User Experience Perception Algorithm for Long and Short Videos Based on Multiple Nonlinear Regression

Zuojun Dai, Ying Zhou, Hui Tian, Nan Ma, and Yuexia Zhang

1School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
2Key Laboratory of Modern Measurement & Control Technology, Ministry of Education, Beijing Information Science and Technology University, Beijing 100101, China
3Key Laboratory of Information and Communication Systems, Ministry of Information Industry, Beijing Information Science and Technology University, Beijing 100101, China

Correspondence should be addressed to Hui Tian; tianhui@bupt.edu.cn

Received 8 August 2022; Accepted 5 September 2022; Published 26 September 2022

Copyright © 2022 Zuojun Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The low correlation of evaluation indices from the current user perception evaluation model and the neglect of the nonlinear relationship between diversified indices and user experience in different duration videos result in low user perception accuracy of long and short videos. To address these issues, we propose a user experience perception algorithm for long and short videos based on multiple nonlinear regression (LSMNR). First, to improve the efficiency and accuracy of modeling, the algorithm involves preprocessing of video data in edge servers and subdivides the videos based on their duration and popularity. Then, we introduce a new multidimensional quantitative evaluation index that fits the user’s subjective experience and further analyze the influence between multiple evaluation indices (video lag, black screen, etc.) and user quality of experience (QoE) for different video types. Moreover, the characteristics of the data in the multiple evaluation indices are extracted; user subjective evaluation experiments are designed using the video quality expert group (VQEG) standard; and sample and test databases were established. Finally, the optimal model parameters were trained by applying the nonlinear least square method and support vector machine (SVM) to fit and cross-verify the sample data. Our simulation results revealed that the Pearson correlation coefficient of the proposed LSMNR algorithm acquires a value of 0.9810. Compared with algorithms based on multinomial linear regression (MLR), linear SVM, and neural network (NN), the perceptual accuracy of the proposed algorithm is improved by at least 4.0%, and it is applicable to a wider range of video types.

1. Introduction

At present, owing to the extremely limited attention span of users, short video streaming services with short time and fast sharing speed and broad market prospects have become the norm of social media. Meanwhile, with the rapid development of short video platforms such as TikTok, Bilibili, and YouTube and fierce competition from service providers [1, 2], the demand for an enhanced user experience has increased. Particularly, a massive good video viewing experience determines whether video providers and platforms can win over new consumers as well as retain the loyalty of existing consumers, thereby providing a competitive advantage to the platform [3–5]. Therefore, it is of great significance to study the user experience perception algorithm of long and short videos.

To accurately describe users’ video perception experience, the international telecommunication union telecommunication (ITU-T) standardization sector released the user quality of experience (QoE) standard, using mean opinion score (MOS) to reflect users’ subjective video experience [6]. On this basis, most studies on video user perception relied on mobile edge computing (MEC) to localize the ability to deal with massive data and the advantages of high computing and storage data, thus analyzing and processing a large number of data streams to extract multiple network...
characteristic indicators (such as transmission delay, jitter, etc.). And they used a neural network, linear regression, and other methods to analyze the correlation between network characteristics and QoE [7–9]. Subsequently, the user QoE perception model was established to accurately perceive the video quality. However, with the increase in extraction index types, the newly introduced redundant index exhibits a low correlation with QoE, and the linearity in the relationship between the multivariate index and user subjective experience is not obvious. In addition to the high complexity of its algorithm, the expected fine-grained perception accuracy could not be achieved using the traditional multiple linear model, which was used to explain the relationship between multiple indicators and QoE. Meanwhile, a large number of redundant new index data collected from the network are difficult to process and analyze quickly, which leads to an increase in algorithm complexity and other problems.

Moreover, most of the current online video streaming media user QoE perception algorithms are designed for long videos [10, 11], which cannot be directly applied to evaluate the current short video user perception. Contrary to researching the relationship between user QoE and evaluation indices such as picture quality and network fluctuation in a poor network environment on 3D or live long video data, short video data are more sensitive to some features such as lag and black screen but less sensitive to indicators such as picture quality and frame rate. In fact, when users watch short videos online (16 s or less in length), they tend to choose to watch nonlagging popular videos with little regard to the video quality. Thus, the current perception algorithms are limited to evaluating the user experience of traditional long videos. Therefore, it is urgent to further explore the relationship between new evaluation indicators and user QoE for videos with varying duration and popularity.

