering time tolerance for ordinary and interesting videos respectively.
- Self-evaluation: to rank one's own patience levels as "5 (very impatient)", "4 (impatient)","3 (moderate)","2 (patient)", and "1 (very patient)".

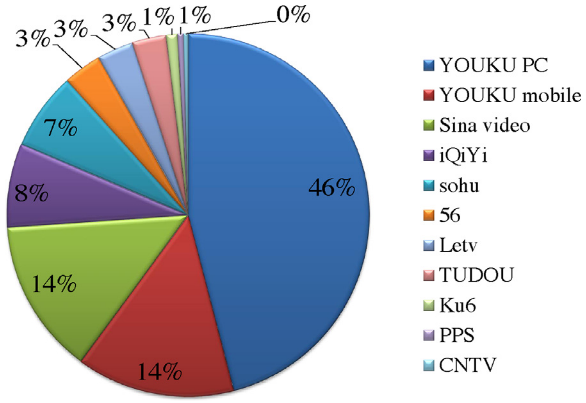


Fig. 4. Video website distribution.

3.4. Dataset statistics

Now we present some general statistics of the video dataset. Of all the video sessions, 5102 are watched by male viewers, 1539 by female viewers, and the rest are unknown. Of all the video sessions, 1952 are watched by undergraduate 4th year (UG4) viewers, 5691 by postgraduate 1st year (PG1) viewers, 649 by postgraduate 2nd year (PG2) viewers, and the rest are unknown[5].

The video sessions come from almost all major video websites which provide VOD service in China. Fig. 4 shows the proportion of the video sessions from each website. We can see that the majority of the videos come from YOUKU, the most popular video website in China. Access to YOUKU from the PC and the mobile devices consists more than half of the total video access in our dataset. The traffic rank of our dataset generally conforms with that of Alexa ranks [21].

4. Correlation and information gain analysis

In this section, we leverage correlation and information gain analysis to quantify the impact of different factors on viewer engagement. We first analyze the entire dataset; then, we separate video sessions based on different viewer groups; finally, we look at different video types and device types.

4.1. Overview

Fig. 5 (a) shows the CDF of viewer engagement across the entire dataset. In general, the viewer engagement is low,

[5] We infer the gender and the grade information of the video session from the location of the router. For instance, if the router is deployed in a PG4 female dorm, we assume that the video sessions are watched by PG4 female viewers. If the router is deployed in an office with mixed viewer groups, we label the viewer demography information as unknown. Though this method may introduce some noise into the dataset, this is by far the only way to get the viewer demography information without intruding the privacy.

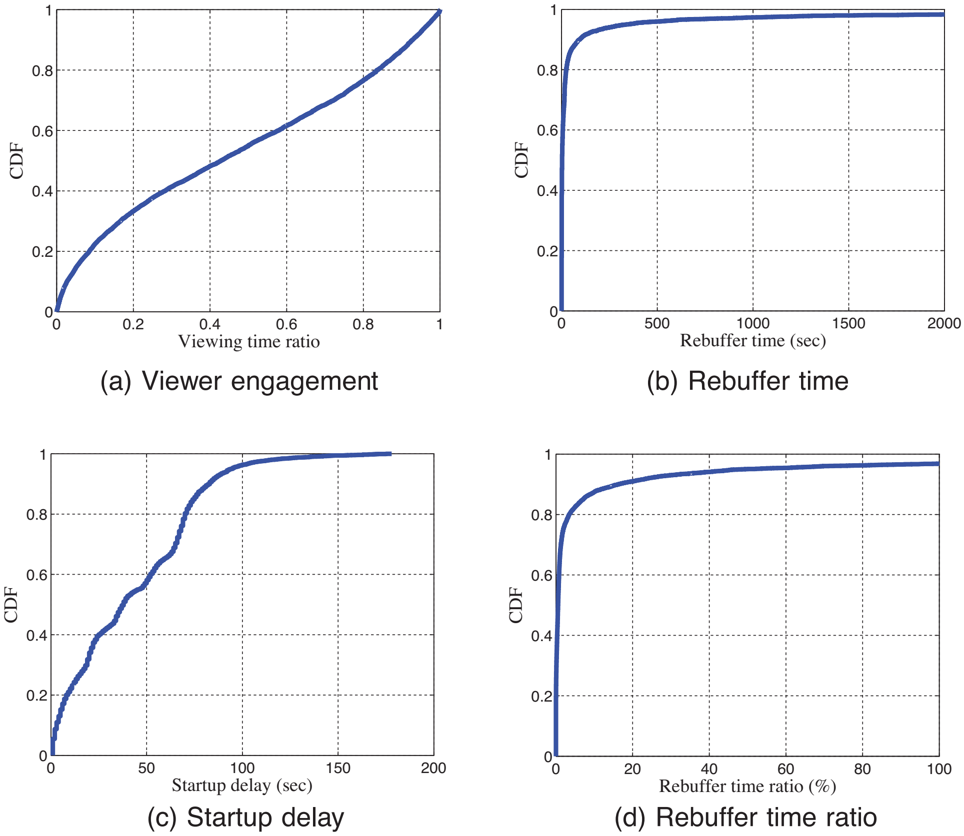


Fig. 5. CDF of viewer engagement, rebuffer time, startup delay and rebuffer time ratio.

which means that there is great room for improvement. We can see that more than 50% of the video sessions are shorter than half of the entire video; more than 20% of the video sessions are even shorter than 10% of the entire video; only top 20% of the video sessions are longer than 80% of the entire video, having satisfactory viewer engagement.

We further show the CDF of three service quality metrics: rebuffer time, startup delay and rebuffer time ratio. Fig. 5(b) shows that the distribution of rebuffer time has a long tail. While 90% of the video sessions have a rebuffer time of fewer than 50 s, the rest of the video sessions experience long rebuffer time ranging from 100 s to 2000 s. However, the rebuffer time ratio is relatively stable as shown in Fig. 5(d), i.e., longer videos have longer rebuffer time. Over 80% of the video sessions have a rebuffer time ratio of less than 3%. Fig. 5(c) shows that most of the startup delay is not long since it is only affected by initial network conditions[6]. More than 45% of the video sessions have a startup delay of fewer than 30 s.

4.1.1. Correlation analysis

We consider three types of correlation analysis to evaluate whether the relationship between viewer engagement and other factors is linear and monotonic or not.

- *Linearity* is characterized by Pearson linear correlation coefficient.
- *Monotonicity* is characterized by Spearman's rank correlation coefficient, where $-1/1$ means the viewer engagement can be represented as a monotonically decreasing/increasing function of a certain factor; 0 means that the viewer engagement and a certain factor are independent. One merit of the Spearman correlation coefficient is that, it does not require prior knowledge of the relationship (e.g., linear, logistic) between the viewer engagement and other factors (referred to as nonparametric).
- *Rank Similarity* is characterized by Kendall rank correlation coefficient, where $-1/1$ means that the ranks of viewer engagement perfectly disaccord/accord with the ranks of a certain factor; 0 means that the viewer engagement and a certain factor are independent. Kendall correlation is also nonparametric.

