Quality of experience prediction model for video streaming in SDN networks

Tasnim Abar*, Asma Ben Letaifa and Sadok El Asmi

COSIM Research Lab, MEDIATRON Lab,
Higher School of Communications of Tunis (SUP'COM),
University of Carthage,
El Ghazala, 2083 Ariana, Tunis, Tunisia
Email: tasnim.abar@supcom.tn
Email: asma.benletaifa@supcom.tn
Email: elasmi@supcom.tn
*Corresponding author

Abstract: To evaluate the network performance, network operators rely on quality of service. This measure has shown limits and great deal of effort has been put into putting in place a new metric that more accurately reflects the quality of service offered. This measure is known as Quality of Experience (QoE). QoE reflects the user's satisfaction for a service. Today, evaluating the QoE has become paramount for service providers and content providers. This necessity pushed us to innovate and design new methods to estimate the QoE. This paper comprises two parts: the first part defines our subjective method which evaluates the video quality over SDN networks. In the second part we try to cover the impairments of subjective methods by a novel method that predicts the QoE (MOS) based on machine learning, so we employ ML-classifiers, then we calculate the performance metrics to measure the performance of each algorithm to deduce the best algorithm.

Keywords: SDN; QoS; QoE; video quality; subjective evaluation; machine learning; MOS; performance metrics; network operators; SDN controller.

Reference to this paper should be made as follows: Abar, T., Ben Letaifa, A. and El Asmi, S. (2020) 'Quality of experience prediction model for video streaming in SDN networks', *Int. J. Wireless and Mobile Computing*, Vol. 18, No. 1, pp.59–70.

Biographical notes: Tasnim Abar is a PhD student at a final stage at Sup'Com Tunisia. She is a teacher at the Higher Institute of Information and Communication Technologies in Tunis. She received her engineer degree in Network and Communication in 2016 from National Engineering School of Gabes. Since 2016, she has been doing research in SDN networks, quality of experience (QoE) and ICN networks. She has published four papers in refereed conference proceeding. Her current research area focuses on D2D communication, IoT, Fog computing, and wireless networks.

Asma Ben Letaifa is an Assistant Professor and member of the MEDIATRON research lab at the Higher School of Communications, SUPCOM, University of Carthage, Tunisia. She holds a Telecom Engineering degree from SUPCOM, University of Carthage, Tunisia, and a PhD jointly from SUPCOM, University of Carthage and UBO, Université de Bretagne Occidentale, Brest, France. Her research activities focus on telecom services, cloud and mobile cloud architectures, service orchestration, big data and quality of experience in an SDN/NFV environment. She is author and co-author of several articles on these subjects. She is also author of several courses on telecommunications services, network modelling with queuing theory, web content, cloud architectures and virtualisation, massive big data content and machine learning algorithms. She is also co-author of the 'Linux practices' MOOC on the FUN platform.

Sadok El Asmi is a Professor and member of Cosim research lab at Sup'Com, University of Carthage, Tunisia. He holds his master degree in Applied Mathematics from University of Paris Sud and his PhD in Science from University of Paris Sud, Laboratory of Signals and Systems, CNRS-ESE. His intention is to identify estimation and detection strategies. Targeted applications are biomedical signal processing, QRS complex detection and R-R interval analysis. He has also initiated a research axis combining the algebraic approach with a stochastic approach, called extreme value theory, for the detection of rare events.

This paper is a revised and expanded version of a paper entitled 'Machine learning based QoE prediction in SDN networks' presented at the '13th International Wireless Communications and Mobile Computing Conference, IWCMC 2017', Valencia, Spain, 26–30 June 2017.

1 Introduction

Over the years, multimedia applications have conquered many segments of the telecommunications industry. Today we are dealing with multimedia services in many areas, starting with the different digital television systems (e.g., DVB), video telephony, Video on Demand (VoD), Internet Protocol television services (IPTV), Voice over IP (VoIP) or simply, video streaming. The development of these services and the end-to-end optimisation of these systems are closely linked to the perception of quality by the user and his satisfaction with the service rendered. In this sense, there is a strong need for a measure that reflects user satisfaction and perception. Indeed, media service providers are increasingly interested in evaluating the performance of their services as perceived by end-users, in order to improve and better understand the needs of their clients. Network operators are also interested in this measure to optimise network resources and even reconfigure network settings to increase user satisfaction. There are several ways to get information about perceived quality. On the one hand, there are subjective evaluations carried out in well-equipped laboratories investigate the perception of the end-use. On the other hand, there are objective measures of quality, which are often used to study the measurable parameters of the whole system, describing the quality of services QoS. However, these parameters cannot describe all the variables that influence the perception of quality on the end-user side. For this reason, a new measure, called QoE, was defined to reflect the quality perceived by end users. The definition of QoE is closely related to the subjective perception of the end user. QoE is described by the ITU-T as “the overall acceptability of an application or service as perceived subjectively by the end-user”, which “may be influenced by user expectations and context” (ITU-T SG12, 2007). Research on Quality of Experience (QoE) is often based on subjective studies. In these subjective studies, users note the perceived quality of a service or application. As a general rule, these studies are carried out in a specialised laboratory. However, these subjective tests are tedious and costly. In addition, this kind of test cannot be applied in a real-time system, such as most media services. In this context, research has focused on new methods that try to approximate and estimate the QoE in an objective way and that can be used in real-time contexts. The main disadvantage of existing solutions lies in the fact that they are not correlated with the subjective tests and therefore cannot adequately reflect the perception of the end-user.

