International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

A machine Learning Model for QoE Prediction of Dynamic Video Streaming

Niveditha. B. S, Jayanthi K Murthy

Abstract*: Quality of Experience (QoE) is one of the important factors in deciding the efficiency of a network provider. This work discusses about the methodology in which machine learning can be incorporated to predict QoE of a dynamic streaming video or a webcam capture. The prediction of QoE is done using the combination of simpler neural network and Deep learning technique. The neural network trained with the H.264 encoded bit-streams is the Non-Linear Autoregressive Exogenous (NARX) model. Along with this unsupervised phase of training, Deep learning of Restricted Boltzmann Machine (RBM) is done keeping in mind the time-series changes of a streaming video. The NARX model acts as the training model that is capable of capturing the events from the database. Its feature of feedback network is extracted to make the prediction more apt. RBM deep learning extracts information from the trained neural network to predict PSNR, RMSE and SSIM which is in par with the ground truth. The combination of these algorithms gives prediction near to ground truth. The dataset used is encoded in the H.264 codec format. The presented work has very effective application in meeting the customer gratification by the network provisionary.

Keywords — QoE, RBM, NARX model, exogenous variable, Unsupervised technique.

I. INTRODUCTION

Quality of Experience (QoE) is a primary gauge of an internet service provider (ISP). It is the measure of overall satisfaction of the customer with ISP. It results from the degree of fulfillment of expectations with respect to the utility. The measurement of this metric score is the primary study over the years. QoE researchers have devised various techniques to define the score thereby ensure video services at its best. In early works, researchers were trying to increase user perceptual video quality by genuinely selecting the QoS parameters like video compression optimization, network bandwidth allocation etc. In few of the papers, authors have studied the relationship between the peak signal-to-noise ratio and quantization parameter.

Quality of Experience is a way to quantify the experience of a user using the service. Although there are a variety of methods in accuracy and complexity to estimate the QoE of a service, the simpler one is the subjective studies. On one hand objective studies rely on the signal distortion at the encoding or transport stage. On the other hand, subjective studies incorporate a complex and expensive process of rating the multimedia quality by test subjects in a tightly controlled environment. Mean Opinion Score (MOS) is one such way of calculating the QoE. It is a measure used in this domain representing the overall quality of a stimulus or system. It is the arithmetic mean over all individual values on a predefined scale that a subject assigns to his opinion on the performance. It is broadly classified under subjective quality evaluation and objective quality evaluation. Methods that involves least human intervention, unlike subjective evaluation, is necessary. To make this possible we present a Machine Learning technique that builds models in an online/offline learning fashion.

Machine learning is a methodology that provides the systems an ability to automatically study and improvise the experience without being explicitly programmed. Its important stages remain the collection of dataset, devising an algorithm to train, fine tune the datasets in order to provide the algorithm with all possible information to learn from. Overall the process can be summed as: gathering data, preparing the data, choosing a model, training the model, evaluation of the inputs and the outputs, tuning the algorithm as well as data to ensure proper learning and finally the prediction. Machine learning has been the most powerful way of determining the QoE score of a video. The neural networks are trained with collected extensive datasets and is made to distinguish between different qualities of videos. There has been several ways in which the types of machine learning modelling has been utilized to obtain the expected results [1]. The different modelling includes, deductive versus inductive, supervised, semi-supervised versus unsupervised learning, offline versus online modelling, batch versus incremental learning, passive versus active learning. One of the widely used machine learning analysis is the regression analysis. This gives the relationship between the dependent and independent variable. With the usage of two variables, prediction and forecasting becomes easier and prominent. Most of these models are built in offline batch manner using regression analysis and is considered to be inductive supervised learning techniques. However, along with choosing a proper learning model algorithm, the video encoding method becomes the first step to collect datasets. Different approaches of streaming the video over internet has been expounded [2]. Having known this, the suitable encoding algorithm can be judged.

Revised Manuscript Received on 30 May 2019.

***** Correspondence Author

Niveditha. B. S*, Student, Master of Technology, Department of Electronics and Communication, B.M.S College of Engineering, Bengaluru, India.

Dr. Jayanthi K Murthy, Associate Professor, Department of Electronics and Communication, B.M.S College of Engineering, Bengaluru, India.

As it is the known fact that video is a collection of frames formed of each images which in turn is the collection of pixels. H.264 encoding has been used in the current work to ensure the extraction of extensive datasets in the form of bit-streams. H.264 video encoding is beneficial in a way that it increases the rapidity of coding around an average of 17% and also the complication of coding is condensed. However, there is little or no effect on the image superiority and bit frequency [3].

