29/4/2019

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

Huyen Tran[*], Duc Nguyen[*], Duong Nguyen[†], Nam Ngoc[‡], Truong Thang[*]

[*] The University of Aizu, Japan

[†] Hanoi University of Science and Technology, Vietnam

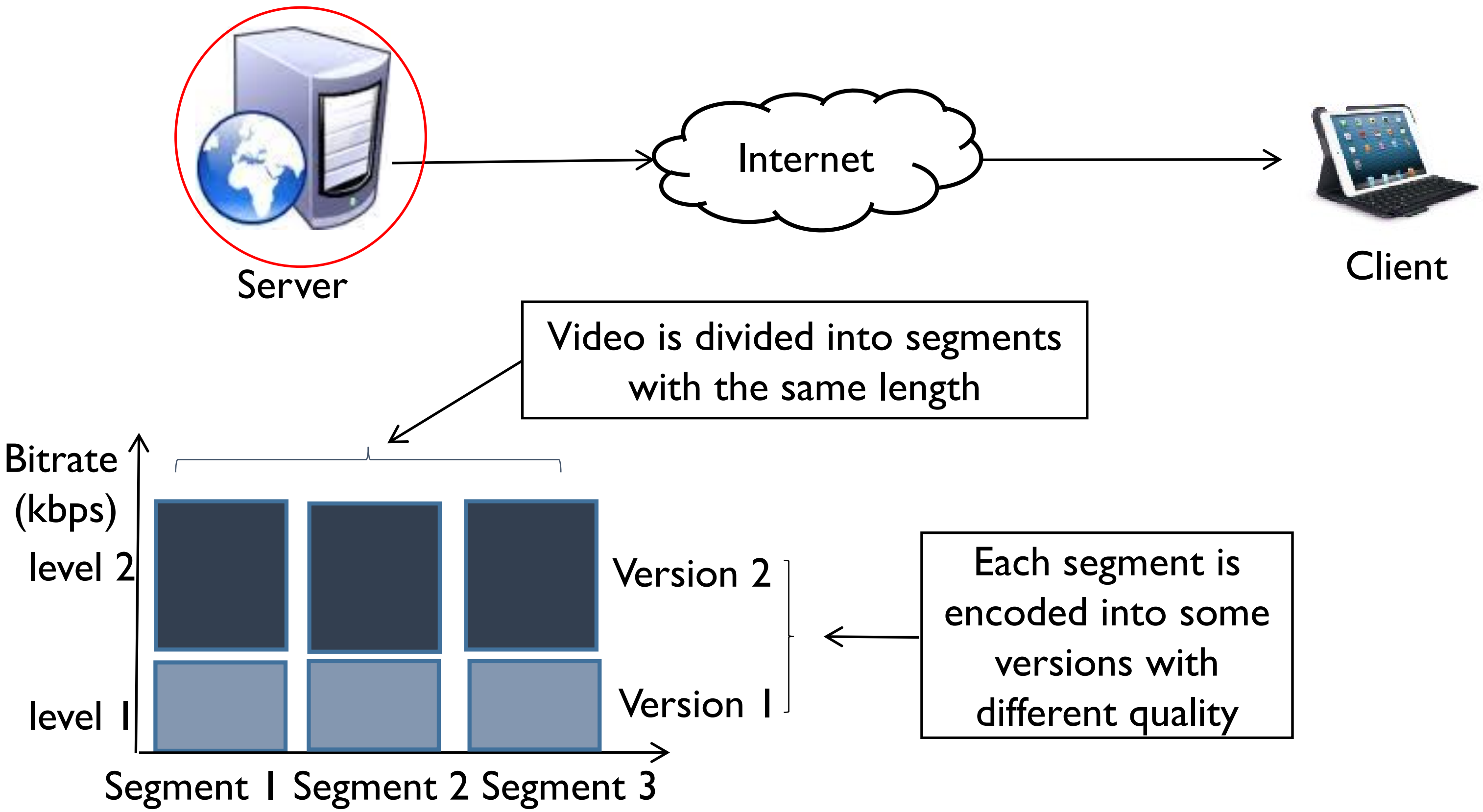
[‡] Vin University Project, Vietnam

Outline

- Introduction
- Related work
- Proposed approach
