IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal
SPECIAL SECTION ON INTELLIGENT AND COGNITIVE TECHNIQUES FOR INTERNET OF THINGS

Received May 26, 2019, accepted June 24, 2019, date of publication July 10, 2019, date of current version July 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928135

Peripheral Sensing: Monitoring Quality of Experience for Video Services Based on Mobile Terminals

XIWEN LIU[1,2], XIAOMING TAO[1], (Member, IEEE), YAFENG ZHAN[2], AND JIANHUA LU[1,2], (Fellow, IEEE)

[1]Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China
[2]Tsinghua Space Center, Tsinghua University, Beijing 10084, China

Corresponding author: Xiaoming Tao (taoxm@mail.tsinghua.edu.cn)

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 61622110, Grant 61471220, and Grant 91538107.

ABSTRACT Quality of experience (QoE), which directly relates to both technical evolvement and profit promotion, is a vital concern for mobile video services. However, the wireless network operators have long been troubled by the problem of lacking effective QoE monitoring approaches since the traditional evaluation methods of communication quality are objective metrics oriented. Considering that mobile terminals are the network elements closest to users, it is promising to realize a real-time QoE estimation for video services by fully utilizing the sensing capabilities of mobile terminals. As the first step, we specifically develop a mobile video testing application. With support from China Unicom, one of the three major wireless network operators in China, over 80 000 data records are collected under the real-world conditions. The collected data consist of four types of subjective scores and 13 objective parameters concerning video attributes, network performance, device capability, playback events, and external factors. After preprocessing the data set through correlation analysis, we establish the two QoE estimation models based on the C4.5 method and the gradient boosting decision tree (GBDT) method, respectively. The experimental results demonstrate that the proposed models can achieve remarkable estimation performances and outperform the baseline models. Specifically, the overall estimation accuracy of the GBDT-based model is approximately 80% for a five-level scale and approaches 90% when a more practical 3-level scale is adopted. Finally, we comprehensively discuss the estimation performances based on characteristics of the data and validate the feasibility of estimating QoE based on mobile terminals—the ‘‘peripheral sensors’’ of the mobile networks.

INDEX TERMS Decision tree, mobile terminals, estimation model, QoE, video services.

I. INTRODUCTION

With the explosive development of wireless networks, mobile video services have been rapidly popularized. According to the Cisco Visual Networking Index [1], the mobile video traffic will account for 79 percent of the world's total mobile data traffic by 2022. When it comes to the era of 5G, the ultimate goal of technological evolvement is no longer simply to approach the theoretical upper limit but to provide better quality of experience (QoE) for mankind [2]–[4]. This development concept has inspired various technological improvements for mobile video services [5]–[10]. In addition, user experience directly determines the viability of a technology or service in the competitive market. Therefore, QoE of wireless video services has become a core concern for wireless network operators from both technical and commercial perspectives.

However, it is still difficult for the wireless network operators to evaluate and monitor user experience about mobile video services effectively. Considering that user rating is much too costly, sometimes even infeasible due to user reluctance, establishing QoE estimation models based on measurable objective parameters is the most promising approach. Some early works proposed to estimate QoE directly through a fundamental communication performance metric such as system throughput and communication delay [11], [12], however, the estimation performances are not satisfactory. For a better estimation performance, some

The associate editor coordinating the review of this manuscript and approving it for publication was Min Jia.

researchers further considered this problem from the perspective of causality within the generation process of user experience and established an estimation framework containing three successive stages, i.e. network quality of service (QoS), application QoS and user QoE [13]. However, the feasibility of this framework is challenged by the difficulties of inferring application QoS from network QoS. First, it is difficult to estimate application QoS at the network side since the information of the upper layers is transparent to the wireless network operators who act as information pipelines today. Even if the wireless network operators are able to overcome the legal issues and parse the packets by some means, such as deep packet inspection (DPI), they are still not able to directly derive the application-specific quality parameters since the main video source providers adopt hyper text transfer protocol over secure socket layer (HTTPS) to encrypt video traffic [14]. In addition, some application parameters are not fully known priori, such as the variable bit rate. Therefore, there is a strong demand for the wireless network operators to extend their sensing range beyond the core network and acquire informative knowledge from the upper layers.

Another challenge for QoE estimation results from its multidisciplinary attributes [15]. In addition to technological factors, various non-technological factors, such as user-specific factors (preference, psychological and behavioral states) and context factors (location, time and environmental conditions), need to be taken into consideration and we call them external factors in this paper. In order to establish a total understanding of QoE, the paper [16] proposed the concept of communication ecosystem that involved four factor domains, i.e., technical aspects, business models, human behavior and context. Though knowing the significant influences of external factors, few researchers successfully introduced them into QoE estimation and commonly regarded them as the introducers of randomness. This was mainly due to the fact that the external factors were beyond network monitoring scope and information about them was usually not available in real scenarios. In addition, the external factors influence QoE in a very complicated way, how to introduce them into estimation models appropriately remains unknown.

Mobile terminals are the carriers of the applications and they can provide various QoE-related information comprehensively, accurately and timely at the application level. Therefore, mobile terminals have unique advantages in QoE estimation, which has been proved by our previous work [17]. However, the abilities of mobile terminals go beyond that. Nowadays, mobile terminals, the network elements closest to users, are all embedded with various powerful sensors. They form a ubiquitous Internet of Things accompanying all mobile users, which makes it possible to extend the sensing range of the network to the physical space around human [18] and introduce external factors into QoE estimation. In this paper, we further propose a data-driven QoE estimation scheme for mobile video services in virtue of mobile terminals and validate the feasibility based on data from real-world scenarios. Different from previous QoE estimation models that just focus on technical factors, the proposed estimation scheme still takes non-technical factors into account. In order to acquire sufficient real-world data, we specially develop a mobile testing application and collect a mass of testing data with the help from China Unicom, the second-largest wireless operator in China. The collected data contain subjective scores about user experience and objective parameters including technical and non-technical parameters. To cope with the complicated unknown relationships among the data, we utilize decision tree to establish two estimation models that map the objective parameters to the subjective scores. The experimental results verify that the proposed scheme can realize an accurate QoE estimation, making it possible for the wireless network operators to monitor QoE through the massive ''peripheral sensors'' of the mobile networks. The main contributions of this work are listed as follows.

1) We comprehensively analyze the causal mechanism in mobile video communications and propose a QoE estimation scheme that takes full advantage of mobile terminals.
2) According to the proposed QoE estimation scheme, we set up a QoE data set for mobile video services through real-world collection.
3) Based on decision tree, we propose two estimation models with remarkable performances. Moreover, we obtain some quotable knowledge about QoE estimation in virtue of a series of statistical analyses.

In the rest of this paper, related works are introduced in Section II. Then, the data collection is described in Section III. In Section IV, we introduce the data preprocessing. Section V describes the estimation models and the corresponding training algorithms. In Section VI, we conduct performance evaluations and provide necessary discussions. Finally, we conclude this paper in Section VII

II. RELATED WORKS

Existing works related to various QoE estimation methods are reviewed in this section. The related works are classified into three categories, i.e., methods based on human response, methods based on human behavior and methods based on estimation model.

A. HUMAN RESPONSE BASED METHODS

There are two types of methods based on human response, which are distinguished by whether users are conscious with the response. Subjective test is usually regarded as synonymous with the method based on human conscious response. In a subjective test, users deliberately report their opinion scores about a specific service according to a predefined standard, such as mean opinion score (MOS) and difference mean opinion score (DMOS) [12]. Though subjective tests are costly and intrusive, they are still widely used to calibrate subjective quality for various QoE researches in the lab settings. In addition, various unconscious physiological responses are also used for QoE measurement. Electroencephalogram (EEG), which can record brain electrophysiological activities and decipher user instantaneous perception and cognitive processes, has been utilized to evaluate subjective quality of video services in both auditory domain [19] and visual domain [20], [21]. In the paper [22], multiple physiological signals, such as EEG, electrocardiograph (ECG) and respiration, were utilized for an alternative evaluation of human experience about immersive multimedia content. More details about physiological QoE assessment methods can be found in the survey [23]. Though physiological methods are informative to QoE estimation, they are rarely applied in practice since the measurement of most physiological responses needs specialized devices and testing environments.

