Personalized User Engagement Modeling for Mobile Videos

Lin Yang[a], Mingxuan Yuan[b], Yanjiao Chen[c,*], Wei Wang[d], Qian Zhang[a], Jia Zeng[b]

[a]*Department of Computer Science and Engineering Hong Kong University of Science and Technology, Hong Kong*
[b]*Huawei Noah's Ark Lab, Hong Kong*
[c]*State Key Lab of Software Engineering, Wuhan University*
[d]*School of Electronic Information and Communications Huazhong University of Science and Technology*

Abstract

The ever-increasing mobile video services and users' demand for better video quality have boosted research into the video Quality-of-Experience. Recently, the concept of Quality-of-Experience has evolved to *Quality-of-Engagement*, a more actionable metric to evaluate users' engagement to the video services and directly relate to the service providers' revenue model. Existing works on user engagement mostly adopt uniform models to quantify the engagement level of all users, overlooking the essential distinction of individual users. In this paper, we first conduct a large-scale measurement study on a real-world data set to demonstrate the dramatic discrepancy in user engagement, which implies that a uniform model is not expressive enough to characterize the distinctive engagement pattern of each user. To address this problem, we propose PE, a personalized user engagement model for mobile videos, which, for the first time, addresses the user diversity in the engagement modeling. Evaluation results on a real-world data set show that our system significantly outperforms the uniform engagement models, with a 19.14% performance gain.

Keywords: User Engagement, User Modeling, Mobile Video

*Corresponding author. Tel.: +86 (027) 68773612
Email address: chenyanjiao@whu.edu.cn (Yanjiao Chen)

1. Introduction

The increasing prevalence of mobile devices has triggered an exponential growth in mobile video services. It is estimated that, by the end of 2018, mobile video will account for over two-thirds of the world's mobile data traffic [**?**]. In the wake of the development of screen size and computation power of mobile devices, users have a higher demand on the viewing experience. To cater for such needs, it is essential to accurately assess video quality.

The assessment of video quality has been widely studied by the multimedia community for a long time. Pioneer researchers have tried to quantify and improve users' viewing experience by optimizing quality-of-service (QoS) parameters [1, 2, 3, 4]. Although such QoS parameters are objective and easy to measure, their relationships to users' viewing experience are hard to quantify. To evaluate video viewing experience from the user's perspective, the concept of Quality-of-Experience has been proposed. A plethora of works try to solicit users' opinion evaluation score by conducting subjective tests [5, 6, 7, 8]. However, such subjective tests inevitably involve lots of human participation, thus are often in small scale due to the high cost.

In recent years, the concept of Quality-of-Experience has involved to *Quality-of-Engagement*. The user engagement, compared with the subjective and hard-to-measure user perceptual experience, is a more actionable metric to evaluate user's satisfaction with the video service and directly related to the service providers' revenue model [9]. As various parties are involved in the video service ecosystem, the user engagement can be evaluated from different angles. As a pioneer, Dobrian *et al.* collected a large-scale data set via client-side instrumentation and investigated how the video quality parameters affect the user engagement from the content provider's perspective [10]. The authors in [11, 12] developed a decision-tree-based engagement model to quantify the relationship between video-delivering QoS parameters and user engagement, which can help the design of content providers. Also, the authors in [13] examined the causal relationship between video quality and viewer behavior from the perspective

of content delivery network (CDN) owner, while another study utilized massive network-provider-side data to measure the impact of network dynamics on users' engagement in mobile video services [14]. Generally, these works leverage the power of machine learning and big data to reveal the complicated relationship between user engagement and confounding factors. Nevertheless, all of the existing models are built upon the entire user data set, averaging the effect of confounding factors on all users. When applied to individual users, such a uniform model may fail to characterize the distinctive patterns of personal user engagement.

To investigate users' differences in their engagement patterns, we collect a large-scale video streaming data set from the core network of a tier-1 cellular network in China. We first study the impact of the downlink throughput, which is an important factor from the perspective of network provider, on the user engagement in mobile video services. The result indicates that the same factor may have distinctive effects on different users. To further investigate the effect of user diversity on engagement modeling, we employ a wildly-used machine learning algorithm, *i.e.*, gradient boosted regression tree (GBRT) [15] to build a uniform user engagement model with data of all users, and individual user engagement models for selected users. Comparing individual models with the uniform model, we find that the model parameters of a specific user are considerably different from those of other users, as well as the uniform model. This implies that a uniform model is insufficient to comprehensively characterize the engagement level of individual users.

To gain a more accurate and fine-grained insight into user engagement, we need a personalized user engagement model which can comprehensively capture the user diversity. To achieve such a goal, there are several challenges: (1) The data set consists of millions of users, and building personalized engagement models for such a large user population is quite difficult. (2) The number of videos watched by each user is rather small compared with the total number of videos in the data set, insulting in a highly sparse viewing record, which makes it hard to build accurate models for each user. (3) While soliciting

information from accessory data sources is a potential solution to the sparsity problem, seamless integration of the information from various data sources is a non-trivial problem.

To tackle the above challenges, in this paper, we propose PE, a personalized quality of user engagement model for mobile videos from the perspective of mobile network provider, which takes user diversity into account and thus can provide a more accurate and fine-grained modeling. PE collaboratively learns the individual model for each user via matrix factorization and exploits the side information from other data sources to alleviate the data sparsity problem. The evaluations on a real-world data set show that PE significantly outperforms state-of-the-art user engagement models with a 19.14% performance gain.

With our system, mobile network providers can gain a more accurate understanding of user engagement with their services. Such knowledge can help them better invest network resources and perform case-by-case optimization [9]. Moreover, though our current implementation serves the need of mobile network providers, PE can easily be extended to meet the requirement of other service providers, *e.g.*, video content provider and CDN owner.

Our key contributions lie in three aspects:

- Our experiment on a large-scale video streaming data set demonstrates a significant user diversity in user engagement, which implies that the uniform model is insufficient for accurate engagement modeling.

- To the best of our knowledge, we are the first to propose a personalized user engagement model for mobile videos from the perspective of mobile network operators. This model can comprehensively capture the dramatic user diversity and provide a more accurate assessment of user engagement.

- We collect a massive video-related data set from a tier-1 network operator in China and perform a thorough evaluation of our system. The experiment results indicate our system can bring a 19.14% performance gain with respect to state-of-the-art user engagement models.

The rest of this paper is organized as follows. Section 2 reviews the related work and Section 4 defines the problem scope and validates the user diversity on a real-world data set. Section 5 formulates the problem and introduces the architecture of our system and Section 6 discusses the design of our personalized user engagement model. The evaluation results are reported in Section 7. Several piratical issues and future exploration are discussed in Section 8, followed by a conclusion in Section 9.

2. Related Work

Video quality assessment has long been studied in academia. Early works on this area mainly focus on objective QoS metrics, *e.g.*, video encoding rate [1, 16], bitrate [17, 18] or network bandwidth [19, 20], and try to improve user's experience by better QoS provision. However, as the video service is highly user-centric, the practical improvement brought by these works is hard to be validated [9]. To evaluate video quality from the user perspective, many researchers have started to evaluate video quality-of-experience via subjective tests in a controlled environment [5, 6, 9]. The high cost and human participation in subjective tests are inevitable for such works and thus limits the scale of their experiments.