- Evaluation
- Conclusion

Introduction

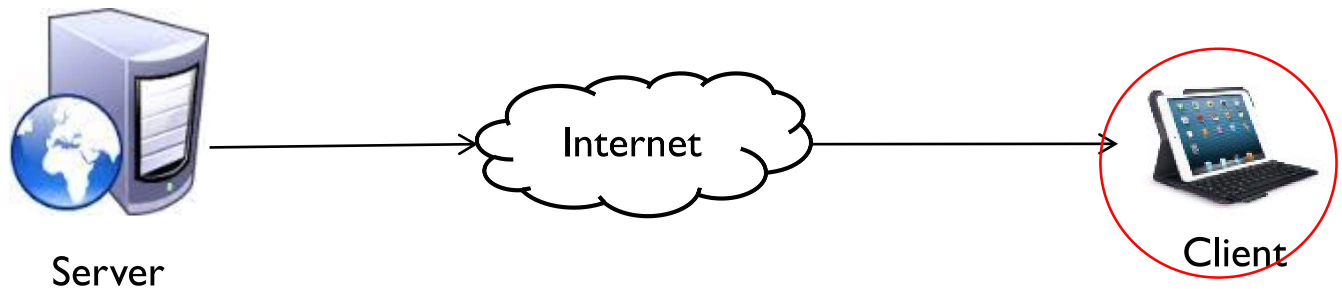
- HTTP Adaptive Streaming (HAS)
  - Server



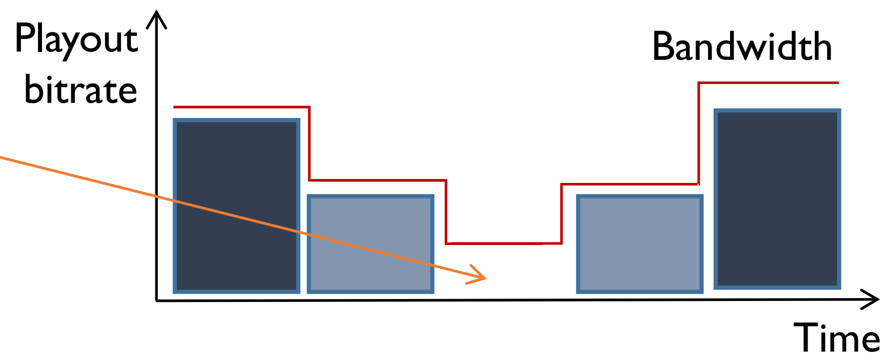
Introduction

HTTP Adaptive Streaming (HAS)

– Client



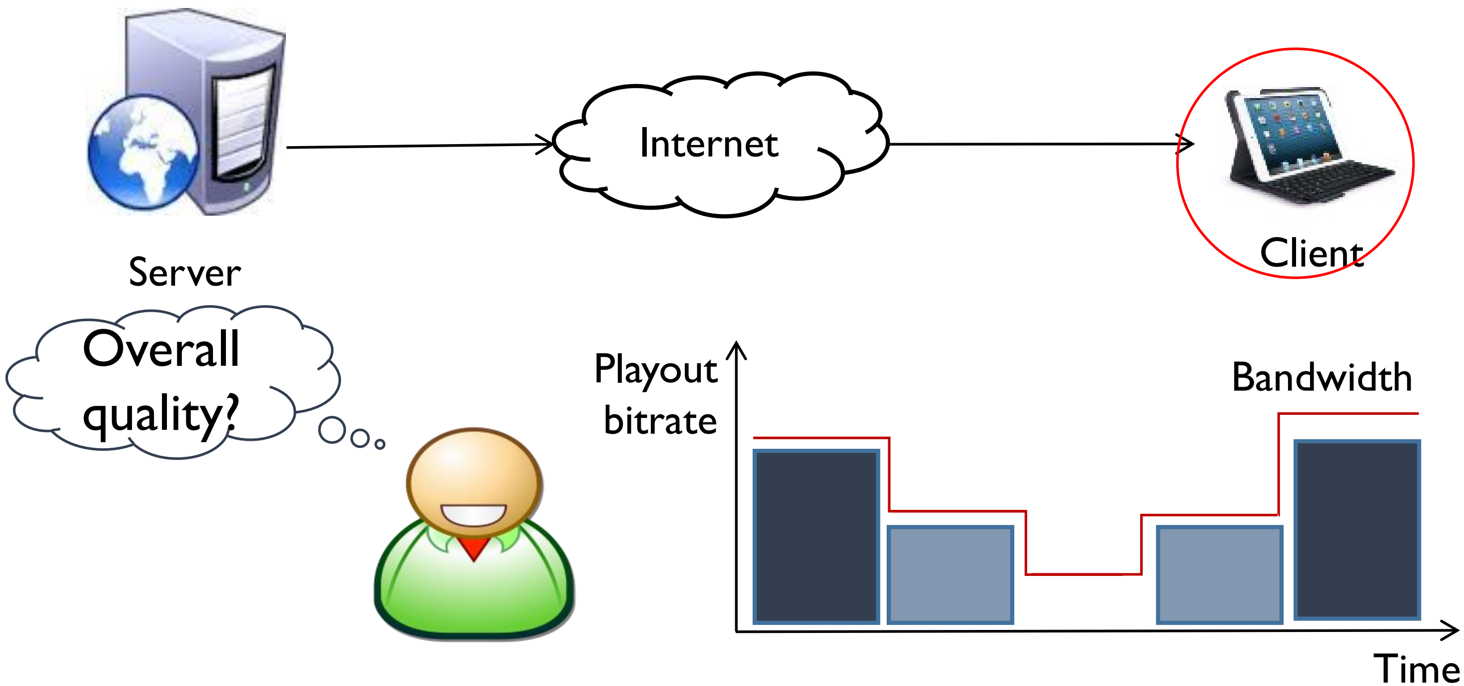
Because of bandwidth fluctuations ➔ Video quality is unstable, and some stalling events can appear in video.