To address the above-mentioned limitations of existing algorithms, this study proposes a user experience perception algorithm for long and short videos based on multiple nonlinear regression (LSMNR). First, to improve the efficiency and accuracy of modeling, the raw video data are mined and analyzed to classify the videos based on their duration and popularity in edge servers. Subsequently, a new multivariate quantitative evaluation index fitting the subjective experience of users is introduced to analyze the relationship between the user experience perception and the key quality indicator (KQI); success rate of video playback, average times of delay, etc.), quality of service (QoS) variable indicators (video playback buffering duration, black screen duration, etc.), and quantitative parameters (such as jitter and packet loss rate, etc.) for different types of videos. Thus, the data characteristics of multiple evaluation indices are extracted; user subjective experiments are designed; and the sample and test databases are established. Furthermore, the nonlinear least square method and support vector machine (SVM) are used to fit and cross-verify the sample data; the model parameters are obtained by the training model; and the perception results are further optimized. Finally, the validity and accuracy of the perception model are verified using test data.

The main contributions of this study are summarized as follows:

1. This study subdivides video types according to duration and popularity and establishes an original video database containing 6 different video types to improve the precision and accuracy of modeling.

2. This study designs a user subjective evaluation experiment based on short and long videos. In this experiment, 20 experimenters scored the videos they watched according to the MOS standard, and 6,000 sets of sample data and test data were generated. The relationship between video evaluation indicators and user experience quality under different duration and popularity was further analyzed to realize the fine-grained experience perception of long and short video users.

3. This study proposes an LSMNR algorithm and introduces a new multivariate quantitative evaluation index that fits the subjective experience of users. On this basis, combined with user perception data in subjective experiments, multiple perceptual models with different model parameters are obtained by training the sample data and test data of different video types. Then, the nonlinear relationship between these indicators and user experience in different types of videos is analyzed. And methods such as nonlinear least square method and SVM are used to train sample data to achieve optimal perception results.

4. Our simulation results revealed that compared with algorithms based on multinomial linear regression (MLR), SVM, and neural network (NN), the perceptual accuracy of the proposed algorithm is greater by approximately 4.0%, and it is applicable to a wider range of video types.

The remainder of this study is organized as follows. In Section 2, we discuss some related work, and in Section 3, we describe the system model and its constituent modules. In Section 4, we introduce the specific process and parameter setting of the user subjective evaluation experiment. In Section 5, we propose the LSMNR algorithm and introduce how to train the perceptual model. In Section 6, we present the simulation results and discuss them. Finally, in Section 7, we discuss the significance of the research results and summarize the study.

2. Related Work

In recent years, combined with MEC technology, researchers have proposed various data analysis methods to study evaluation indices that predominantly affect the QoE of users [12–14] and establish the user experience perception model of the high-precision video business. García-Pineda et al. [15] proposed an MLR algorithm, with a high perception accuracy, to analyze the relationship between evaluation indicators such as delay jitter and packet loss rate and QoE. Shi and Huang [16] proposed a video perception
algorithm based on a decision tree, which comprehensively considered the mapping relationship between grid and video parameters (such as noise standard deviation, ambiguity and block effect, etc.) and users’ subjective feelings, and designed a fuzzy logic system to give objective scores, which greatly improved the accuracy of user experience perception. Considering the low accuracy and easy overfitting of traditional methods such as linear regression and decision tree, researchers began to construct QoE perception models for videos using machine learning methods. For example, Hao and Qingbing [17] proposed a new video quality perception evaluation method based on a convolutional neural network, which extracted spatial and time domain information of distorted videos, realized the evaluation of end-to-end video quality, and solved the problem of insufficient training samples of distorted videos and their long training time, which were difficult to converge. Bulkam and Dagiuklas [18] proposed a linear-based SVM and NN algorithm. They conducted a subjective survey on users’ viewing experience, and a sample database was constructed, and indicators with high correlation with QoE were extracted from videos such as total pause duration and initial buffer duration for training tests, which effectively reduced modeling time and improved perceptual accuracy. In addition, to improve perceptual accuracy, some researchers try to extract the coding data of the original video and compare and analyze the images frame by frame by preprocessing the video information, to objectively evaluate the user QoE based on the characteristics of the damaged images. However, it is important to note that there are many factors that can influence the QoE evaluation index; the impact of these numerous factors tends to have a complex relationship, such as some of the features and user QoE exhibit a simple linear independent relationship, whereas some have an irregular nonlinear relationship. The above work only considers the linear relationship between these factors, and a direct analysis of these factors increases the model training time substantially and limits the model accuracy and other problems.