In recent years, the concept of Quality-of-Experience has evolved to the Quality-of-Engagement. The data-driven user engagement analysis for video services has been boosted by the availability of massive data traces from service providers and the fast development of big data processing techniques. Recent literature on data-driven user engagement analysis mainly focuses on understanding the influence of different factors on user engagement. In these works, user engagement is quantified from the different perspectives. For example, content providers can quantify user engagement via the viewing time ratio [10], while network service provider may employ the video download ratio [14] as a metric. These metrics also conform with the business models of subscription-based or advertisement-based video services, which is very important from the perspec-

tive of service providers. In [10], the authors studied the impact of start-up delay, rebuffer time and encoding bitrate on user engagement. As an extension, in [11, 12], the authors further investigated the impact of types of video, device, and connectivity on user engagement and proposed a decision tree-based prediction model to characterize the complicated relationship between user engagement and confounding factors. In [14], Shafiq and *et al.* studied how cellular network metrics affect the video download ratio, and predict the download ratio with a regression tree model. In [21], Jiang *et al.* observe that the video quality is mainly determined by a subset of critical features and propose a novel Critical Feature Analysis (CFA) system to predict video QoE by examining the QoE of similar sessions. However, these existing user engagement models only quantify the *average* engagement of all users, while user diversity in the engagement pattern has been overlooked. As a remedy, our work propose to personalize the user engagement modeling to capture the diversity of user behaviors. We believe our personalize model can serve as a complement to these works.

3. Data Set

To comprehensively study the engagement behavior at a large scale, we collect a massive data set from a tier-1 cellular network provider in China [22], which contains more than 8 million users and covers a large metropolitan area in one of the biggest cities in China from August 1st, 2014 to September 2nd, 2014.

This dataset contains information from two data sources. One data source is the raw IP flow trace captured from the links between the serving GPRS support nodes (SGSN) and the gateway GPRS support nodes (GGSN) in the core network of a 3G cellular network. It contains the flow-level information of all the IP traffic carried in the packet data protocol (PDP) context tunnels, that is, flows that are sent to and from mobile devices. This trace includes: start and end timestamps, anonymized user identifiers, traffic volume in terms of bytes, packet numbers for each flow, application information and location information.

All user identifiers are anonymized to protect privacy without affecting our analysis. The other data source is the information from the user profile database, which is managed by cellular network operators to better understand users' needs. These information consists of user's demographic information, *e.g.*, age, gender, address, and data usage behavior, such as current data plan and data usage in the last month. To protect user privacy, all user-related identifiers are strictly anonymized and robust to de-anonymization.

4. Problem Definition

Existing user engagement models are built upon the entire user data set, averaging the effect of factors on all users [9]. However, as users' viewing behavior diversifies, we expect the effects of these confounding factors to be disparate for different users and such diversity would affect the modeling of user engagement. To validate this, two natural questions follow: **(1)** ***Is there diversity in user engagement patterns?*** **(2)** ***If yes, how does such user diversity affect the user engagement modeling?***

In this section, we first define the engagement metric, then provide answers to the above two questions with experiments on a real-world data set.

4.1. Quantifying User Engagement

To quantify user engagement from the perspective of mobile network provider, we collect a large-scale anonymized IP flow trace from a tier-1 cellular network provider in China [22], which contains information of more than 8 million users and covers a large metropolitan area in one of the biggest cities in China from August 1st, 2014 to September 2nd, 2014 (the city name is anonymized for privacy issues). Through filtering and combining raw IP flow traces and signaling messages, we can obtain a fine-grained view of all mobile video sessions.

From the perspective of mobile network operators, fewer abandoned-video sessions and more downloading traffic are desirable, since a higher data usage results in a higher profit according to most revenue models of mobile network

providers. Therefore, the *download ratio* is often used as a metric to measure the user engagement from the perspective of network provider [14]:

$$\text{Download ratio } r = \frac{\text{downloaded bytes}}{\text{video file size in bytes}}. \tag{1}$$

As our work is in part of a large on-going service optimization project for the cellular operator, we closely cooperate with many front-line engineers to perform on-site monitoring and measurements. With a full support from the cellular operator, the file size, downloaded bytes, and the corresponding bit rate changes during the video session can be captured by employing an internal deep packet inspection (DPI) system, which is deployed by the cellular operator at the IP layer for the purpose of network QoS/QoE analysis and security monitoring. By inspecting the IP packet, examining the payload (*e.g.*, the Media Presentation Description (MPD) data in MPEG-DASH protocol [23]), and applying a variety of protocol-specific rules and machine-learning-based algorithms on both packet payload and network traffics, the DPI system can get the detailed video session information from IP packet level measurement, even the traffics are encrypted [24, 25, 26].

Note that, considering various video service providers are contained in out dataset, albeit their streaming techniques are different, the adaptive streaming is commonly used. Therefore, the byte-range request and sudden video quality adjustment are common practices in our dataset. To address this issue, the download ratio of the video is computed "adaptively" by monitoring the video quality change. For example, at the beginning, the video is streaming at quality level A, which corresponds to a video size of y_A. If no video quality change happens, the download ratio can be easily computed as $r = \frac{x_0}{y_A}$, where x_0 is the downloaded bytes. Later, at time t_1, the video quality changes to level B of size y_B and the downloaded bytes at this moment is x_1. Suppose that, at time t_2, the user abandons this video-streaming session. The total downloaded bytes from t_0 to t_2 are x_2. Therefore, the download ratio $r = \frac{x_1}{y_A} + \frac{x_2 - x_1}{y_B}$.

In this paper, we employ the download ratio to quantify the user engagement, as it can be accurately measured from the network provider side. We

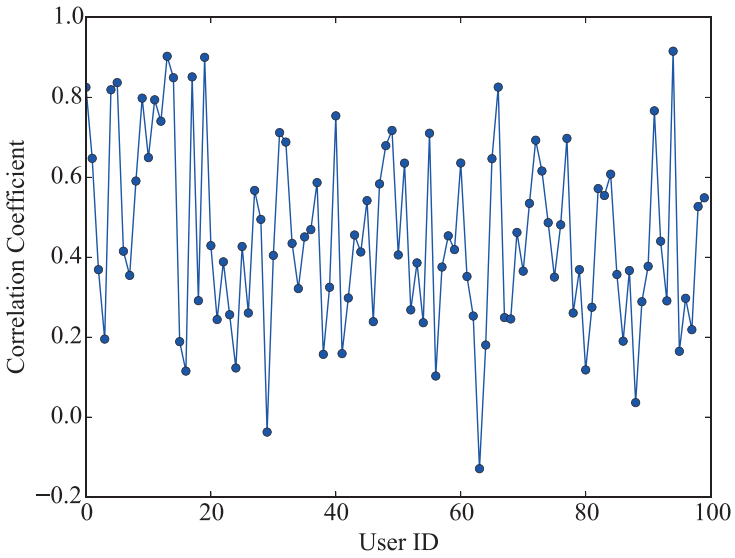

Figure 1: The correlation between video download ratio and average downlink throughput during the video session for 100 random users.

understand that the download ratio can only capture the downloading phase of video streaming, but not users' behaviors after video download, *e.g.*, users may not watch the whole downloaded video due to lack of interest. Nevertheless, such events are out of the control of mobile network providers. Besides, other user engagement metrics suffer a similar problem, *e.g.*, video-played time can not reflect user engagement if the video is played in the background [11].