4.1.2. Information gain

Information gain helps to uncover non-monotonic relationships that cannot be revealed by correlation analysis. Information gain quantifies how the knowledge of a certain factor reduces the uncertainty of the viewer engagement. Let Y denote the viewer engagement, and X denote a certain factor. The (normalized) information gain for Y, given X, is $[I(Y) - I(Y|X)]/I(Y)$, in which $I(\cdot)$ is the entropy of the metric. The entropy represents how much information is

[6] Since our computation of the startup delay includes the advertisement time, it is longer than that in existing works such as [5].

Table 2
Correlation coefficient & information gain.

		Pearson	Spearman	Kendall	Information gain
Startup delay	all	0.001	−0.145	−0.095	1.8%
	skimmed	−0.01	−0.1	−0.06	1.5%
Rebuffer time	all	0.016	0.099	0.070	0.34%
	skimmed	0.000	0.014	0.01	0.35%
Rebuffer time ratio	all	0.013	0.20	0.14	–
	skimmed	0.01	0.11	0.08	–
Average data rate	all	0.011	0.006	0.004	0.45%
	skimmed	−0.008	−0.005	−0.003	0.42%
Data rate variance V_1	all	0.088	0.112	0.076	12.9%
	skimmed	0.07	0.08	0.05	11.8%
Data rate variance V_2	all	0.107	0.212	0.144	18.57%
	skimmed	0.06	0.11	0.07	16.94%
Data rate variance V_3	all	0.205	0.235	0.158	14.00%
	skimmed	0.12	0.13	0.09	12.22%
Signal strength	all	0.098	0.097	0.064	0.47%
	skimmed	0.06	0.05	0.03	0.23%
Video length	all	−0.153	−0.428	−0.285	7.73%
	skimmed	−0.30	−0.31	−0.20	4.90%
Video popularity	all	−0.116	−0.086	−0.056	0.42%
	skimmed	−0.07	−0.02	−0.01	0.20%

known of a random variable. It can be calculated as $I(X) = -\sum_i p(x_i) \ln p(x_i)$, in which $p(x_i)$ is the probability that the value of the random variable X is x_i. High information gain means that a certain factor has a significant impact on the viewer engagement.

4.1.3. Analysis results

Table 2 shows the results of correlation and information gain analysis both on the entire dataset and the skimmed dataset. For the skimmed dataset, we remove the viewers whose viewing time ratio is less than 5% [5]. The absolute values of correlation coefficient are extremely small, showing that no single factor has obvious linear or monotonic relationship with viewer engagement. The skimmed dataset strengthens the correlation for video length, and weakens the correlation for other factors. We also tried to skim the early quitters with other viewing time ratio thresholds, but on the whole, there is no significant change. The highest absolute value is the Spearman correlation coefficient between video length and viewer engagement. As video length increases, viewer engagement will decrease, because on the one hand, viewers get impatient as the video drags on; on the other hand, viewers are more likely to experience quality degradation during a long video session. Information gain is also low for all factors. It is interesting that, although average data rate has a low information gain, all the data rate variances have relatively high information gains and correlation coefficients. This implies that the raw data rate at the physical layer may not be an accurate indicator of the network throughput.

4.2. Gender difference

The subjective survey comparison of male and female viewers is summarized in Table 3[7]. In general, male and female viewers do not have much difference in general viewing behavior, such as the time spent on each video website visit. But male and female viewers do have different patient levels, which affect their tolerance for rebuffer time and startup delay.

Table 3
Subjective survey results: male vs female.

Days per week to visit video websites				
	1 ∼ 2	3 ∼ 4	5 ∼ 6	≈ 7
Male	12%	13%	8%	67%
Female	8%	33%	9%	50%
Time spent for each visit (min)				
	<10	10 ∼ 30	30 ∼ 60	>60
Male	17%	23%	28%	33%
Female	17%	15%	35%	33%
Average viewing time ratio (%)				
	<20	20 ∼ 60	60 ∼ 80	>80
Male	20%	18%	13%	49%
Female	13%	13%	11%	65%
Rebuffer time tolerance for ordinary videos (sec)				
	<30	30 ∼ 60	>60	
Male	53%	38%	9%	
Female	31%	53%	17%	
Rebuffer time tolerance for interesting videos (sec)				
	<30	30 ∼ 60	>60	
Male	34%	38%	28%	
Female	23%	45%	32%	
Startup delay tolerance (sec)				
	<5	5 ∼ 10	10 ∼ 30	>30
Male	28%	27%	31%	15%
Female	10%	21%	52%	18%
Self evaluation				
	Very impatient	Impatient	Moderate	Patient
Male	8%	36%	53%	3%
Female	3%	52%	39%	6%

Most of the viewers claim that they will watch a majority part of the video. This contradicts the result in Fig. 5(a), which shows that more than half of the video sessions have a viewing time ratio of less than 50%. The reason for this discrepancy between viewers' subjective perception and

[7] Since the number of viewers who chose "Patient" and "Very patient" are quite small, we combine the two results as "Patient".

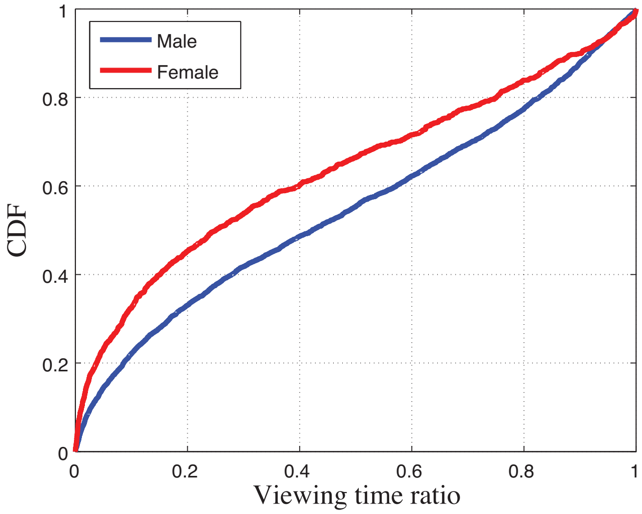


Fig. 6. CDF of viewer engagement for male and female viewers.

objective behavior may be that, most of the people will browse several videos before they dedicate to one long video session, but they may do so subconsciously without paying much attention.