B. HUMAN BEHAVIOR BASED METHODS

User online behaviors have been proved to be informative for various video services [24], [25]. Some of them are correlated with user experience and can reflect QoE to some extent. Therefore, it was proposed to derive behavior based metrics, such as viewing time [26], number of views [27] and abandonment rate [28], from the recorded service data for QoE evaluation. Such behavior based metrics enable us to infer user experience without the risk of annoying users, which makes them optional low-cost indicators of QoE. However, the evaluation accuracy of the behavior based metrics cannot be guaranteed and these metrics are obviously not suitable for real-time estimation.

C. ESTIMATION MODEL BASED METHODS

Estimating QoE through a model based on objective parameters can lead to a real-time and non-intrusive monitoring of QoE. Therefore, it is widely favored and studied. In early works, researchers attempted to establish a model of explicit mapping equation from objective parameters to QoE. Fiedler *et al.* [11] proposed a generic formula with a negative exponential form to characterize the quantitative relationship between QoE and a single QoS parameter. In the series of studies [29], [30], Hossfeld et al. studied the impacts of initial delay and stalling to user experience concerning Youtube video services. Similarly, they also discovered the exponential relationships between MOS and the two performance factors, respectively. These exponential models can reflect the marginal diminishing effect of user perception; however, a univariate model is inadequate for QoE estimation in practice. Gómez *et al.* [31] proposed a linear model containing initial buffering time, rebuffering frequency and mean rebuffering time, and further developed a YouTube QoE evaluation tool for Android terminals. In the paper [32], an explicit model was developed through regression analysis of three performance parameters at both the application and network levels. However, explicit models are usually generated by fitting a predefined expression to training data, which makes it difficult for the model to cover more influence factors without sufficient prior information about the complicated model structure.

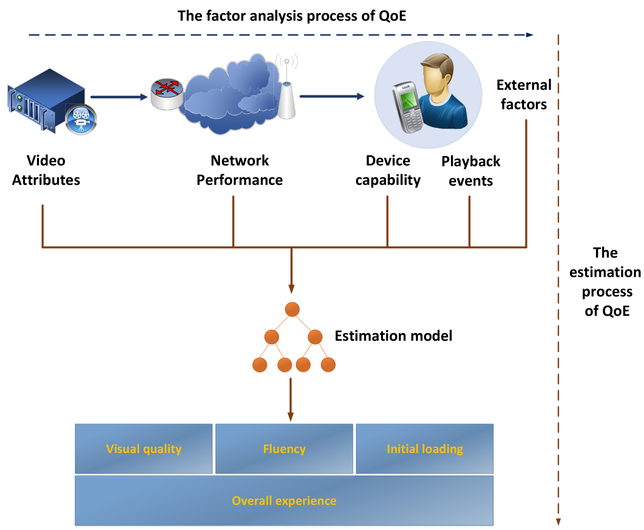

FIGURE 1. The mind map of the proposed approach, which includes the factor analysis process and the estimation process of QoE.

When more influence factors are taken into consideration, the models based on machine learning methods embody their advantages in dealing with the complex relationships between QoE and various QoS factors. In [33], thirteen QoS parameters were used for QoE prediction and the nearest neighbor approach was adopted due to the high-dimensionality of feature space. Owing to the outstanding ability of representation and prediction, some deep learning methods were successfully utilized for quality feature representation [34] and QoE prediction [35]. Moreover, some advanced machine learning methods such as boosting support vector regression [36] and adaptive network based fuzzy inference system (ANFIS) [37] were also utilized for QoE estimation.

III. DATA SET

In order to take full advantage of mobile terminals to obtain as much beneficial information as possible, we comprehensively analyzed the causal mechanism in mobile video communication and factorized the QoE generation process into technical factors and external factors (Fig. 1). The technical factors contained 11 objective parameters that could be classified into four types: video attributes, network performance, device capability and playback events. For the external factors, we utilized the acceleration sensors and light sensors embedded in mobile terminals to record the parameters about user movements and ambient illuminance condition. The two external parameters were selected since user perception of the video services can be greatly influenced by user movement state [38] and ambient illuminance [39]. In addition, both external parameters may contain informative scene knowledge. For instance, lighting information is tightly related with the user environmental conditions, such as outdoor or indoor. All the recorded objective parameters are listed and described in Table 1.

TABLE 1. **List of the recorded objective parameters.**

Number	Type	Parameter	Description
1	Video attributes	Video definition	240P, 360P, 720P and 1080P
2	Video attributes	Video bitrate	In kbps
3	Video attributes	Video data size	In kB
4	Network performance	Server response time	From the time the application sends out the data request to the time it receives the first piece of data, in ms.
5	Network performance	Initial loading peak rate	The peak downloading rate during the initial loading, in kbps.
6	Network performance	User perceived rate	The averaged downloading rate during the whole playback process (including stalling), in kbps.
7	Network performance	Global peak rate	The peak downloading rate from the playback of the first frame to the end of the playback, in kbps.
8	Device capability	Screen resolution	Number of pixels on the long side of the screen.
9	Playback events	Join time	The time from a user's request for playing a video to the playback of the first frame, in ms.
10	Playback events	Playback duration	The total time of the whole playback process (including stalling), in ms.
11	Playback events	Stalling ratio	The ratio of the total stalling time to the time of the whole playback process.
12	External factors	Ambient light	The time series of ambient illuminance intensity, in lx.
13	External factors	Movement	The time series of acceleration data in three dimensions, in m/s^2.

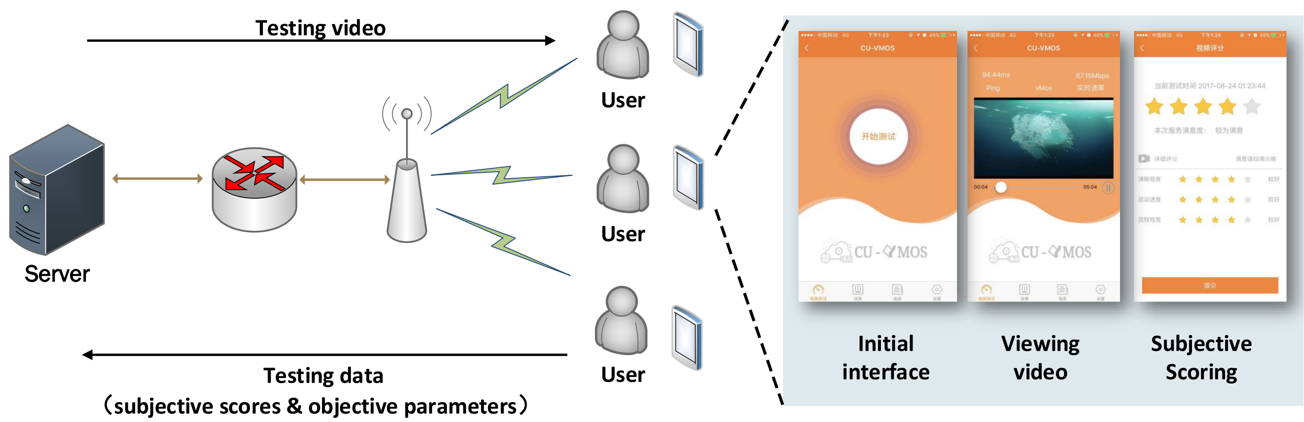

FIGURE 2. **The diagrammatic drawing of data collection.**

Moreover, the subjective ratings about the video quality in terms of four aspects, i.e., *visual quality*, *playback fluency*, *initial loading* (the satisfaction degree about the starting delay due to initial loading) and *overall experience*, were also recorded. In addition to the overall experience, the other three aspects were taken into account since the overall experience of video service is mainly determined by the three aspects [27]. Therefore, using the subentry scores as inter-layer components can help us to have a better understanding of the relationships between the objective parameters and QoE.

For clarity, the diagrammatic drawing of data collection is demonstrated in Fig. 2. We specially developed a testing application (app) called CU-vMOS which could be installed on both the Android system and the iOS system. A user started a test by pressing the starting button on the initial interface. Then, the app randomly requested a video from our pre-established server for the user. The video was played in a classic manner. First, the user experienced a starting delay caused by the initial loading until the first frame appeared. Over the course of the playback, the visual quality of the video maintained consistent and stalling might occur if the network performance was unable to meet the demand. Meanwhile, this app automatically recorded the 13 objective parameters in a non-intrusive way.