II. RELATED WORK

Many works have been carried out by QoE researchers to determine the ways in which QoE can be judged. Video streaming is done either in the downloaded mode or the streaming mode. In streaming mode, the video content is not available at once, but is received and decoded. Hence this is a real-time nature and which has bandwidth, delay and loss requirements. Video encoding methods and the decoding efficiency plays a vital role in this case [2].

The evolving of QoE from QoS factor has been explained in multiple works till date. The major purpose here was to identify the objective QoS parameters that contribute to the perceptual quality, and map these to QoE aware parameters. Data-driven QoE assessment model has been devised by the author [4]. As and when the advancements on the Machine learning algorithms proved right, many correlation models were designed. The international telecommunication union (ITU) has categorized the types based on what a model focuses on. These classification includes parametric, bit-stream, media layer and hybrid models. QoE determination based on available QoS data, and improvement over time with more and more feedback availability has been discussed in [5]. Therefore, changes to the environment such as terminal types, content types will be detected via user feedback. This feedback will gradually help the model adapt to the new situation. The strengths and shortcomings of the then existing techniques have been discussed in [6]. Also they have discussed the necessary features and parameters to define and predict a QoE score. The techniques such as decision trees (DTs), random neural networks (RNNs), hidden Markov models (HMMs), Bayesian networks (BNs) and dynamic Bayesian networks (DBNs) were applied successfully to predict users' QoE in both laboratory and real-life environments. The main reason for the success of AI- and ML-based methods is due to attributed solid mathematical models for QoE modelling and prediction. This QoE prediction system designed for video streaming, were generally uni- or low-dimensional and modelled the impact of single video descriptors independently. Considering this, a high-dimensional input space to model the impact of buffering and initial delay on QoE has been described in [7]. The Bayesian Network model is expounded and proposed in [8]. Here the machine learning technique is used to exploit QoS factors. From the QoS score, QoE was estimated and predicted. Apart from the above mentioned machine learning techniques, there has been experiments made with No-Reference and Full-Reference algorithms [9]. No-Reference (NR) methods have low computation and assists in real-time quality measurement. This particular approach is not tied to any particular type of video, compression, or transmission means. However, the estimation is more accurate if the dataset is reduced from 80% to 20%.

III. IMPLEMENTATION SETUP

As described earlier, the first step in a machine learning algorithm is collection of datasets. In this work the inputs to the machine learning training model is the parameters of a video. This has been provided from the YouTube and Netflix collected database which is encoded in the H.264 codec format. The database has been collected and is expounded in [10]. The parameters of video is extracted from each frame of the video considering it as an image. The training model used here is the NARX model and the learning technique used is deep learning through Restricted Boltzmann Machine. The below sections has each of the steps explained. The outline of the work is described in the block diagram below:

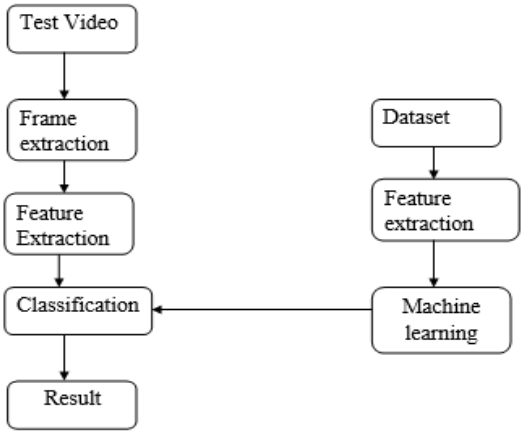

Figure 1. Block diagram of the implementation outline

A. Video encoding

H.264 advanced video coding is a motion-compensation based video compression standard. This standard was established by ITU-T Video Coding Experts Group. The H.264 video format has a broad application range that covers all forms of digital compressed video from low bit-rate Internet streaming applications to HDTV broadcast and Digital Cinema applications with nearly lossless coding. In the current work the same format has been used to encode the video frames to bit-streams. Encoding outputs are initialized. These include PSNR for each frame, BitRate for each frame and output bit-stream variable. Initial frame is defined as the I-frame and is followed by consecutive P-frames. A prototypic way in which the encoding is done has been implemented. The database used has collected information from the videos on YouTube and Netflix. It has been referred with different database labels.