Introduction

❑ HTTP Adaptive Streaming (HAS)

– Client



Key challenge: How to assess overall quality of a streaming session considering the impacts of quality variations and stalling events?

Related Work and Motivation

Approach	RNN [1]	ATLAT [2]	P.1203.3 [3]	Proposed
Video duration (seconds)	16	74	60~300	60 ~76

[1] 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 CCNC, Las Vegas, Jan. 2012, pp. 127–131.
[2] C. G. Bampis and A. C. Bovik, "Learning to Predict Streaming Video QoE: Distortions, Rebuffering and Memory," submitted to Signal Processing: Image Communication.
[3] 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," 2017.

Related Work and Motivation

RNN [1]	ATLAT [2]	P.1203.3 [3]	Proposed
+ **Quality variation** • 1 feature: average of SQVs + **Stalling events** • 3 features: total number of stalling events, maximum and average of stalling durations.	+ **Quality variation** • 3 features: average of SQVs, total time of quality-decrease events, time since the last impairment + **Stalling events** • 2 features: total number of stalling events, sum of stalling duration	+ **Quality variation** • 8 features: average of SQVs in each interval, etc. + **Stalling events** • 5 features: total number of stalling events, sum of stalling duration, frequency, etc.	+ **Quality variation** • SQV of each segment + **Stalling events** • Duration of each stalling event + **Content feature** • Spatial complexity of each segment • Temporal complexity of each segment

*SQVs: segment quality values

→ Previous approaches: two factors of quality variations and stalling events

→ Proposed approach: three factors of quality variations, stalling events, and content features