All these new skills need a new programmable context, allowing a global centralised view of the network and management dynamic resource. Software-Defined Networking (SDN) is one of recent solutions. It is based on a complete separation in equipment network between the control plane and the plane routing data. This separation offers operators the ability to schedule and automate control of their networks. SDN’s goal is to offer network flexibility and programmability to make its management simple.

In this paper, we propose a QoE prediction model based on five different learning algorithms to estimate the QoE for video streaming over SDN network. The rest of this work is organised as follows. In Section 2, we define SDN, MOS and QoE. In Section 3, we briefly clarify our problematic and contributions. We present some related work that talks about QoE and estimated models using machine learning in Section 4. Section 5 details our subjective and predictive methodology and experiments and Section 6 analyses and discusses the results. We sum up the work and share the future work in the conclusion section.

2 Context

In this section, we try to present SDN network, QoE and MOS.

2.1 SDN network

Software Defined Networking (SDN) is a new networking paradigm in which the forwarding hardware is decoupled from control decisions. It promises to dramatically simplify network management and enable innovation and evolution (Boutaba and Da Fonseca, 2015). By separating the control plane, SDN allows to introduce the concept of the programmability data plan, in fact, with this separation the switches become a simple device for transmission and all the intelligence of the network is implemented in a logic controller, programmable and centralised. In fact, in a typical network, when a packet arrives at a port on a switch or router, the packet applies the routing or switching rules that are written to its operating system. Generally, all packets that have the same destination follow the same path. In high-end models, hardware is able to recognise the type of application and apply specific rules to it. But this programming is rigid. It can only be changed manually by the administrator, which obviously takes time and does not lend itself to rapid context changes. With the SDN, these changes are automated and even programmable. The administrator defines the rules in the controller (system brain), and these are instantly transmitted to the network devices.

SDN uses OpenFlow as protocol; the development of OpenFlow began in 2007 as part of a collaboration between the worlds of university and business. Originally established by Stanford University and the University of California in Berkeley, this standard is now defined by the Open Networking Foundation (ONF). HP has been a leader in OpenFlow technology since its inception and is a founding member of the ONF (SDN, 2012). OpenFlow allows simplified programming via a standard interface for network devices. The ease of programming makes it possible to design a robust control layer, in order to centralise the intelligence in the network and to provide the programmability promised by SDN.

Unfortunately, it is expensive and not easy to configure and test an SDN environment composed of real devices. However, a few SDN network emulators have been developed to make these tasks simpler. One of the most widespread SDN emulators is Mininet (Mininet Simulation, 2016).

2.2 QoE and MOS (mean opinion score)

QoE is composed of a subjective dimension, alongside with objective measurements. It determines the level of end-user satisfaction by taking into account all the factors that affect not only the perception but also expectations. In fact, satisfaction reflects the degree of concordance between the user's experiences during the usage of a service (Martinez-Yelmo and Guerrero, 2010). QoE is an important indicator for network operators and service providers to help them assessing the user acceptability towards a particular service or a particular application. In many researches the output QoE scores are represented in terms of MOS (Mean Opinion Score) (see Table 1).

Table 1 MOS

MOS value	*Impairment*	*Video quality*
5	Imperceptible	Excellent
4	Perceptible but not annoying	Good
3	Slightly annoying	Acceptable
2	Annoying	Bad
1	Very annoying	Poor

3 Problematic and contributions

Today, user experience is a right indicator for network operators and service providers. Moreover, to compete for an important market part, different network operators and service providers should maintain and increase the clients' subscription. So they require an efficient QoE monitoring and estimation. QoE is a subjective metric which deals with user satisfaction and can vary due to the user expects, context and the state of the network. Moreover, subjective QoE evaluation is expensive and time consuming since it requires human participation. Therefore, there is a need for a tool that can objectively measure the QoE with reasonable accuracy in real time.

To fulfil these requirements, first, we choose to work with an SDN thanks to the intelligence and automation of the architecture (Boutaba and Da Fonseca, 2015) since this type of architecture uses hardware network equipment that is configured and controlled by a centralised software program called controller hence the term "software-defined". The network is configured and controlled by software, not hardware or equipment. The network benefits from an optimal allocation of its resources. Besides, we divide our work in two parts, the first which is subjective aims to study the effect of objective parameters on user satisfaction, the second is based on Machine Learning to build our model for estimating the quality of experience QoE Video streaming within SDN networks. In fact, subjective tests are very expensive, take a lot of time and have no real time applications (Machado et al., 2011). These methods cannot give an estimation of the user's satisfaction based on network and application parameters before the reception of the service in order to be improved if the prediction is poor. For this we need a machine learning algorithm to learn automatically from the current environment, the quality of experience more easily.

The next sections describe in details our approach and model.

4 Related work

Several researches talk about the QoE of video streaming and the parameters that affect this value, most of them use subjective methods to determine this influence on user perception.