Additionally, some studies consider that the correlation between influencing factors in different types of videos and QoE is not the same [19, 20]; therefore, the design of corresponding models can effectively reduce the training time and improve perceptual accuracy. For example, by subdividing video types (virtual reality video, live video, etc.), special perception models can be designed to meet specific requirements [21, 22]. Guzmán et al. [23] proposed a user experience evaluation algorithm suitable for 3D video streaming media services to test the performance of adaptive coding for 2D and 3D videos and to objectively evaluate the 3D videos according to different parameters such as the resolution, duration, and average bit rate, to further improve the accuracy of model evaluation. Fei et al. [24] proposed a virtual reality (VR) panoramic video user experience perception algorithm based on four indicators, namely, the quality, immersive index, nonrotation, and an overall score, for subjective evaluation, and subsequently designed an objective evaluation method using the input parameters of bit rate, delay, and packet loss. Simulation results show that it can effectively improve perception accuracy. Moreover, Chen et al. [25] proposed a user experience perception algorithm suitable for live videos, which uses multiscale temporal relation reasoning to discover the internal connection between ordered frames, capture short- and long-term distortion perception changes, and fine-grained estimation of QoE from the perspective of video understanding. Thus, the above work analyzes the mapping between different factors and user perception based on video type, without considering the influence of time and popularity on QoE. Therefore, the new video user perception algorithm should also incorporate the influence of these two parameters in addition to the other important factors and further analyze their impact on QoE of short and long videos.

Different from existing research work, this study mainly studies the nonlinear relationship between user experience and new multiple evaluation indicators (video lag, black screen, jitter, packet loss rate, etc.) that fit the subjective experience of users under different video types such as duration and popularity. To improve the accuracy and accuracy of modeling, subjective user perception experiments are designed, and the original video sample database and test database are established. Meanwhile, this study proposes a user experience perception algorithm based on multiple nonlinear regression for long and short videos, further analyzes the nonlinear relationship between new evaluation indicators and user experience under different types of videos, and uses the nonlinear least square method and SVM to train sample data, thus achieving optimal perception results.

3. System Model

In this study, we propose an LSMNR algorithm. First, we establish a user experience perception system model of long and short videos based on multiple nonlinear regressions, as shown in Figure 1. In the figure, the model is mainly composed of four parts, including the edge server, big data preprocessing module, feature extraction module, and user perception training module. The edge server is responsible for providing and collecting raw video and generating the video database. The big data preprocessing module is responsible for preprocessing video service data in the 5G network. The feature extraction module extracts the key information that affects the user perception evaluation accuracy from the original data. The training module of the user perception model is responsible for training data and objective user experience evaluation of the original video.

In the system model, video service data in a 5G network is preprocessed firstly, followed by subdivision of video types based on duration and popularity, and subsequently, the key information affecting user perception accuracy (KQI, QoS, and quantified indicators) in the original data is collected and processed.

Thereafter, the eigenvalues of multiple evaluation indices in the video services are extracted, and the sample database and test database are generated based on the results of the subjective tests. Then, using the multivariate nonlinear regression algorithm, the nonlinear relationship between the
multiple variable indices and the user QoE is established by adopting the method of least squares and SVM, fitting the sample data, and cross-validation. Finally, the output presents the user experience perception evaluation results of short and long videos.

3.1. Big Data Preprocessing Module. It is necessary to consider the subjective and objective factors such as network lag, buffering delay, and so on, which affect the accuracy of user perception in long and short video services. During actual online viewing of long and short videos, traditional abstract network quantization parameters cannot further fine-grain the perception of subjective user experience in videos of different duration (such as short videos of less than 16 s and long videos of more than 30 min). Conversely, users’ tolerance for the video quality of popular video content, unpopular content, and other content is quite different. Therefore, this study considers mining and analysis of video business big data, further refining video classification based on duration and popularity in edge servers, and establishing different sample databases to further optimize user perception accuracy in developing the subsequent training model.

3.2. Feature Extraction Module. Additionally, this study comprehensively considers the influence of multiple index variables such as KQI, QoS, and the quantitative index on perception accuracy and ensures that the variable index conforms to the user experience perception law by extracting characteristic information of all the variables. First, in the video business, the KQI mainly reflects the subjective feelings of users’ perceived experience, including the success rate of video playback and the average number of video playback freezes. The success rate of video playback is the percentage of videos that can be successfully played when the same type of video samples are collected on a 5G network. Its expression is as follows:

$$KQIS = \frac{T_{ca}}{T_{to}} \times 100\%,$$

where $T_{ca}$ indicates the number of times that a user can successfully watch the same type of video and $T_{to}$ indicates the total number of times that a user requests to play the same type of video.