Related Work and Motivation

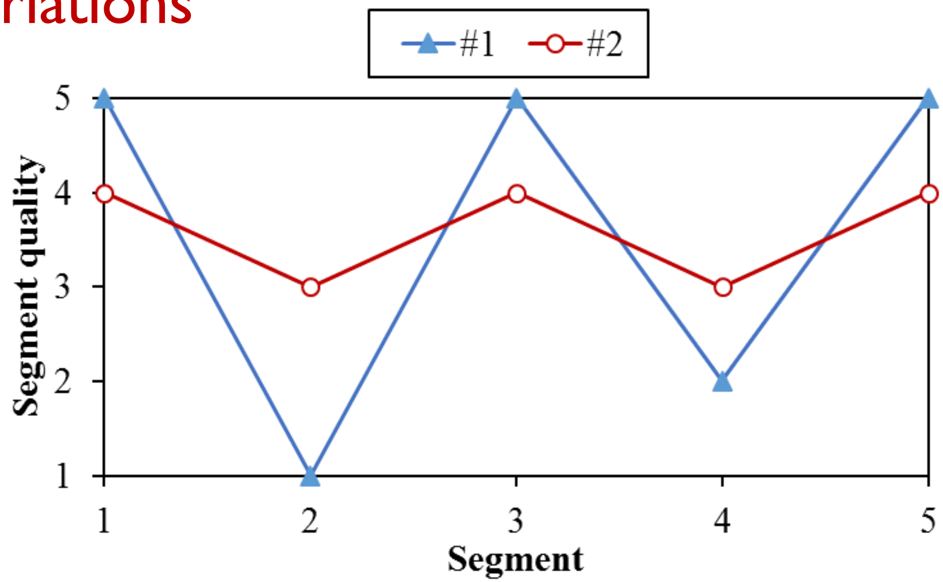
RNN [1]	ATLAT [2]	P.1203.3 [3]	Proposed
+ Quality variation • 1 feature: average of SQVs **+ Stalling events** • 3 features: total number of stalling events, maximum and average of stalling durations.	**+ Quality variation** • 3 features: average of SQVs, total time of quality decreases, and time since the last impairment **+ Stalling events** • 2 features: total number of stalling events, sum of stalling duration	**+ Quality variation** • 8 features: average of SQVs in each interval, etc. **+ Stalling events** • 5 features: total number of stalling events, sum of stalling duration, frequency, etc.	**+ Quality variation** • SQV of each segment **+ Stalling events** • Duration of each stalling event **+ Content feature** • Spatial complexity of each segment • Temporal complexity of each segment

*SQVs: segment quality values

→ Previous approaches: Inputs are statistics on a session basis such as average of segment quality values, the total number of stalling events, and average of stalling durations.

Related Work and Motivation

- Quality variations



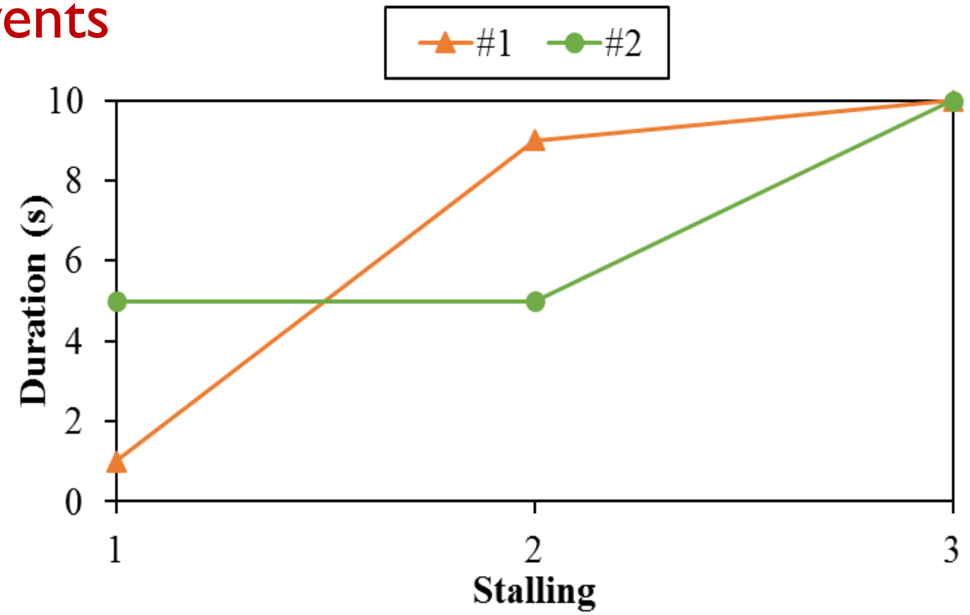
An example of two sessions with the same some statistics

- The same average of segment quality values (=3.6)
- The same total time of quality decreases (=2 segments)
- The same time since the last quality decrease (=1 segment)

➔ Such statistics can not fully reflect quality variations in a streaming session

Related Work and Motivation

- Stalling events



An example of two sessions with the same some statistics

- The same number of stalling events
- The same maximum of stalling durations
- The same average of stalling durations

➔ Such statistics can not fully reflect stalling events in a streaming session

Related Work and Motivation

RNN [1]	ATLAT [2]	P.1203.3 [3]	Proposed
+ Quality variation • 1 feature: average of SQVs **+ Stalling events** • 3 features: total number of stalling events, maximum and average of stalling durations.	**+ Quality variation** • 3 features: average of SQVs, total time of quality-decrease events, time since the last impairment **+ Stalling events** • 2 features: total number of stalling events, sum of stalling duration	**+ Quality variation** • 8 features: average of SQVs in each interval, etc. **+ Stalling events** • 5 features: total number of stalling events, sum of stalling duration, frequency, etc.	**+ Quality variation** • SQV of each segment **+ Stalling events** • Duration of each stalling event **+ Content feature** • Spatial complexity of each segment • Temporal complexity of each segment

*SQVs: segment quality values

→ Proposed approach: Inputs are taken on a segment-by-segment basis.

