An LSTM-based Approach for Overall Quality Prediction in HTTP Adaptive Streaming

Conference Paper · April 2019

DOI: 10.1109/INFCOMW.2019.8845041

Tran Huyen
RIKEN
45 PUBLICATIONS **517** CITATIONS
SEE PROFILE

Nguyen Duc
Tohoku Institute of Technology
45 PUBLICATIONS **317** CITATIONS
SEE PROFILE

Nam Pham Ngoc
Hanoi University of Science and Technology
102 PUBLICATIONS **913** CITATIONS
SEE PROFILE

Truong Cong Thang
The University of Aizu
184 PUBLICATIONS **2,136** CITATIONS
SEE PROFILE

An LSTM-based Approach for Overall Quality Prediction in HTTP Adaptive Streaming

Huyen T. T. Tran*, Duc V. Nguyen*, Duong D. Nguyen†, Nam Pham Ngoc‡,Truong Cong Thang*
*The University of Aizu, Aizuwakamatsu, Japan
†Hanoi University of Science and Technology, Hanoi, Vietnam
‡Vin University Project, Vietnam

***Abstract*—HTTP Adaptive Streaming (HAS) has become a popular solution for multimedia delivery nowadays. In HAS, video quality is generally varying in each streaming session. Therefore, a key question in HTTP Adaptive Streaming is how to evaluate the overall quality of a streaming session. In this paper, we propose a machine learning approach for overall quality prediction in HTTP Adaptive Streaming. In the proposed approach, each segment is represented by four features of segment quality, stalling durations, content characteristics, and padding. The features are fed into a Long Short Term Memory (LSTM) network that is capable of exploring temporal relations between segments. The overall quality of the streaming session is predicted from the outputs of the LSTM network using a linear regression module. Experiment results show that the proposed approach is effective in predicting the overall quality of streaming sessions. Also, it is found that our proposed approach outperforms four existing approaches.**

***Index Terms*—Quality of Experience, Machine Learning Approach, Long Short Term Memory**

I. Introduction

HTTP Adaptive Streaming (HAS) has become a cost effective means for multimedia delivery nowadays. In HAS, a video is firstly encoded into multiple versions with different quality levels. Each version is further divided into a sequence of short segments [1], [2]. Based on network statuses, suitable versions of segments are selected. Due to network bandwidth fluctuations, selected versions of segments may vary strongly during a streaming session, causing quality variations [3], [4]. Also, stalling events may occur if a segment cannot arrive at the client before its playback deadline. Quality variations and stalling events are known to cause negative impacts to the user viewing experience [3]. Therefore, a main challenge in HAS is how to evaluate the overall quality of a streaming session considering the impacts of these factors.

Most existing approaches for overall quality prediction are analytical model-based approaches, in which the impacts of the factors are modeled using some analytical functions (e.g., a linear function) of some statistics such as the average of segment quality values and the average of stalling durations [5]–[8]. Among these approaches, only a few took into account the impacts of both quality variations and stalling events [5], [6].

To the best of our knowledge, the work in [9] is the first study that proposed an advanced machine learning approach for overall quality prediction. In the approach, a random neural network is employed. The inputs consist of the average of quantization parameters over all macro blocks of all video frames, the number of stalling events, the average and maximum of stalling durations. The approach was evaluated using 118 streaming sessions with the duration of 16 seconds.

In [10], the authors proposed an advanced machine learning approach using support vector regression. The inputs of the approach are the average of segment quality values, the time over which segment quality decreases took place, the time since the last impairment event (i.e., either a stalling event or a segment quality decrease), the number of stalling events, and the sum of stalling durations. The approach was evaluated using 112 sessions with the duration of approximately 72 seconds. It should be noted that the two studies of [9], [10] did not take into account the impacts of video content characteristics.