The average number of video playback freezes KQIP represents the times of video playback freezes owing to various factors in the 5G network, and its expression is as follows:

$$KQIP = \frac{T_{pa}}{T_{su}},$$

where $T_{pa}$ represents the number of video stalling from the beginning to the end and $T_{su}$ represents the total number of successes in playing a video of the same type.

The QoS index in video services can also reflect users’ objective perception of experience to a certain extent, including the black screen and the buffering durations. QoSBC refers to the duration of a black screen in a poor network environment when the same type of videos is collected on a 5G network. Its expression is as follows:

$$QoSBC = \frac{T_{BC}}{T_{su}} T_{B},$$

where $T_{BC}$ indicates the number of black screen times and $T_{B}$ indicates the duration of each black screen.

The buffer duration of video playback QoS_L is the estimated time from users clicking the video link to the beginning of video playback when the 5G network collects the same type of video samples. The expression is as follows:

$$QoS_L = t_{ca} - t_{re},$$
where \( t_{ca} \) represents the time when sufficient cached data is transmitted to start playing the video and \( t_{rc} \) represents the time when the user requests the same video. In addition, in case there is an advertisement in the video, the playing time of the advertisement is not counted in \( t_{ca} \).

Furthermore, network-layer QoS parameters of video services, such as jitter delay \( t_{ja} \) and packet loss rate \( t_{sl} \), also affect users’ perceived experience quality and can be collected using Wireshark.

### 3.3. User Perception Model Training Module

On account of extracting KQI, QoS, and the quantitative parameters, a corresponding database was generated. Then, based on the subjective evaluation experiments carried out in the follow-up work, we established sample and test databases. A user perception evaluation model based on a multivariate nonlinear regression algorithm is built to further analyze the nonlinear relationship between multivariate variables and user experience quality, thus enabling the perception of a fine-grained, highly perceived user experience. Among them, the basic expression of the user perception evaluation model of the multiple nonlinear regression algorithm is as follows:

\[
Y = \Phi(x_1, x_2, \ldots, x_n, \theta_1, \theta_2, \ldots, \theta_n) + \delta,
\]

where \( \theta = [\theta_1, \theta_2, \ldots, \theta_n] \) is the model parameter, \( X = [x_1, x_2, \ldots, x_n] \) is an \( n \)-dimensional model vector, and \( x = [\text{KQI}_t, \text{QoS}_p, \text{QoS}_{bc}, \text{QoS}_{st}, t_{ja}, t_{sl}] \). \( \delta \) represents other possible disturbances in the model, and \( \delta \)'s mathematical expectation \( E(\delta) \) is usually set to 0.

Moreover, it is assumed that the training sample data is \((x_i, y_i)\), where \( x_i \) represents the \( i \)-th input parameter Deleted of the \( j \)-th sample, and \( y_j \) represents the output result of the \( j \)-th sample. The basic expression of the model can be further simplified as follows:

\[
y_j = g_j(x, \theta) + e_j,
\]

where \( e_i \) denotes the random error in the model. Also, the error function obeys independent identically distributed. In the process of evaluating user QoE, six groups of model parameters were trained according to different video types to establish different perception models.

Subsequently, it is considered that the parameter estimation problem of nonlinear regression models is mainly to estimate the values of model parameters through known input data observations. Therefore, it is necessary to estimate the model parameter \( \theta \), and the square sum of its standard deviation is as follows:

\[
f(\theta) = \sum_{j=1}^{M} (y_j - g_j(x, \theta))^2,
\]

where \( M \) represents the total number of training samples. Considering that the purpose of training an evaluation model is to minimize \( f(\theta) \) on the training sample set, let the function \( g \) be continuously differentiable to the model parameter \( \theta \), and obtain the model parameter by taking the partial derivative of \( \theta_j \):

\[
\frac{\partial f(\theta)}{\partial \theta_j} \bigg|_{\theta_j = \hat{\theta}_j} = -2 \sum_{j=1}^{M} (y_j - g_j(x, \hat{\theta})) \frac{\partial g_j}{\partial \theta_j} \bigg|_{\theta_j = \hat{\theta}_j} = 0.
\]

### 4. User Subjective Evaluation Experiment

To evaluate the validity and accuracy of the proposed length video user perception model, at the same time, we consider that there are currently no published data and corresponding databases on subjective experiments of long and short videos. In this study, 300 of the original videos from TikTok, Bilibili, and other platforms were downloaded, and the video database was further established for the subsequent subjective user QoE evaluation experiment. The database contains 150 short videos of less than or equal to 16 s and 150 long videos of more than 30 min. Table 1 presents the specific parameter settings of the long and short videos.