Viewers' reported tolerance for startup delay and rebuffer time conforms the distribution of startup delay and rebuffer time in Fig. 5(a) and (b). It seems that the current video service quality can well satisfy viewers' demands. However, in fact, this may be the results of viewers' choice: viewers quit the video sessions that exceed their rebuffer time or startup delay tolerance, resulting in the rebuffer time and startup delay distribution in Fig. 5(a) and (b).

Fig. 6 shows the viewing time ratio of male and female viewers based on the dataset. Contradictory to the subjective survey, female viewers actually have a higher viewing time ratio than male viewers. Correlation analysis shows one significant change for female viewers: the Spearman correlation coefficient between video popularity and engagement becomes −0.20, compared with −0.086 in Table 2. This means that videos with high popularity actually have lower viewing time ratio. The possible reason is that, upon visiting the video website, most of the viewers will first "skim" several popular videos recommended by the website, then dedicate to one video that interests them. This may lead to many video sessions with small viewing time ratio for the popular videos. Female viewers may be more involved in such video browsing activities.

4.3. Grade difference

The subjective survey comparison of PG1 and PG2 viewers[8] shows two major differences[9]. First, PG2 viewers are more tolerant of rebuffer time and startup delay than PG1 viewers, even though more PG2 viewers claim to be impatient than PG1 viewers. Second, PG2 viewers claim to have a longer viewing time ratio than PG1 viewers. However, according to our dataset, the PG2 viewers actually have a lower viewing time ratio than the PG1 viewers as shown in Fig. 7. Such interesting contradictions between subjective perception and real engagement also exist for male and female viewers. Further investigations may be needed to seek for the psychological reason and verify whether the current viewer engagement metrics indeed reflect real subjective viewer experience.

Fig. 7 shows that as the year of study increases, viewer engagement decreases, probably due to more work pressure. Correlation analysis shows that, like female viewers, video popularity has a more significant influence on PG2 viewers (with Spearman correlation coefficient as −0.25) but not PG1 and UG4 viewers.

4.4. Video length

As video length has a relatively high influence on viewer engagement according to Table 2, in this section, we compute correlation coefficients for videos with different lengths. We classify videos as "Short" (<10 min), "Mediate" (≥10 but <30 min), and "Long" videos (≥30 min). Fig. 8(a) shows a significant decrease in viewer engagement when the video length increases. The correlation analysis shows that:

- Rebuffer time becomes increasingly influential for long videos. The Spearman and Kendall coefficients become 0.34 and 0.25 for absolute rebuffer time; 0.36 and 0.27 for rebuffer time ratio. Surprisingly, the coefficients are positive, which means that the longer the rebuffer time is, the higher the viewer engagement. One possible reason is that the streaming video with a higher bit rate has longer rebuffer time [5], and the viewer engagement is high due to better video quality.
- Data rate variances become increasingly influential, especially for mediate and long videos. For mediate videos, the Spearman and Kendall coefficients for V_2 are 0.46 and 0.33. For long video, the Spearman and Kendall coefficients for V_2 are as high as 0.54 and 0.38.

[8] We do not compare the results of UG4 viewers since our subjective survey covered only a small number of UG4 viewers.

[9] Due to page limit, we do not show the detailed table here.

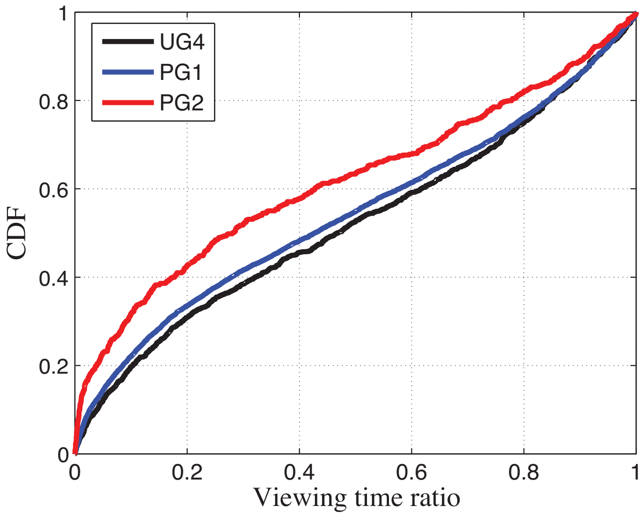


Fig. 7. CDF of viewer engagement for UG4, PG1 and PG2 viewers.

