ISSN- 2394-5125 VOL 06, ISSUE 04, 2019

MACHINE-LEARNING BASED QOE PREDICTION FOR DASH VIDEO STREAMING

Dr. P. Malathi, Asso. Professor, Dhanalakshmi Srinivasan College of Engineering and Technology
Dr. V. Janakiraman, Professor, Dhanalakshmi Srinivasan College of Engineering and Technology
Selventhiran S, Student, Dhanalakshmi Srinivasan College of Engineering and Technology

ABSTRACT: (Quality of experience (QoE) is an essential metric for video service platforms such as Youtube and Netflix to monitor the service perceived by their end-users. Driven by the popularity of MPEG-Dynamic Adaptive HTTP Streaming (DASH) format among service providers, a plethora of QoE prediction models have been proposed for MPEG-DASH video streaming. However, conventional models are established based on machine learning techniques, which are unable to extract high-level features from low-level raw inputs via a hierarchical learning process. The capabilities of deep learning have paved the new way for more powerful QoE prediction models. The aim of this project is to propose a deep-learning-based QoE prediction method. The starting point of the project is a state-of-the-art framework called DeepQoE, which encompasses three phases: feature pre-processing, representation learning and QoE predicting phase. The framework is further improved by integrating ensemble learning in the prediction phase. Extensive experiments are conducted to evaluate the performance of the proposed QoE prediction model as compared to conventional algorithms. By using a publicly available LIVE-NFLX-II dataset, the newly trained model outperforms not only conventional methods

but also the DeepQoE by 0.226% and 0.06% in terms of Spearman Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient LCC), respectively

1. INTRODUCTION

In the past few years, there is a rapid increase in the usage of mobile devices and multimedia applications. Mobile data traffic is also increasing significantly, with mobile video being one of the main contributors. Based on the Cisco Visual Networking Index (VNI), mobile video traffic contributed 59 per cent of the global mobile data traffic in the year 2017. It is also predicted that the global mobile data traffic will increase up to 7-fold from the year 2017 to 2022 while having 79 per cent of the traffic is in video. To combat the bandwidth fluctuations, MPEG-Dynamic Adaptive HTTP Streaming (DASH) is now widely adopted in modern major video streaming platforms such as Youtube and Netflix. In MPEG-DASH format, the videos will be separated into different durations of temporal segments, where each segment is encoded at different qualities, resulting in different sizes (Vasilev et al., 2018). This enables DASH clients to have flexible optimization strategies, where they can scale up or scale down the video quality by selecting the most optimum segment for different end-users.

All video streaming services need to find a balance between the operating cost and the quality of service perceived by their end-users. To cope with the increasing video traffic, they

ISSN- 2394-5125 VOL 06, ISSUE 04, 2019

need to plan their resource allocation and bandwidth to ensure that the experience of end-users is not affected greatly while being cost-effective. Starting from this premise, it is clear that video quality of experience (QoE) opens the possibilities for content providers to optimize their streaming service strategies.

QoE can be measured as the level of satisfaction or displeasure with the quality of service provided to the end-users. In recent years, QoE has become an important metric for companies to provide insight on their resource allocation and bandwidth provision. The factors that affect the QoE are called influence factors which can be further categorized into system, context and human influence factors. The resolution, bitrate, and demographic of end-users are some of the examples of influence factors

There are many publications and research on QoE prediction models. This project aims to apply deep learning for QoE prediction model to address the disadvantages of conventional QoE prediction models. To overcome the requirements of large training datasets for deep neural networks, transfer learning on pre-trained models was utilized. Finally, the proposed framework introduces ensemble learning to seek further improvement, and the results are compared with the existing QoE prediction algorithms in terms of the collinearity improvement between the predicted QoE and the real QoE.