In each of the databases, we divided the videos into three types: popular, common, and unpopular video content according to the number of videos played and the number of likes. The resolution of the videos is 1,280 x 720, and they are royalty-free. We used the Premiere and MATLAB software to simulate the effects of buffering, stalling, and black screen in the video playback process and encapsulate the original video based on the H.264 standard while exporting the videos. We also used the QNET software to imitate the video playback effect in the differential network environment and set different delay jitter and packet loss rates to deal with the original video.

In this study, 20 subjects from the Beijing University of Information Science and Technology participated in the subjective evaluation experiment. All the subjects were between 20 and 30 years of age and had normal eyesight. The subjective evaluation experiment was carried out according to the environment and light setting specified by the video quality experts group (VQEG) [26]. The subjects rated the videos they watched according to the MOS standard, which is presented in Table 2.

Moreover, VQEG gives the root mean square error (RMSE) and Pearson correlation coefficient (PCC) as criteria to measure the effectiveness and accuracy of the perceptual algorithm. Among them, the value of RMSE reflects the deviation between the evaluation value calculated by the objective model and the subjective user perception. The closer the value is to 0, the more accurate the evaluation of the model. The expression is as follows:

\[
\text{RMSE} = \sqrt{\frac{1}{M} \sum_{j=1}^{M} (Y_j - y_j)^2},
\]

where \( M \) is the total number of original video samples and \( Y_j \) is the evaluation value after watching the \( j \)-th video in the user’s subjective experiment.

In addition, most algorithms use PCC to judge the accuracy of the video quality perception model and use the linear correlation between the evaluated value and the real value as the evaluation criterion. The size of the PCC is...
between \([-1, 1]\), and the closer its value is to 1, the better the effect of the evaluation model. The PCC expression is as follows:

\[
PCC = \frac{\sum_{j=1}^{M}(Y_j - \overline{Y})(y_j - \overline{y})}{\sqrt{\sum_{j=1}^{M}(Y_j - \overline{Y})^2} \times \sqrt{\sum_{j=1}^{M}(y_j - \overline{y})^2}},
\]

where \(\overline{Y}\) is the mean value of the user’s subjective experiment score set and \(\overline{y}\) is the model to calculate the mean value of the evaluation set.

5. LSMNR Algorithm

In this study, based on the subjective experimental results of users to generate sample and test databases, a short and long video user experience perception algorithm based on multiple nonlinear regression (LSMNR) is proposed. In the process of training the perceptual model, we establish an independent perceptual model by subdividing the video types. Similar to the traditional multiple linear regression, the optimal model parameters of the perceptual model are obtained by mapping the input variables to high-dimensional space and fitting them by applying the nonlinear least square method and the SVM method, respectively.

First, the optimal model parameters of the perceptual model are obtained using nonlinear least square method. In Figure 2, video sample data numbered from 1 to 50 are selected. Among them, we assume that every five videos are a group, and in the multivariate independent variable \(X\) of video samples with different numbers in this group, all other variables are kept unchanged, except for one variable that changes linearly. And then observe the relationship between the multivariate independent variable \(X\) and the dependent variable \(Y\). For example, in the multivariate independent variable \(X\) of video samples numbered 1 to 5 (abscissa value), only the packet loss rate increases linearly, and the remaining multivariate independent variables do not change. However, the corresponding dependent variable \(Y\) (ordinate MOS value) in Figure 2 does not increase linearly, exponentially, or logarithmically. Similarly, among the multivariate independent variables \(X\) of video samples