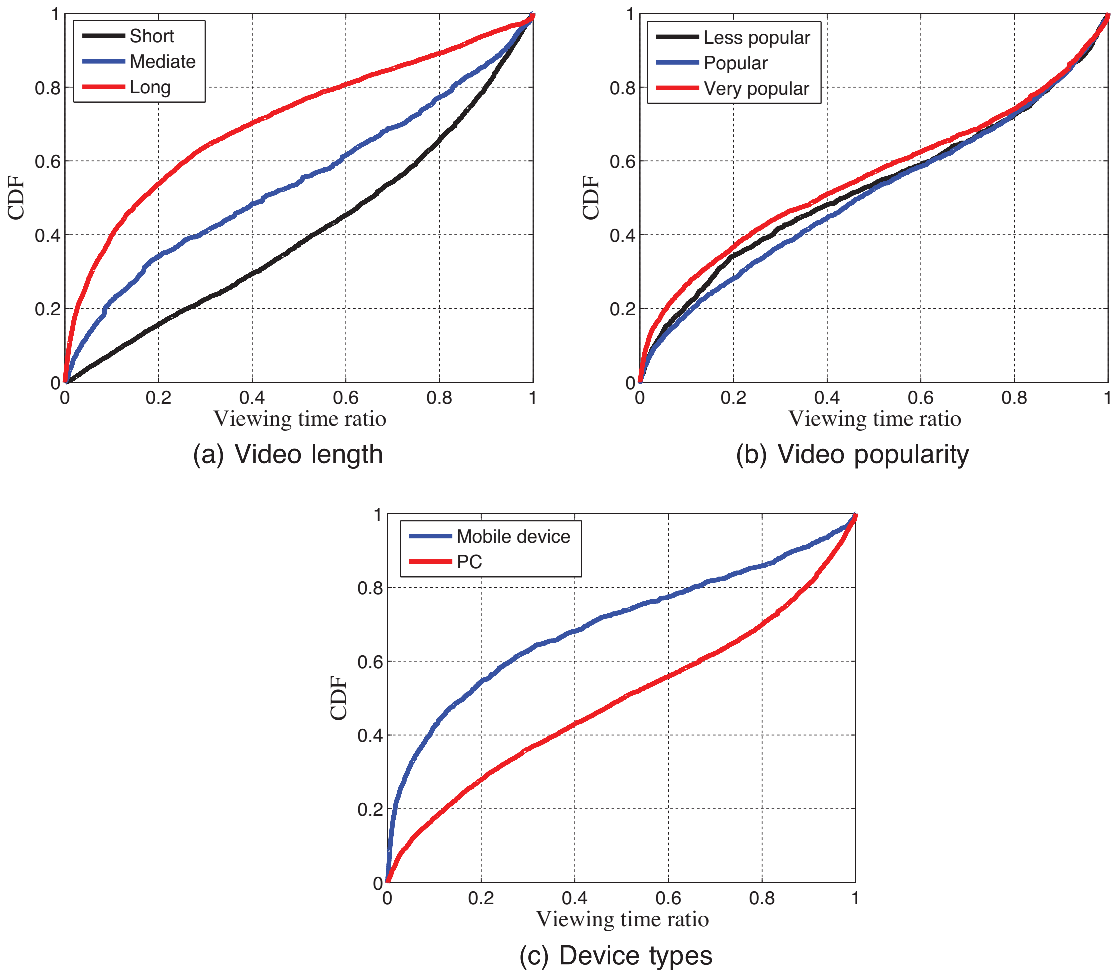
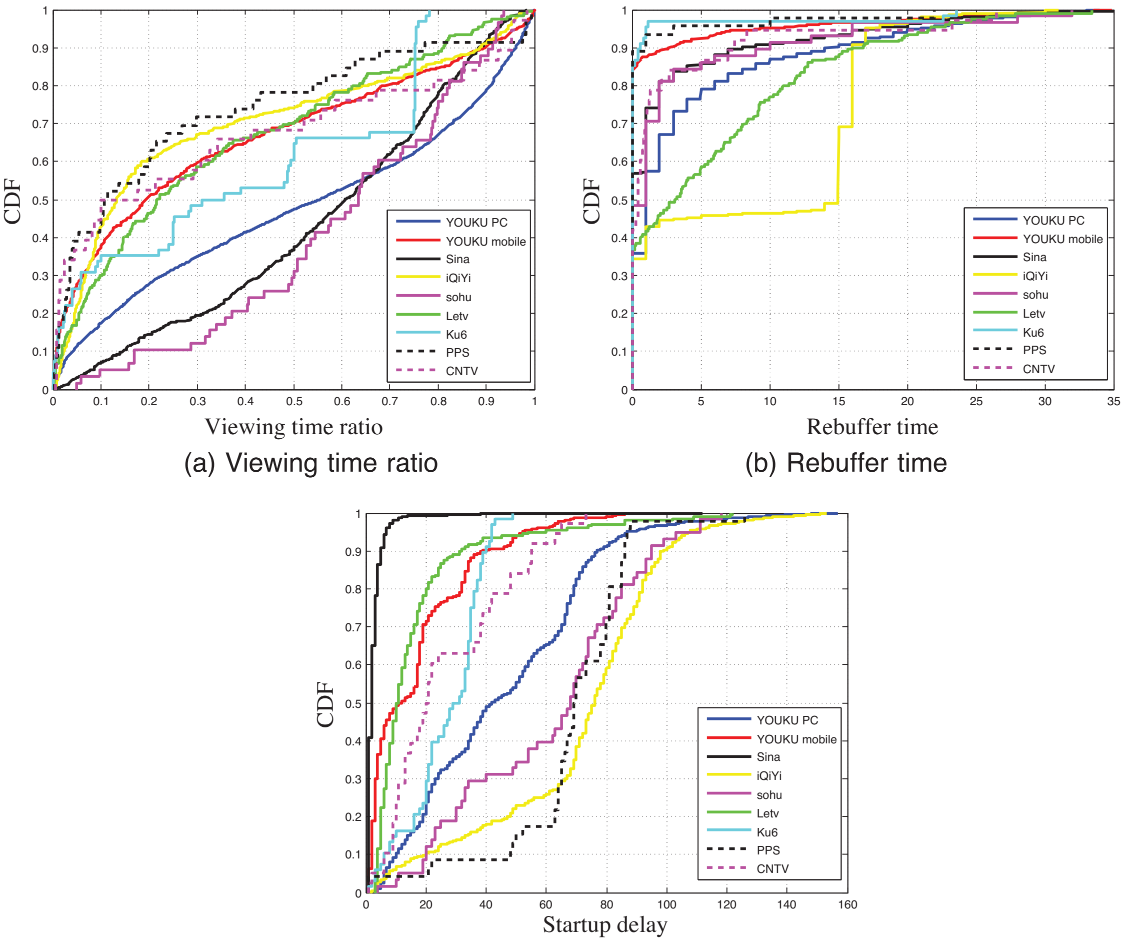


Fig. 8. CDF of viewer engagement for different video types and device types.



(a) Viewing time ratio (b) Rebuffer time (c) Startup delay

Fig. 9. Different video websites.

4.5. Video popularity

We classify videos as “Less popular” ($< 10^5$), “Popular” ($\geq 10^5$ but $< 10^6$) and “Very popular” ($\geq 10^6$). It is shown in Fig. 8(b) that the most popular videos have the lowest viewer engagement. This partly explains the negative correlation between video popularity and viewer engagement. But the discrepancy in viewer engagement for videos with different popularity is quite small. Therefore, video popularity may only affect certain viewer groups, such as female viewers and PG2 viewers.

4.6. Device types

Fig. 8(c) shows that viewer engagement on mobile devices is much lower than that on PC, indicating that viewing experience on mobile devices needs to be improved. The correlation analysis shows that, for mobile devices, the influence of average data rate becomes more significant, but the correlation coefficients are negative (the Pearson and Spearman coefficients both become -0.28). This is surprising as we expect that high average data rate will lead to high throughput, and therefore high viewer engagement. One possible explanation is that, for the mobile devices, the channel estimation (for choosing appropriate data rate) is adversely influenced by the fast-changing channel condition due to high mobility. In this case, high (raw) data rate may instead result in low throughput, which leads to lower viewer engagement.

4.7. Video website

Fig. 9 shows the comparison of viewing experience in different video websites. We can see that the difference in QoE, rebuffer time and startup delay is significant. However, the relationship between QoE and QoS metrics are complicated. For example, “PPS” website has high QoE, low rebuffer time, but its startup delay is high; “iQiYi” website also has high QoE, but its rebuffer time and startup delay are both high. The correlation results for most websites are similar to Table 2. However, for “Ku6” and “iQiYi”, the Spearman correlation coefficient between QoE and video length are as significant as -0.75 and -0.63, respectively; for “CNTV”, the Pearson and Spearman correlation coefficients between QoE and data variance V_3 are as high as 0.67 and 0.69, respectively. The

Table 4
Confusion table.