When the video was over, each user was asked to rate each of the four aspects. All the scores were on a 5-level scale according to the absolute category rating (ACR), which mapped the 1-5 points into the "Bad", "Poor", "Fair", "Good" and "Excellent" levels, respectively. After the user finished the rating, the opinion scores together with the recorded objective parameters were all transferred to the pre-established server.

The contents of testing videos were diversified, including cartoons, documentary films, variety shows, sport videos and so on. Each type of content had four different versions in terms of definition, i.e., 240P, 360P, 720P and 1080P. According to the statistics collected at Youtube, the majority of videos watched by the mobile users are short: nearly 90% of the videos are shorter than 5 minutes, and the average length is 160 seconds [31]. Therefore, our prepared videos were of three lengths, i.e., 30 seconds, 5 minutes and 10 minutes, which could be regarded as short, middle and long videos for the mobile scenarios, respectively. In total, 240 videos were prepared for the test.

Over six hundred professional mobile network inspectors from China Unicom took part in the data collection during their routing inspections. Nevertheless, the signal providers were not limited to China Unicom. The participants were elaborately instructed about the principles of the test and received appropriate reward for this task. From January to February 2018, a data set containing over eighty thousand data records was established for further study.

IV. DATA PREPROCESSING AND ANALYSIS

The data preprocessing consists of two steps, i.e., unifying the data form and cleaning the outliers.

A. DATA TRANSFORMATION

In the raw data set, the 11 technical parameters are measured by a single scalar quantity while the two external parameters are in the form of time series. Instead of using the two time series directly, we transform each time series into a single scalar quantity in advance.

The parameters of movement are in the form of time series of acceleration data, which are characterized in three dimensions. In order to transform them into a single scalar that can well reflect user movement intensity, the acceleration data are processed as the following steps. Denoting the three-dimensional acceleration at time t as $\vec{a}(t)$, we decompose $\vec{a}(t)$ by the following expression

$$\vec{a}(t) = \vec{g} + \vec{a}_{bg}(t) + \vec{a}_{tb}(t), \tag{1}$$

where $\vec{g}$, $\vec{a}_{bg}(t)$ and $\vec{a}_{tb}(t)$ are the gravitational acceleration, the acceleration of user's body relative to the ground and the acceleration of user's mobile terminal relative to user's body, respectively. $\vec{a}_{bg}(t)$ corresponds to the holistic motion intensity of user body while $\vec{a}_{tb}(t)$ can reflect the slight waggling of mobile terminal caused by the swing of the holding arm. Accordingly, the time average of $\vec{a}(t)$ over the course of video playback can be expressed as

$$\vec{a}_{mean} = \vec{g} + \overline{\vec{a}_{bg}} + \overline{\vec{a}_{tb}}, \tag{2}$$

where $\overline{\vec{a}_{bg}}$ and $\overline{\vec{a}_{tb}}$ are the time average of $\vec{a}_{bg}(t)$ and $\vec{a}_{tb}(t)$, respectively. Calculating the variance for $\vec{a}(t)$, we have

$$\begin{aligned} Var(\vec{a}(t)) &= \frac{1}{T}\sum_{t=1}^{T}(\vec{a}_{bg}(t) + \vec{a}_{tb}(t) - \overline{\vec{a}_{bg}} - \overline{\vec{a}_{tb}})^2 \\ &\approx \frac{1}{T}\sum_{t=1}^{T}[(\vec{a}_{bg}(t) - \overline{\vec{a}_{bg}})^2 + (\vec{a}_{tb}(t) - \overline{\vec{a}_{tb}})^2] \\ &= Var(\vec{a}_{bg}(t)) + Var(\vec{a}_{tb}(t)), \end{aligned} \tag{3}$$

where the approximation is based on the reasonable assumption that user holistic movement relative to the ground is uncorrelated or mildly correlated with the random arm swing. According to the basic principles of kinematics, it can be inferred that the variance of $\vec{a}_{bg}(t)$ and the variance of $\vec{a}_{tb}(t)$ are both positive correlated with the intensity and variability of user motion. Therefore, the approximate value of $Var(\vec{a}(t))$, which is the sum of $Var(\vec{a}_{bg}(t))$ and $Var(\vec{a}_{tb}(t))$, is used as an integrated scalar indicator of user motion intensity. In addition, considering that the length our testing video are all less than 10 minutes, we directly take the time average of the ambient light parameter to represent the overall ambient illuminance condition.

B. DATA CLEANING

As the first step of data cleaning, the fragmentary parts of the raw data are restored in virtue of the information redundancy in order to preserve more data for the subsequent analysis. After the unrecoverable data are deleted, the outliers are removed through correlation analysis. Specifically, we first denote the four subjective scores of overall experience, initial loading, playback fluency and visual quality by S_i $(i = 1, 2, 3, 4)$, respectively. Similarly, the 13 objective parameters are denoted by P_j, where the subscripts are in accordance with Table 1. Then, calculating the Spearman's rank correlation coefficients for each pair of subjective scores and objective parameters, we obtain a correlation coefficient matrix **CM**, whose elements are defined by the following equation [40]

$$\mathbf{CM}(i,j) = 1 - \frac{6\sum_n (rg(S_i^n) - rg(P_j^n))^2}{N(N^2-1)}, \tag{4}$$

where $rg(S_i^n)$ and $rg(P_j^n)$ denote the ranks of S_i^n and P_j^n, respectively. The superscript n indicates the index of all the samples and N is the total number of samples. The correlation coefficients of the 13 selected parameters and the 4 subjective scores are demonstrated in Fig. 3. It can be seen that most of the objective parameters demonstrate weak correlations with the four subjective scores. This should be unsurprising. First, a single objective parameter can just partly determine the perceived performances of the video services. In addition, unlike the tests conducted in a laboratory environment, the tests conducted in the real-world scenarios are inevitably influenced by unpredictable factors, which further reduce the correlations between the subjective scores and the considered objective parameters.

Then, the box plots based on the correlation coefficients are used to identify the outliers. The validity of a sample needs to be verified by each subjective score. To be specific, we first select four objective parameters that are most correlated with S_i according to the correlation coefficients as the primary parameters of S_i. Then, the data set is divided into five subsets D_i^l $(l = 1, 2, 3, 4, 5)$ according to the values of S_i and four box plots corresponding to the four primary parameters are depicted for each subset D_i^l. A sample belonging to D_i^l is verified to be valid for S_i if the sample is located within the interval $[Q_1 - 1.5IQR, Q_3 + 1.5IQR]$ in at least three box plots, where Q_1, Q_3 and IQR represent the first quartile, the third quartile and the interquartile range, respectively. The samples