### Table 1: Long and short video parameter settings.

<table>
<thead>
<tr>
<th>Video case number</th>
<th>Video playback success rate</th>
<th>Average number of video playback freezes</th>
<th>Buffering time (second)</th>
<th>Video black screen duration (second)</th>
<th>Delay jitter</th>
<th>Packet loss probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–50 short videos (popular content)</td>
<td>100%, 90%, 80%, 70%, 60%</td>
<td>0, 1, 2, 3, 4</td>
<td>0, 1, 2, 3, 4</td>
<td>0, 1, 2, 3, 4</td>
<td>0 ms, 0%; 100 ms, 2%; 200 ms, 4%; 300 ms, 6%; 400 ms, 8%</td>
<td>0%, 2%, 4%, 6%, 8%</td>
</tr>
<tr>
<td>51–100 short videos (common content)</td>
<td>100%, 90%, 80%, 70%, 60%</td>
<td>0, 1, 2, 3, 4</td>
<td>0, 1, 2, 3, 4</td>
<td>0, 1, 2, 3, 4</td>
<td>0 ms, 0%</td>
<td>0%</td>
</tr>
<tr>
<td>101–150 short videos (unpopular content)</td>
<td>100%, 90%, 80%, 70%, 60%</td>
<td>0, 1, 2, 3, 4</td>
<td>0, 1, 2, 3, 4</td>
<td>0, 1, 2, 3, 4</td>
<td>0 ms, 0%</td>
<td>0%</td>
</tr>
<tr>
<td>151–200 long videos (popular content)</td>
<td>100%, 90%, 80%, 70%, 60%</td>
<td>0, 2, 4, 6, 8</td>
<td>0, 2, 4, 6, 8</td>
<td>0, 2, 4, 6, 8</td>
<td>0 ms, 0%; 100 ms, 2%; 200 ms, 4%; 300 ms, 6%; 400 ms, 8%</td>
<td>0%, 2%, 4%, 6%, 8%</td>
</tr>
<tr>
<td>201–250 short videos (common content)</td>
<td>100%, 90%, 80%, 70%, 60%</td>
<td>0, 2, 4, 6, 8</td>
<td>0, 2, 4, 6, 8</td>
<td>0, 2, 4, 6, 8</td>
<td>0 ms, 0%</td>
<td>0%</td>
</tr>
<tr>
<td>251–300 short videos (unpopular content)</td>
<td>100%, 90%, 80%, 70%, 60%</td>
<td>0, 2, 4, 6, 8</td>
<td>0, 2, 4, 6, 8</td>
<td>0, 2, 4, 6, 8</td>
<td>0 ms, 0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

### Table 2: Mean opinion score (MOS).

<table>
<thead>
<tr>
<th>MOS</th>
<th>QoE</th>
<th>Extent of damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bad</td>
<td>There is a degree of damage that you find annoying</td>
</tr>
<tr>
<td>2</td>
<td>Pass</td>
<td>There is a certain degree of damage that is annoying</td>
</tr>
<tr>
<td>3</td>
<td>General</td>
<td>There is a degree of damage that is slightly annoying</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
<td>There is some damage, but it is acceptable</td>
</tr>
<tr>
<td>5</td>
<td>Excellent</td>
<td>No damage to the quality of the video was detected, and the viewing quality was excellent</td>
</tr>
</tbody>
</table>
numbered 6 to 10 (abscissa value), only the delay jitter increases linearly, while the other multivariate independent variables do not change. However, the corresponding dependent variable \( Y \) (ordinate MOS value) does not increase linearly, exponentially, or logarithmically. Therefore, it can be seen that the relationship between multivariate independent variable \( X \) and dependent variable \( Y \) is not exponential or logarithmic and cannot be directly expressed by common functional relations. At this time, the nonlinear least square problem cannot construct a system of matrix linear equations. Therefore, we tried to transform it into a linear least square problem by first-order Taylor expansion. The process of model training is to solve the following problems:

\[
\min \sum_{j=1}^{M} (Y_j - y_j)^2 = \min \sum_{j=1}^{M} \left( Y_j - g_j(x_k, \theta) - \nabla g_j(x, \theta)(x - x_k) \right)^2 \\
\min \sum_{j=1}^{M} \left( e_j - f_j(x - x_k) \right)^2 = \min \| e - f(x - x_k) \|,
\]

where \( \nabla g_j(x, \theta) \) is the first derivative of the function \( g_j(x, \theta) \), and it can also be expressed as the \( j \)-th row of the Jacobian matrix \( f_j \), \( e_j = Y_j - g_j(x_k, \theta) \).

In addition, the optimal model parameters of the perceptual model are obtained using the nonlinear SVM method. We introduced appropriate kernel functions to map the low-dimensional inseparable space vector to the high-dimensional feature space so that the issue of optimizing the nonlinear solution obtained after its training is also transformed into solving the dual problem. Let \( \omega^T x + b = 0 \) be the separation hyperplane, \( \omega \) is the normal vector, and \( b \) is the hyperplane intercept. The model training process requires solving the following problem:

\[
\min \frac{1}{2} \omega^T \omega + \lambda \sum_{j=1}^{M} \zeta_j,
\]

where \( \lambda > 0 \) is the penalty tradeoff factor and \( \zeta_j \geq 0 \) is the relaxation variable, indicating tolerance of moderate classification errors. Let the kernel function be \( K(x_{m'}, x_{n'}) = \exp (-\gamma \| x_{m'} - x_{n'} \|^2) \), where \( \gamma \) is a radial basis function parameter, which usually adopts the value of 0.125. Thereafter, (12) can be converted into

\[
\min_{\alpha} \frac{1}{2} \sum_{m'=1}^{M} \sum_{n'=1}^{M} \alpha_{m'} \alpha_{n'} y_{m'} y_{n'} K(x_{m'}, x_{n'}) - \sum_{j=1}^{M} \alpha_j \\
\text{s.t.} \sum_{j=1}^{M} \alpha_j y_j = 0 \\
0 \leq \alpha_j \leq C,
\]

where \( \alpha \) is the Lagrange multiplier vector and \( C \) is the penalty factor.