Vranješ et al. (2007) introduced in their work some objective metrics "full reference" which determine in some way the visual quality of the video compared to the original video like Peak Signal to Noise Ratio (PSNR) SSIM (structural similarity), they used the subjective test Double Stimulus Impairment Scale (DSIS) to define the user's perception MOS and concluded their paper by a table representing the relation between SSIM, PSNR, VQM and MOS for example for static videos and with PSNR 38.06 and SSIM 0.989 they obtain MOS as 4.58. Then, Joskowicz and Ardao (2011) studied the influence of frame rate, bit rate, display size and video content in the perceived video quality. Also, Pande (2013) did a study on no-reference video quality metrics to evaluate mobile quality of video streaming over LTE network such as blocking, blurring, BRISQUE, the author found that blocking metric is able to quantify the degradation caused by packet losses and the blurring metric is able to quantify the losses in quality due to source code. Finally, Wu and Yuen (1997) used a weighted mean-squared difference along block boundaries to measure blocking effect, where the weights are obtained according to human visual masking effects. However, they cannot distinguish how much of the grey level difference between block boundaries is due to a real blocking discontinuity or the oscillation of the original signal itself.

So the most accurate approach to assess perceived quality is the subjective assessment because there is no better indicator of video quality than the one given by humans, but those methods have some weaknesses like they are expensive in terms of manpower, time consuming and cannot work in real time. That's why several researches develop various objective tools which predict the quality of experiences MOS from many parameters such as network parameters (packet loss, delay…) and application parameters (bit rate, resolution…). Those methods are usually based on machine learning, so we take the example of Menkovski et al. (2009)where authors estimated the QoE of video streaming from bit rate, frame rate, spatial and temporal information using three learning algorithms decision tree (DT), Support Vector Machine (SVM) and discriminate analysis, after comparison for the accuracy of each method they concluded that DT is the better (with a mobile phone they obtained 93.55 with J48 while with SVM they obtained 88.59 and 76.9 with discriminate analysis). Rodriguez et al. (2013) studied by VOIP and they decided to predict the MOS (calculated previously with subjective method PESQ) using decision tree, neural network, and Bayesian. They concluded also that DT is the best predictor J48: 0.98 while MLP: 0.92 and Naives: 0.78. Finally, Machado et al. (2011) estimated QoE metrics based on Quality of Service (QOS) metrics (throughput, packet loss, jitter, delay) in WiMAX networks using Artificial Neural Network (ANN). They found that ANN had a very good prediction and that errors of testing and validation optimal had satisfactory values. Also, there are other methods to optimise problems because of their excellent performances: the bat algorithm (BA) with triangle-flipping strategy (Cai et al., 2018), Ant Colony Optimisation (ACO) (Dorigo and Stützle, 2019) have been used to solve complex computational problems. It became popular because of its superior ability, which deals with a variety of complex issues.

In this paper, our work is concentrated in SDN network, it is divided in two. One part is for the subjective test to build our dataset that will be used in the second part which is the prediction test. The next part tries to predict the user perception based on some different intelligent algorithms and to compare the result of each method to deduce the best learning algorithm for our prediction model.

5 Our proposed model

The proposed method is a methodology for the prediction of QoE. It can be considered as a hybrid method between subjective and objective evaluation techniques. The main idea, illustrated in Figures 1 and 2, is to have several deformed samples evaluated subjectively, and passed by the MSU tool to calculate the objective metrics (VQM, SSIM ....), this step allows, on the one hand to build a well-defined data set to be used in the machine learning process and also to analyse the impact of several network and application parameters on video quality. In the second part of our method, we will build the predictive model that will estimate the user satisfaction by avoiding human interventions that are expensive and not in real time based on machine learning. For this and in order to build the most powerful model, one must select among several algorithms by analysing the metrics of performances: Pearson correlation, RMSE, MAE.

We will now describe in more details the role of each sub-module of our model.

5.1 Subjective test

Subjective assessments are the most accurate way to measure the quality of a video. In subjective experiments, a number of subjects (observers or participants) are invited to attend a set of tests and give a judgment on the quality of the videos or the inconvenience caused by distortions. The average of the values obtained for each test sequence is known under name Mean Opinion Score (MOS). In general, subjective assessments are costly and time-consuming. In consequence, the number of experiments that can be carried out is limited and, consequently, a methodology should be used to make the best use of resources. The International Telecommunication Union (ITU) has made recommendations for subjective test procedures. The two most important documents are ITU-R Recommendation BT.500-11 (2002), for television applications, and ITU-T Recommendation P.910 (2008), for multimedia applications.