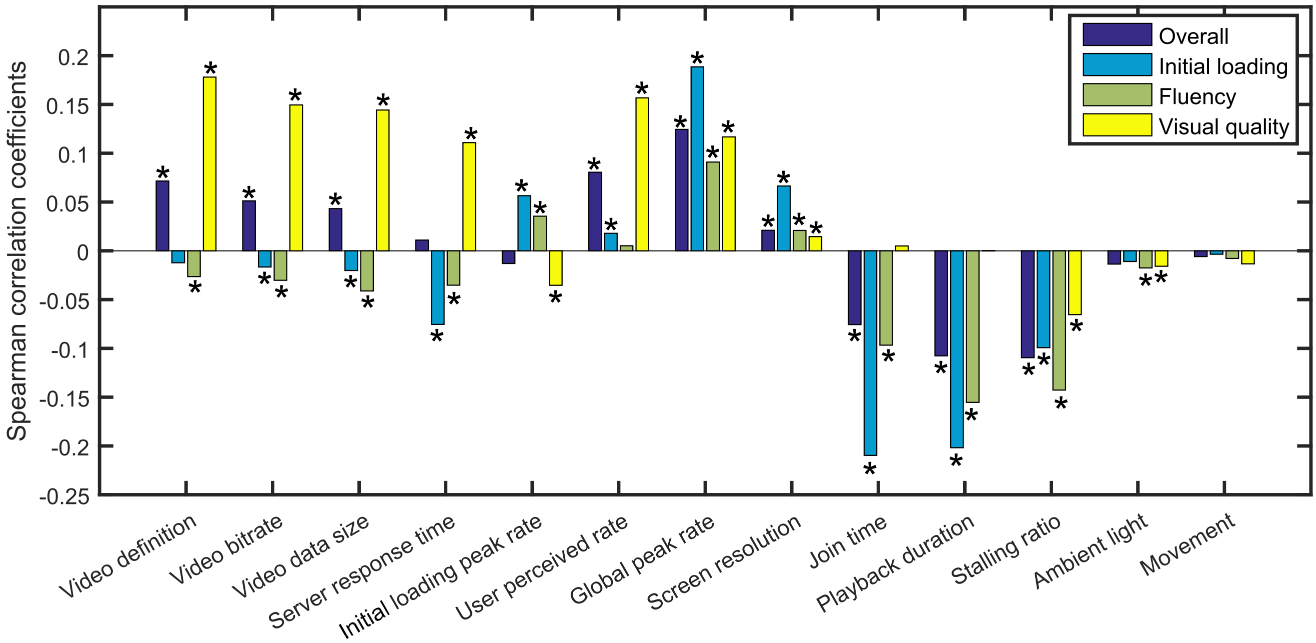

FIGURE 3. The correlation coefficients between the 4 subjective scores and the 13 objective parameters. The asterisks indicate result significance at $p < 0.01$.

that are valid for all four subjective scores are regarded as the valid samples. Finally, 29,778 valid samples are retained in the training data set D_{train} for further modeling.

V. ESTIMATION MODELS

In this paper, the decision tree method is utilized to develop the QoE estimation models. Among various machine learning approaches, we select the decision tree method for the following reasons. First, non-parametric models are better choices for QoE estimation considering that the relationships between the objective parameters and the subjective scores are so complex that it is difficult to assume a valid parametric form in advance. Second, the model needs to be able to cope with the interdependence between the objective parameters. In addition, decision tree has long been applied in decision analysis to help identifying a strategy most likely to reach a goal. Thus, a model based on decision tree may provide the wireless network operators with rewarding information to the benefit of root cause analysis and network optimization [41].

Regarding the 5 discrete values of each subjective score as the tags of 5 categories, we cast the estimation problems as classification problems. In the following subsections, we propose two estimation models that are based on classification tree and regression tree, respectively. Considering that the training processes of the estimation models of the four subjective scores are identical and independent, we omit the indices of different subjective scores in the following expressions of this section.

A. CLASSIFICATION TREE BASED MODEL

In this subsection, the estimation models of the subjective scores are established based on the classification tree method C4.5 [42], which is suitable for the problems whose input variables have both discrete and continuous forms. In this work, video definition is regarded as discrete since it has only four different values while the other objective parameters are regarded as continuous. The training algorithm starts at a root node, which contains all the training samples and is denoted by $\mathcal{R}$. Then, the information gain ratio (IGR) is calculated for each parameter as follows.

When an objective parameter P_m is continuous, each sample in the root node can be classified into one of the two child nodes $\mathcal{C}_{m,q}^1$ and $\mathcal{C}_{m,q}^2$ depending on whether P_m of the sample is smaller or larger than a specific splitting value $v_{m,q}$, where q indexes all possible splitting values. The IGR of this splitting scheme is defined as:

$$IGR(v_{m,q}) = \frac{H(S) - \sum_{r\in 1,2} \frac{|\mathcal{C}_{m,q}^r|}{|\mathcal{R}|} H(S|P_m \in \mathcal{C}_{m,q}^r)}{H(P_m)}, \quad (5)$$

where $H(S)$ and $H(P_m)$ are the entropies of all the samples at the root node with respect to the subjective score S and the objective parameter P_m, respectively. $H(S|P_m \in \mathcal{C}_{m,q}^r)$ is the conditional entropy of S given P_m belonging to $\mathcal{C}_{m,q}^r$. The largest value of all possible $IGR(v_{m,q})$ is selected as the IGR of P_m at the root node, i.e., $IGR_m = \max_q\{IGR(v_{m,q})\}$.

When an objective parameter P_m is discrete, the samples in the root node are directly divided into multiple child nodes corresponding to each value of P_m. Denoting the rth child node by $\mathcal{C}_m^r$, the IGR of P_m is directly defined as:

$$IGR_m = \frac{H(S) - \sum_r \frac{|\mathcal{C}_m^r|}{|\mathcal{R}|} H(S|P_m \in \mathcal{C}_m^r)}{H(P_m)}, \quad (6)$$

whose notations are similar to those of the equation (5).

The parameter with the largest IGR is selected to be the splitting parameter of the root node and the root node is split into multiple child nodes based on the corresponding splitting scheme. Then, regarding each child node of $\mathcal{R}$ as a new root node, the training algorithm further splits the corresponding subtrees in the same way as introduced above. The algorithm keeps on expanding the classification tree recursively until one of the following conditions is satisfied at the current node.

- All the samples have the same subjective score.
- The number of the samples is no larger than a predefined value N_{min}.
- All the samples have the same value of the splitting parameter.

We call the nodes that stop further splitting the leaf nodes. At a certain leaf node, the subjective score corresponding to the largest sample proportion is selected to be the estimated score for any sample belonging to the leaf node.

The training procedures of the classification tree based estimation model are listed in Algorithm 1. The algorithm builds a classification tree by a recursive function BUILD_CT whose inputs are D_{train} and N_{min}. The algorithm controls the complexity of the classification tree by the model parameter N_{min} since it determines the minimum sample number requirement in a node for further splitting. The optimal value of N_{min} is determined by cross-validation. The output of BUILD_CT is a multi-layer nested struct T and the struct layers from the outside to the inside correspond to the tree-structure layers from the root to the leaves. N_s is the number of discrete values of P_s and the function Maj means taking the score with the largest proportion.

Algorithm 1 Training Algorithm of the Classification Tree Based Estimation Model

Input: the training data set D_{train}; minimum sample number for splitting N_{min};
Output: the classification tree T: the child nodes $T.child$, the splitting parameter index $T.dim$, the splitting value $T.spl$;

function $T = \text{BUILD\_CT}(D_{train}, N_{min})$
if $|D_{train}| \leq N_{min}$ **or** the values of S in D_{train} are identical **then**
  return $T.child = \text{Maj}(D_{train})$, $T.dim = 0$, $T.spl = \text{inf}$;
end if
calculate IGR_m for each objective parameter P_m; determine the splitting parameter P_s and the splitting value v_s by $IGR_s = \max_m\{IGR_m\}$;
if the values of P_s in D_{train} are identical **then**
  return $T.child = \text{Maj}(D_{train})$, $T.dim = 0$, $T.spl = \text{inf}$;
else
  $T.dim = s$;
end if
if P_s is discrete **then**
  $T.spl = \text{inf}$;
  divide D_{train} into N_s subsets SD_i $(i = 1, 2, \ldots, N_s)$ according to the values of P_s, calculate the child nodes recursively $T.child(i) = \text{BUILD\_CT}(SD_i, N_{min})$;
else
  $T.spl = v_s$;
  divide D_{train} into 2 subsets SD_j $(j = 1, 2)$ by splitting the values of P_s by v_s, calculate the child nodes recursively $T.child(j) = \text{BUILD\_CT}(SD_j, N_{min})$;
end if
return $T.child, T.dim, T.spl$;
end function

B. REGRESSION TREE BASED MODEL

In addition to the classification tree C4.5, we utilize the gradient boosting decision tree (GBDT) method [43], which conducts multi-class logistic regression by multiple boosting trees, to establish another estimation model. In this model, 5 estimation trees, which are denoted by T_l $(l = 1, 2, 3, 4, 5)$, are built for each of the four subjective scores. When being used to estimate the subjective score S of a new data sample $\mathbf{x}$, each of the five estimation trees produces a classification score $T_l(\mathbf{x})$ related to the possibility that $S(\mathbf{x}) = l$. Finally, $S(\mathbf{x})$ is estimated to be k $(k = 1, 2, 3, 4, 5)$ if $T_k(\mathbf{x}) = \max_l\{T_l(\mathbf{x})\}$.