4.2. Validation of User Diversity

To rigorously validate the user diversity, several issues need to be considered. The first one is the effect of user interest. As we can image, the user engagement level deeply depends on whether the user is interested in the video content. To focus on the quantification of user engagement to network service, the effect of user interest should be eliminated. Therefore, we need to filter out the video sessions abandoned due to the interest mismatch, despite the lack of any quality issues during the session. To this end, many existing works [12, 27, 28] point out that the users tend to sample the video to check whether its content meets

his/her interest. This results in many of them abandoning the video session early at the very beginning. In light of this idea, we drop out all the video sessions which are abandoned before 10% of the whole video is downloaded. We understand such filtering can not guarantee that the issue caused by the user interest is fully addressed, but at least its effect can be obviously reduced. After this filtering, there are 2,074,965 video sessions left in our dataset.

Apart from this, we drop out the users whose video-viewing records are less than 100 to ensure each user left in the dataset has sufficient records to provide a statistically meaningful result. After these two filters, we randomly select 100 of them as target users.

For each of these selected users, we examine the Pearson correlation coefficient [29] between the *download ratio* and an important network quality feature, *i.e.*, the downlink throughput. Here, the Pearson correlation coefficient is a measure of the linear correlation between two variables. Its value is between $+1$ and -1, where $+1$ is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. Given two series of variables $X = \{x_1, x_2, ..., x_n\}$ and $Y = \{y_1, y_2, ..., y_n\}$, the Pearson correlation coefficient r is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{Y})^2}}, \tag{2}$$

where the $\bar{X}$ and $\bar{Y}$ is the mean value of X and Y, respectively.

The result are presented in the Figure 1. We can see that the correlations between download ratio and downlink throughput change dramatically, spanning from -0.15 to 0.9. This variation indicates that the impact of downlink throughput on user engagement is quite diversified.

To further examine the second question of how such user diversity affects the engagement modeling, we build two kinds of models: one is the *uniform model* which is built upon the entire data set and averages the effect of confounding factors on all users. The other is the *individual model* which only uses the data records of a specific user u_i as the training set and can be regarded as

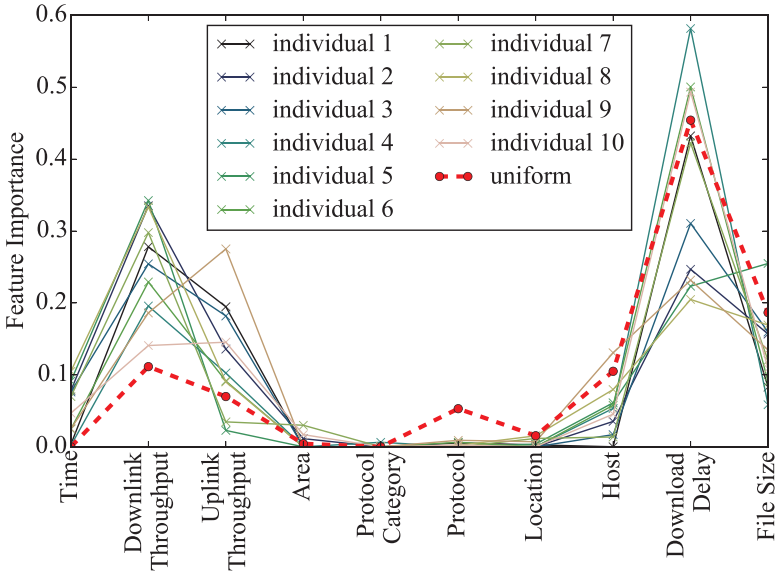

Figure 2: Feature importance of individual models and uniform model.

a precise engagement modeling for u_i. By comparing the uniform model with these individual models, we can examine how the uniform model deviates from these users.

In light of this idea, we adopt a mature machine learning algorithm—Gradient-Boosting Regression Trees (GBRT) [15]—to model the relationship between the user engagement and video-streaming-related features. We first build a uniform model by inputting all the users' data into GBRT. Then, we randomly select 10 users and train a separate individual model for each user. To ensure the result is statistical reliable, each individual model should be built on a user with sufficient data records. To this end, these 10 users are randomly selected from the top 10% users who have watched the most number of videos in our dataset. The average number of watched videos is 324.4 for these 10 users.

The comparison of the individual models and the uniform model is reported in Figure 2. Two interesting observations can be made: First, although the distributions of feature importances are somewhat similar, some feature importances can be obviously different. For example, in the 8-th individual model,

the feature importance of *download delay* is smaller than 0.2, but this feature is weighted more than 0.5 in the 4-th individual model. This corresponds to our findings in the previous experiment on downlink throughput. Another observation is that the uniform model does not fit individual models well. The relative feature importances of the uniform model actually diverge from individual models, which implies that the uniform model only captures the average viewing patterns of users, but overlooks the diversity among individual users.

5. Design Overview

In this section, we first outline the architecture of our system, then introduce our formulation of personalized user engagement model.

5.1. System Architecture

In this work, we leverage a collaborative approach to build a personalized user engagement model, which is *de facto* a set of collaborative individual models. By collaboratively learning individual models for each user, this approach would make a better use of individual historical data and discover the latent connections hidden in users' viewing traces. To alleviate the data sparsity problem, we utilize the collective matrix factorization [30] to learn side information from the user feature matrix and the video feature matrix.

As shown in figure 3, our system comprises six major components:

(1) Data input. Our system mainly exploits two major data sets. One is the user profile database, which is operated by the network operator and includes rich user-side information. The other one is the IP flow traces collected from the core network of a 3G cellular network. It contains all the traffic traces at the IP layer, which can be used to extract video downloading records and the network quality during each video session. More information about this data set is presented in section 3.

(2) Raw data processing. The process of raw data is dataset-dependent and involves many engineering works. In general, it contains the following steps:

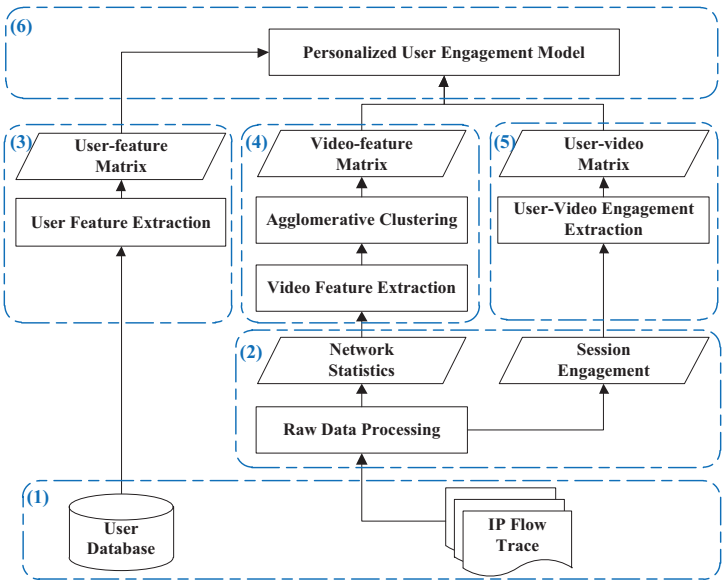

Figure 3: System overview of PE.

first, as the raw IP flow traces includes the traffic of various types of contents, we identify the video-streaming-related flows by the source IP/port, content-type header, as well as protocol type and traffic pattern. Then, these video-streaming flows are aggregated into sessions by combining the signaling messages and src/dest address pair. For each video session, the corresponding network quality statistics in this session can be computed from its IP-layer traffics. After that, we associated each video-streaming session with the corresponding user profile by using the anonymized International mobile Subscriber Identity (IMSI) number.