		Predicted class									
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
True class	C_1	**369**	10	5	7	16	5	9	5	8	15
	C_2	12	**227**	4	4	9	6	7	7	3	13
	C_3	13	6	**146**	2	9	5	0	10	8	11
	C_4	15	4	4	**133**	5	5	6	6	8	6
	C_5	9	3	3	3	**132**	4	2	5	3	8
	C_6	8	4	5	4	2	**112**	2	10	6	7
	C_7	7	4	5	5	12	6	**110**	8	5	10
	C_8	13	1	6	5	8	6	6	**169**	9	13
	C_9	8	8	3	4	12	1	5	14	**261**	20
	C_{10}	21	14	5	10	15	11	13	19	17	**497**

difference between various video websites may due to their customized services. Fig. 9(a) indicates that websites with poor QoE have great room for improvement compared with websites with satisfactory QoE.

5. Significance of influence of different factors on viewer engagement

In this section, we first build a Bayesian classification model, in which we use posterior probability to represent the dependency of viewer engagement on a certain factor. Then, we analyze whether a certain factor is significant in determining the class of viewer engagement.

5.1. Bayesian classification model

Let Y denote the viewer engagement, $X = (x_1, x_2, \dots, x_n)$ denote the vector of potential factors that may influence viewer engagement, referred to as viewing condition. $P(Y)$ is the prior probability of Y. For example, $P(Y = 0.1)$ is the probability that any viewer's engagement is 0.1. $P(X)$ is the prior probability of X. For example, $P(x_{startup} = 10\text{ s}, x_{rebuffer} = 20\text{ s}, \dots, x_{popularity} = 10^6)$ is the probability that a video session has a startup delay of 10 s, a rebuffer time of 20 s, ..., and a popularity of 10^6. $P(X|Y)$ is the posterior probability of X conditioned on Y. That is, it is the probability that a video session belongs to viewing condition X, given that we know the viewer engagement of this video session is Y. The major goal of Bayesian classifier is to find $P(Y|X)$, the posterior probability that the viewer engagement is Y, given that the viewing condition is X. According to the Bayes' theorem, we have:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}. \tag{3}$$

We divide viewer engagement into 10 classes: $C_1, C_2, \dots, C_{10}$ denote $[0, 10\%), [10, 20\%), \dots, [90, 100\%]$, respectively. Given viewing condition X, the posterior probabilities for each class $P(C_i|X)$ are calculated. We predict that X belongs to C_i if and only if the posterior probability $P(C_i|X)$ is the highest among all possible classes, that is, $P(C_i|X) > P(C_j|X), \forall j \in [1, 10], j \neq i$.

In (3), $P(X)$ is the same for all classes, so we only have to derive $P(X|C_i)$ and $P(C_i)$ for each class. If the class prior probabilities are unknown, the common practice is to assume that $P(C_1) = P(C_2) = \dots = P(C_{10})$ [22]. In fact, we find that this yields the best prediction accuracy for our dataset. To differentiate the influence of individual factors on viewer engagement, we make the simplified assumption that different factors are independent[10]:

$$P(X|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \cdots P(x_n|C_i), \forall C_i. \tag{4}$$

Table 4 shows the confusion table of the prediction results. The prediction accuracy is 76.1%.

5.2. Significance of factors in determining viewer engagement

In this section, we ask such a question: is factor x_j significant in determining the class of viewer engagement? According to (4), the contribution of factor x_j to $P(C_i|X)$ is $P(C_i|x_j)$. If $P(C_i|x_j)$ is very dissimilar for different classes C_i, $i \in [1, 10]$, then x_j will be important in determining which class the viewer engagement belongs to. On the contrary, if $P(C_i|x_j)$ is almost the same for different classes C_i, $i \in [1, 10]$, then x_j cannot differentiate which class the viewer engagement belongs to.

To characterize the significance of factor x_j in determining the viewer engagement class, we define the following significance indicator SD_j:

$$SD_j = \frac{\sum_{m=1}^{N-1} \sum_{n=m+1}^{N} |P(C_m|x_j) - P(C_n|x_j)|}{N(N-1)/2 \times \max_k P(C_k|x_j)}. \tag{5}$$

in which $N = 10$ is the number of all classes, and $N(N-1)/2$ is the number of sums in the numerator. Fig. 10 shows the significance indicator of different factors[11].

- The peak in Fig. 10 indicates that a factor has a significant influence on viewer engagement. Some factors may have significant influence at a single point, such as rebuffer time and data rate variance, as shown in Fig. 10(b) and (c). Other factors may have significant influence at multiple ranges. For example, startup delay affects viewer engagement at $90 \sim 110$ s and $145 \sim 160$ s as shown in Fig. 10(a).

[10] This simplified assumption cannot capture the interdependency among different factors. In the future, we will adopt more complicated models such as Bayesian belief networks to address this problem.

[11] Startup delay, rebuffer time, average data rate and data rate variance are continuous; video length and video popularity are discrete.