In this study, we propose a new advanced machine learning approach to predict the overall quality of HAS sessions. In the proposed approach, we employ a Long Short Term Memory (LSTM) network because of two reasons. First, it can exploit temporal relations between video segments by using a memory [11]. Second, LSTM network has been successfully employed in various video-related tasks such as video summarization [12] and video action recognition [13].

Our main contributions in this study are summarized as follows. First, we propose a new advanced machine learning approach using an LSTM network to predict the overall quality of HAS sessions. The proposed approach takes into account the impacts of quality variations, stalling events, and content characteristics. To the best of our knowledge, this is the first study using an LSTM network in overall quality prediction for HTTP adaptive streaming. Second, the proposed approach is evaluated using a dataset consisting of 515 sessions with the durations from 60 to 76 seconds. Experiment results show that the proposed approach achieves a high prediction performance and outperforms four existing approaches.

The rest of this paper is organized as follows. The proposed approach is presented in Sect. II. In Sect. III, we evaluate the performances of the proposed approach and four existing approaches. Finally, conclusions are provided in Sect. IV.

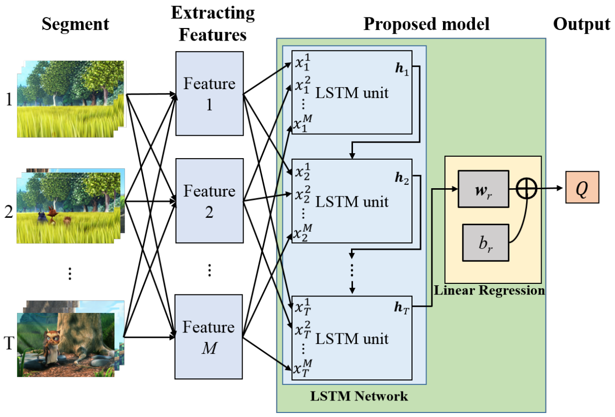


Fig. 1: Architecture of the proposed approach.

II. Proposed Approach

In this section, we first present the architecture of the proposed approach. Then the four segment features are described in detail.

A. Architecture

Figure 1 shows the architecture of the proposed approach. In particular, each streaming session is considered as a series of segments. Each segment is attributed by a set of features. The features are then fed into an LSTM network. The outputs of the LSTM network are used to predict the overall quality of streaming sessions through a linear regression module.

Let bold capital letters (e.g., **X**), bold lowercase letters (e.g., **x**), and italic letters (e.g., X) denote matrices, vectors, and scalars, respectively. T denotes the number of segments in a streaming session. Let

$$\mathbf{x}_t = \begin{bmatrix} x_t^1 \\ x_t^2 \\ \vdots \\ x_t^M \end{bmatrix} \tag{1}$$

be the feature vector of segment t ($1 \leq t \leq T$) with M is the number of features per segment.

Each vector $\mathbf{x}_t$ is connected to a hidden state $\mathbf{h}_t$ via an LSTM unit [14] as shown in Fig. 2. Specifically, the hidden state $\mathbf{h}_t$ is calculated using the following equations.

$$\mathbf{i}_t = sigm(\mathbf{W}_{ix}\mathbf{x}_t + \mathbf{V}_{ih}\mathbf{h}_{t-1} + \mathbf{b}_i), \tag{2}$$

$$\mathbf{f}_t = sigm(\mathbf{W}_{fx}\mathbf{x}_t + \mathbf{V}_{fh}\mathbf{h}_{t-1} + \mathbf{b}_f), \tag{3}$$

$$\mathbf{o}_t = sigm(\mathbf{W}_{ox}\mathbf{x}_t + \mathbf{V}_{oh}\mathbf{h}_{t-1} + \mathbf{b}_o), \tag{4}$$

$$\mathbf{g}_t = tanh(\mathbf{W}_{gx}\mathbf{x}_t + \mathbf{V}_{gh}\mathbf{h}_{t-1} + \mathbf{b}_g), \tag{5}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t, \tag{6}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot tanh(\mathbf{c}_t), \tag{7}$$