Figure 1 Our framework

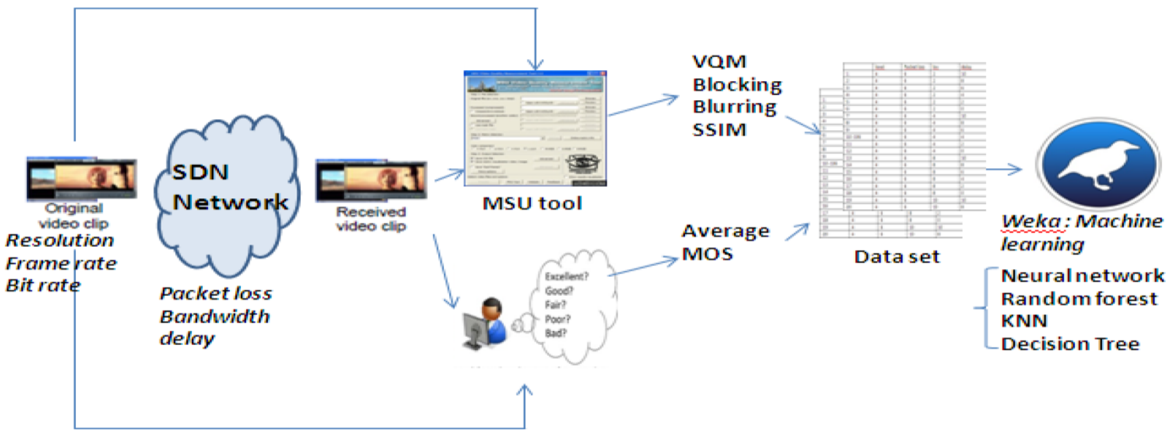
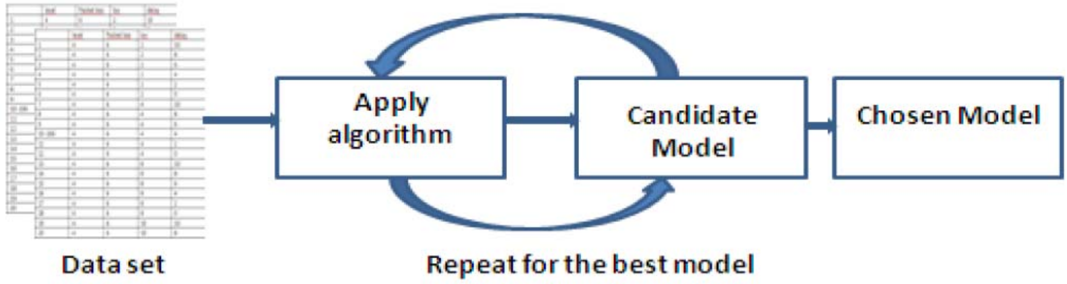


Figure 2 Machine learning process



Among the subjective test procedures proposed in ITU-R Rec. BT.500-11, we can cite: Double Stimulus Impairment Scale (DSIS) where the reference sequence is always displayed before the test sequence and the pair does not repeat. Observers are asked to judge the level of depreciation for each test sequence, using a five-level scale. This method is appropriate for assessing visible artefacts. Single Stimulus Continuous Quality Assessment (SSCQE) where pairs of multiple sequences (containing the reference and a randomly degraded sequence) are presented to the observers. DSCQS is useful when it is not possible to provide test conditions that show the full range of quality, Pair Comparison (PC), Stimulus Comparison Adjectival Categorical Judgment (SCACJ), Absolute Category Rating (ACR) and Degradation Category Rating (DCR) used in this work when participants will see short video sequences on the screen. Each sequence will be presented twice in rapid succession: within each pair only the second sequence is processed. At the end of each paired presentation, they should evaluate the impairment of the second sequence with respect to the first one. They will express their judgment by using a scale from 1 to 5 (Table 1) (ITU-R Recommendation BT.500-11, 2002; ITU-T Recommendation P.910, 2008).

5.2 MSU tool and objective parameters

It is a free software (MSU Codes, 2016) for objective video quality assessment. It provides functionality for both full-reference (two videos are examined) and single-reference (one video is analysed) comparisons. This application allows us to evaluate the video quality objectively by calculating different metrics: full-references metrics: VQM and SSIM and no-reference metrics: blocking effect.

1 *Full-references metrics*: NTIA VQM (2016) developed by Boulder Colorado Institute of Telecommunications Science. It is a standardised method for the objective measurement of video quality. VQM makes a comparison between the original video sequence and the distorted video sequences based only on a set of features extracted independently of each video. The algorithm used by VQM measures the perceptual effects of several distortions, such as blurring, jerky/non-natural movement, noise, blocks distortion and colour distortion. These measurements are combined into a single measure that gives a prediction of overall quality.

   The SSIM metric measures the structural similarity based on the HVS model. This method uses the measurement of structural distortion instead of error. SSIM is based on the fact that the HVS is more sensitive to structural changes in video than to luminance and contrast. It believes the quality of the video by extracting from the image information such as structure and contrast, and comparing the values of this information instead of directly comparing the pixels. Studies on the performance of SSIM have shown that this simple metric offers good results (Sheikh et al., 2006).

2 *No-reference metrics*: Most modern algorithms of video compression, including MPEG-2, MPEG-4 ASP, MPEG-4, AVC/H.264 etc., divide each frame into blocks of predefined size. The motion compensation technique is applied to each block after transformation of predicted residual. The goal of transformation is to minimise dependencies between block's pixels. Resulting coefficients are quantising and coding using lossless compression. Information loss during quantisation produces a number of artefacts in compressed video such as blocking effect, blurring effect, etc. (MSU Documentation, 2016).

Blocking metric: This metric was created to measure subjective blocking effect in video sequence. For example, in contrast areas of the frame blocking is not appreciable, but in smooth areas these edges are observable. This metric also contains heuristic method for detecting object's edges, which are placed to the edge of the block. In this case, metric value is pulled down, allowing measurement of blocking more precisely. We use information from previous frames to obtain better accuracy.

