A Linear Regression Framework For Assessing Time-Varying Subjective Quality in HTTP Streaming

Nagabhushan Eswara
IIT Hyderabad
November 14, 2017

Introduction

- Increased mobile data traffic[1] : Data traffic growth of 63% in 2016 and is estimated to increase 7-fold between 2016 and 2021
- Exorbitant rise in video traffic[1] : 60% of total mobile data traffic in 2016 and is estimated to increase 9-fold between 2016 and 2021
- More than three-fourths of the world's mobile data traffic will be videos by 2021[1]
- Need for *careful* and *efficient* design of networks for video streaming

---

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021.

HTTP Streaming

- DASH: the ISO standard developed by MPEG for video streaming over HTTP-based networks

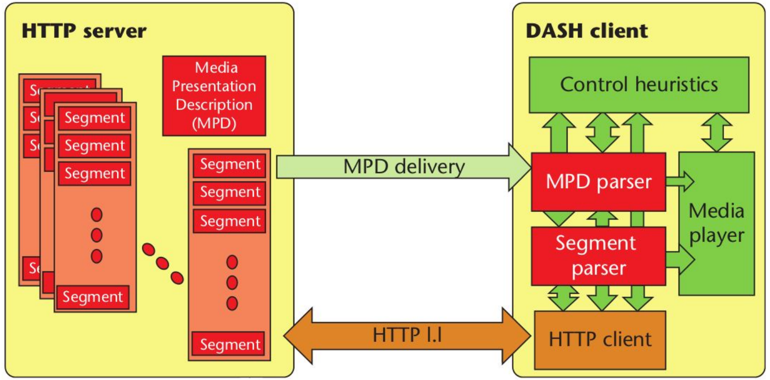


Figure: DASH Framework[2]

- Rate adaptation - a key feature of DASH

---

[2] Iraj Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet", *IEEE MultiMedia*, vol.18, no.4, pp. 62-67, October-December 2011.

Time-Varying Quality

- Rate adaptation leads to the time-varying quality (TVQ)
- TVQ affects the quality-of-experience (QoE) of users
- QoE is defined as *the overall acceptability of an application or service, as perceived subjectively by the end user*[3]
- Continuous monitoring of TVQ is essential in HTTP streaming for QoE maximization
- Need to quantify the dynamic perceptual TVQ, also known as the *time-varying subjective quality* (TVSQ)

---

[3] "Quality of Experience Requirements for IPTV Services", document *ITU-T G.1080*, Dec. 2008.

Quantifying Time-Varying Subjective Quality

- Network based measurements as TVQ proxies - video bitrate, user throughput etc.

Quantifying Time-Varying Subjective Quality

- ➢ Network based measurements as TVQ proxies - video bitrate, user throughput etc.

Q: Which of these video clips has got a better quality?

Quantifying Time-Varying Subjective Quality

- Network based measurements as TVQ proxies - video bitrate, user throughput etc.

A: Both video clips have an average bitrate of 250kbps!

Quantifying Time-Varying Subjective Quality

- ➢ Network based measurements as TVQ proxies - video bitrate, user throughput etc.

A: Both video clips have an average bitrate of 250kbps!

- ➢ Clearly, network based TVQ measures fail to capture the perceptual TVQ

Measuring TVSQ

- Video quality assessment (VQA) is the basis for measuring TVSQ
- A good number of VQA metrics proposed in the literature, broadly categorized into 3 categories –
  - Full-Reference (FR)
  - Reduced-Reference (RR)
  - No-Reference (NR)
- Using VQA, we propose a framework for measuring TVSQ based on linear regression

Proposed Framework for TVSQ Prediction

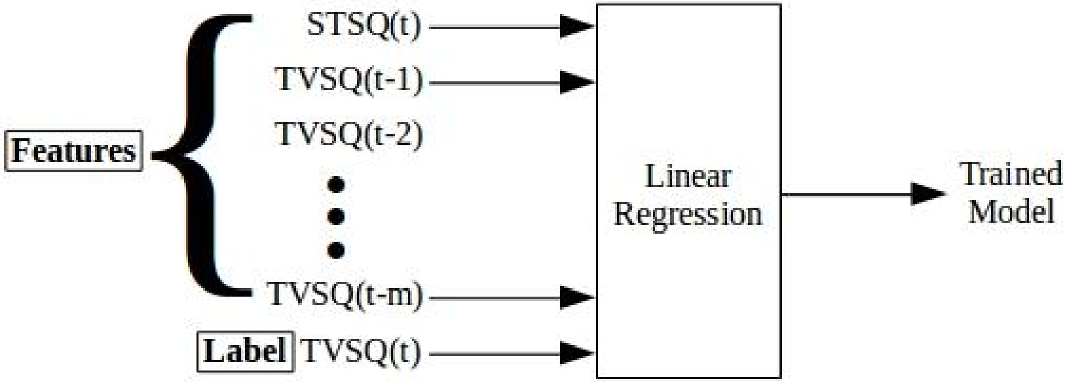


Figure: The proposed linear regression framework for TVSQ prediction

- Short Time Subjective Quality (STSQ): Current STSQ(t) evaluated using an efficient VQA metric
- Past TVSQs: TVSQ(t-1), TVSQ(t-2), · · · TVSQ(t-*m*)
- *m* represents the feedback order for TVSQ prediction

Selection of Feedback Order

- ➢ The correlation between STSQ and TVSQ on LIVE QoE[4] database is 0.61!
- ➢ Boost in the correlation performance with TVSQ feedback
- ➢ No significant improvement beyond II order feedback (*m*=2)

Figure: Significance of different feedback orders illustrated in terms of LCC.