where $\odot$ denotes the element-wise product, and the parameters of $\mathbf{W} \in \mathbb{R}^{d \times M}$, $\mathbf{V} \in \mathbb{R}^{d \times d}$, and $\mathbf{b} \in \mathbb{R}^{d}$ are learned during the training process and shared across LSTM units. $\mathbf{i}_t, \mathbf{f}_t, \mathbf{o}_t$, and $\mathbf{c}_t$ are respectively the output vectors of the input gate, forget gate, output gate, and memory cell. They are important components to enable the LSTM unit to exploit temporal relations between segments. In particular, the input gate elects whether or not to add new information from the current inputs to the memory cell. The forget gate selects and removes old information from the memory cell. The output gate selects useful information from the memory cell to update the hidden state $\mathbf{h}_t$.

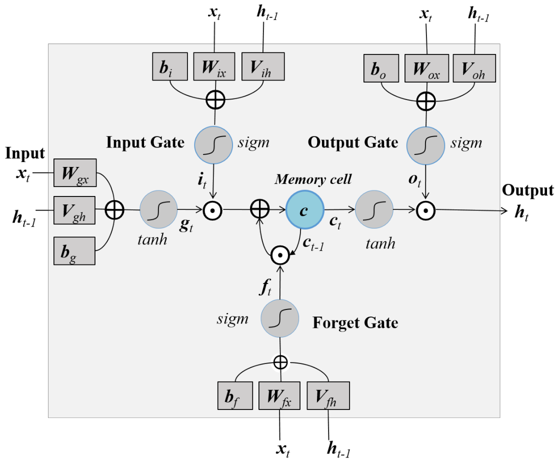


Fig. 2: LSTM unit architecture.

At the linear regression module, the overall quality Q of the session is predicted from the hidden state $\mathbf{h}_T$ corresponding to the last segment as follows.

$$Q = \mathbf{w}_r\mathbf{h}_T + b_r, \tag{8}$$

where $\mathbf{w}_r$ and b_r are also parameters to be learned.

B. Segment Features

In this part, we will describe the four segment features used in the proposed approach, namely segment quality, stalling durations, content characteristics, and padding.

1) Segment Quality: The segment quality feature represents visual quality of video segments. In this study, we use one of three metrics, namely bitrate (*BR*), Peak Signal-to-Noise Ratio (*PSNR*), and segment-MOS (*S-MOS*) [15]–[17] to represent this feature.

2) Stalling Durations: The stalling duration feature (denoted *SD*) of a segment represents the amount of time that an user has to wait since the playback of the previous segment ends until the playback of that segment begins. If that segment arrives at the client before the playback of the previous segment finishes (called the playback deadline), then *SD* is set to 0. Otherwise, a stalling event occurs and *SD* is a positive number.

3) Content Characteristics: It is well known that the overall quality of a session may be affected by video content characteristics [18]. Similar to [18], two dimensions of the content characteristic feature, namely spatial complexity and temporal complexity, are taken into account in the proposed approach.

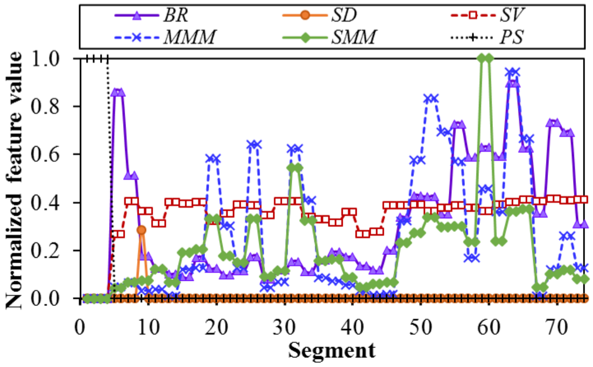


Fig. 3: An example of segment feature values

To represent the spatial complexity of a segment, we use a metric of Spatial Variance (*SV*) in [18]. This metric is calculated from MPEG-7 edge histogram descriptor. Specifically, each frame is firstly divided into 4x4 sub-blocks, and then histograms of 5 edge types (vertical, horizontal, 45°, 135°, and non-direction) are calculated for all sub-blocks [19]. Let S_{qp} denote the average histogram value of edge type p for all sub-blocks in frame q. Finally, the *SV* value of a segment is derived by