B. *Training Model*

The training model used is Non-Linear Auto Regressive Exogenous model. In a NARX model, the two types of inputs are: past outputs which are fed back as future inputs to the dynamic model and external (or "exogenous") variables.

Hence it is a recurrent dynamic network, with feedback connections enclosing several layers of the network. The NARX model is a subset of ARX model, which is generally used in time-series modeling. The defining equation for the NARX model is

$y(t)=f(y(t-1),y(t-2),\ldots,y(t-ny),u(t-1),u(t-2),\ldots,u(t-n_u))$--- 1
In the above equation the next value of the dependent output signal $y(t)$ is regressed on previous values of the output signal and previous values of an independent (exogenous) input signal. We can implement the NARX model by using a feed-forward neural network in order to approximate the function f. A diagram of the resulting network is shown below, figure 2, where a two-layer feed-forward network is used for the approximation. This implementation also subsides for a vector ARX model, in which the input and output can be multidimensional.

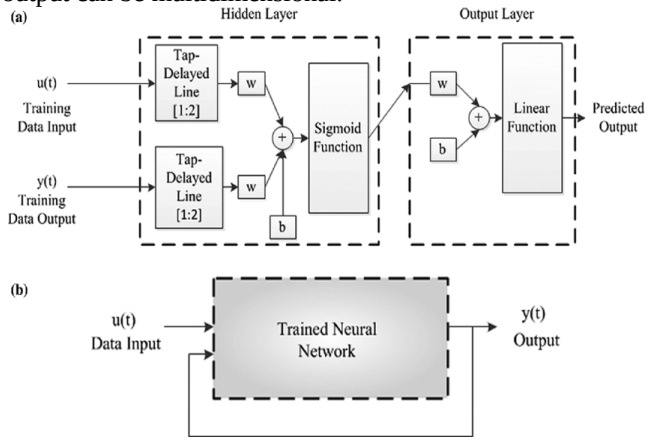

Figure 2. Two layer feed-forward network of a NARX model.

The datasets collected is fed into the NARX model by defining db_test variable. Different labels of the dataset is chosen by this variable. The database read module is defined in MATLAB. Random selection of the video label is chosen by rand module. This particular video frame/image is the first test index for the training model. The train indices is stored and noted each time the test index proceeds in the array. This training is repeated for possible quality models. The plot of these NARX training results is plotted against the ground truth which will be later described in the results section. However, the observation with this training is that, the RMSE (root mean square error) does not completely trace back the ground truth, due to which a deep learning technique has been employed.

C. Deep learning

Restricted Boltzmann Machine is the Deep learning algorithm and is the unsupervised method of machine learning. The aim of RBMs is to find patterns in data by reconstructing the inputs using only two layers namely visible layer and the hidden layer. The visible input variables is where the number of rows is number of data and number of columns is the number of visible input nodes. Hidden output variables are where the number of rows is the number of data and number of columns is the number of hidden output nodes. A linear mapping module calculates the linear mapping matrix between the input and the output data.

IV. EXPERIMENTAL RESULTS

As explained in the implementation setup section, training model i.e. NARX reads along the indices and records the quality model. For example, a video with the quality read of a news channel, will be identified with a same set of PSNR and RMSE thereby QoE prediction values. The first record of the training model is considered as the test index and the next as train indices. Likewise, every random index picked up by the rand module gives the preceding prediction. In order to ease the explanation, one of the database labels has been chosen i.e. STALL DB. The video number is noted and the corresponding ground truth is plotted with number against the parameter to be studied (i.e. SSIM, PSNR or RMSE). The prediction of the trained NARX is observed for all the collected video streams in the database. The plot of video number v/s RMSE for ground truth and NARX model is as shown in the figure 3. However, the prediction as observed from the graph is more deviated. This is due to the fact that NARX requires more collection of videos in database and stronger feedback inputs from the algorithm themselves and no regressive learning is possible to greater extent. As a result we also employ the deep learning into the prediction scenario. Due to this the prediction is not all dependent on the inputs from the coded algorithm, but has flexibility of hidden layer inputs which the model itself calculates in the iterations. The RBM captures the latent features of the input data, thus providing a high-level representation of the video segments at different compression levels. The approach here relies on a family of generative models which can be implemented as stochastic recurrent neural networks. They can be interpreted as probabilistic graphical models. With equal weight in either direction, it can be said that the connections between units are symmetric. The input to the network is given through units of a visible layer (i.e., observed), and these are fully connected to another layer of hidden units. This in turn is used to model the latent features of the data. The deep learning benefits are applied and exploited on the NARX-trained model. The RMSE and QoE (combination of RMSE and PSNR is considered to be the QoE deciding factors) are fed into the algorithm in order for it to recognize the quality of the video under test. The deep learning plot of RMSE against Video number using RBM learning is shown in figure 4. The simpler equation of RMSE calculation is:
$RMSE=sqrt(sum(Orginal\_Image)–(Estimated\_Image))^{2/N}$---2
PSNR value which is treated as the direct QoE prediction is calculated as below: $PSNR = 10\log_{10}(peak\ value^2/RMSE)$ ----- 3 Along with the above parameters being captured, SSIM is also studied. The Structural Similarity (SSIM) Index, a quality assessment index, is based on the computation of three terms, namely the luminance term, the contrast term and the structural term. The overall index is a multiplicative combination of the three terms.
$SSIM(x,y)=[l(x,y)]\alpha\cdot[c(x,y)]\beta\cdot[s(x,y)]\gamma$ ---- 4