Related Work and Motivation

Approach	RNN [1]	ATLAT [2]	P.1203.3 [3]	Proposed
Learning algorithm	Random neural network (RNN)	Support Vector Regression (SVR)	Random Forest (i.e., an ensemble of 20 decision trees)	Long-short term memory (LSTM)

- Long short term memory (LSTM)
  - Can exploit temporal relationships between segment features to generate the output

    → it can be more effective to reflect temporal quality variations and stalling events in a streaming session
  - Has been successfully used in multiple temporal sequence tasks such as video summarization, video classification, and video action recognition.

Proposed approach

- Key features
  - Three factors
    - Quality variations
    - Stalling events
    - Content features
  - Inputs on a segment-by-segment basis
  - Learning algorithm
    - Long short term memory (LSTM)

Proposed approach

- Each segment is represented by 4 features
  1. Quality feature
     - Calculated using one of three quality metrics: bitrate (BR), Peak signal to noise ratio (PSNR), and Mean Opinion Score (S-MOS).
  2. Stalling feature
     - Stalling duration (SD): Time from when the preceding segment is completely displayed until when the current segment starts being played.
  3. Content feature
  4. Padding status (PS)

Proposed approach

- Each segment is represented by 4 features

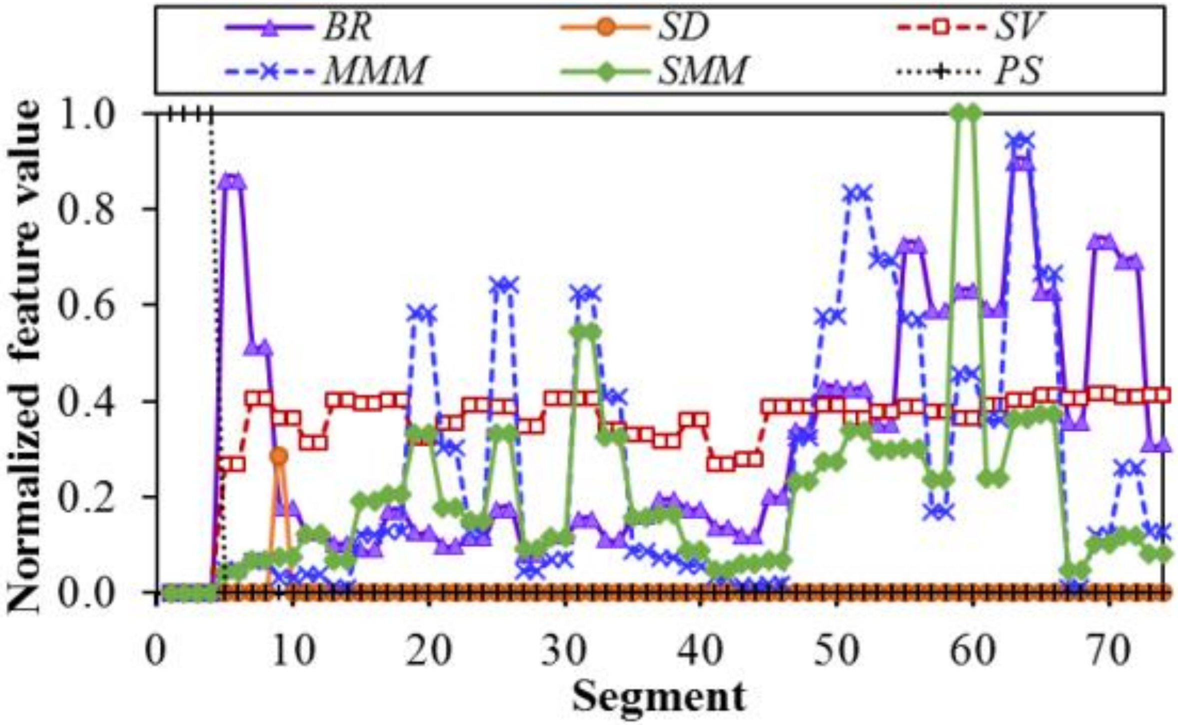
3. Content feature
   - Temporal complexity: using a metric of Spatial Variance (SV) calculated from MPEG-7 edge histogram descriptor.
   - Spatial complexity: using two metrics of mean (MMM) and standard deviation (SMM) of motion vector magnitudes
4. Padding status (PS)