5.3 Machine learning

Intelligent machine learning algorithms are used in many domains because they are cheap, flexible, and more accurate than humans and can be performed automatically and in real time (Pokhrel, 2014). They learn automatically from the past observations to make accurate predictions in the future. They are mostly used in classification models. In our work, our framework uses machine learning to establish the correlation between subjective metric MOS and objective metrics VQM, SSIM, blocking. We chose as software to do this learning WEKA tool which is a collection of machine learning algorithms (WEKA Tool, 2017). We define in the following paragraph the five algorithms used in this work:

1 *Decision Tree (M5P)*: Decision trees are an important predictive algorithm modelling machine learning, this technique is fast to learn, easy to understand and has a high variance and can yield more accurate predictions. The representation of the tree model is binary. This is your binary tree from algorithms and data structures, nothing too imagination. Each node represents an input variable and a share point on that variable. The leaf nodes of the tree contain an output variable which is used to make a prediction. The estimated value is made by walking the partitions of the tree until arriving at a leaf node and output the class value at that leaf node. We have many important algorithms such as C4.5, ID3, and M5P defined as follows: it is used for numeric estimation and in each leaf it builds a linear regression model. This algorithm uses the splitting criterion to define the root node. The standard deviation of the estimated value in S is used as a measure of the error at that node and each attribute at that node is tested by calculating the expected reduction in error. The chosen attribute for splitting maximises the expected error reduction at that node. The Standard Deviation Reduction (SDR) equation (1) is the expected error reduction.

$$SDR = sd(S) - \sum \frac{|S_i|}{|S|} . sd(S_i) \tag{1}$$

where *sd* (*S*) function refers to the standard deviation of the values in the set *S*, while $S_1, S_2, \ldots, S_n$ are sets of values resulting from a split in a feature. The term |*S*| refers to the number of observations in set *S*.

2 *Random forest*: This is one of the most well-known machine learning algorithms, its principle is to use a large number of decision trees each constructed with a different sub-sample of the training set, and for each tree construct the decision at a node is made according to the subset of variables drawn randomly. Then, we use the set of decision trees produced to make the prediction, with a majority vote (for classification, predicted factor variable), or an average (for regression, predicted variable of numeric type).

3 *Linear regression*: Linear regression is a well-known and well understood algorithm in machine learning, it is characterised by its simplicity. This technique has existed for more than 200 years and has been widely studied. Some good rules of thumb when using this technique are to remove very similar variables (correlated) and eliminate the noise of your data if possible. It is presented by a linear equation which defines the best relation between the input variables and the estimated value, by finding specific weights for the inputs called coefficients.

4 *Gaussian process*: This is a supervised learning technique designed to solve classification problems, it uses lazy-learning. This method is accurate thanks to the prediction which interpolates the observations and we can compute empirical confidence intervals and decide based on those. The algorithm is also volatile, i.e., several kernels can be used, WEKA uses $\ker\mathrm{nel}(x, y) = \langle x, y \rangle$.

5 *Meta classifier (WeightedInstancesHandlerWrapper)*: Meta learning (Vilalta and Drissi, 2002; Lemke et al., 2015) is a sub domain of machine learning where automatic learning algorithms are applied to metadata on machine learning experiences. Although different researchers hold several views on what the term exactly means, the object is to use such metadata to understand how automatic learning can become flexible to improve the performance of existing learning algorithms.

5.4 Performance metrics

In this subsection, we introduce the four parameters used in the second part of our work to evaluate the performance of our models; these parameters help us to select the best algorithm.

1 *Pearson correlation coefficient*: This coefficient *r* provides quantitative measurements of the performance of the proposed metric. This coefficient measures the accuracy of video quality estimates relative to subjective results, if we have |*r*| = 1, we have $\mathit{Subjective\,MOS} \approx \mathit{Predicited\,MOS}$.The Pearson correlation coefficient is presented as follows:

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{2}$$

where x_i and y_i represent subjective MOS and estimated MOS, respectively. $\bar{x}$ and $\bar{y}$ represent the average subjective MOS and average estimated MOS.

2 *RMSE*: Root-Mean-Square-Error (RMSE) is also used to calculate the estimated accuracy of the objective models. We have better performance when RMSE tends to 0.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \tag{3}$$

3 *Mean absolute error*: MAE compares the deviation between the actual and the estimated value quantitatively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i| \tag{4}$$

MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average.

4 *Root relative squared error*: RRSE is the mean absolute error divided by the corresponding error of the ZeroR classifier on the data (i.e., the classifier predicting the prior probabilities of the classes observed in the data). Lower values are better and when we overtake 100%, the scheme is doing worse.