As the first step, the training algorithm translates the subjective score of the nth training sample $\mathbf{x}_n$ to a 5×1 vector $\mathbf{h}_n$ according to the one-hot coding scheme, where the lth element $\mathbf{h}_n(l)$ equals one while the other elements equal zero. Then, according to the multi-class logistic regression, the probability that $S(\mathbf{x}_n) = l$ can be defined as

$$c_l(\mathbf{x}_n) = \frac{e^{T_l(\mathbf{x}_n)}}{\sum_{k=1}^{5} e^{T_k(\mathbf{x}_n)}}. \tag{7}$$

The corresponding likelihood function can be expressed as

$$p(\mathbf{H}|T_1, T_2, \ldots, T_5) = \prod_{n=1}^{N}\prod_{k=1}^{5} c_k(\mathbf{x}_n)^{\mathbf{h}_n(k)}, \tag{8}$$

where $\mathbf{H}$ is a $N \times 5$ matrix with the nth row being $\mathbf{h}_n^{\top}$. Taking the negative logarithm for the equation (8), we have the cross-entropy error function

$$E(T_1, T_2, \ldots, T_5) = -\sum_{n=1}^{N}\sum_{k=1}^{5} \mathbf{h}_n(k) \ln c_k(\mathbf{x}_n). \tag{9}$$

Calculating the gradient of the equation (9) with respect to the classification score $T_l(\mathbf{x}_n)$, we obtain

$$\nabla_{T_l(\mathbf{x}_n)} E(T_1, T_2, \ldots, T_5) = c_l(\mathbf{x}_n) - \mathbf{h}_n(l). \tag{10}$$

Then, the estimation trees are trained iteratively. First, the training data set D_{train} is randomly divided into a training subset D_1 and a test subset D_2 according to the proportions $1 - p_{test}$ and p_{test}, respectively. The training process starts from an initial state $T_1^{(0)}, T_2^{(0)}, \ldots, T_5^{(0)}$. In the uth round $(u = 1, 2, 3, \ldots)$, the residual error between the target value $\mathbf{h}_n(l)$ and the estimated value $c_l^{(u-1)}(\mathbf{x}_n)$ is calculated for each $\mathbf{x}_n$, i.e.,

$$r_{n,l}^{(u)} = \mathbf{h}_n(l) - c_l^{(u-1)}(\mathbf{x}_n), \tag{11}$$

Algorithm 2 Training Algorithm of the Regression Tree Based Estimation Model

Input: the training data set D_{train}; the training parameters ρ_l, d_{max}, p_{test}, r_{stop}, r_{max};

Output: Regression tree T_1, T_2, T_3, T_4, T_5;

```
Initialization: u = 1, T_1^(0), T_2^(0), T_3^(0), T_4^(0), T_5^(0);
randomly divide D_train into a training subset D_1 and a test
    subset D_2 by the ratio p_test;
while u ≤ r_max do
    for l = 1 to 5 do
        calculate the residual data set R_l^(u) on D_1, d = 1;
        function T̂_l^(u) = BUILD_RT(R_l^(u), d, d_max)
        if d ≥ d_max or r_{n,l}^(u) are identical for all x_n or N_tr ≤ 1
        then
            return T̂_l^(u).child = mean_n(r_{n,l}^(u)), T̂_l^(u).dim = 0,
                T̂_l^(u).spl = inf;
        end if
        search for the optimal splitting scheme (P_t, v_t);
        if the values of P_t in D_1 are identical then
            return T̂_l^(u).child = mean_n(r_{n,l}^(u)), T̂_l^(u).dim = 0,
                T̂_l^(u).spl = inf;
        else
            T̂_l^(u).dim = t, T̂_l^(u).spl = v_t;
            divide R_l^(u) into 2 subsets R_l^(u)(1) and
                R_l^(u)(2) by the splitting scheme (P_t, v_t),
                recursively calculate the child nodes
                T̂_l^(u).child(j) = BUILD_RT(R_l^(u)(j), d + 1, d_max)
                (j = 1, 2);
        end if
        return T̂_l^(u).child, T̂_l^(u).dim, T̂_l^(u).spl;
        end function
        T_l^(u) = T_l^(u-1) + ρ_l T̂_l^(u);
    end for
    if the test performance on D_2 has not been promoted
        for r_stop rounds then
        break;
    end if
end while
```

where $c_l^{(u-1)}(\mathbf{x}_n)$ is calculated by $T_l^{(u-1)}$ according to the equation (7). Then, the residual data set of T_l in this round is defined as

$$\mathbf{R}_l^{(u)} = \{[\mathbf{x}_n^\top, r_{n,l}^{(u)}] | n = 1, 2, \ldots, N_{tr}\}, \tag{12}$$

where N_{tr} is sample number of D_1.

The algorithm fits a boosting tree $\widehat{T}_l^{(u)}$ to $\mathbf{R}_l^{(u)}$, where $\widehat{T}_l^{(u)}$ is a binary regression tree with the maximal depth d_{max}. The training process of $\widehat{T}_l^{(u)}$ starts at a root node containing all the training samples. Then, the samples in the root node are divided into two child nodes $\mathbf{R}_l^{(u)}(1)$ and $\mathbf{R}_l^{(u)}(2)$ according to a splitting scheme that are characterized by an objective parameter P_m and a splitting value $v_{m,q}$. The splitting scheme (P_t, v_t) that minimizes the sum of intra-class variations of the S in the two child nodes is selected as the optimal scheme.

Treating each child node of the root node as a new root node, the training algorithm keeps on expanding the corresponding subtrees by splitting the new root nodes in the same way. The boosting tree is expanded recursively till the leaf nodes, at which one of the following conditions is satisfied.

- The depth of the current node has reached the maximal depth d_{max}.
- All the samples have the same residual error.
- The number of the samples is no more than 1.
- All the samples have the same value of the splitting parameter.

For the samples in the same leaf node, the residual errors are all set to be the mean value of themselves. Finally, each of the five estimation trees is updated according to the equation

$$T_l^{(u)} = T_l^{(u-1)} + \rho_l \widehat{T}_l^{(u)}, \tag{13}$$

where ρ_l is the learning step size for T_l.

At the end of each round, the algorithm evaluates the model performance on the test subset D_2. The iteration process stops when there is no performance improvement for over r_{stop} rounds or the total number of iterations reaches r_{max}, where the early stoping round number r_{stop} and the maximal number of iterations r_{max} are predefined to avoid overfitting. The training procedures of the regression tree based estimation model are listed in Algorithm 2, where each boosting tree is built by the recursive function BUILD_RT in a similar way as BUILD_CT in Algorithm 1.

VI. PERFORMANCE EVALUATION AND DISCUSSION

In this section, the performances of the two proposed estimation models are evaluated. In addition, we provide necessary discussions on the characteristics of the data for a more comprehensive understanding of the performances.

A. ESTIMATION PERFORMANCES

For the classification tree based model, the parameter N_{min} needs to be determined first. As demonstrated in Fig. 4, the relationships between the estimation performances and model complexity are obtained through 10-fold cross-validation, where R_{min} is the percentage of N_{min} to the total number of training data N, i.e., $N_{min} = \frac{N \cdot R_{min}}{100}$. It can be seen that as the model complexity increases, the estimation performances on the training data get promoted significantly while the estimation performances corresponding to the test data tend towards the opposite direction. It is evident that an overly complicated classification tree model will lead to the problem of overfitting. In order to avoid overfitting, N_{min} is assigned the value corresponding to the best estimation performance on the test data. Therefore, we set $R_{min} = 10, 2, 10, 10$ for the estimation of overall experience, initial loading, playback fluency and visual quality, respectively.