Figure 5 shows the plot of SSIM obtained from the RBM learning model against the ground truth.

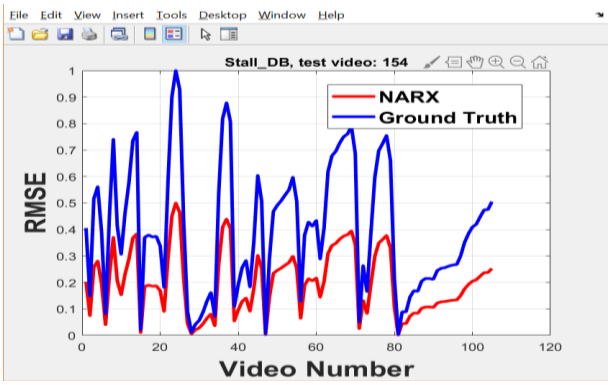

Figure 3. Plot of ground truth and NARX training model for RMSE against Video Number studied.

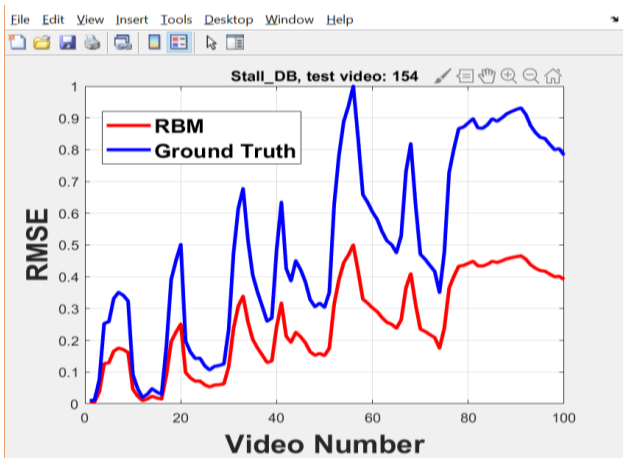

Figure 4. Plot of ground truth and RBM learning model for RMSE against Video Number studied.

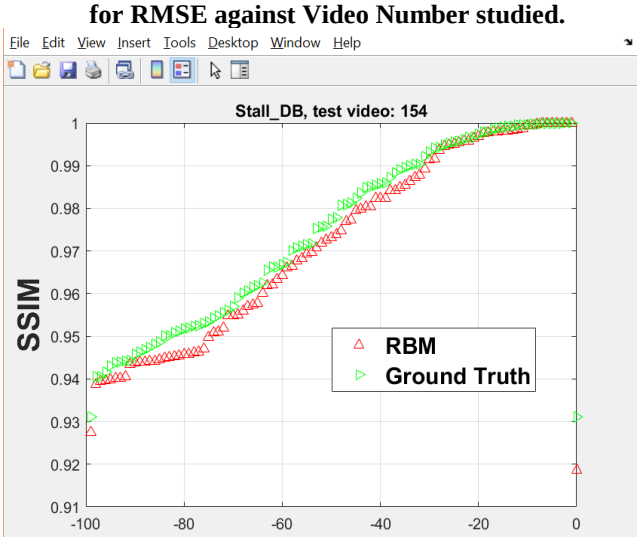

Figure 5. Plot of SSIM showing the RBM model learning prediction against the ground truth.