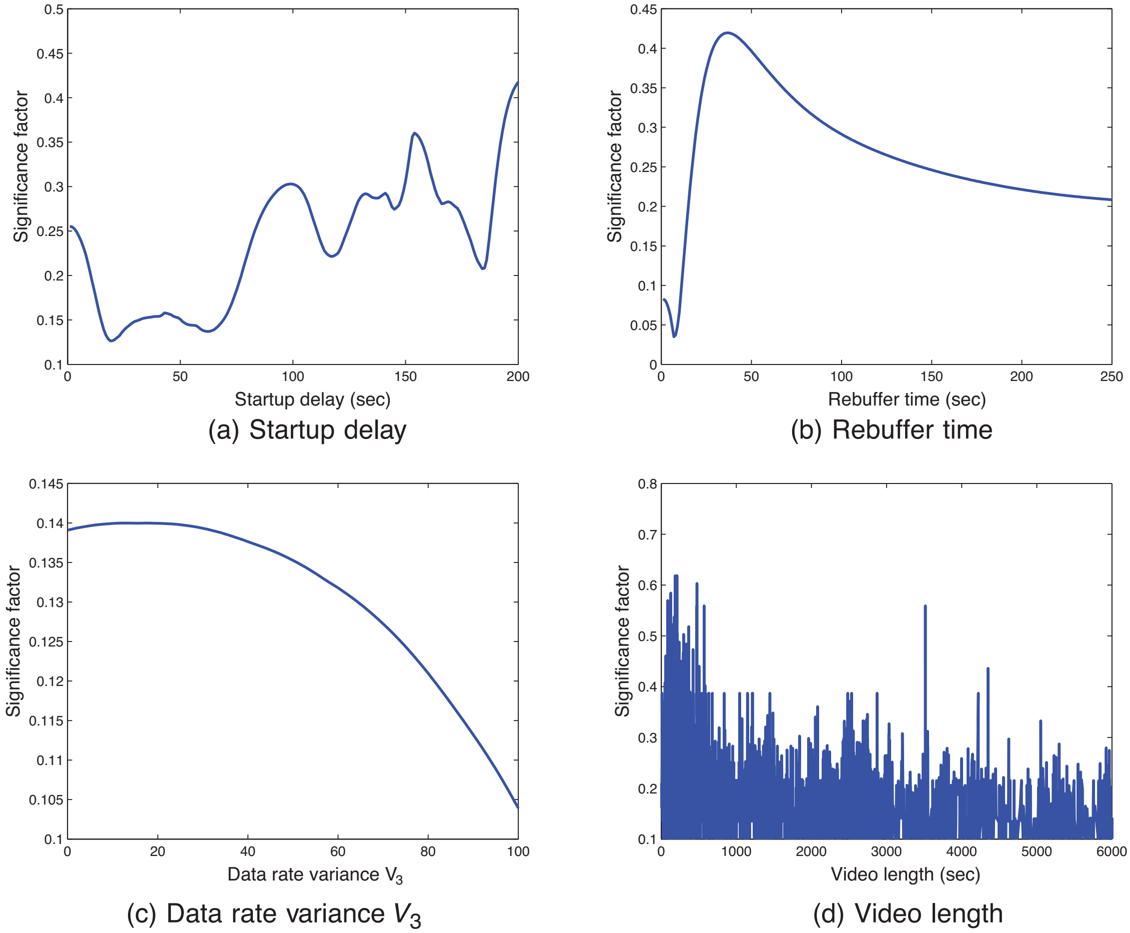


Fig. 10. Significance of factors in determining viewer engagement.

- The most significant factor in determining viewer engagement is video length (with the highest significance indicator), especially for short videos, as shown in Fig. 10(d). The least significant factor in determining viewer engagement is data rate variance V_3, as shown in Fig. 10(c), though it has relatively high correlation coefficients.
- The significance of a certain factor is different at different ranges. For example, as shown in Fig. 10(b) the rebuffer time is highly influential at around 40 s. But when the rebuffer time exceeds 100 s, it has little influence on determining the viewer engagement (threshold effect). Similar trend can be found in other factors.

6. Viewer engagement prediction

In the previous section, we classify viewer engagement at a coarse-grained level. In this section, we build more fine-grained models to predict viewer engagement based on various factors. The prediction models can help network operators and video service providers to monitor viewer engagement and make corresponding adjustment to improve viewer engagement.

6.1. General prediction model

To model the non-linear and non-monotonic relationship between viewer engagement and other factors, we select two candidate models: Classification and Regression Tree (CART) and neural network (NN)[12]. In CART, we use 10-fold cross-validation method: the data are partitioned into 10 groups; each time, 9 groups are used for training, and 1 group is used for validation. In NN, we use 70% of the data for training, 15% for validation, and 15% for testing. We train the NN for 100 times and average the results. Fig. 11 shows the performance of the two models for all viewers and different viewer groups[13]. Skimming the viewers will improve the prediction results by as high as 16% for CART, and 26% for NN. In comparison, NN has a poorer performance, probably due to its “black box” nature [23], which makes it difficult to interpret the

[12] We have done pilot experiments to try other algorithms, and find that CART and NN work the best for our dataset.

[13] Compared with existing works [5,6], the RMSE in Fig. 11 is relatively high. This is because our dataset has a much smaller scale, and the data for training the model is quite limited. However, in this paper, we focus on building personalized model to improve prediction accuracy for viewer engagement.

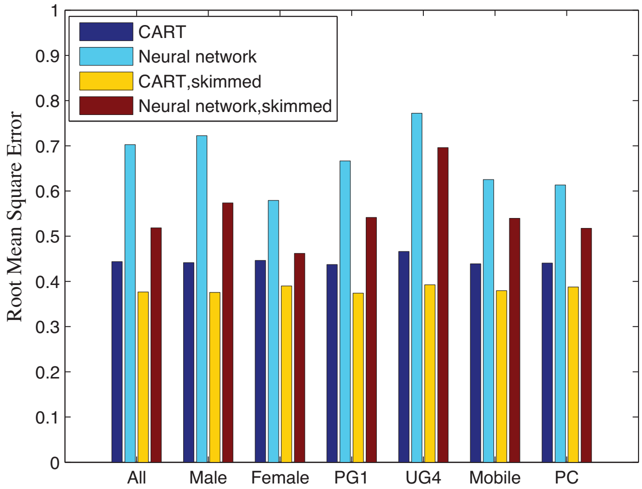


Fig. 11. RMSE of general prediction model.

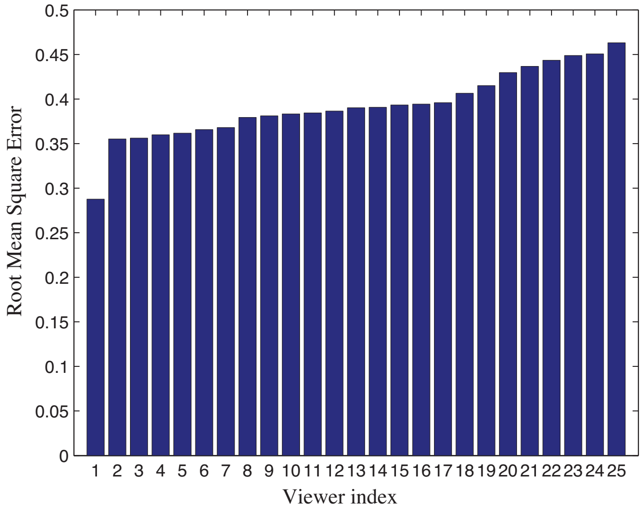


Fig. 12. RMSE of personalized prediction model.

relationship between inputs and outputs. Therefore, we choose CART as the basis for building personalized viewer engagement prediction model.

6.2. Personalized prediction model

Viewer engagement is closely related to personal disposition and perception. Therefore, the best way to predict a viewer's engagement is to build a personalized model based on his/her own historical data. We choose the top 25 viewers with the most video sessions, and build personalized prediction models with each viewer's individual data[14]. We sort the RSME and present the results in Fig. 12. The average RMSE is 0.39, lower than that of the general prediction model (0.45). The personalized model works especially well for some viewers (the lowest RMSE is 0.28). Although some viewers still have high RMSE (the maximum RMSE is 0.49) mainly due to limited training data, we believe that with enough data, the personalized model will significantly outperform the general model.

6.3. Customized prediction model

While the personalized prediction model is the most ideal choice, it suffers from insufficient training data when a new viewer comes (*cold start problem*). Therefore, we attempt to form a customized prediction model based on a group of viewers who share similar attributes, e.g., female PG1 student. If successful, this model can be used as an initial model for a new user, bootstrapping the training process.