(3) User feature extraction. In order to provide better services, network operators have collected rich user information, including personal profiles (*e.g.*, age, gender) and usage behaviors (*e.g.*, current data plan, bills, and historical data usage). To protect user privacy, all user-related identifiers are strictly anonymized.

The user information can be very helpful for building a personalized model,

but there is an abundance of user features, and we need to filter out "minor" features, which contain less information about users' preferences, to prevent the curse of dimensionality. This feature selection can be done via information gain analysis, which is a standard approach to uncover relationships between variables [31]. In the feature selection, we use the download ratio as the target variable. Since it is a continuous variable, we first discretize it into 10 bins with 0.1 granularity. Then, we employ the information gain to select features. The underlying idea of information gain analysis is the entropy, which represents the informative level of a feature. The entropy of a random variable Y is defined as $I(Y) = -\sum_i P(Y = y_i) log P(Y = y_i)$, in which $P(Y = y_i)$ is the probability of $Y = y_i$, and the conditional entropy of Y given another random variable X, *i.e.*, $I(Y|X)$, can be computed as $\sum_j P(X = x_j) I(Y|X = x_j)$. The information gain then can be defined as $G = \frac{I(Y)-I(Y|X)}{I(Y)}$. To filter out unnecessary features, we can compute the information gain of each feature and select top features in terms of information gain. As a result, many less-important features, *e.g.*, IP address/port of video content provider, TCP reconnection times, and last network error, are filtered out and the selected features are shown in Table 1.

Through the feature selection, we select 19 out of 70 user features. These user features can be further categorized into two groups: (I) demographic information, which characterizes user population, *e.g.*, age, gender. (II) usage behavior, including current data plan, data traffic generated in last month, and so on. Table 1 gives some examples of user features.

Domain	**Feature**	**Description**
User Features	age	Age of user
	gender	Gender of user
	area	Active area of user
	addr_id	Living area index
	credit_value	User's credit value in the operator's credit system
	page_rank	User importance in pagerank

	product_id	User's current data plan ID
	product_kind	User's data plan specification
	product_price	Price of user's current data plan
	sale_id	Product selling area ID
	total_charge	Total cost in a month
	flux_last_month	Data traffic generated in last month
	flux_current_month	Data traffic generated in current month
	streaming_time	Total access time of the streaming service
	voice_cnt	Count of voice call
	voice_dura	Duration of voice call
	innet_dura	Duration of using the network service
	balance	Account balance
	total_recharge	Total recharge values
Video features	TIME	Start/end time of video session
	CELL_COUNT	Cells the user roamed during the session
	STREAMING_URL	The video's URL address
	STREAMING_FILESIZE	File size of the video in bytes
	CDN	IP address of the CDN
	STREAMING_SERVER	Video service provider
	PROT_TYPE	Streaming protocol name
	DEVICE_TYPE	End-device Type
	APN	User's access point name
	L4_UL_THROUGHPUT	Uplink throughput
	L4_DW_THROUGHPUT	Downlink throughput
	TCP_RTT	Round-trip time in seconds
	GET_STREAMING_DELAY	Video service response delay
	TCP_DW_RETRANS	TCP-level retransmission count

Table 1: The user-side & video-side features

(4) Video feature extraction. A major purpose of studying user engagement from the perspective of a network operator is to understand how it is affected by the network quality variation. Thus, apart from primitive attributes of videos (*e.g.*, file size, CDN server host), we also exploit the network quality statistics during a video session as video-session-associated features of this video. We select 14 important network quality statistics via information gain analysis, and generally categorize them into three groups, *i.e.*, *video attributes*, *session-associated features* and *context*. Table 1 illustrates these video features.

The main concern of using the network quality as video feature is that a video can be streamed under various network qualities and results in multiple video-quality tuples in our data set, each of which corresponds to a specific network quality combinations. This will result in an explosive size of the video feature matrix. To reduce the computation complexity, we leverage agglomerative clustering [32] to aggregate video-feature tuples that have the same URL and are streamed in a similar network quality condition into clusters. The number of clusters is a trade-off between the computation complexity and the granularity of network quality. After that, we merge videos that belong to the same cluster together and define the *video template* as the mean of all video-feature tuples in this cluster. Then, we represent the feature value of the video templates in a matrix format and incorporate it into our model.

(5) User-video engagement extraction. In our system, we quantify user engagement from the perspective of network operator via a continuous variable, the download ratio, which ranges from 0 to one to represent the fraction of video downloading. The data is transformed in a user-video matrix R, in which r_{ij} is the download ratio of the video j by user i.

(6) Personalized user engagement model. To address the user diversity, we choose to quantify the user engagement with a collaborative filtering model. Also, to alleviate the data sparsity problem, we employ a collective matrix factorization to integrate user- and video-feature data set. An in-depth discussion is given in section 6.

5.2. Data Formulation

By extracting and aggregating these raw data traces, we can obtain rich information about each video session, including the network statistics during the video streaming, the viewer personal information, and their past behavior patterns. We can formulate the processed data as follows:

- m users, each of whom has l features. Let $D_{m\times l}$ denote the user feature matrix.
- n videos, each of which is associated with h video features. Let $S_{n\times h}$ denote the video feature matrix.
- $R_{m\times n}$ is the user-video matrix, in which r_{ij} is user i's download ratio for video j. $R_{m\times n}$ is a highly sparse matrix (sparse rate $\approx 99\%$). Since there are billions of videos and each user has only watched a tiny fraction of them, many items of r_{ij} are unknown.

In this context, modeling the user engagement is equivalent to building a model which can accurately predict the missing values in the user-video matrix R, based on the user feature matrix D and the video feature matrix S. Considering the significant user diversity in the user-video matrix R, this can be a challenging problem.

5.3. Limitations

We acknowledge that there are two potential limitations in our current formulation and implementation:

- **Download ratio as user engagement.** Although the download ratio is directly related to the revenue model of mobile network provider, it constrains our analysis within the downloading phase. Some user behaviors after downloading, unobservable from the network side, can not be captured. However, as the download ratio can be accurately and objectively measured from network side and other analysis also employ the same metric [14], we use it as a start point for our analysis and our system can be easily applied to other metrics of engagement.

- **Data coverage over confounding factors.** As our data set is collected from the network provider, several confounding factors that affect engagement are not captured in our dataset (*e.g.*, video content and its popularity). As a result, our current implementation of PE only provides a baseline performance and other data sources can be further integrated to provide a more comprehensive and accurate engagement assessment.

6. Personalized User Engagement Model

After data modeling, we have the user feature matrix D, video feature matrix S and user-video matrix R. Our goal is to build a personalized model to predict the missing values in R, with the help of D and S.

To build a personalized engagement model, one intuitive solution is to build an individual model for each user separately. However, this is impractical as there are millions of users. Another possible alternative is to first cluster users into groups, then build an independent model for each user group. This sounds like a reasonable solution, but it is based on a strong assumption that users with similar user features would behave analogously. This assumption poses a high requirement for the quality of user features. If the user feature does not fully capture the similarity of users on their engagement level, the clustering quality will be poor and eventually degrade the model's performance. Apart from that, this approach also understates the behavior patterns hidden in the historical data. For example, user i and user j are quite similar according to their user features, but indeed they behave in quite different ways (this may happen when the similarity of user features does not perfectly reflect user behaviors). In this case, even if there are adequate historical data records for both users, they will still be clustered into the same user group and thus share a comprised model which does not fit either of them.