$$RRSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \tag{5}$$

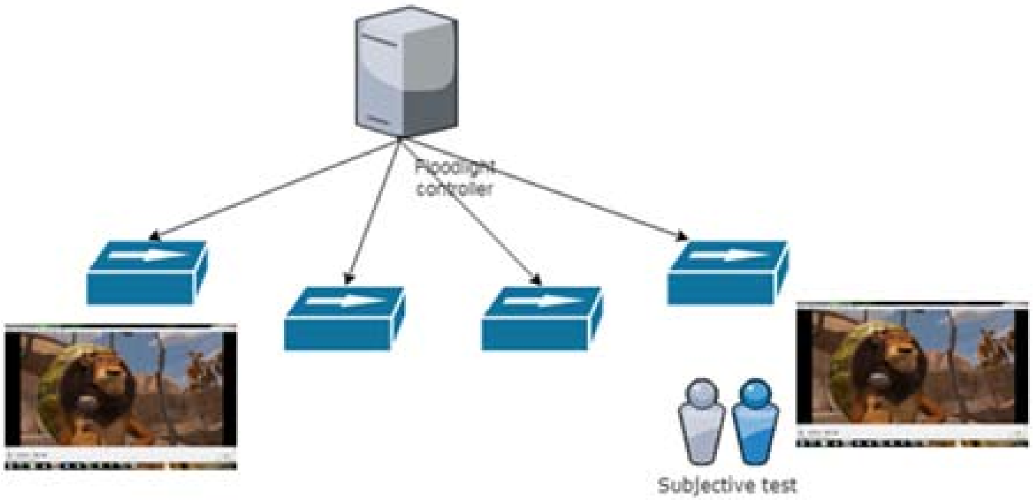
6 Methodology and test bed

6.1 Subjective evaluation

1 *Experiment setup*: In this part, we will use mininet as noted in the previous section for the simulation of SDN network, floodlight as a controller (www.floodlightcontroller/Floodlight+VM) to implement the control plan indeed it is an open source project sponsored by the Global constructor Big Switch. Floodlight can interact with equipment (physical or virtual) must necessarily implement the protocol OpenFlow, VLC for transmission and reception of videos streaming and finally we use MSU tool for the computation of objective parameters. Figure 3 shows our network architecture.

2 *Scenario and workflow*: Several works in the context of video streaming and QoE have chosen the tested videos according to the movement (static, in motion) (Vranješ et al., 2007; Tan et al., 2006) but in this work, we will use a new type, the subtitled videos, since the text in video gives another view on the quality. Then so the users won't feel uncomfortable if the video is boring, we try to choose comedy sequences.

Figure 3 Our network architecture



We define 152 levels (the worst 1 (packet loss = 6%,bandwidth = 2 Mbits/s, delay = 10, Triple (resolution, bit rate, frame rate) =1) and the best is 152 (packet loss = 0%,bandwidth=10 Mbits/s, delay = 0, Triple (resolution, bit rate, frame rate) =2)), varying each time the packet loss or bandwidth or delay, and we define also two types of triple (resolution, bit rate, frame rate) as shown in Table 2.

Table 2 QoS parameters and application parameters

Packet loss	*0,1,3,6 %*
delay	0, 2, 4, 6, 8, 10 secs
Bandwidth	2, 4, 8, 10 Mbits/s
Triple(resolution, bit rate, frame rate)	1 = (320 × 240, 768 kbps, 15 fps) 2 = (1280 × 720, 3000 kbps, 50 fps)

Each video is sent over SDN network, at the receiver we both save the video with the original in our database that will be used in the subjective test DCR (described in the next subsection) and we calculate with MSU tool our objective parameters (SSIM,VQM, blocking) to be compared after with subjective QoE.

3 *Data collection*: Total of 20 users were made part of our experiment, there are students, professor, staffs. Out of 20 users 10 viewers were females and 10 were males with age group of 20 to 35. To get the MOS we have used a rating application which exists at the client, thus the participant writes his name, age, profile, then he looks at the original video and the received video and at the end he is asked to note the quality of the second sequence with respect to the first one using a scale from 1 (poor) to 5 (excellent).

6.2 Predictive evaluation

Subjective evaluation is the most accurate method because it depends directly on the opinion of the user who is the goal of the service providers but these methods are very expensive, time-consuming, and often impractical, for this we try to estimate this opinion (MOS) in real time during the stream of video. The best way to do this is machine learning.

In this paper, we try to predict the user perception based on five different intelligent algorithms (three type of classifiers define in WEKA) and compare the result of each method to deduce the best learning algorithm for our prediction model, we chose Tree Classifiers (Decision Tree(M5P), Random Forest), Meta functions (WeightedInstancesHandlerWrapper), Function Classifiers (Gaussian Process, Linear Regression).

Our data set is built in the subjective evaluation we note levels, VQM, SSIM and blocking as attributes and MOS as the class which will be predicted by our model. In this part we start by calculating the performance parameters described in the previous section for each algorithm in different values of cross validation which can be described theoretically as follows: we divide our base into k subsets, then we leave the subset number *k* for the test and use *k*–1 subsets for the training, after we calculate the average of the errors of all *k* Subsets. So logically when *k* is increased the variance of the result is diminished.

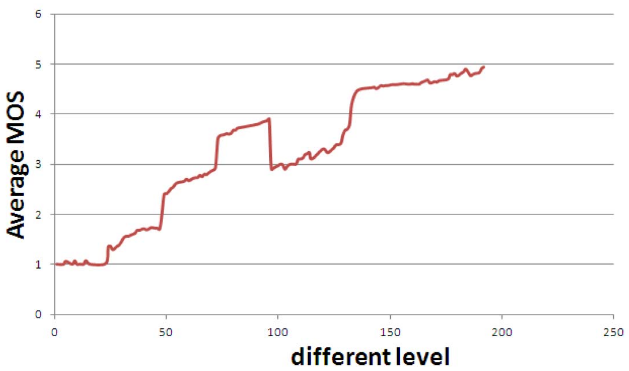
Figure 4 Frame from the test sequence

7 Analyse and discuss

7.1 Subjective evaluation

Figure 5 gives a view of the effect of the QoS parameters on the end users' perception. This curve has a growing gait, i.e. on the last level (packet loss = 0%, delay = 0, bandwidth = 10, resolution = 1280 × 720, bite rate = 3000 and frame rate = 30) we have MOS = 4.94 very high and an excellent quality. The drop presented in the same curve corresponds to levels having a triple type 2 and high packet loss = 6% which means that the packet loss is the parameter most affecting the quality of video and user' opinion especially if the video is in standard quality (triple type1) which verifies our work (Abar et al., 2017): {MOS(level 2 = (packet loss = 6%, delay = 8, bandwidth = 2, resolution = 360 × 240, bite rate = 768 and frame rate = 15) =1 and MOS (level 97 = (packet loss = 6%, delay = 8, bandwidth = 2, resolution = 1280 × 720, bite rate = 3000 and frame rate = 30) = 2.9}. Examine the levels defined in Table 3

Figure 5 Average MOS in different scenarios



The parameter that varies here is the bandwidth, from this list we can conclude that with standard quality (triple type 1) the bandwidth has no effect and it cannot improve the quality so the video remains very annoying. Or users are more comfortable when using high resolution (triple type 2) even the bandwidth decreases by 6 Mbits/s but MOS ≈ 3.