2. MATERILAS AND METHODOLOGY

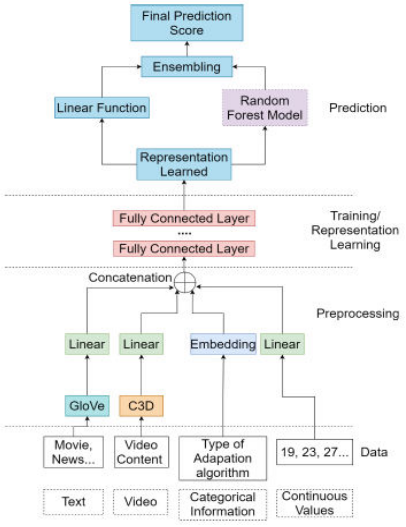

Fig 2.1: Proposed QoE Prediction

ISSN- 2394-5125 VOL 06, ISSUE 04, 2019

Model Framework

Figure 3.2 shows the proposed deep learning model framework in this project. The deep learning model architecture is based on the DeepQoE model, with an additional step to perform ensemble learning. The framework integrates a random forest model to perform ensembling with a linear layer together. The purpose of performing ensembling is to further enhance the accuracy of the deep learning model by exploiting the advantages of different algorithms used. Random forest is great with high dimensional data, which is suitable for the framework as it takes in the high-dimension representation from the deep learning model as input. Besides, the random forest model has the versatility to perform both classification and regression problems. With the ensembling of random forest model, the model will have a lower chance of overfitting and can achieve better performance.

A. Dataset Evaluation – LIVE-NFLX-II

The LIVE-NFLX-II database contains 420 video streams that are derived from 15 different original videos of diverse content, which includes action, documentary, sports, animation and video games. The video sources were rendered under different lighting conditions. Besides, the videos in the database were derived using seven different network traces to simulate the real-life effects of network variability during the HTTP-based adaptive video streaming

B. Dataset Pre-processing

There are a total of 36 features available in the data files of the LIVE-NFLX-II dataset. The metrics are provided in .pkl and .mat format. Hence, to simplify the training process, only 14 useful independent features are chosen and will be converted to a CSV file for easier processing

C. Feature Extraction (C3D & GloVe)

The raw footage of the video will be extracted by a pre-trained model called C3D. First, the mp4 raw video will be extracted into frames by using FFmpeg in the Linux command terminal. After all the frames of each video have been extracted, C3D will extract every 16 frames of a video into a 512-dimension vector. Eight frames are overlapped between each extraction so that the temporal feature of the video will not be lost. The illustration of the feature extraction for C3D is shown in Figure 3.4.

ISSN- 2394-5125 VOL 06, ISSUE 04, 2019

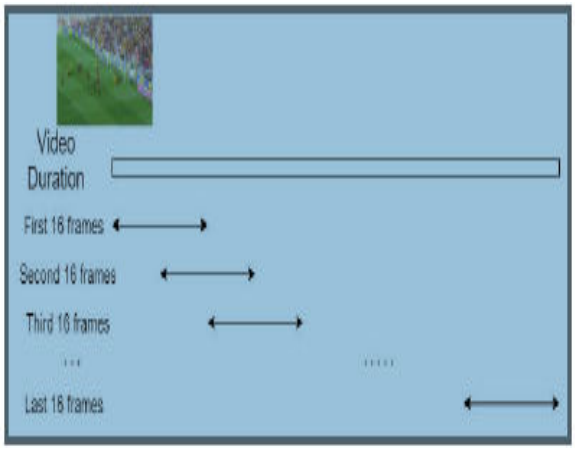

Fig. 2.2. Example Illustration of Feature Extraction of C3D

D. Extracting into CSV file

The information for all 420 videos are provided in individual .pkl files and contains redundant information such as the video file name. Hence, to simplify the training process, all features chosen are extracted to a .csv file instead. A python code is utilized to organize all of the training data needed, which includes the 512-dimension feature extracted from the C3D pre-trained model.

As mentioned in the previous sub-chapter, the 512-dimension feature is extracted from every 16 frames in each video. Hence, when compiling into the CSV file, the average data for the 16 frames are obtained. Then, the z-scored continuous MOS is also averaged to serve as the true value for the dependent variable. All data for every 16 frames of the videos are saved in the CSV file.

E. Training for Deep Learning Model

Before the training process begins, the dataset is split into a training set and a testing set, which will be a 90% and 10% ratio in this project. Then, the training data will be fed for training, while the test set will be used for model evaluation. This process is called the hold-out test.

During the deep learning model training, the features are mapped accordingly before being fed into three fully connected layers in the training phase. Firstly, the adaptation algorithm is mapped to a 10-dimension vector after an embedding layer. Features such as frame rate, ‘rebuffer bool’, scene cuts, playout bitrate are each transformed into 5-dimension vectors by a

ISSN- 2394-5125 VOL 06, ISSUE 04, 2019

linear layer. Metrics such as MSSIM, VMAF, STRRED, SSIM and PSNR are encoded into five 10-dimension vectors. The C3D extracts raw video footage into a 512-dimension vector, which will be encoded into a 412-dimension vector.