To conquer the user diversity and data sparsity, we establish a model based on the collective matrix factorization framework [30]. The basic idea is that, we can first model each matrix via a low-rank approximation:

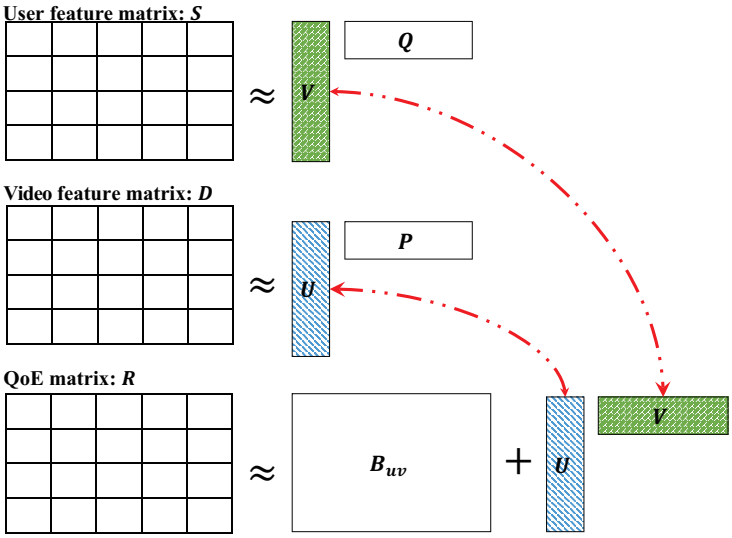

Figure 4: Overview of personalized user engagement model.

$$D = U \cdot P^T \tag{3}$$

$$S = V \cdot Q^T \tag{4}$$

$$R = U \cdot V^T + \underbrace{\mu J_{mn} + B_u \cdot J_n^T + J_m \cdot B_v^T}_{B_{uv}}, \tag{5}$$

where U, V, P and Q are latent factors, B_{uv} is the baseline predictor, μ is the average download ratio of all users, B_u, B_v are user bias and video bias matrices, and J_* are matrices of all the ones in different dimensions as suggested by their subscripts.

In this model, the user-video matrix R is approximated by the summation of baseline predictor B_{uv} and the product of latent factors U and V. The baseline predictor B_{uv} captures the basic engagement pattern and the bias introduced by each user and video, with which we can alleviate the cold-start problem [33]. In other words, given a previously unseen user u_i, as the corresponding video-viewing records are unavailable, his/her download ratio r_{ij} of a video v_j is dominated by the average download ratio of all users μ, as well as the download

ratios of this video by other similar users, which are captured by the vector $B^T_{v_j}$. A similar approach can be applied for newly added videos. Furthermore, the lower-rank approximation part, $U \cdot V^T$, characterizes the fluctuation caused by the user and video diversity.

The rationale is that, by transforming both user and video to the same latent factor space, we can estimate users' *interest* in these latent factors (*i.e.*, U) and the video's extent of these factors (*i.e.*, V). Each user i and each video j correspond to a preference vector $\vec{u_i}$ and score vector $\vec{v_j}$, and the learning procedure is conducted in a collaborative approach. We iteratively use all the video data of user i to help the training of user i's preference vector, and feed all data of users who watched video j into the model to learn the video score vector v_j.

This collaborative approach may still suffer from data sparsity problem as only users or videos which have common interactions (*i.e.*, users who have watched the same video, or videos that are watched by the same user) would collaborate. To further alleviate data sparsity, we also simultaneously factorize user feature matrix D and video feature matrix S and intentionally let latent factor U and V be shared among these factorizations. As a result, information from D and S can be propagated to R, and thus help gain a better performance. Figure 4 provides an intuitive overview of our model.

According to this model, we can formulate our objective function as:

$$
\begin{aligned}
& L(B_u, B_v, U, V, P, Q) = \\
& \parallel I_1 \circ (R - U \cdot V^T - \mu J_{mn} - B_u \cdot J_n^T - J_m \cdot B_v^T) \parallel_F^2 \\
& + \frac{\alpha_1}{2} \parallel I_2 \circ (D - U \cdot P^T) \parallel_F^2 + \frac{\alpha_2}{2} \parallel I_3 \circ (S - V \cdot Q^T) \parallel_F^2 \\
& + \frac{\lambda_1}{2} (\parallel U \parallel_F^2 + \parallel V \parallel_F^2) \\
& + \frac{\lambda_2}{2} \parallel P \parallel_F^2 + \frac{\lambda_3}{2} \parallel Q \parallel_F^2 \\
& + \frac{\lambda_4}{2} \parallel B_u \parallel_F^2 + \frac{\lambda_5}{2} \parallel B_v \parallel_F^2, \qquad (6)
\end{aligned}
$$

where α_1 and α_2 are the reconstruction weights which control the degree of reconstruction and information sharing. The larger α is, the more important the corresponding term is in loss function and propagates more information to the others. $\lambda_i, i = 1, 2, ..., 5$ are the regularization parameters, and $I_i, i = 1, 2, 3$ are the indicator matrices where $I_{ij} = 0$ if the corresponding value is missing. Let the operator $\circ$ denote the element-wise product of two matrices and $\|\cdot\|_F$ be the Frobenius norm. Note that we do not restrict the download ratio to be a non-negative value between 0 and 1. Such restriction, albeit reasonable, renders the problem a non-negative matrix factorization (NMF) problem. Due to the consideration of the modeling complexity and solving time, we relax this restriction.

In general, this objective function is not jointly convex, and we cannot get a close-form solution for minimization of this objective function. Therefore, we turn to search for a practical local optimal solution by gradient descent. More specifically, the gradients of loss function are:

$$\begin{cases} \nabla_{B_u} L &= E_r \cdot (-J_n) + \lambda_4 B_u \\ \nabla_{B_v} L &= E_r^T \cdot (-J_m) + \lambda_5 B_v \\ \nabla_U L &= E_r \cdot (-V) + \alpha_1 E_D \cdot (-P) + \lambda_1 U \\ \nabla_V L &= E_r^T \cdot (-U) + \alpha_2 E_S \cdot (-Q) + \lambda_1 V \\ \nabla_P L &= \alpha_1 E_d^T \cdot (-U) + \lambda_2 P \\ \nabla_Q L &= \alpha_2 E_s^T \cdot (-V) + \lambda_3 Q \end{cases} \tag{7}$$

where E_r, E_d and E_s are the residual errors with respect to R, D and S.

$$E_r = I_1 \circ (R - U \cdot V^T - \mu J_{mn} - B_u \cdot J_n^T - J_m \cdot B_v^T), \tag{8}$$

$$E_d = I_2 \circ (D - U \cdot P^T), \tag{9}$$

$$E_s = I_3 \circ (S - V \cdot Q^T). \tag{10}$$

With the gradients, we can resort to the gradient descent approach to iteratively minimize the objective function. The detail is described in Algorithm 1.

Although gradient descent can be quite straightforward, more efficient approaches, *e.g.*, the stochastic approximation approach or a parallel version of

Algorithm 1 PE Solver

Require:

Maximum iteration number S, convergence threshold ϵ;

User feature matrix D, video feature matrix S;

Sparse user-video matrix R.