$$SV = \frac{1}{Q \times P} \sum_{q=0}^{Q-1} \sum_{p=0}^{P-1} S_{qp}, \tag{9}$$

where Q and P are respectively the number of frames in the segment and the number of edge types.

The temporal complexity of a segment is represented by two metrics calculated from the motion vectors of the segment. Specifically, the mean (denoted *MMM*) and standard deviation (denoted *SMM*) of the magnitudes of the motion vectors are used.

4) Padding: In practice, streaming sessions usually have different durations, and so the lengths (i.e., the number of segments). In this study, we employ zero-padding method to ensure that all sessions have the same length. In particular, some segments, called *padded segments*, are appended to the beginning of every session so that its length is the same as the length of the longest session. Note that, for all padded segments, their features consisting of segment quality, stalling durations, and content characteristics take a value of 0.

To differentiate the padded and actual segments, we define a boolean variable *PS* as follows.

$$PS(t) = \begin{cases} 1, & \text{if segment } t \text{ is a padded segment} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Figure 3 shows an example of normalized segment feature values in a streaming session. As can be seen in Fig. 3, the segment quality (i.e., the *BR* metric) varies strongly during the session. Also, it can be seen that there is one stalling event occurring at the ninth segment where $SD > 0$. Regarding the content characteristic feature, this session does not have significant changes of the spatial complexity, whereas the temporal complexity varies drastically. The *PS* values indicate that the first four segments are padded segments and the remaining segments are actual segments.

III. Evaluation

In this section, we first present experiment settings used to evaluate the prediction performance of the proposed approach. Next, we give some discussions on the roles of the segment features in the proposed approach. Finally, a comparison in terms of prediction performance between the proposed approach and four existing approaches will be presented.

A. Experiment Settings

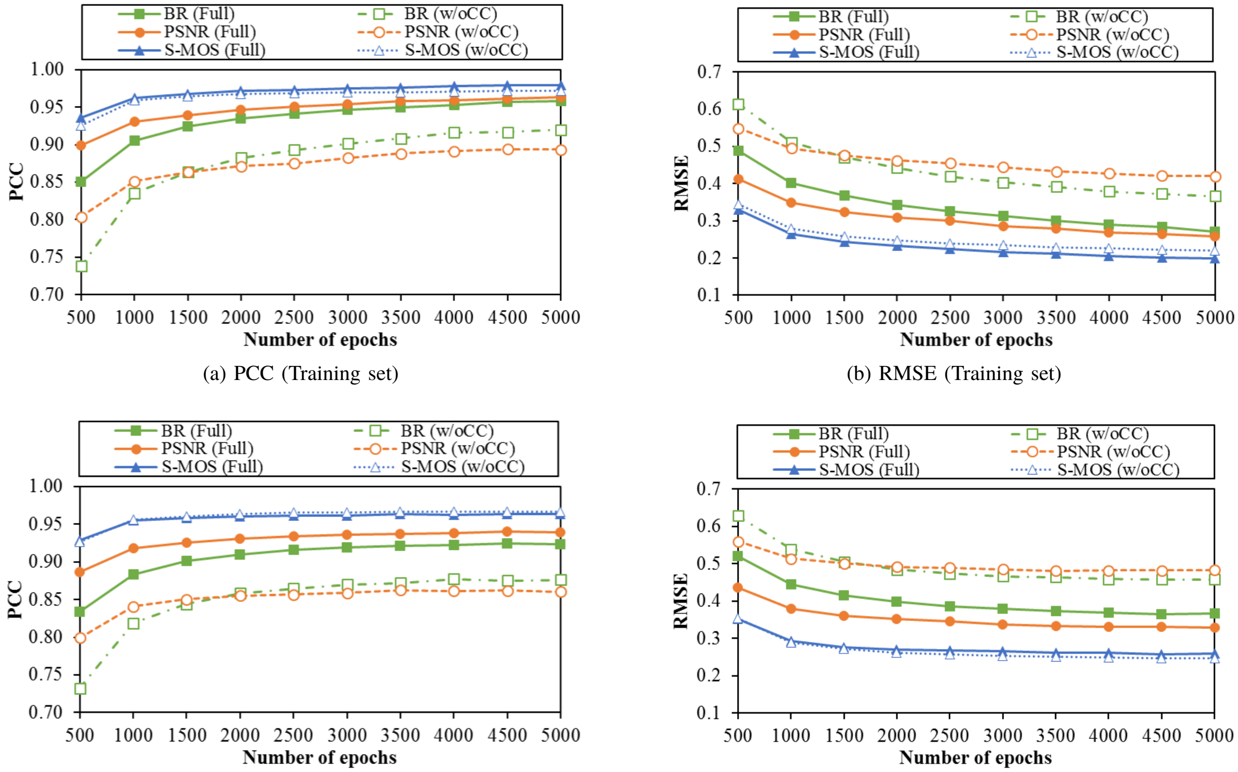
1) Dataset: To address the problem of lack of training data, the dataset used in this study was combined from three datasets. Two datasets were from the previous work of [5], [15]. The remaining dataset was newly created by conducting a subjective test. There were totally 144 sessions rated in the subjective test. These sessions were generated from two videos different from videos used in [5], [15].