Table I

VALUES COLLECTED FROM RANDOM VIDEO FRAMES

Video Number	NARX	Ground Truth	RBM	Ground Truth
1	48.7579	53.8941	61.0268	61.9472
2	49.2040	57.6457	61.2996	63.1172
3	54.2614	59.3272	61.5946	63.6076
4	52.6642	60.0034	61.9742	63.3670
5	50.3832	60.5984	62.3247	63.5498
6	49.7119	61.9472	61.4892	63.4140
7	50.6245	63.1172	50.4729	49.2977

V. CONCLUSION AND FUTURE SCOPE

This work has developed a learning model with a combination of regression analysis model and deep learning. The main aim with which the work was initiated has been to develop a more robust video quality learning methodology. This has been achieved and the deep learning of the neural network is tested on various databases. The prediction is more efficient than previously devised methods due to the fact that deep learning algorithm designed is not based only on static frames of a video. Rather it is concentrated on re-buffering events as well. The datasets chosen are in such a manner that the real-time comportment of a video is captured and tuned. Network and buffer conditions, along with server-client related adapting capabilities during streaming of a video has been taken into account. RBM has more appropriate prediction and lesser deviation from the ground truth. It talks about the QoE in terms of PSNR without having the previously used subjective method but with rather a mathematical results. As mentioned in equations 2 and 3, these parameters are well extracted leading to a more statistical prediction rather than only random data. NARX model being able to capture the non-linear data has confirmed advantageous as against earlier linear models. Non-linear models get trained over real-time nature of a video.

Various set of database can be experimented with these training and learning models. Global efficiency of the methodology is approximately observed to be between 80-90%. (Calculated w.r.t the ground truth alongside the RBM learnt values).

Furthermore, the deviation from the ground truth can be lessened by using more varied characteristic video database. Deep learning algorithm can be further fine-tuned to give more appropriate results on the chosen QoE parameters. Enhancement on the machine learning models has to be done so as to capture the re-buffering events. In the present work, these have been captured as parameter in the database.

REFERENCES

1. Sana Aroussi, Abdelhamid Mellouk, “Survey on Machine Learning-based QoE-QoS Correlation Models”. 2014 International Conference on Computing, Management and Telecommunications.
2. Dapeng Wu, Thomas Hou, Wenwu Zhu Ya-Qin Zhang “Streaming Video over the Internet: Approaches and Directions” IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 3, MARCH 2001
3. Yufeng Li, Jufei Xiao, Wei Wu College of Electronic information engineering Shenyang Aerospace University Shenyang, China “Motion Estimation Based On H.264 Video Coding” 2012 5th International Congress on Image and Signal Processing (CISP 2012)
4. Yanjiao Chen, Kaishun Wu, “From QoS to QoE: A Tutorial on Video
5. Quality Assessment” IEEE COMMUNICATION SURVEYS & TUTORIALS, VOL. 17, NO. 2, SECOND QUARTER 2015.
6. Vlado Menkovski, Georgios Exarchakos, Antonio Liotta, “Machine Learning approach for Quality of Experience aware networks” 2010 International Conference on Intelligent Networking and Collaborative Systems.
7. Karan Mitra, Arkady Zaslavsky, and Christer Åhlund, “QoE Modelling, Measurement and Prediction: A Review” arXiv:1410.6952v1 [cs.NI] 25 Oct 2014
8. Pedro Casas, Sarah Wassermann, “Improving QoE Prediction in Mobile Video through Machine Learning”, Conference Paper · November 2017

9. Vladislav Vasilev, J´er´emie Leguay, Stefano Paris, Lorenzo Maggi, M´erouane Debbah, "Predicting QoE Factors with Machine Learning" IEEE 201
10. Maria Torres Vega , Decebal Constantin Mocanu , Antonio Liotta "Predictive No-Reference Assessment of Video Quality" arXiv:1604.07322v2 [cs.MM] 27 Apr 2016
11. C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of Temporal Effects on Subjective Video Quality of Experience," IEEE Trans. Image Process., vol. 26, no. 11, pp. 5217–5231, 2017

AUTHORS PROFILE

Niveditha. B. S, Student, pursuing Master of Technology in the department of Electronics and Communication at B.M.S College of Engineering, Bengaluru, India.

Dr. Jayanthi K Murthy is presently serving as an Associate Professor in Department of Electronics and Communication, B.M.S College of Engineering, Bengaluru, India with 24 years of experience. She has authored more than 20 research papers in international conferences and repute journals. Her research interests span Wireless Communication, Computer Networking and its application in AI/ML.