Parameters set $P = \{p_*|p_* = \{\mu, B_u, B_v, U, V, P, Q\}\}$

Ensure:

Completed user-video matrix $\hat{R}$.

$s \leftarrow 1$;

while $s \leq S$ **and** $L^{(s)} - L^{(s+1)} > \epsilon$ **do**

  $\gamma \leftarrow 1$;

  Compute current residual error by Eq. 8;

  Compute the gradients $\nabla_* L$by Eq. 7;

  while $L(p_* - \gamma \nabla_{p_*} L) \geq L(p_*)$ **do**

    $\gamma = \frac{\gamma}{2}$;

  end while

  $B_u^{(s+1)} = B_u^{(s)} - \gamma \nabla_{B_u}^{(t)}$, $B_v^{(s+1)} = B_v^{(s)} - \gamma \nabla_{B_v}^{(t)}$

  $U^{(s+1)} = U^{(s)} - \gamma \nabla_U^{(t)}$, $V^{(s+1)} = V^{(s)} - \gamma \nabla_V^{(t)}$

  $P^{(s+1)} = P^{(s)} - \gamma \nabla_P^{(t)}$, $Q^{(s+1)} = Q^{(s)} - \gamma \nabla_Q^{(t)}$

  $s = s + 1$

end while

Predict with: $\hat{R} = U \cdot V_s^T + \mu J_{mn} + B_u \cdot J_n^T + J_m \cdot B_v^T$

return $\hat{R}$

the gradient descent [34, 35, 36] can be adopted to improve training efficiency. We will not discuss these algorithms in detail, as it is out of the scope of this paper.

7. Evaluation

In this section, we evaluate our system using a real-world dataset. We start by comparing the model performance with three state-of-the-art baselines. Then, we investigate our system by studying how the different parameter settings affect our system's performance.

To conduct the experiments, we have implemented our model in Python 2.7 and ran it on an enterprise server machine with 24 Intel Xeon E5-2420@1.90GHz CPUs, 120 GB memory, and 100 TB hard disk.

7.1. Model Performance

To evaluate the effectiveness of personalized models, we employ three widely-used and high-performance machine learning models as baselines:

- the **Decision-Tree Regressor** is a predictive model widely-adopted in statistics, data mining, and machine leanring. Many existing data-driven user engagement/QoE analysis employ this model due to its simplicity and explainability [10, 11, 12].
- The **Random forest** is a mature and high-accuracy ensemble learning models, which is proved to be robust for the overfitting problem [37].
- To compare with previous work based on regress tree [14], we also employ a regression-tree-based model in our comparison, *i.e.*, the **Gradient-Boosting Regression Tree (GBRT)**, which is an ensemble learning model based on regression tree. It incorporates the gradient descent and boosting techniques to achieve a better performance [15]. It builds a single strong learner by combining multiple weak "learners" in an iterative fashion: at the m-th stage of gradient boosting, an imperfect model F_m

can be improved by constructing a new model $F_{m+1} = F_m(x) + h(x)$, which corrects its error residual $h(x) = y - F_m(x)$. By fitting the h to the residual $y - F_m(x)$, each F_{m+1} learns to corrects its predecessor F_m. More details of this algorithm is added in the revision.

For these baseline models, we transform all the user-side and video-side features listed in Table 1 into flat-table format and use the *download ratio* as the variable to predict. Also, the best parameters for each model is determined via a 10-fold cross validation. That is, the dataset is randomly partitioned into 10 equal-sized subsets. Of these 10 subsets, a single subset is used for testing the model, while the remaining 9 subsets are used as training data. The cross-validation process is then repeated for 10 times, with each subset used exactly once as the testing data. After that, the test errors of these 10-fold testing results are averaged to produced a single error estimation. In our evaluation, the test error of each model is evaluated in terms of *Mean Absolute Error* (MAE) and *Root-Mean-Square Error* (RMSE). Given n tuples, let r_i and $\hat{r_i}$ be the real value and predicted value for the i-th tuple, the MAE and RMSE are defined as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|r_i - \hat{r_i}|, \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \hat{r_i})^2}. \tag{12}$$

Although both MAE and RMSE are metrics for measuring error rate, there are some subtle differences between them. As their names imply, the MAE is a linear metric which means that all the individual errors are weighted equally in the average, while the RMSE gives a relatively high weight to large errors and thus it is more useful when large deviations are particularly undesirable. The MAE and RMSE are used together to measure the variance in the individual errors in the prediction. The greater the difference between them, the larger the variance is [31].

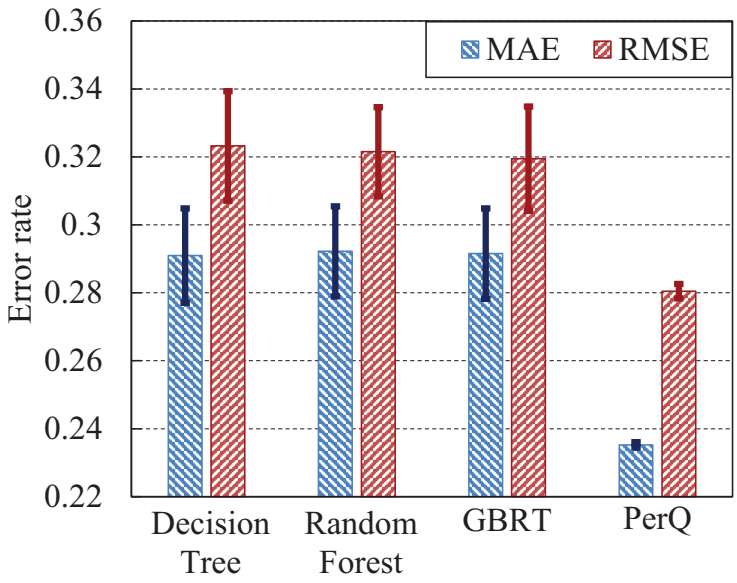

Figure 5: Model performance comparison.

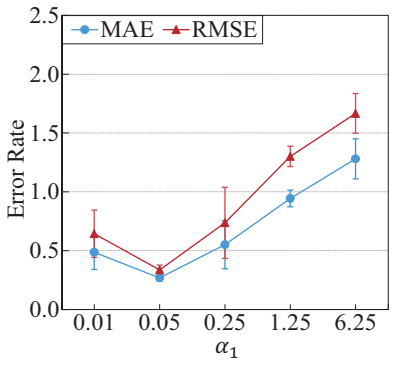

Figure 6: Impact of reconstruction weight α_1 on model performance. The larger α_1 is, the more information is learned from user-feature matrix D, and less emphasis is putted on the approximation of user-video matrix R

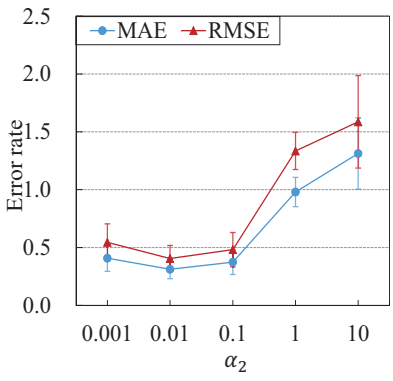

Figure 7: Impact of reconstruction weight α_2. Each value is computed via 10-fold cross validation while fixing α_1 to 0.05

Figure 5 shows the comparison of prediction performance of PE and baselines. We observe that, as the *gradient boosted regression trees* and *random forest* leverage the ensemble power of a set of weak learners to build a better model, they slightly outperform the basic *decision tree* model. However, as these baseline models treat all users as a uniform group, they neglect the diversity of user behavior. On the contrary, our model learns the individual model for each user with the help of information from user-video interactions, and rich side information from other data sources. When compared with the best of baseline models (*i.e.*, gradient boosted regression trees), the performance improvement is 19.14% and 12.20% in terms of MAE and RMSE, respectively.