Table 3 Some levels from our data set

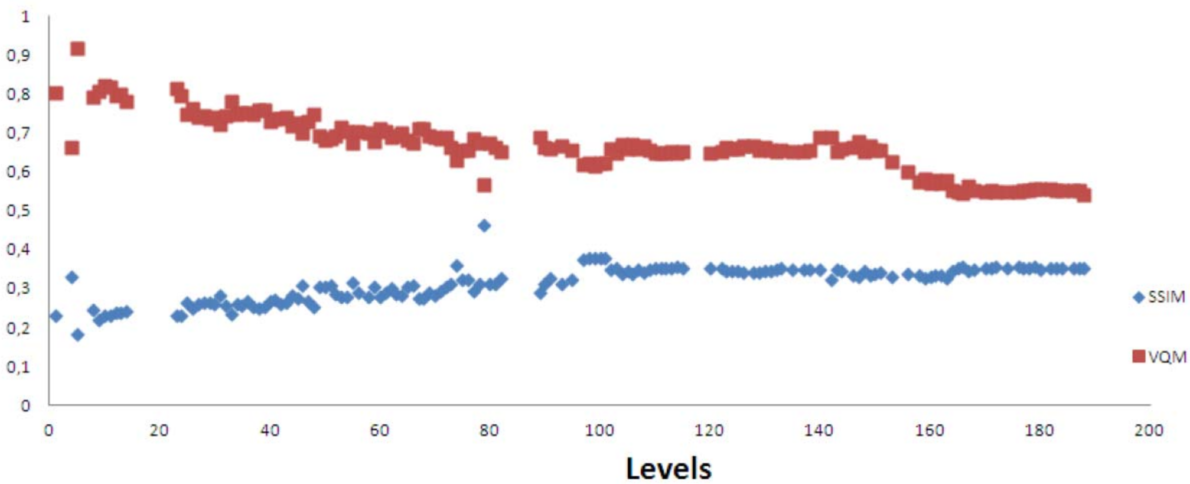
Level	*Packet loss %*	*delay*	*bandwidth*	*Triple type*	*MOS*
8	6	8	4	1	1
20	6	8	10	1	1.17
97	6	8	4	2	2.9
104	6	8	4	2	2.95
116	6	8	4	2	3.17

Based on Figure 5 also, it can be seen that the MOS is almost unchanged in each six successive levels in the first part but there is a small variation of MOS at the second part of the curve quality, which implies that the delay has no effect on the user's perception and the video is annoying whereas the delay has an effect for a high resolution (triple type 2) because of slowdown, especially at the beginning of the video and the speed of this slowdown increases with the delay so the quality of video decreases from good to slightly annoying (MOS (level 126 (delay = 10)= 3.39, MOS (level 131 (delay = 0)=3.7)).

After this discussion about QoS parameters we can conclude that all parameters are interesting for the video quality and therefore to user's perception but the packet loss is the most dangerous in a network (gives poor quality and users are not satisfied)

Figure 6 presents the relation between full-reference objective metrics (SSIM, VQM) and the different levels. We can notice that VQM is a decreasing function and SSIM is an increasing function.

Figure 6 Evolution of SSIM (blue) and VQM (red)



For the first 60 levels, where the triple type is 1 and we have a lot of packet loss in the network, we observe that SSIM is less than 0.32 and VQM is above 0.72, that means a colour distortion, jerkiness, and a problem in contrast and luminance; MOS (see Figure 1) is between 1 and 2.5 (very annoying video quality and users are not satisfied at all)

For levels between 60 and 96 where VQM between 0.66 and 0.7, SSIM is less than 0.32 and MOS is between 2.8 and 3.9, we mark again from the MOS value that the quality of video is improved but is still slightly annoying. This improvement is due to the decrease of packet loss and increase of bandwidth (from 2 to 10 Mbits/s)

For videos between 97 and 192 with high resolution (triple type 2): The blue curve shows that SSIM is about 0.39, this value remains small, which means that there are again problems of luminance and contrast, but if we look the MOS values (min = 3.9 and max = 4.9) we can conclude that these problems have no effect on the user's perception. Now looking at to the red curve we will discuss the values of VQM: VQM is between 0.66 and 0.57 in the first slice of the videos (97…140) this value is because the packet loss number is elevated although when the packet loss = <1% colour distortion and jerkiness decrease and VQM = < 0.55 and the video will be perceptible but not annoying and sometimes excellent.

So when SSIM tends to 1 and VQM tends to 0 the video quality will be better and the MOS achieves 4.99. Finally making the comparison with MOS we can plot Table 4 and we conclude a small correspondence between MOS (user 'perception), VQM, QoS parameters and application parameters SSIM for example when we obtain 0.79 as VQM we deduce that we have many packet loss (≈ 6%), bandwidth is low (1 or 2 Mbits) and MOS is between 1 and 1.2.