In summary, different model parameters are calculated based on the above two methods. Consequently, the test samples are imported, and the effectiveness of the training output is compared with the subjective evaluation results according to different types of videos, as shown in Figure 3. By selecting the model parameters with high precision, the user perception evaluation model corresponding to the video type is established followed by the realization of a fine-grained perception.

6. Simulations and Discussion

Based on the MATLAB platform and combined with the subjective evaluation experimental data, this study also performed simulations regarding the perception of the QoE of long and short videos for users. Meanwhile, the accuracy of the perception model in this algorithm is compared with that in MLR, SVM, and NN algorithms, and the impact of different factors on the quality of user experience under different video types is simulated and analyzed.

The simulations compare the changes in the user experience with different types of videos under different evaluation models, as shown in Figures 4 and 5. The horizontal axis in Figure 4 represents popular video, common video, and unpopular video. The horizontal axis in Figure 5 represents the video with a long duration and the video with a short duration. In Figures 4 and 5, the vertical axis represents the Pearson correlation coefficient (PCC). As can be seen from Figure 4, the user perception accuracy of all algorithms for popular videos is much lower than that of ordinary and unpopular videos. This is because there are many factors that affect the perceived user experience in different popularity videos, and the correlation between the evaluation indicators selected in the existing algorithms and the user perceived experience is low. Thus, when users watch popular videos, they have higher expectations in terms of the quality of the video content and lower tolerance for video damage, which makes it difficult for the algorithm to accurately evaluate the subjective experience of actual users based on objective evaluation indicators. As can be seen
from Figure 5, the user perception accuracy of all algorithms for long videos is higher than that of short videos, while the accuracy of the LSMNR algorithm for video users is much higher than that of MLR, SVM, NN, and other algorithms. This is because the current online video streaming media perception algorithm is designed for long videos, and the evaluation indicators they study have a higher correlation and higher perception accuracy with the user perception experience of a long video. However, the LSMNR algorithm proposed in this paper refines different evaluation indicators for long as well as short videos and introduces evaluation indicators with high correlation such as video lag and black screen, which can train different model parameters according to different video types to achieve fine-grained user perception accuracy. Moreover, as can be seen from Table 3, the perceptual accuracy of LSMNR, MLR, SVM, and NN for popular videos is 0.9645, 0.9125, 0.8962, and 0.9140, respectively, 0.9851, 0.9250, 0.9594, and 0.9590 for ordinary videos and 0.9771, 0.8957, 0.9481, and 0.9432 for unpopular videos, respectively. The LSMNR, MLR, SVM, and NN algorithms developed for short videos exhibit a perceptual accuracy of 0.9870, 0.9037, 0.9459, and 0.9401, respectively. For long videos, the perceptual accuracy is 0.9912, 0.9912, 0.9556, and 0.9592, respectively. To sum up, the average perception accuracy of this algorithm is 0.9810, which is improved by 7.7%, 4.2%, and 4.0%, compared with other
algorithms. Thus, it can be inferred that this algorithm has higher perception accuracy and is suitable for a wide range of video types.

The simulation compares the correlation between the subjective evaluation of the user and the objective evaluation of the perceptual model, as well as the performance of the model, as shown in Figure 6. In the figure, the horizontal axis represents the MOS benchmark value, and the vertical axis represents the objective evaluation score of the perceptual algorithm model after training. The distance from the purple

<table>
<thead>
<tr>
<th>Video types/Evaluation models</th>
<th>Popular video</th>
<th>Common video</th>
<th>Unpopular video</th>
<th>Long video</th>
<th>Short video</th>
<th>Average perception accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSMNR</td>
<td>0.9645</td>
<td>0.9851</td>
<td>0.9771</td>
<td>0.9912</td>
<td>0.9870</td>
<td>0.9810</td>
</tr>
<tr>
<td>MLR</td>
<td>0.9125</td>
<td>0.9250</td>
<td>0.8957</td>
<td>0.9912</td>
<td>0.9037</td>
<td>0.9256</td>
</tr>
<tr>
<td>SVM</td>
<td>0.8962</td>
<td>0.9594</td>
<td>0.9481</td>
<td>0.9556</td>
<td>0.9459</td>
<td>0.9410</td>
</tr>
<tr>
<td>NN</td>
<td>0.9140</td>
<td>0.9590</td>
<td>0.9432</td>
<td>0.9592</td>
<td>0.9401</td>
<td>0.9431</td>
</tr>
</tbody>
</table>