In particular, each video was used to generate 72 sessions consisting of 42 hand-crafted sessions and 30 real streaming sessions. The hand-crafted sessions were generated from 5 patterns having no quality variation (i.e., selected versions of segments fixed during sessions) and no stalling event, 10 patterns having periodic quality variations with the period of 10 seconds and no stalling event, and 27 patterns containing from 1 to 6 stalling events with the durations of 0.25s, 0.5s, 1s, 2s, 3s, and 4s and no quality variation. The 30 real streaming sessions were generated by running two adaptation methods of [20], [21] in a streaming test-bed using bandwidth traces from a mobile network [22]. The real streaming sessions consist of both quality variations and stalling events.

Similar to prior studies [5], [15], the test conditions were designed following Recommendation ITU-T P.913 [23]. In order to minimize subjects' fatigue, the subjective test was divided into four parts that were conducted in different days. The duration of each part was approximately 50 minutes. Every 20 minutes there was a break of 10 minutes. Each subject took part in at most two test parts. Before doing actual subjective tests, subjects were trained to get accustomed to the rating procedure and the range of video quality scores. The sessions were randomly displayed on a 14-inch screen with a resolution of 1,366×768 and a black background. At the end of each session, each subject gave a rating score with the score range from 1 (worst) to 5 (best).

There were totally 53 subjects taking part in the subjective test with ages ranging from 18 to 41. The total time of the subjective test was approximately 78 hours. A screening analysis of the test results was performed following Recommendation ITU-T P.913 [23], and two subjects were rejected. After eliminating the scores of the rejected subjects, each session was rated by 21 valid subjects. The subjective overall quality value of each session is calculated as the average score of the valid subjects.

The combined dataset consists of totally 515 sessions with 183 hand-crafted sessions and 332 real streaming sessions.



(a) PCC (Training set)

(b) RMSE (Training set)

(c) PCC (Test set)

(d) RMSE (Test set)

Fig. 4: Prediction performance of the proposed approach for the *Full* and *w/oCC* cases.

The durations of the sessions are from 60 to 76 seconds. These sessions are randomly divided into a training set of 412 sessions and a test set of the 103 remaining sessions. The division is repeated 100 times, resulting in 100 pairs of training and test sets. The results presented in the following sections are the average values over the 100 pairs of training and test sets.