We group the viewers according to their gender ("male" or "female") and grade ("PG1", or "UG4"). Each time, we remove a viewer from his/her group, and train a customized prediction model on the rest of the viewers. Then we use this model to predict the engagement of the previously removed

[14] For personalized models, we use 3-fold cross validation due to data size limitation.

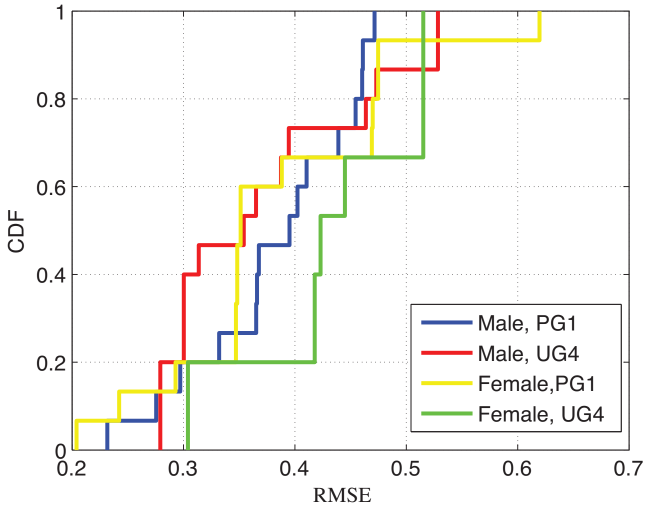


Fig. 13. RMSE of customized prediction model.

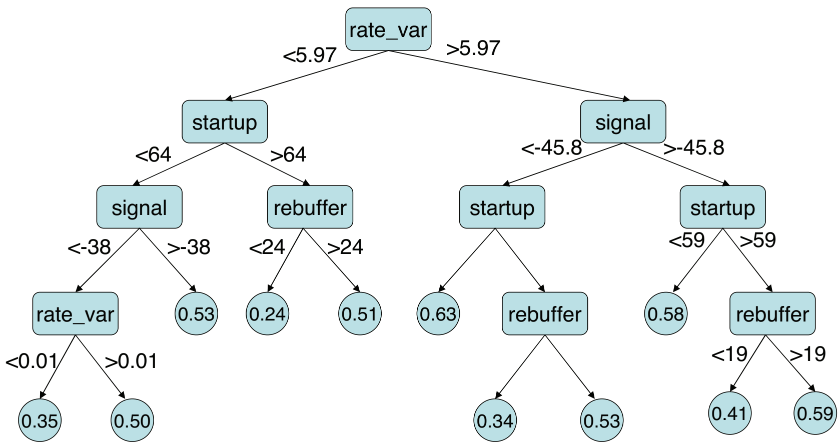


Fig. 14. Pruned CART tree.

viewer, and calculate the RMSE. We select 100 viewers from each group.

Fig. 13 shows the CDF of the RMSE of the customized prediction model. The customized model works surprisingly well for some viewers, with RMSE as low as 0.2. However, for many other viewers, the RMSE is relatively high. We recommend that the customized model be used as the initial model for a new viewer, and be replaced by personalized model as soon as enough data are accumulated for the viewer.

7. Discussion

Viewer engagement vs viewer experience: Viewer experience is subjective, depending not only on video quality, but also on individual viewer's characteristics and viewing environment. Viewer engagement, while providing an accessible way to quantify viewer satisfaction (the more engaged a viewer is, the better the viewer experience is), can only partially reflect viewer experience. The relationship between viewer engagement metrics (e.g., viewing time ratio) and real viewer experience needs to be verified. In addition, we may further explore metrics that can better represent viewer experience, e.g., metrics that involve viewer interactivity with the video system (such as pause, fast forward and mouse click).

Dynamic/online prediction model: Current prediction models are built offline, based on full information of the video session. However, real-time viewer engagement control requires decision-making based on partial information. For example, during the video session, instead of average data rate and total rebuffer time of the entire session, we only know the average data rate and total rebuffer time up till the present moment. The interaction between real-time control and final viewer engagement is uncertain. Another problem is that we don't know whether the influence of a certain factor is temporary or long-lasting. For example, although startup delay may initially affect viewer engagement, its influence may attenuate once the video starts playing.

Actionable insights: Video service providers and network operators can leverage the analysis results for improving viewer engagement with limited resource in the most effective way. For example, when the rebuffer time is within the range [30, 50 s], the video service provider knows that the viewer engagement is largely influenced by rebuffer time, and may prioritize certain measures regarding rebuffer time. We show the pruned CART tree on the entire dataset in Fig. 14. The network operator or the video service provider can leverage this tree to improve viewer engagement. For example, data rate variance is on the root note, showing its great influence on viewer engagement. The path from the root to the leaf node shows one possible way to reach a target viewer engagement level. For example, if it is achieved that the data rate variance is less than 5.97, the startup delay is less than 64 s, and the signal strength is larger than −38 db, the viewer engagement is estimated to be 0.53.

8. Conclusion

With the ever-increasing video traffic over the Wi-Fi networks, a good understanding of viewer engagement is essential for video service providers and network operators. In this paper, we build an experimental platform to collect video information on service quality metrics, network quality metrics, video content and viewer demography. We analyze the influence of each factor on viewer engagement, first through the entire dataset, then for different viewer groups and video types. Then, we analyze the significance of different factors in determining viewer engagement. Finally, we propose to build personalized models to predict viewer engagement for individual viewers, which have a better performance than general prediction models.

Acknowledgment

This work was supported by grants from 973 project 2013CB329006, RGC under the contracts CERG 622613, 16212714, HKUST6/CRF/12R, M-HKUST609/13, the grant from Huawei-HKUST joint lab, NSFC under Grant No. 61322104, Program for New Century Excellent Talents in University (NCET-13-0908), Guangdong Natural Science Funds for Distinguished Young Scholar (No. S20120011468), the Shenzhen Science and Technology Foundation (Grant No. JCYJ20140509172719309) and China NSFC Grant 61472259.