**Figure 6:** Comparison between user subjective and model objective evaluation: (a) LSMNR, (b) MLR, (c) SVM, and (d) NN.
discrete point to the black diagonal is closely related to the accuracy of the perceptual model, that is, the smaller the distance from the discrete point to the blue diagonal, the higher the accuracy of the corresponding model. As can be seen from Figure 6, the perceptual model of an LSMNR algorithm shows better convergence and evaluation performance than the perceptual model of other algorithms. This is because the algorithms such as MLR, SVM, and NN are mainly based on a single long video and perceive the user experience according to the network fluctuation index. The correlation between these indicators and the subjective feelings of users for short videos with different popularity videos is low, which leads to the low accuracy of the objective evaluation results of the model. The evaluation indices such as black screen and stalling times considered by the LSMNR algorithm are more suitable for estimating the subjective feelings of users, and the correlation of these indicators is higher; however, by subdividing video types, the algorithm establishes a mutually independent perception model to fine-grained user perception of long and short videos and further expands the perception video scene and improves the perception accuracy. It can be seen from Figure 6 that the RMSE of the LSMNR algorithm is 0.28511, which is much smaller than 0.41352, 0.33093, and 0.31514 of the MLR, SVM, and NN algorithms, respectively. Thus, it is proved that the evaluation error and perceptual accuracy of the proposed algorithm is the least and maximum, respectively.

Our simulations compare the correlation between different impact parameters and user subjective evaluation for different types of videos, as shown in Figures 7–10. In these figures, the horizontal axis represents the video playback success rate, the average number of video playback delays, the buffering time, and the black screen time, respectively, and the vertical axis represents the average opinion score (MOS) of users. As can be seen from Figures 7 and 8, the MOS increased rapidly with an increase in the success rate of video playback and decreased rapidly with the increase in the average number of times a video stalled. This is because there is a high correlation between the quality of user experience and the success rate of video playback and the average number of playback times selected. At this point, changes in these indicators can easily affect the user’s viewing experience, which further proves the rationality and effectiveness of the selection of these evaluation parameters in this model. As can be seen from Figures 9 and 10, MOS decreases with the increase of buffering duration and black screen duration, and the amplitude decreases gradually. This is because users are sensitive to buffering and black screens. Meanwhile, when actual users watch videos, they pay more attention to the number of times the screen buffers or goes black and pay less attention to the duration so that the length continues to increase at any time, and the rate of decrease in MOS gradually slows down. In addition, it can be seen from Figures 7–10 that MOS curves of popular videos are easily affected by the success rate of video playback, average number of video playback delays, and other indicators, whereas the fluctuations in MOS curves of common and unpopular videos are relatively small. This is because, when actual users watch videos, they pay more attention to the popular videos with high attention and are more sensitive to the damage degree of the videos. However, users have a low concentration for common and unpopular videos and can only perceive the videos with large damage but cannot perceive the videos with small damage. This also indicates that users have different perception evaluation criteria for different types of videos. The algorithm in this paper is more widely applicable to establish independent perception evaluation criteria.
models by refining video types, and the perception results are more ideal.

7. Conclusion

In this study, a user experience perception algorithm for long and short videos based on multiple nonlinear regression (LSMNR) is proposed. We analyzed the mapping relationship between user QoE and multiple evaluation indicators (video stalling, black screen, jitter, packet loss rate, etc.) under different video types based on their duration and popularity. Consequently, the data features of multiple evaluation indicators were excavated, followed by the establishment of the sample and test databases, combined with the subjective experimental results of users. Subsequently, the perception model was trained by the nonlinear least square method and SVM method to further study the nonlinear relationship between multiple evaluation indicators and the quality of user experience for the optimization of the results. Our simulation results show that compared with the MLR, SVM, and NN algorithms, the LSMNR algorithm can effectively improve perceptual accuracy and is suitable for a wide range of videos. In our future studies, we aim to develop a more accurate video user experience perception model, improve the ability of edge servers to analyze and process massive video data features, and provide users with a high-precision and fine-grained viewing experience.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Beijing Science and Technology Project (Z211100004421009) and National Natural Science Foundation of China (61971048).

References