2) Training Parameters: For the training process in the proposed approach, we apply a loss function of root mean squared error. The loss function is minimized using stochastic gradient descent method based on Adam optimization algorithm [24]. The parameters of the Adam algorithm are set as follows: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e - 08$. The learning rate, the number of hidden units, and the number of epochs are set to 0.01, 5, and 5000, respectively.

3) Input Features: To investigate the roles of the segment features in the proposed approach, we consider four cases of input features. In the first case (denoted *Full*), each segment is represented by all the four features described in Subsect. II-B. For the three remaining cases, only three of the four features are used. In particular, the content characteristic feature is excluded from the inputs in the second case (denoted *w/oCC*). In the third case (denoted *w/oSQ*), the segment quality feature is not considered. For the last case (denoted *w/oSD*), the stalling duration feature is not used as the inputs of the proposed approach.

4) Evaluation Metrics: To evaluate the prediction performance of the proposed approach, we use two metrics of Pearson Correlation Coefficient (PCC) and Root Mean Squared Error (RMSE) which are averaged over the 100 test sets. Note that a higher PCC and a lower RMSE mean a better prediction performance.

B. Roles of Segment Features

In this subsection, we will investigate the roles of the segment features in the proposed approach. For this purpose, we evaluate the prediction performance of the proposed approach in the four cases of input features as presented in Subsect. III-A3.

Figure 4 shows the PCC and RMSE values of the proposed approach in the *Full* and *w/oCC* cases when the number of epochs e is from 500 to 5000 with a step size of 500. Note that, the segment quality feature is represented by one of the three metrics mentioned in Subsect. II-B1. From Fig. 4, it can be seen that, given a segment quality metric, the training set always has higher PCC values and lower RMSE values than the test set.

For both the training and test sets, the PCC values increase quickly and the RMSE values reduce rapidly when the number of epochs e first increases. When the number of epochs increases further, the PCC and RMSE values become stable. It can be noted that the stable state is reached much quicker

TABLE I: Prediction performance of the proposed approach using the *S-MOS* metric for the *Full*, *w/oSQ*, and *w/oSD* cases.

Case	**Training set**		**Test set**	
	PCC	***RMSE***	***PCC***	***RMSE***
Full	0.98	0.20	0.96	0.26
w/oSQ	0.66	0.70	0.58	0.78
w/oSD	0.90	0.42	0.81	0.56

TABLE II: Prediction performance of the proposed approach and four existing approaches.

Approach	**Test set**	
	PCC	***RMSE***
Proposed	0.96	0.26
Tran's	0.90	0.40
P.1203.3	0.91	0.38
Singh's	0.72	0.65
ATLAS	0.88	0.45

for the *S-MOS* and *PSNR* metrics than for the *BR* metric. Specifically, the optimal number of epochs e for both the *Full* and *w/oCC* cases is 1500 for the *S-MOS* and *PSNR* metrics, and 2500 for the *BR* metric.

For the two metrics of *BR* and *PSNR*, the *Full* case achieves a significantly higher prediction performance than the *w/oCC* case. For the *S-MOS* metric, the *Full* and *w/oCC* cases have similar prediction performances. This result implies that the additional use of the content characteristic feature does not bring significant improvements to the proposed approach when *S-MOS* is used as the segment quality metric. Meanwhile, for the metrics of *BR* and *PSNR*, it is necessary to include the content characteristic feature. In other words, the role of the content characteristic feature depends on the metric used to represent the segment quality feature.

In the *Full* case, the prediction performance is highest when the segment quality metric is *S-MOS*. This means that *S-MOS* is the best metric to represent the segment quality feature. Meanwhile, using the *BR* metric results in the lowest prediction performance. Note that, in the rest of this paper, the *S-MOS* metric is used to represent the segment quality feature in the proposed approach because of the best prediction performance.