---

[4] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, and A. C. Bovik, "Modeling the time-varying subjective quality of http video streams with rate adaptations", *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 22062221, May 2014.

RR-TVSQ

- Using the framework, we propose a RR method for predicting TVSQ called RR-TVSQ
- In RR-TVSQ, we employ STRRED[5] for computing STSQ
- STSQs can be pre-computed and be made available to video users apriori to facilitate TVSQ evaluation
- However, availability of the reference video may not be guaranteed in some scenarios

---

[5] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing", *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 23, no. 4, pp. 684694, Apr. 2013.

No-Reference TVSQ

- Using the framework, we also propose a NR method for predicting TVSQ called C3D-TVSQ
- We employ a pre-trained 3D convolutional neural network C3D[5] to extract spatio-temporal features from the video

Figure: C3D Architecture[6]

- Input spatio-temporal resolution = 171×128×16 (width×height×frames)
- 4096-dimensional feature vector

----

[6] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3d convolutional networks”, in *Proc. IEEE International Conference on Computer Vision,* pp. 4489–4497, Dec. 2015.

C3D-TVSQ

- ➢ Resizing the video resolution affects STSQ! Hence, videos are fragmented rather than resizing
- ➢ The extracted features are then used to train the linear regressor in the framework

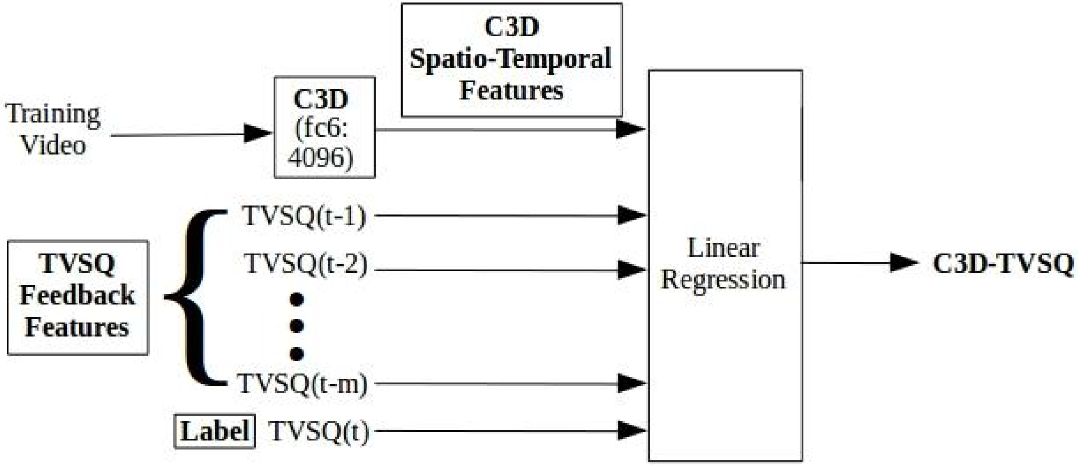


Figure: C3D-TVSQ

TVSQ Evaluation

- The proposed approach is trained and evaluated on LIVE QoE database[4]
- The database consists of 15 videos having TVQs due to rate adaptation and have a duration of 5 minutes each
- The database is divided into non-overlapping training and test sets with a split ratio of 2:1
- Thus, there are 10 videos in the training set and 5 videos in the test set

----

[4] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, and A. C. Bovik, “Modeling the time-varying subjective quality of http video streams with rate adaptations”, *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 22062221, May 2014.