In the following subsections, we will study the performance of our system under different parameter settings.

7.2. Impact of Reconstruction Weights

The reconstruction weights, α_1 and α_2, control the importance of corresponding matrix approximation in the loss function and the degree of information propagation. For example, a large α_1 not only implies that more emphasis is placed on the approximation of user feature matrix D in loss function but also suggests that more information should be learned from D.

To study the impact of these reconstruction weights, we vary the value of one reconstruction weight each time and plot the dynamics of the model performance in terms of the average MAE/RMSE and the standard deviation obtained from the 10-fold cross-validation.

Figure 6 demonstrates how the system performance changes as we vary the value of α_1 while fixing $\alpha_2 = 0.01$. We notice that as the value of α_1 increases, the error rate first decreases and then starts to rise. The reason is that, when α_1 is small, our model cannot fully exploit the user-side information to understand the similarity between users and therefore degrades the performance. However, if the value of α_1 is too large, the contribution of the user feature matrix D would dominate in the loss function. This would restrain the approximation of the user-video matrix, which will eventually downgrade the model performance.

The best value of α_1 in our experiment is 0.05.

A similar pattern can also be observed in the analysis of α_2. As shown in Figure 7, the error rate starts to decrease as we enlarge the value of α_2, and more information is propagated from the video feature matrix S. However, if α_2 keeps growing, more errors would be introduced as the approximation of the video feature matrix S dominates in the loss function and thus reduces the importance of filling the missing value in R. The optimal value of α_2 in our experiment is 0.01.

7.3. Impact of Video Clustering Granularity

As we aim to understand the impact of network statistics on user engagement in mobile video services, we utilize the network quality statistics during a video session as features of this video. However, the fluctuation in network quality may lead to a dramatic data explosion and therefore introduce a negative effect on the overall performance. To alleviate this problem, we aggregate video-feature tuples into video templates via agglomerative clustering. The extent of this aggregation is controlled by a clustering granularity c. To understand the impact of video clustering granularity on our system, we validate our system under different cluster granularity settings and present the results in Figure 8.

A small clustering granularity c aggregates more video-feature tuples into a single video template. This can significantly reduce the data size, but also introduce large deviations inside a video template. As a consequence, there will be large prediction errors for video-feature tuples which are *far away* from the video templates. On the other hand, if the value of c is too large, the video-feature tuples are barely aggregated, which again exposes the problem of network quality variations. According to Figure 8, we set $c = 0.5$ for our data set.

7.4. Impact of Dimensionality of Latent Factor Space

As our model factories each user and video into two vectors in a latent factor space of dimensionality f, a small f can significantly compress information into

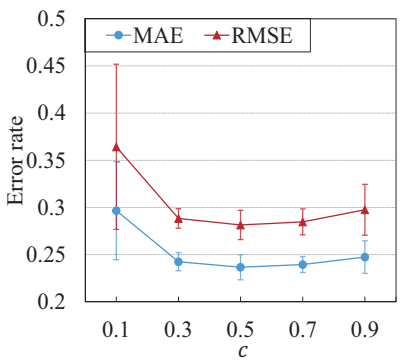

Figure 8: Impact of video clustering granularity c. A larger c enables more video-feature tuples to be aggregated into a single video template and reduce the data sparsity, but also introduce large deviations inside single video template.

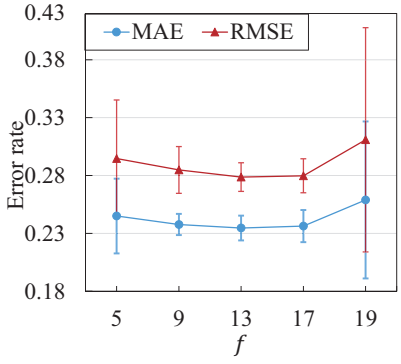

Figure 9: Impact of latent factors f. A small f can significantly compress information into a compact latent space with information loss, while a large f can better capture the underlying pattern in the data, but may aggravate the data sparsity problem.

a compact latent space, which can help conquer the data sparsity problem. However, such information compression may also suffer from information loss and thus cannot achieve the optimal performance. On the other hand, a large f can help better capture the underlying pattern in the data, but bring about a more serious data sparsity problem, which would introduce many errors for users with a small number of watching records.

We plot the average and standard deviation of our system under different values of f in Figure 9. We can observe that the error rate first decreases when the value of f increases and achieves an optimum when $f = 13$. If we continue to enlarge the value of f, the data sparsity problem will lead to performance fluctuation.

8. Discussion

Our system is a first step toward the personalized user engagement modeling for mobile videos. While the evaluation results demonstrate that it is promising, there are still some limitations and open problems as below.

The Open of database. The dataset we used is a highly-sensitive property of the cellular network operator, which consists of massive IP flow traces captured from the core network of a city-level cellular network. Since much sensitive information about the network infrastructure can learn from such dataset, many network operators concern that publishing such dataset may pose potential security threats to their network service. Furthermore, this dataset also contains abundant personal information of millions of users, albeit they are well-anonymized, the user privacy is still a key concern which hinders the open of this dataset.

System Reliability. Although our evaluation shows a promising result with a 19.14% performance gain, some may concern whether such personalized user engagement can be adopted in the cellular system as it presents higher complexity than uniform models. In our vision, personalization is a trend. First, to provide better service, it is an inevitable that more and more user-side information will be collected by the cellular operator. This provides a great opportunity for the cellular operator to personalize their network service. Also, unlike other wireless communication systems, the cellular network allocates a dedicated channel to each user, which also enables the potential possibility for the per-user optimization of resource allocation. Most importantly, the immense user base of cellular network operator implies that a small optimization in user engagement can affect millions of users and may lead to whopping changes in monetization opportunities. Existing user engagement assessment systems employ the uniform models to quantify the user engagement at a coarse level and can only archive a moderate performance as the essential distinction of individual users is overlooked. As a remedy, we propose a personalized model to quantifies the user engagement at a fine-grained level (as a continuous value from 0 to 1). We understand our current implementation of personalized user engagement model still needs further improvement, but our evaluation demonstrates that the personalized model has a great potential to bring significant improvement for the large-scale user engagement assessment.

Possible Improvements. To further improve the system performance,

several directions are worth exploring. First, we believe incorporating features from different control panel can help our system performs better. For example, user's preference can help us better characterize the user viewing behavior, and the operation log (video skip/stop) from end device enables us to understand user's engagement level during the video viewing. Meanwhile, some researchers point out that only a small set of the critical features are determining the video-streaming quality [21]. In light of this idea, a more reliable feature selection process can be performed to enhance data quality. We leave these improvements for further exploration.

Yet another possible improvement direction is the online algorithm. For now, our model works in a batch-processing way. That is, we need to retrain the model periodically, *e.g.*, every month. This is indeed not efficient in both terms of complexity and time. One promising solution is to leverage the online streaming algorithms to update the model "continuously". There are many existing works on this topic [38, 39]. In the future, we plan to incorporate these updating algorithms to our system.