Table I shows the prediction performance of the proposed approach for the three cases of *Full*, *w/oSQ*, and *w/oSD* when the number of epochs is 5000. We can see that the PCC value significantly reduces and the RMSE value substantially increases when either the segment quality feature or the stalling duration feature is excluded from the inputs of the proposed approach. This indicates that quality variations and stalling events have significant impacts on the overall quality of sessions.

C. Comparison with Existing Approaches

In this part, we will compare the proposed approach with four existing approaches, namely *Tran's* [5], *P.1203.3* [25]–[27][1], *Singh's* [9], and *ATLAS* [10][2]. For the proposed approach, the *Full* case and the segment quality metric of *S-MOS* are used.

Table II shows the PCC and RMSE values of the proposed and existing approaches for the test set. We can see that the proposed approach outperforms the existing approaches by achieving the highest prediction performance. In particular, the PCC and RMSE values of the proposed approach are 0.96 and 0.26, respectively. This result indicates that LSTM network is more effective than random neural network and support vector regression in predicting the overall quality of HAS sessions. Consequently, temporal relations between segments in a session are essential to overall quality prediction. This explains why the existing approaches, which use the statistics over all segments such as the average of segment quality values and the sum of stalling durations, have lower prediction performances than that of the proposed approach.

IV. Conclusion

In this study, we have proposed a new advanced machine learning approach using an LSTM network for predicting the overall quality of HTTP Adaptive Streaming sessions. The proposed approach took into account the four segment features of quality, stalling durations, content characteristics, and padding. Based on the experiment results, it was shown that LSTM network is effective in predicting the overall quality of HTTP Adaptive Streaming sessions. Also, temporal relations between segments in sessions are essential to overall quality prediction. In addition, segment-MOS is found to be the best metric to represent the segment quality feature. For future work, we plan to employ the proposed approach in performance evaluations of adaptation strategies for HTTP Adaptive Streaming.

References

[1] T. C. Thang, Q. D. Ho, J. W. Kang, and A. T. Pham, "Adaptive Streaming of Audiovisual Content using MPEG DASH," *IEEE Transactions on Consumer Electronics,*, vol. 58, no. 1, pp. 78–85, Feb. 2012.

[2] H. T. T. Tran, N. P. Ngoc, T. C. Thang, and Y. M. Ro, "Real-time quality evaluation of adaptation strategies in VoD streaming," in *2016 Digital Media Industry & Academic Forum (DMIAF)*, Santorini, Greece, Jul. 2016, pp. 217–221.

[3] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[4] H. T. T. Tran, T. H. Le, N. P. Ngoc, A. T. Pham, and C. T. Truong, "Quality improvement for video on-demand streaming over HTTP," *IEICE Transactions on Information and Systems*, vol. E100-D, no. 1, pp. 61–64, 2017.

[5] H. T. T. Tran, N. P. Ngoc, A. T. Pham, and T. C. Thang, "A Multi-Factor QoE Model for Adaptive Streaming over Mobile Networks," in *2016 IEEE Global Communications Conference (IEEE GLOBECOM)*, Washington DC, USA, Dec. 2016, pp. 1–6.

[6] W. Robitza, M.-N. Garcia, and A. Raake, "A modular HTTP adaptive streaming QoE model-Candidate for ITU-T P. 1203 ("P. NATS")," in *2017 Nineth international workshop on Quality of multimedia experience (QoMEX)*, Erfurt, Germany, Jul. 2017, pp. 1–6.

[7] 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 *International Workshop on Quality of Multimedia Experience*, Yarra Valley, VIC, Australia, Jul. 2012, pp. 1–6.

[8] H. T. Tran, T. Vu, N. P. Ngoc, and T. C. Thang, "A novel quality model for HTTP Adaptive Streaming," in *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*, Ha Long, Vietnam, Jul. 2016, pp. 423–428.