Proposed approach

- Each segment is represented by 4 features

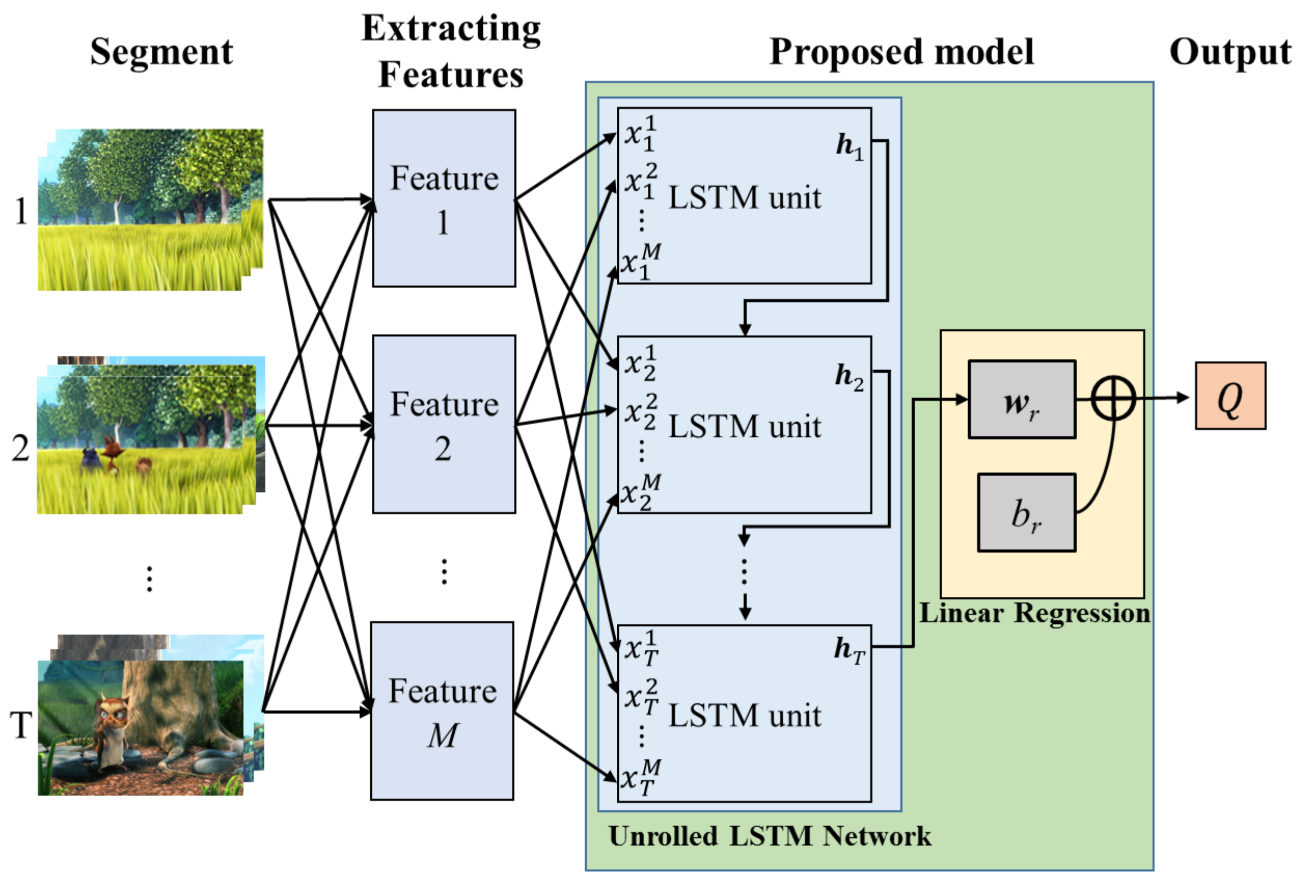
4. Padding status (PS)
    - In practice, streaming sessions usually have different durations → zero-padding method
    - Some segments, called padded segments, are added to the beginning of every session so that its length is the same as the length of the longest session.
    - $PS(t) = \begin{cases} 1, & \text{if t is a padded segment} \\ 0, & \text{otherwise} \end{cases}$