For the GBDT based model, the model parameters are set as $\rho_l = 0.01$, $D_{max} = 8$, $p_{test} = 0.1$, $r_{stop} = 80$ and $r_{max} = 4000$. For comparison, we also evaluate the estimation performances of three classical machine learning methods: linear discriminant analysis (LDA), K nearest

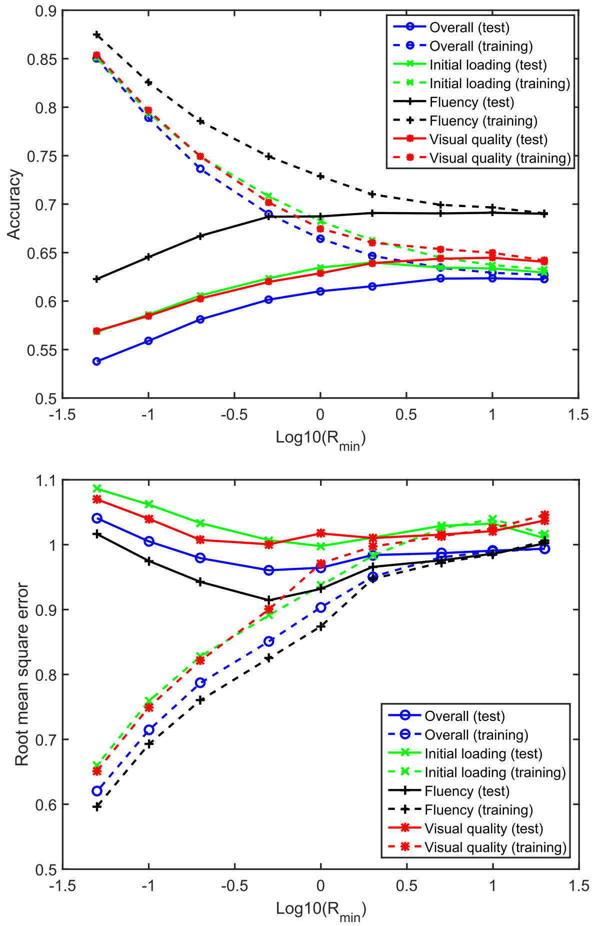

FIGURE 4. **The quantitative relationship between estimation performance (the upper: Classification accuracy, the lower: Root mean square error) and complexity of the classification tree based model by 10-fold cross-validation.**

neighbor (KNN) and the support vector machine (SVM). All the experimental results are obtained through 10-fold cross-validation.

As shown in Fig. 5, both of the proposed models outperform the other three baseline models, which indicates that the models with tree-formed structure are indeed more suitable for QoE estimation. Specifically, the estimation accuracies of the C4.5 based model range between 60% and 70% for the four subjective scores, while the GBDT based method demonstrates better performances that the overall estimation accuracy approaches 80% and the accuracy for playback fluency reaches 82.4%. In addition, considering that wireless network operators typically adopt a 3-level scale (good, medium and bad) to evaluate user experience in practice, we transform the score ranges of 1 point, 2-3 points and 4-5 points into the "bad", "medium" and "good" levels, respectively. As demonstrated in Fig. 6, the estimation performances of both proposed models corresponding to the 3-level scale are raised above 80% (except for the visual quality of C4.5) and the overall performance of the GBDT based model approaches 90%. Considering the large variance in subjective opinions among different users (cf. Section VI-D) and the complicated conditions in

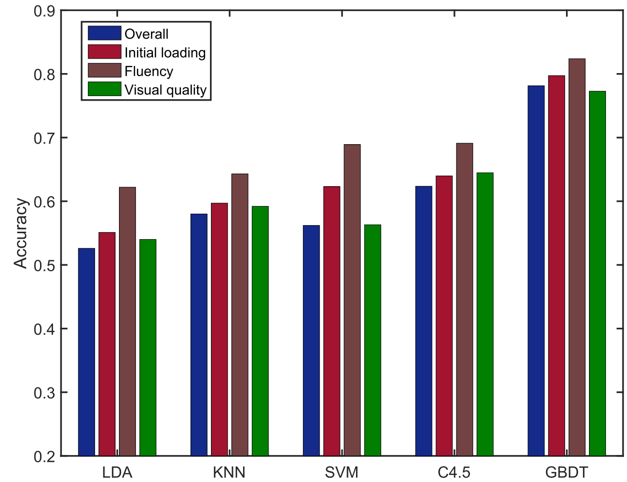

FIGURE 5. **The estimation performances of the proposed decision tree based methods and the other baseline methods on the 5-level scale.**

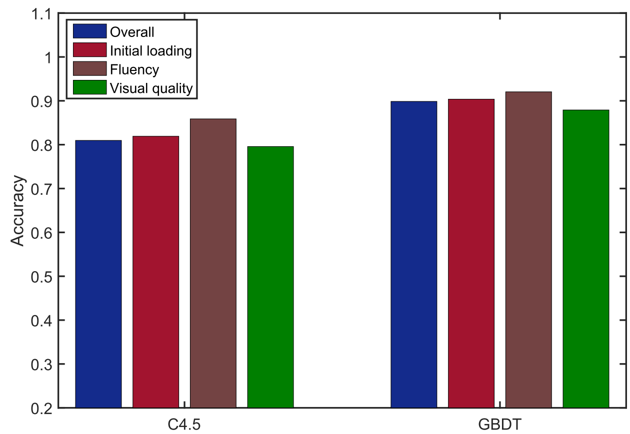

FIGURE 6. **The estimation performances of the proposed decision tree based methods on the practical 3-level scale.**

real-world scenarios, such performance level is remarkable, which validates the feasibility of the proposed QoE estimation approach based on mobile terminals.

It should be noted that the performance advantage of the GBDT based model compared to the C4.5 based model results from the power of boosting trees. However, we should not blurt out that the GBDT based model absolutely overmatches the C4.5 based model. The intuitional structure of the C4.5 based model preserves important information about the statistical relationships between the subjective scores and the objective parameters, which makes the C4.5 based model more explicable and analyzable than the GBDT based model. When detecting a QoE degradation in a certain area, the wireless network operators can conduct root cause analysis based on the structure of the decision tree and seek for an effective solution. Comparing the merits and demerits of both models, wireless network operators can select an appropriate model according to practical needs.

B. PARAMETER SELECTION

To have a better understanding of the estimation performance, the statistical relationships between the objective parameters

TABLE 2. The main parameters of each subjective score.

Order	1	2	3	4	5	6
Overall	Global peak rate	Stalling ratio	Playback duration	User perceived rate	Join time	Video definition
Initial loading	Join time	Playback duration	Global peak rate	Stalling ratio	Server response time	Screen resolution
Fluency	Playback duration	Stalling ratio	Join time	Global peak rate	Video data size	Initial loading peak rate
Visual quality	Video definition	User perceived rate	Video bitrate	Video data size	Global peak rate	Server response time

and the subjective scores are discussed in this subsection. As demonstrated in Fig. 3, most of the correlation coefficients are in accordance with our common sense, which confirms the validity of the collected data. For instance, a longer join time, which may be attributed to a higher video definition or a poorer network performance, is likely to lower the score of initial loading. Therefore, the join time is negatively correlated with the scores of initial loading and playback fluency but positively correlated with the score of visual quality. Moreover, the global peak rate directly indicates the network capacity. It is straightforward that the global peak rate should be positively correlated with the scores of initial loading, playback fluency and overall experience, respectively. Interestingly, there also exists a positive correlation with the global peak rate and the visual quality score, which may be explained by the correlations among the three subscores (cf. Section VI-C).

As demonstrated in Table 2, we list the six most correlated parameters, which are called the main parameters, for each subjective score. Not surprisingly, the scores of initial loading and playback fluency mainly depend on the parameters related to network performance and playback events, while the score of visual quality mainly depends on video attributes. It is worth noting that the first five main parameters of the overall experience are all related to network performance and playback events, which means that network performance is still the main restricting factor of QoE.

In this paper, we specially introduce external factors into QoE estimation. Though the external parameters show weak correlations with the subjective scores, it is necessary to find out how much performance improvement the external factors can contribute to the QoE estimation. To this end, we first remove the movement parameter and the ambient light parameter, and then conduct model training based on the abridged data in the same way as introduced above. As shown in Fig. 7, the GBDT based model obtains considerable performance improvement from the external factors, which demonstrates the significance of external factors for QoE estimation. Meanwhile, there is little performance improvement for the C4.5 based model and this can be explained by the characteristics of classification tree. As introduced in Section V, C4.5 makes each split only based on the parameter with the maximal information gain ratio. Therefore, the external factors which show slight correlations with the subjective scores can hardly be selected and the beneficial information can not be fully utilized by the C4.5 based model.