[1] https://github.com/itu-p1203/itu-p1203/

[2] http://live.ece.utexas.edu/research/quality/VideoATLAS_release.zip

[9] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of experience estimation for adaptive HTTP/TCP video streaming using H. 264/AVC," in *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, USA, Jan. 2012, pp. 127–131.

[10] C. G. Bampis and A. C. Bovik, "Feature-based prediction of streaming video QoE: Distortions, stalling and memory," *Signal Processing: Image Communication*, vol. 68, pp. 218–228, 2018.

[11] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, pp. 11 – 24, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865514000221

[12] B. Zhao, X. Li, and X. Lu, "Hierarchical recurrent neural network for video summarization," in *ACM International Conference on Multimedia*, New York, USA, Oct. 2017, pp. 863–871. [Online]. Available: http://doi.acm.org/10.1145/3123266.3123328

[13] A. H. Nguyen, H. T. T. Tran, C.-T. Truong, and Y. M. Ro, "Fast Recognition of Human Actions Using Autocorrelation Sequence," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, Nara, Japan, Oct. 2018, pp. 114–115.

[14] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," *arXiv:1409.2329*, 2014. [Online]. Available: http://arxiv.org/abs/1409.2329

[15] H. T. Tran, N. P. Ngoc, Y. J. Jung, A. T. Pham, and T. C. Thang, "A Histogram-Based Quality Model for HTTP Adaptive Streaming," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 100, no. 2, pp. 555–564, 2017.

[16] Z. Guo, Y. Wang, and X. Zhu, "Assessing the visual effect of non-periodic temporal variation of quantization stepsize in compressed video," in *2015 IEEE International Conference on Image Processing (ICIP)*, Quebec City, Canada, Sept. 2015, pp. 3121–3125.

[17] J. D. Vriendt, D. D. Vleeschauwer, and D. Robinson, "Model for estimating QoE of video delivered using HTTP adaptive streaming," in *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, Ghent, Belgium, May 2013, pp. 1288–1293.

[18] H. Sohn, H. Yoo, W. D. Neve, C. S. Kim, and Y. M. Ro, "Full-reference video quality metric for fully scalable and mobile svc content," *IEEE Transactions on Broadcasting*, vol. 56, no. 3, pp. 269–280, Sept. 2010.

[19] C. S. Won, D. K. Park, and S.-J. Park, "Efficient use of mpeg-7 edge histogram descriptor," *ETRI journal*, vol. 24, no. 1, pp. 23–30, 2002.

[20] T. C. Thang, H. T. Le, H. X. Nguyen, A. T. Pham, J. W. Kang, and Y. M. Ro, "Adaptive video streaming over HTTP with dynamic resource estimation," *Journal of Communications and Networks*, vol. 15, no. 6, pp. 635–644, 2013.

[21] P. Juluri, V. Tamarapalli, and D. Medhi, "SARA: Segment aware rate adaptation algorithm for dynamic adaptive streaming over HTTP," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*, London, UK, Jun. 2015, pp. 1765–1770.

[22] C. Müller, S. Lederer, and C. Timmerer, "An evaluation of dynamic adaptive streaming over HTTP in vehicular environments," in *Proceedings of the 4th Workshop on Mobile Video*, Chapel Hill, North Carolina, Feb. 2012, pp. 37–42.

[23] Recommendation ITU-T P.913, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment," *International Telecommunication Union*, 2014.

[24] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980*, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[25] Recommendation ITU-T P.1203.3, "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport-Quality integration module," *International Telecommunication Union*, 2017.

[26] A. Raake, M.-N. Garcia, W. Robitza, P. List, S. Göring, and B. Feiten, "A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1," in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, May 2017, pp. 1–6.

[27] W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M.-N. Garcia, K. Yamagishi, and S. Broom, "HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 - Open Databases and Software," in *Proceedings of the 9th ACM Multimedia Systems Conference*, Amsterdam, Netherlands, Jun. 2018, pp. 466–471.