F. Anticipated Problems and Solutions

There are several problems that are anticipated. One of the problems is that different datasets will have different metrics available, while the impact of the metric cannot be fully known. Thus, deep evaluation of datasets must be carried out before choosing for feature extracting or QoE model training. Besides, it is expected that the physical memory or processing power of the personal computer alone is not sufficient to run the training processes. However, for this project, the CUDA version of the code is avoided, and the CPU is being utilized for the training and evaluation process instead

3. RESULTS AND DISCUSSION

To evaluate the deep learning model framework proposed in this project, several tests were done. As the deep learning model is performing regression tasks, the SROCC and LCC values will be used as evaluation metrics. Besides, RMSE is also included during evaluation. Measures such as hyper-parameter tuning, the feature extraction ability of the deep learning model and the overall predicting results of the deep learning model are evaluated.

A. Hyper-parameter tuning

Hyper-parameter analysis was conducted to find the best parameter for the proposed deep learning QoE prediction model framework. Besides, it can show the overall effectiveness of the proposed framework. In this project, the analysis of the learning rate and the training ratio was conducted. 10-fold cross-validation was conducted during the analysis to get the best hyper-parameter.

B. Enhancement to the Deep Learning Model with Ensemble Learning

The effectiveness of the deep learning model is evaluated. The deep learning framework proposed is compared with four different algorithms in a hold-out test. Due to computational

ISSN- 2394-5125 VOL 06, ISSUE 04, 2019

limitation, a hold-out test is performed in this section. Hold-out test can evaluate the performance of different algorithms on unseen data.

Besides, the representation generated from the deep learning model is proved to be useful. When other shallow learning uses the representation to predict QoE, it can be observed that the values obtained improved compared to using pre-processed data from the dataset.

The deep learning model achieved the best results when compared to other shallow learning algorithms. Lastly, it can be observed that the implementation of ensemble learning is able to enhance the overall accuracy of the deep learning model for QoE prediction. SROCC, LCC and the RMSE values are improved after adding an ensembling process with the Random Forest model.

4. CONCLUSION

This project has managed to identify several problems in the conventional machine learning QoE prediction model based on the literature review done. These problems include the ML conventional algorithms, which tend to rely on hand-crafted features and need different feature extraction methods. Thus, this project successfully addresses the issues by achieving the aim and objectives as follows.

This project has implemented a deep learning QoE prediction model for MPEG-DASH video via PyTorch and Scikit-learn. The deep learning prediction model was trained with the LIVE-NFLX-II dataset. The hyper-parameter of the model is then analyzed to obtain the default parameters for the training process. Besides, techniques to improve deep learning such as stacked generalization ensemble can be integrated to further investigate the corresponding prediction performance of the model. Lastly, it is suggested that the deep learning QoE model can be applied in a real-time DASH video streaming system to further evaluate the usefulness of the deep learning model in real-time QoE prediction, helping maximize the QoE for the end-users.

REFERENCES

1) Aung, W.T. and Hla, K.H.M.S., 2009. Random Forest Classifier for Multi-category Classification of Web Pages. *2009 IEEE Asia-Pacific Services Computing Conference (APSCC)*, pp.372–376.

ISSN- 2394-5125 VOL 06, ISSUE 04, 2019

2) Ayodele, T.O., 2010. Types of Machine Learning Algorithms. *New Advances in Machine Learning*, pp.19–49.

3) Bampis, C.G., Li, Z., Katsavounidis, I., Huang, T.Y., Ekanadham, C. and Bovik, A.C., 2018. Towards perceptually optimized end-to-end adaptive video streaming. *arXiv*, pp.1–16.

4) Berrar, D., 2018. Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(January 2018), pp.542–545.

5) Casas, P., D'Alconzo, A., Wamser, F., Seufert, M., Gardlo, B., Schwind, A., Tran-Gia, P. and Schatz, R., 2017. Predicting QoE in cellular networks using machine learning and in-smartphone measurements. *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, (May).

6) Cicco, L. De, Mascolo, S. and Palmisano, V., 2019. QoE-driven resource allocation for massive video distribution. *Ad Hoc Networks*, [online] 89, pp.170–176. Available at: <https://doi.org/10.1016/j.adhoc.2019.02.008