9. Conclusion

Accurate and reliable user engagement assessment is the key to optimize the video service quality. Our experiments on real-world data set reveal a significant diversity in user engagement and a uniform user engagement model is not sufficient to characterize the distinctive engagement patterns of different users. To deal with this problem, we propose PE, a personalized user engagement model for mobile videos from the perspective of cellular network providers, in which the user diversity is well addressed. The evaluation results on a real-world data set show that our personalized user engagement model outperforms uniform models with a 19.14% performance gain.

Acknowledgment

The research was supported in part by grants from 973 project 2013CB329006, RGC under the contracts CERG MHKUST609/13 and 622613, China NSFC under Grant 61502114, 61373092, 61272449, 61572339 and 61202029.

References

[1] G. J. Sullivan, T. Wiegand, Rate-distortion optimization for video compression, IEEE signal processing magazine 15 (6) (1998) 74–90.

[2] M. Wu, R. A. Joyce, H.-S. Wong, L. Guan, S.-Y. Kung, Dynamic resource allocation via video content and short-term traffic statistics, IEEE Transactions on Multimedia 3 (2) (2001) 186–199.

[3] J. Shin, J. W. Kim, C.-C. Kuo, Quality-of-service mapping mechanism for packet video in differentiated services network, IEEE Transactions on Multimedia 3 (2) (2001) 219–231.

[4] Q. Zhang, W. Zhu, Y.-Q. Zhang, End-to-end qos for video delivery over wireless internet, Proceedings of the IEEE 93 (1) (2005) 123–134.

[5] I. Recommendation, 800, methods for subjective determination of transmission quality, International Telecommunication Union.

[6] P. ITU-T RECOMMENDATION, Subjective video quality assessment methods for multimedia applications.

[7] H. R. Wu, K. R. Rao, Digital video image quality and perceptual coding, CRC press, 2005.

[8] S. Tao, J. Apostolopoulos, R. Guérin, Real-time monitoring of video quality in ip networks, IEEE/ACM Transactions on networking 16 (5) (2008) 1052–1065.

[9] Y. Chen, K. Wu, Q. Zhang, From qos to qoe: A tutorial on video quality assessment, IEEE Communications Surveys & Tutorials 17 (2) (2015) 1126–1165.

[10] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, H. Zhang, Understanding the impact of video quality on user engagement, in: ACM SIGCOMM Computer Communication Review, Vol. 41, ACM, 2011, pp. 362–373.

[11] 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 2013 Conference on SIGCOMM, SIGCOMM '13, ACM, New York, NY, USA, 2013, pp. 339–350.

[12] 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 11th ACM workshop on hot topics in networks, ACM, 2012, pp. 97–102.

[13] S. S. Krishnan, R. K. Sitaraman, Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs, IEEE/ACM Transactions on Networking 21 (6) (2013) 2001–2014.

[14] 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: The 2014 ACM International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS '14, ACM, New York, NY, USA, 2014, pp. 367–379.

[15] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2001) 1189–1232.

[16] Z. Chen, K. N. Ngan, Recent advances in rate control for video coding, Signal Processing: Image Communication 22 (1) (2007) 19–38.

[17] G. J. Sullivan, T. Wiegand, Rate-distortion optimization for video compression, IEEE signal processing magazine 15 (6) (1998) 74–90.

[18] Z. Chen, K. N. Ngan, Recent advances in rate control for video coding, Signal Processing: Image Communication 22 (1) (2007) 19–38.

[19] S. Chong, S.-q. Li, J. Ghosh, Predictive dynamic bandwidth allocation for efficient transport of real-time vbr video over atm, IEEE Journal on Selected Areas in Communications 13 (1) (1995) 12–23.

[20] M. Wu, R. A. Joyce, H.-S. Wong, L. Guan, S.-Y. Kung, Dynamic resource allocation via video content and short-term traffic statistics, IEEE Transactions on Multimedia 3 (2) (2001) 186–199.

[21] J. Jiang, V. Sekar, H. Milner, D. Shepherd, I. Stoica, H. Zhang, Cfa: A practical prediction system for video qoe optimization, in: 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), USENIX Association, 2016, pp. 137–150.

[22] China unicom.
URL `https://en.wikipedia.org/wiki/China_Unicom`

[23] I. Sodagar, The mpeg-dash standard for multimedia streaming over the internet, IEEE MultiMedia 18 (4) (2011) 62–67.

[24] W. Pan, G. Cheng, H. Wu, Y. Tang, Towards qoe assessment of encrypted youtube adaptive video streaming in mobile networks, in: Quality of Service (IWQoS), 2016 IEEE/ACM 24th International Symposium on, IEEE, 2016, pp. 1–6.

[25] S. Dharmapurikar, P. Krishnamurthy, T. Sproull, J. Lockwood, Deep packet inspection using parallel bloom filters, in: High performance interconnects, 2003. proceedings. 11th symposium on, IEEE, 2003, pp. 44–51.

[26] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, K. Papagiannaki, Measuring video qoe from encrypted traffic, in: Proceedings of the 2016

Internet Measurement Conference, IMC '16, ACM, New York, NY, USA, 2016, pp. 513–526. doi:10.1145/2987443.2987459.
URL http://doi.acm.org/10.1145/2987443.2987459

[27] Y. Chen, F. Zhang, K. Wu, Q. Zhang, Qoe-aware dynamic video rate adaptation, in: 2015 IEEE Global Communications Conference (GLOBECOM), 2015, pp. 1–6. doi:10.1109/GLOCOM.2015.7416940.

[28] J. Erman, A. Gerber, K. K. Ramadrishnan, S. Sen, O. Spatscheck, Over the top video: The gorilla in cellular networks, in: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11, ACM, New York, NY, USA, 2011, pp. 127–136. doi:10.1145/2068816.2068829.
URL http://doi.acm.org/10.1145/2068816.2068829

[29] A. M. Mood, Introduction to the theory of statistics.

[30] A. P. Singh, G. J. Gordon, Relational learning via collective matrix factorization, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08, ACM, New York, NY, USA, 2008, pp. 650–658.

[31] C. M. Bishop, Pattern recognition, Machine Learning 128 (2006) 1–58.

[32] L. Kaufman, P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, Vol. 344, John Wiley & Sons, 2009.

[33] Y. Koren, R. Bell, Advances in collaborative filtering, in: Recommender systems handbook, Springer, 2011, pp. 145–186.

[34] T. Zhang, Solving large scale linear prediction problems using stochastic gradient descent algorithms, in: Proceedings of the twenty-first international conference on Machine learning, ACM, 2004, p. 116.

[35] M. Zinkevich, M. Weimer, L. Li, A. J. Smola, Parallelized stochastic gradient descent, in: Advances in neural information processing systems, 2010, pp. 2595–2603.

[36] B. Recht, C. Re, S. Wright, F. Niu, Hogwild: A lock-free approach to parallelizing stochastic gradient descent, in: Advances in Neural Information Processing Systems, 2011, pp. 693–701.

[37] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[38] A. S. Das, M. Datar, A. Garg, S. Rajaram, Google news personalization: scalable online collaborative filtering, in: Proceedings of the 16th international conference on World Wide Web, ACM, 2007, pp. 271–280.

[39] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, Journal of Machine Learning Research 11 (Jan) (2010) 19–60.