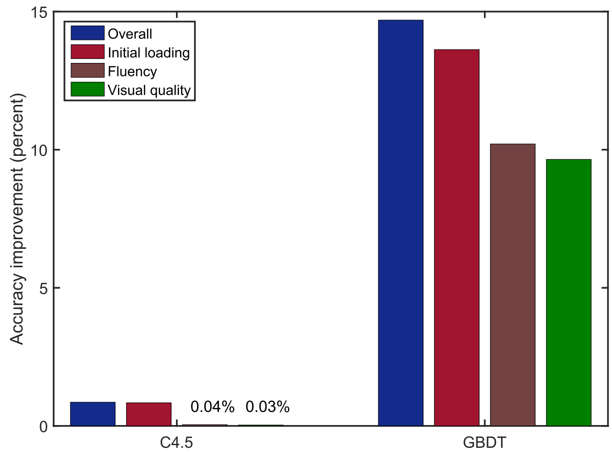

FIGURE 7. The performance improvement due to the external parameters (5-level scale).

C. THE RELATIONSHIPS AMONG THE SUBJECTIVE SCORES

To have an insight into the statistical relationships among the subjective scores, we demonstrate the correlation matrix of the subjective scores in Table 3. It can be seen that the three subscores all have a strong positive correlation with the overall score and the correlation coefficients are close to each other. In addition, the score of visual quality is also significantly correlated with the other two subscores, which does not fit with our common sense. This phenomenon may result from positive interactions among the subjective scores, e.g., a user may be more tolerant to visual quality degradation if the video is played fluently. Moreover, comparing the estimation results demonstrated in Fig. 5, we can find that almost all the models (except for GBDT) provide the lowest estimation accuracy for the overall score. Compared to the overall score, the three subscores are more closely and directly related to the objective parameters and they are less influenced by the variance and fluctuation of subjective opinion, which eventually leads to higher estimation accuracy. This phenomenon verifies the assumption that the three subentries can be regarded as interlayer components that provide auxiliary information leading to a more comprehensive understanding of QoE.

D. USER RATING DIVERSITY

In this paper, we aim at developing a universal QoE monitoring approach for various mobile users. Therefore, user rating diversity inevitably introduces error to the estimation performances. The paper [44] systematically analyzed the standard deviation of opinion scores (SOS) caused by user

TABLE 3. Correlations among the four subjective scores.

Correlations	Overall	Initial loading	Fluency	Visual quality
Overall	1	0.78	0.81	0.83
Initial loading	0.78	1	0.76	0.69
Fluency	0.81	0.76	1	0.69
Visual quality	0.83	0.69	0.69	1

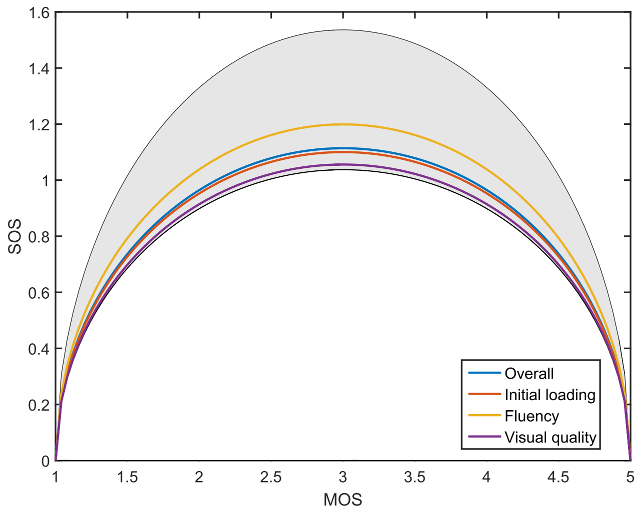

FIGURE 8. The MOS-SOS curves of the four subjective scores in our test.

rating diversity and formulated the SOS hypothesis, which postulates a square relationship between MOS and SOS:

$$SOS^2 = a(-MOS^2 + 6 \times MOS - 5), \tag{14}$$

where a is a characteristic parameter that varies with the tested crowd and rating conditions. The value of a is obtained by data fitting and can be used as a direct measurement of user rating diversity.

In the same way, we divide all the data samples into multiple subsets based on the proposed classification tree model, where each subset corresponds to a leaf node of the tree. Then, the samples in a certain subset are considered approximately congeneric since they have similar objective parameters. Calculating the MOS and SOS values of each subset for each subjective score and fitting the function (14) to them, we obtain the four MOS-SOS curves and depict them in Fig. 8. The shaded area indicates the empirical variation range discovered in [44] and is provided for comparison. It can be seen that the user rating diversity in our test is normal but it indeed significantly contributes to the estimation error.

User rating diversity mainly results from the variations of user specific factors and context factors. Therefore, it is possible to reduce user rating diversity with more information about the related factors. In the future, we attempt to further exploit the capabilities of mobile terminals to sense more beneficial information and develop customized QoE estimation models based on long-term QoE data for a more refined network monitoring.

VII. CONCLUSION

In this paper, we proposed a QoE monitoring approach based on mobile terminals for video services. First, we comprehensively analyzed the causal mechanism in mobile video communications and designed a QoE estimation approach that took full advantage of mobile terminals. To explore the feasibility of the proposed approach, we developed a mobile testing app and established a video service data set containing over eighty thousand data records from various real-world scenarios. In addition to the various technical factors, two external factors related to user movement and environment lighting condition were specially recorded for further exploration. We preprocessed the raw data based on correlation analysis and obtained the training data set. In consideration of the specific requirements of estimation problem, two estimation models were established based on the C4.5 method and the GBDT method, respectively. The experimental results demonstrated that both tree-formed estimation models outperformed other baseline machine learning models. The remarkable estimation accuracies proved that it is feasible for the wireless network operators to monitor QoE efficiently based on mobile terminals. In addition, the discussions demonstrated that the external factors were informative for QoE estimation and the user rating diversity significantly contributed to the estimation error. These discoveries motivate us to further exploit the capabilities of mobile terminals for QoE estimation and develop customized QoE estimation models for a more refined QoE monitoring in the future.

ACKNOWLEDGMENT

This paper received a great support from the China Unicom Research Institute in terms of data collection and technical consultation.

REFERENCES

[1] Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," Cisco Syst. Inc., White Paper 1551296909190103, 2019.

[2] M. Jia, X. Gu, Q. Guo, W. Xiang, and N. Zhang, "Broadband hybrid satellite-terrestrial communication systems based on cognitive radio toward 5G," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 96–106, Dec. 2016.

[3] E. Liotou, H. Elshaer, R. Schatz, R. Irmer, M. Dohler, N. Passas, and L. Merakos, "Shaping QoE in the 5G ecosystem," in *Proc. 7th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, May 2015, pp. 1–6.

[4] M. Jia, Z. Yin, D. Li, Q. Guo, and X. Gu, "Toward improved offloading efficiency of data transmission in the IoT-cloud by leveraging secure truncating OFDM," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4252–4261, Jun. 2019.

[5] J. Liu, X. Tao, and J. Lu, "QoE-oriented rate adaptation for DASH with enhanced deep Q-learning," *IEEE Access*, vol. 7, pp. 8454–8469, 2019.

[6] A. El Essaili, D. Schroeder, E. Steinbach, D. Staehle, and M. Shehada, "QoE-based traffic and resource management for adaptive HTTP video delivery in LTE," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 6, pp. 988–1001, Jun. 2015.

[7] E. Liotou, K. Samdanis, E. Pateromichelakis, N. Passas, and L. Merakos, "QoE-SDN APP: A rate-guided QoE-aware SDN-APP for HTTP adaptive video streaming," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 598–615, Mar. 2018.

[8] J. Yun, M. J. Piran, and D. Y. Suh, "QoE-driven resource allocation for live video streaming over D2D-underlaid 5G cellular networks," *IEEE Access*, vol. 6, pp. 72563–72580, 2018.

[9] T. Zhao, Q. Liu, and C. W. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 285–302, 1st Quart., 2017.

[10] K. Gai and M. Qiu, "Optimal resource allocation using reinforcement learning for IoT content-centric services," *Appl. Soft Comput.*, vol. 70, pp. 12–21, Sep. 2018.

[11] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Netw.*, vol. 24, no. 2, pp. 36–41, Mar./Apr. 2010.