Proposed approach



An example of segment features

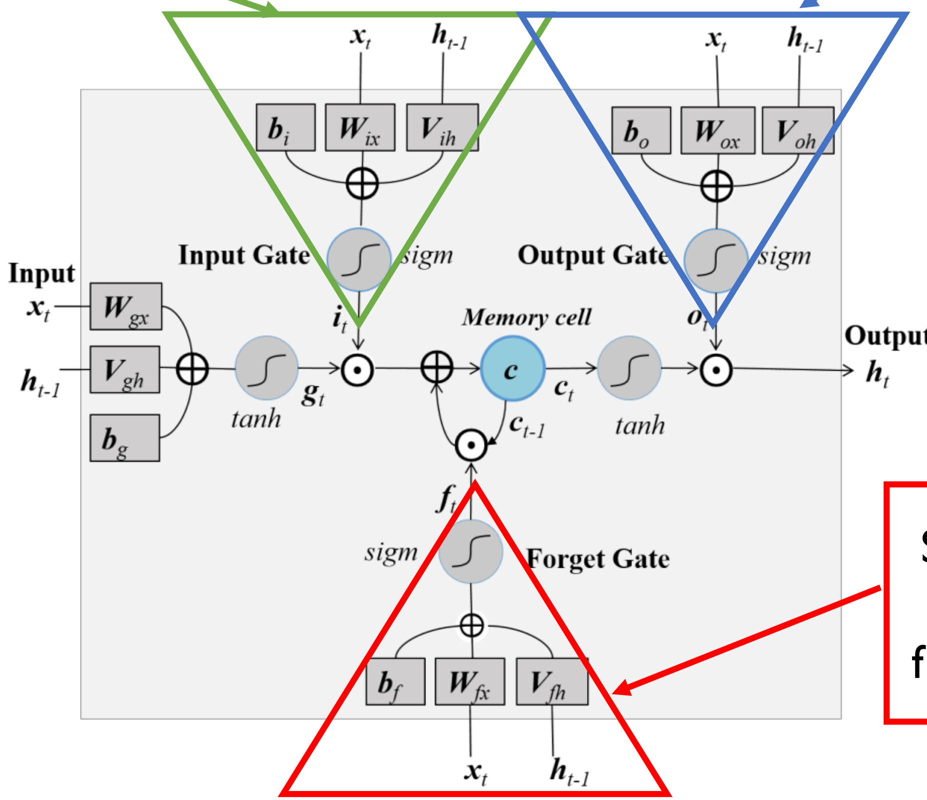
Architecture of the proposed approach



LSTM unit architecture

Select to add new information from current inputs to memory cell

Select useful information from memory cell to update hidden state



Select to remove old information from memory cell

29/4/2019

Evaluation and Analysis

❑ Settings

Number of Epochs	5000
Number of Hidden Units	5
Learning rate	0.01
Loss function	Root mean squared error
Optimization algorithm	Adam optimization algorithm
Number of sessions	515 generated from 5 different videos (412 for training set, 103 for test set)
Duration	60→76 seconds
Performance metrics	Pearson correlation coefficient (PCC) and Root Mean Squared Error (RMSE)

Performance evaluation

Performance of the approaches

Approach	Performance	
	PCC	*RMSE*
RNN [1]	0.72	0.65
ATLAT [2]	0.88	0.45
P.1203.3 [3]	0.91	0.38
Proposed (using S-MOS)	0.96	0.26