References

[1] Cisco visual networking index: forecast and methodology, 2012-2017, May 29, 2013.
[2] Cisco visual networking index: global mobile data traffic forecast update, 2013-2018, February 5, 2014.
[3] The state of the CDN market: video pricing, contract, volume and market sizing trends, http://blog.streamingmedia.com/wp-content/uploads/2015/08/2015CDNSummit-Rayburn-Pricing.pdf (Accessed September 3, 2015).
[4] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D.A. Joseph, A. Ganjam, J. Zhan, H. Zhang, Understanding the impact of video quality on user engagement, SIGCOMM-Comput. Commun. Rev. 41 (4) (2011) 362.
[5] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, H. Zhang, Developing a predictive model of quality of experience for internet video, in: Proceedings of the ACM SIGCOMM Conference, 2013, pp. 339–350.
[6] M.Z. Shafiq, J. Erman, L. Ji, A.X. Liu, J. Pang, J. Wang, Understanding the impact of network dynamics on mobile video user engagement, in: Proceedings of the ACM SIGMETRICS, 2014.
[7] M. Vilas, X.G. Pañeda, R. García, D. Melendi, V.G. García, User behavior analysis of a video-on-demand service with a wide variety of subjects and lengths, in: Proceedings of the IEEE EUROMICRO Conference on Software Engineering and Advanced Applications, 2005, pp. 330–337.
[8] L. Chen, Y. Zhou, D.M. Chiu, Video browsing-a study of user behavior in online vod services, in: Proceedings of the IEEE International Conference on Computer Communications and Networks (ICCCN), 2013, pp. 1–7.
[9] H. Yin, X. Liu, F. Qiu, N. Xia, C. Lin, H. Zhang, V. Sekar, G. Min, Inside the bird's nest: measurements of large-scale live vod from the 2008 olympics, in: Proceedings of the ACM Internet Measurement Conference (IMC), 2009, pp. 442–455.
[10] P. Gill, M. Arlitt, Z. Li, A. Mahanti, Youtube traffic characterization: a view from the edge, in: Proceedings of the ACM Internet Measurement Conference (IMC), 2007, pp. 15–28.
[11] A. Finamore, M. Mellia, M.M. Munafò, R. Torres, S.G. Rao, Youtube everywhere: Impact of device and infrastructure synergies on user experience, in: ACM Internet Measurement Conference (IMC), 2011, pp. 345–360.
[12] A. Balachandran, V. Sekar, A. Akella, S. Seshan, Analyzing the potential benefits of CDN augmentation strategies for internet video workloads., in: Proceedings of the ACM Internet Measurement Conference (IMC), 2013, pp. 43–56.
[13] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, L. Martens, Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching, IEEE Trans. Broadcast. 59 (1) (2013) 47–61.
[14] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, H. Zhang, A quest for an internet video quality-of-experience metric, in: Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets), 2012, pp. 97–102.
[15] S.S. Krishnan, R.K. Sitaraman, Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs, in: Proceedings of the ACM Internet Measurement Conference (IMC), 2012, pp. 211–224.
[16] OpenWrt, https://openwrt.org/ (Accessed September 3, 2015).
[17] TCPDUMP, http://www.tcpdump.org/ (Accessed September 3, 2015).
[18] iw, http://wireless.kernel.org/en/users/Documentation/iw (Accessed September 3, 2015).
[19] S. Shepler, M. Eisler, D. Robinson, B. Callaghan, R. Thurlow, D. Noveck, C. Beame, Network file system (NFS) version 4 protocol, Network (2003).
[20] TSTAT, http://tstat.polito.it/index.shtml (Accessed September 3, 2015).
[21] Alexa, http://www.alexa.com/ (Accessed September 3, 2015).
[22] J. Han, M. Kamber, Data Mining, Southeast Asia Edition: Concepts and Techniques, Morgan Kaufmann, 2006.
[23] J.V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, J. Clin. Epidemiol. 49 (11) (1996) 1225–1231.

YANJIAO CHEN received her B.E. degree of electronic engineering from Tsinghua University in 2010 and Ph.D. degree of computer science and engineering in Hong Kong University of Science and Technology in 2015. She is currently a Postdoctoral fellow in the University of Toronto. Her research interests include spectrum management for Femtocell networks, network economics and Quality of Experience (QoE) of multimedia delivery/distribution.

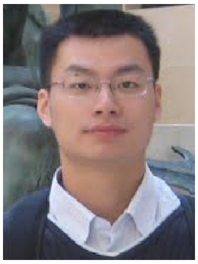
QIHONG CHEN received his MPhil degree of computer science and engineering from Hong Kong University of Science and Technology in 2014. His research interest is on Quality of Experience (QoE) of multimedia delivery/distribution.

FAN ZHANG is currently a MPhil candidate in Hong Kong University of Science and Technology. His research interest is on Quality of Experience (QoE) of multimedia delivery/distribution.

QIAN ZHANG joined Hong Kong University of Science and Technology in September 2005 where she is a full Professor in the Department of Computer Science and Engineering. Before that, she was in Microsoft Research Asia, Beijing, from July 1999, where she was the research manager of the Wireless and Networking Group. Dr. Zhang has published about 300 refereed papers in international leading journals and key conferences in the areas of wireless/Internet multimedia networking, wireless communications and networking, wireless sensor networks, and overlay networking. She is a Fellow of IEEE for "contribution to the mobility and spectrum management of wireless networks and mobile communications". Dr. Zhang has received MIT TR100 (MIT Technology Review) world's top young innovator award. She also received the Best Asia Pacific (AP) Young Researcher Award elected by IEEE Communication Society in year 2004. Her current research is on cognitive and cooperative networks, dynamic spectrum access and management, as well as wireless sensor networks. Dr. Zhang received the B.S., M.S., and Ph.D. degrees from Wuhan University, China, in 1994, 1996, and 1999, respectively, all in computer science.

KAISHUN WU is currently a research assistant professor in Fok Ying Tung Graduate School with the Hong Kong University of Science and Technology (HKUST). He received the Ph.D. degree in computer science and engineering from HKUST in 2011. His research interests include wireless communication, mobile computing, wireless sensor networks and data center networks.

RUOCHEN HUANG is currently a Ph.D. candidate in Nanjing University of Posts and Telecommunications. His research interest is on Quality of Experience (QoE) of multimedia delivery/distribution.



LIANG ZHOU received his Ph.D. degree major at Electronic Engineering from Ecole Normale Suprieure (E.N.S.), Cachan, France, and Shanghai Jiao Tong University, Shanghai, China, respectively in 2009. From 2009 to 2010, he was a postdoctoral researcher in ENSTA-ParisTech, Paris, France. From October 2010 to September 2011, he was a Humboldt Research Fellow in Technical University of Munich, Munich, Germany. In Sep. 2011, he joined in Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests are in the area of network-based multimedia communications and computing, in particular, resource allocation, system scheduling, cross-layer design, multimedia security, multimedia signal processing. He currently serves as an editor for IET Communications, Wireless Communications and Mobile Computing, International Journal of Ad Hoc and Ubiquitous Computing, KSII Transactions on Internet and Information Systems, and guest editor for ACM/Springer Multimedia Systems Journal, and EURASIP Journal of Wireless Communications and Networking, etc. He also serves as Co-Chair and Technical Program Committee (TPC) member for a number of international conferences and workshops (e.g., IEEE GLOBECOM'10-12, IEEE ICC'10-12 etc.). He is a member of IEEE, IEEE ComSoc, IEEE Multimedia Signal Processing Technical Committee, and IEEE Multimedia Communications Technical Committee.