[12] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1126–1165, 2nd Quart., 2015.

[13] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the quality of experience of HTTP video streaming," in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manage. (IM)*, May 2011, pp. 485–492

[14] W. Pan and G. Cheng, "QoE assessment of encrypted YouTube adaptive streaming for energy saving in smart cities," *IEEE Access*, vol. 6, pp. 25142–25156, 2018.

[15] T. Hoßfeld, F. Liers, R. Schatz, B. Staehle, D. Staehle, T. Volkert, and F. Wamser, "Quality of experience management for YouTube: Clouds, fog and the AquareYoum," *PIK-Praxis der Informationsverarbeitung und Kommunikation*, vol. 35, no. 3, pp. 133–143, 2012.

[16] K. U. R. Laghari, N. Crespi, and K. Connelly, "Toward total quality of experience: A QoE model in a communication ecosystem," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 58–65, Apr. 2012.

[17] X. Liu, X. Tao, L. Wang, Y. Zhan, and J. Lu, "Developing a QoE monitoring approach for video service based on mobile terminals," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2019, pp. 480–485.

[18] K. Gai and M. Qiu, "Reinforcement learning-based content-centric services in mobile sensing," *IEEE Netw.*, vol. 32, no. 4, pp. 34–39, Jul./Aug. 2018.

[19] A. K. Porbadnigk, M. S. Treder, B. Blankertz, J.-N. Antons, R. Schleicher, S. Möller, G. Curio, and K.-R. Müller, "Single-trial analysis of the neural correlates of speech quality perception," *J. Neural Eng.*, vol. 10, no. 5, 2013, Art. no. 056003.

[20] S. Bosse, L. Acqualagna, W. Samek, A. K. Porbadnigk, G. Curio, B. Blankertz, K.-R. Müller, and T. Wiegand, "Assessing perceived image quality using steady-state visual evoked potentials and spatio-spectral decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1694–1706, Aug. 2018.

[21] X. Liu, X. Tao, and Y. Zhan, "Calibrating human perception threshold of video distortion using EEG," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3543–3547.

[22] A.-F. N. M. Perrin, H. Xu, E. Kroupi, and M. Řeřábek, and T. Ebrahimi, "Multimodal dataset for assessment of quality of experience in immersive multimedia," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 1007–1010.

[23] U. Engelke, D. P. Darcy, G. H. Mulliken, S. Bosse, M. G. Martini, S. Arndt, J.-N. Antons, K. Y. Chan, N. Ramzan, and K. Brunnström, "Psychophysiology-based QoE assessment: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 6–21, Feb. 2017.

[24] L. Chen, Y. Zhou, and D. M. Chiu, "A study of user behavior in online VoD services," *Comput. Commun.*, vol. 46, pp. 66–75, Jun. 2014.

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

[26] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "A quest for an Internet video quality-of-experience metric," in *Proc. ACM Workshop Hot Topics Netw.*, 2012, pp. 97–102.

[27] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," in *Proc. ACM SIGCOMM Conf.*, 2011, pp. 362–373.

[28] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 2001–2014, Dec. 2013.

[29] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, "Quantification of YouTube QoE via crowdsourcing," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2011, pp. 494–499.

[30] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: Between the devil and the deep blue sea," in *Proc. 4th Int. Workshop Qual. Multimedia Exper.*, Jul. 2012, pp. 1–6.

[31] G. Gómez, L. Hortigüela, Q. Pérez, J. Lorca, R. García, and M. C. Aguayo-Torres, "YouTube QoE evaluation tool for Android wireless terminals," *EURASIP J. Wireless Commun. Netw.*, vol. 2014, no. 1, p. 164, 2014.

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

[33] M. Venkataraman and M. Chatterjee, "Inferring video QoE in real time," *IEEE Netw.*, vol. 25, no. 1, pp. 4–13, Jan./Feb. 2011.

[34] H. Zhang, H. Hu, G. Gao, Y. Wen, and K. Guan, "DeepQoE: A unified framework for learning to predict video QoE," 2018, *arXiv:1804.03481*. [Online]. Available: https://arxiv.org/abs/1804.03481

[35] M. T. Vega, D. C. Mocanu, J. Famaey, S. Stavrou, and A. Liotta, "Deep learning for quality assessment in live video streaming," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 736–740, Jun. 2017.

[36] Y. B. Youssef, M. Afif, R. Ksantini, and S. Tabbane, "A novel QoE model based on boosting support vector regression," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[37] L. Amour, S. Sami, S. Hoceini, and A. Mellouk, "Building a large dataset for model-based QoE prediction in the mobile environment," in *Proc. 18th ACM Int. Conf. Modeling, Anal. Simulation Wireless Mobile Syst.*, 2015, pp. 313–317.

[38] H.-I. M. Copcu and H.-I. Cheng, "The quality of contextual experience of multimedia on the smartphone," *Int. J. Emerg. Sci. Eng.*, vol. 3, pp. 30–33, Jan. 2015.

[39] K. De Moor, B. Andrén, Y. Guo, K. Brunnström, K. Wang, A. Drott, and D. S. Hermann, "Influence of high ambient illuminance and display luminance on readability and subjective preference," *Proc. SPIE*, vol. 9394, Mar. 2015, Art. no. 93941E. doi: 10.1117/12.2078380.

[40] L. A. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Berlin, Germany: Springer, 2013.

[41] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for Internet video," in *Proc. ACM Conf. SIGCOMM*, 2013, pp. 339–350.

[42] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1992.

[43] R. E. Schapire and Y. Freund, *Boosting: Foundations and Algorithms*. Cambridge, MA, USA: MIT Press, 2012.

[44] T. Hoßfeld, R. Schatz, and S. Egger, "SOS: The MOS is not enough!" in *Proc. QoMEX*, Sep. 2011, pp. 131–136.

XIWEN LIU received the B.E. degree in electronic and information engineering and the M.E. degree in communication and information system from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Wireless Multimedia Communication Laboratory, Tsinghua University. His research interests include subjective evaluation of multimedia and the quality of experience of wireless communication.

XIAOMING TAO received the Ph.D. degree in information and communication system from Tsinghua University, in 2008, where she is currently a Professor with the Department of Electronic Engineering. She served as a Workshop General Co-Chair for IEEE INFOCOM 2015, and also the volunteer leadership for IEEE ICIP 2017. She has been an Editor of the *Journal of Communications and Information Networks (JCIN)* and *China Communication*, since 2016. She was a recipient of the National Science Foundation for Outstanding Youth (2017–2019) and many national awards, including the 2017 China Young Women Scientists Award, the 2017 Top Ten Outstanding Scientists and Technologists from China Institute of Electronics, the 2017 First Prize of Wu Wen Jun AI Science and Technology Award, the 2016 National Award for Technological Invention Progress, the 2015 Science and Technology Award of China Institute of Communications.

YAFENG ZHAN received the B.S.E.E. and Ph.D.E.E. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 1999 and 2004, respectively, where he is currently an Associate Professor with the Space Center. His current research interests include communication signal processing and deep space communications.

JIANHUA LU (M'98–SM'07–F'15) received the B.S.E.E. and M.S.E.E. degrees from Tsinghua University, Beijing, China, in 1986 and 1989, respectively, and the Ph.D. degree in electrical and electronic engineering from The Hong Kong University of Science and Technology. Since 1989, he has been with the Department of Electronic Engineering, Tsinghua University, where he currently serves as a Professor. He has published more than 180 technical papers in international journals and conference proceedings. His current research interests include broadband wireless communication, multimedia signal processing, and wireless networking. He is currently a Chief Scientist of the National Basic Research Program (973), China. He received the National Distinguished Young Scholar Fund from the NSF Committee of China, in 2005. He is a fellow of the IEEE Communication Society and the IEEE Signal Processing Society. He has served in numerous IEEE conferences as a member of technical program committees. He has been an active member of professional societies. He was one of the recipients of the Best Paper Awards at the IEEE International Conference on Communications, Circuits and Systems, in 2002, ChinaCom 2006, and the IEEE Embedded-Com 2012. He served as the Lead Chair for the General Symposium of IEEE ICC 2008 and the Program Committee Co-Chair for the 9th IEEE International Conference on Cognitive Informatics, in 2010.