TVSQ Evaluation

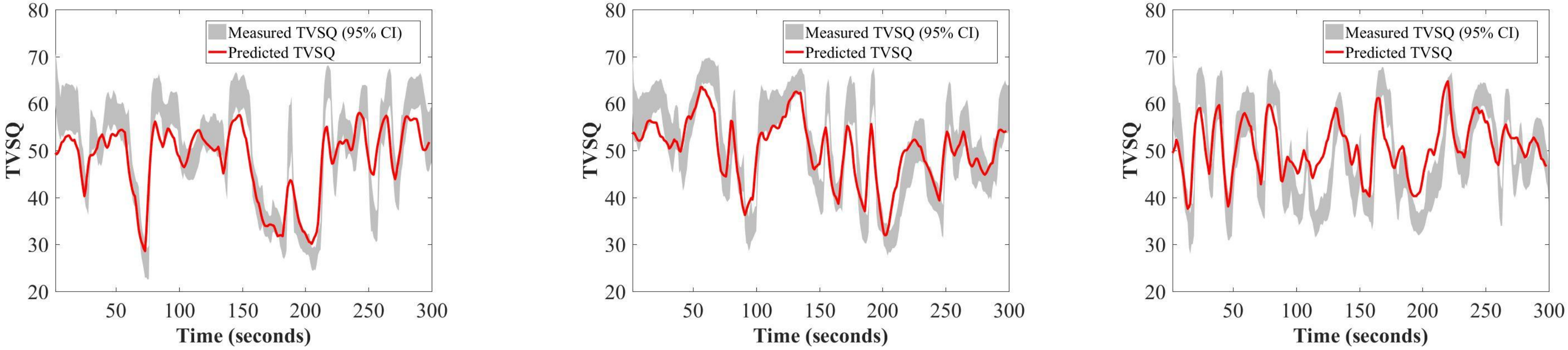
- The TVSQ feedbacks are initialized to 50, which is the average of the TVSQ's operating range [0,100]
- We employed two evaluation methodologies as following:
  - Monte Carlo Cross Validation (MCCV)
  - Leave-p-Out Cross Validation (LpOCV)
- The performance is evaluated using the following four measures -
  - Linear Correlation Coefficient (LCC)
  - Spearman Rank Order Correlation Coefficient (SROCC)
  - Root Mean Squared Error (RMSE)
  - Outage Rate (OR)

TVSQ Prediction Performance

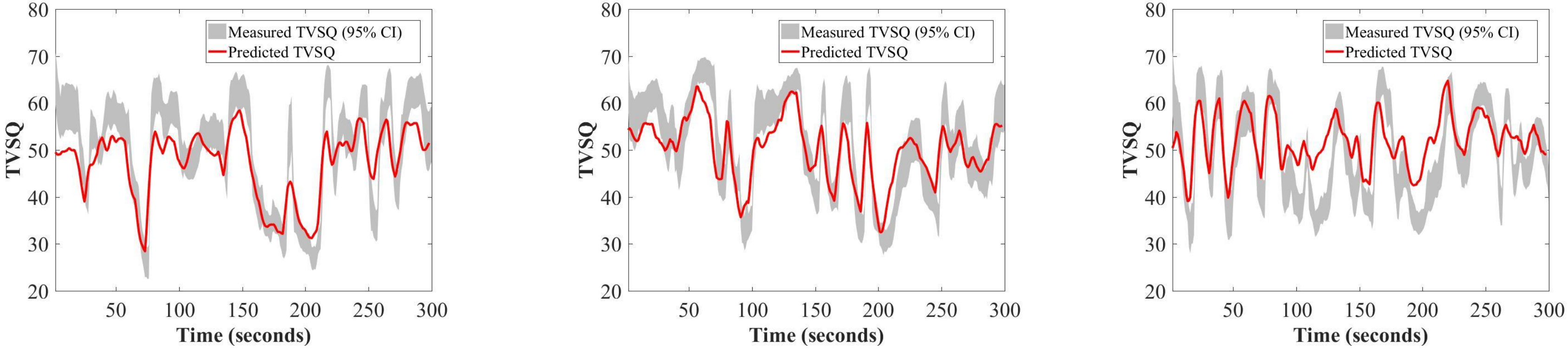
Table: Performance comparison of the proposed RR-TVSQ using STRRED and C3D-TVSQ against Hammerstein-Wiener[4] (HW) model on LIVE QoE

Method	*m*	LCC	SROCC	RMSE	OR(%)
MCCV-HW[4]	12	0.8787	0.8820	4.7675	10.8093
MCCV-RR-TVSQ	**2**	**0.8704**	**0.8729**	**4.6503**	**10.9396**
MCCV-C3D-TVSQ	**2**	**0.9041**	**0.9095**	**4.7468**	**10.2013**
LpOCV-HW[4]	12	0.8776	0.8917	4.6358	10.8000
LpOCV-RR-TVSQ	**2**	**0.8766**	**0.8698**	**4.8282**	**11.4094**
LpOCV-C3D-TVSQ	**2**	**0.9058**	**0.9008**	**4.7916**	**9.7315**

TVSQ Prediction Plots



(a) C3D-TVSQ: MCCV



(b) C3D-TVSQ: LpOCV

Inferences

- **State-of-the-art prediction performance of RR-TVSQ** using STRRED
- **Outstanding performance of NR C3D-TVSQ** demonstrates the effectiveness of C3D spatio-temporal features
- **Robust C3D spatio-temporal features**; could be useful in many potential applications
- Significant **reduction in the model order** from 12 to 2. This implies that the hysteresis dependencies are greatly simplified
- RR-TVSQ and C3D-TVSQ prediction performances demonstrate the **effectiveness of the proposed linear regression framework**
- The proposed framework for TVSQ prediction also provides **flexibility in choosing the VQA of choice** for STSQ with suitable feedback order

THANK YOU