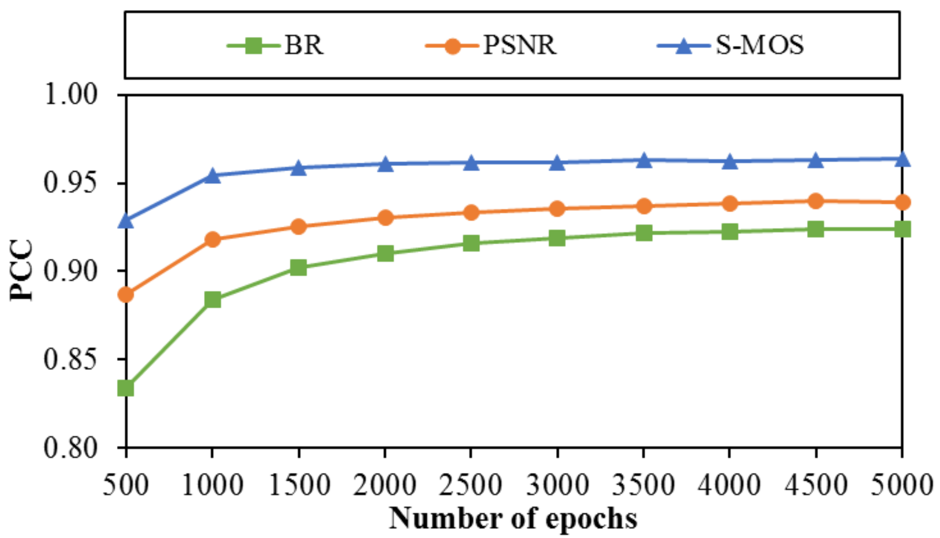
➔ Proposed approach outperforms the existing approaches.

Impact of Segment Quality Metric

- Quality feature
  - S-MOS: highest PCC
    - But, in practical, it is difficult to obtain S-MOS values
  - Interestingly, BR and PSNR perform well (PCC>0.92) when the number of epochs is 5000.

→ BR and PSNR can also used in the proposed approach.

Performance of the proposed approach for different segment quality metrics



Roles of Segment Quality and Stalling Features

Performance of the proposed approach (using S-MOS)
with and without quality and stalling features

Approach	**Test set**	
	PCC	***RMSE***
Full	0.96	0.26
w/o quality feature	0.57	0.78
w/o stalling feature	0.82	0.55

→ Performance is significantly reduced, especially for w/o quality feature.

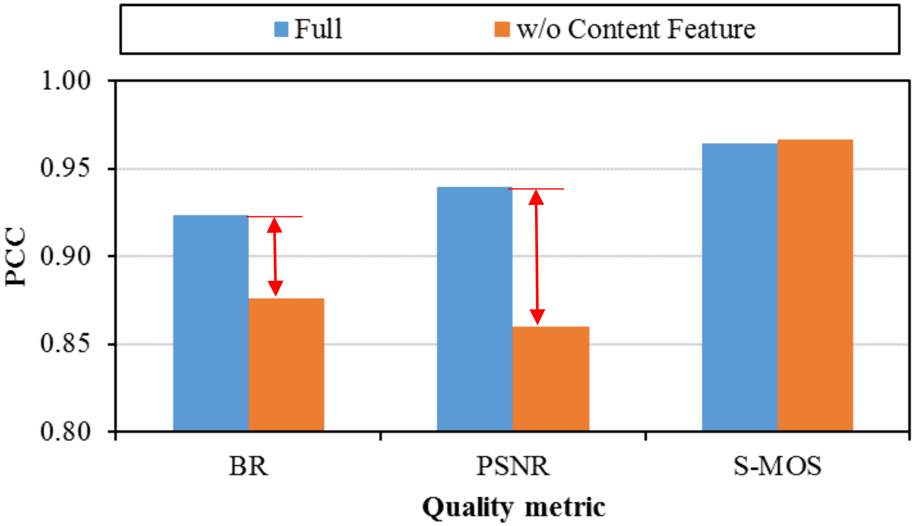
→ Segment quality and stalling features are key features in the proposed approach.

Role of Content Feature

- Content feature
  - For different quality metrics, the impact of content feature is different.
    - S-MOS: negligible
    - BR and PSNR: significant

  ➔ Necessary to feed the content feature into learning-based approaches when using BR and PSNR

Performance of proposed approach w/ and w/o content feature



Summary

- Proposed a learning-based approach for video quality predictions
  - Fed by four segment features
  - Achieving very high performance
  - Outperforming three existing approaches
- It is found that segment quality and stalling features are key features in the proposed approach.
- For different quality metrics to represent segment quality feature, the impact of content feature is different.
- Future work: evaluating the performance of the proposed approach for sessions of durations longer than 1 minute.

Thank you for listening!