Now, we move to discuss the results given by no-reference objective metrics. Figure 7 presents the correlation between MOS (satisfaction of users) and blocking metric calculated by MSU tool. Red curve presents original blocking where we have calculated before the stream of video, and the set of points presented in blue defines the blocking effect in received video. In this figure, we are interested only in the difference between the two curves. When MOS is between 1 and 2, there is a large difference between red and blue curves ≈ 7 . Returning to the levels, we remark that users vote 1 and 2 as MOS when resolution is standard and the network is mediocre, the higher deviation appears when we have most of packet loss (level 1, packet loss = 6%, MOS = 1).

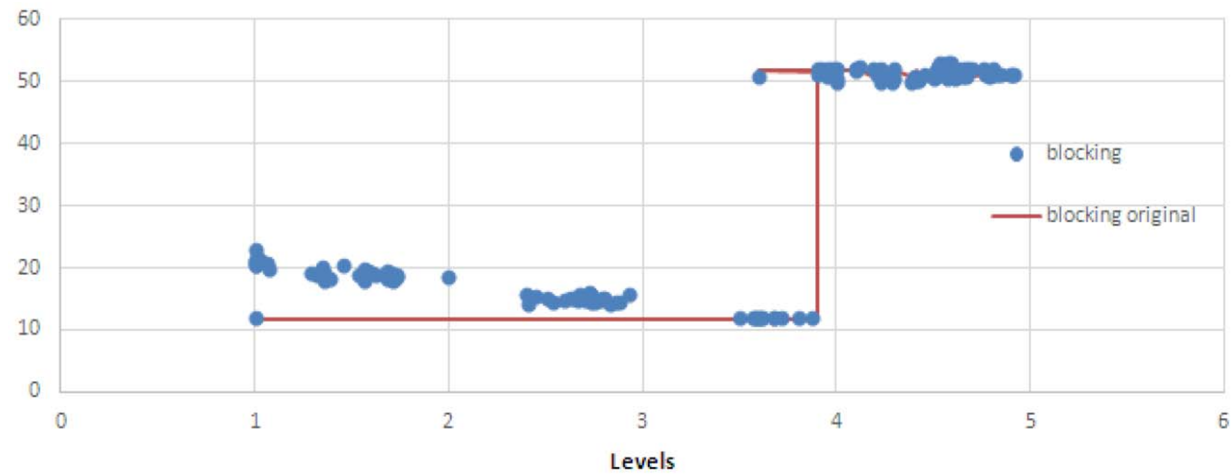
Table 4 Correspondence between MOS and VQM

MOS		*VQM*
1	bad	0.78 < VQM
2	low	0.68 < VQM < 0.78
3	fair	0.66 < VQM < 0.68
4	good	0.55 < VQM < 0.66
4.8	excellent	VQM < 0.55

When MOS is between 2.5 and 3, the values approach the original slowly with an improvement of MOS, i.e., the network is improving according to Figure 5 (packet loss ≈1). Passing to the MOS between 3.5 and 4.94, the two curves are almost confused, i.e., the network does not block in the videos and comparing the different levels that is done when we improve the network parameters (packet loss <1%).

So we can conclude that blocking effect is the most annoying artefact, it plays an important role in the optimisation, and development, it is very useful to evaluate the quality of videos, we can say also that blocking is able to quantify the defacement caused by packet loss strongly.

Figure 7 Evolution of blocking effect



8 Predictive evaluation

In this part we will interpret the performances of the algorithms described before and conclude the best technique to our model. The meta classifier WeightedInstancesHandlerWrapper algorithm gives very low performances and an imperceptible correlation ($|r| = 0.24 \ll 1$, RMSE = 1.3 >> 0, RRSE = 100%) so this algorithm does not fit our model thus it will be eliminated.

Now we vary the *k*-Cross Validation (CV) = 5,..., 14 and we compute *r*, RMSE, RRSE and MAE of the four other algorithms to obtain these four figures (see Figure 8). The four algorithms have a strong correlation between the estimated value and the original $|r| > 0.96$ (Figure 8a) or M5P and Random Forest have the highest values (M5P with CV = 5,…, 9, Random Forest with CV = 9,…, 14). The curves in Figures 8b and 8c have almost the same pace and it is normal since RMSE and RRSE have almost the same physical significance. M5P has the smallest value with all the CV (RMSE ≈0.1, RRSE ≈8%) on the other side Gaussian process has the highest values (RMSE ≈0.3, RRSE ≈25%) which degrades its performances. The last curve defines the variation of MAE as a function of the CVs for the four algorithms. MAE compares the deviation between the actual and the estimated value quantitatively. Random Forest reaches the smallest deviation 0.06 and Gaussian process gets the highest value which deteriorates its performances again. We cannot ignore the linear regression algorithm which is known as the simplest and most understood algorithm in machine learning, we obtain an important $|r| \approx 0.98$ but a large value of RMSE ≈ 0.23 and RRSE $\approx 18.6\%$.

From this analysis it is shown that M5P and Random Forest have the best performances. Table 5 summarises the different parameters between them: besides the two algorithms have more advantages compared to other learning techniques. For this we decide to add another parameter that compares the time required to execute each algorithm or called complexity time of the algorithm.

Comparing those values, we conclude that M5P has the best performances, also the lowest complexity time (20 ms). Figure 9 defines the tree constructed by M5P to calculate the predicted MOS where *r*, RMSE, RRSE and MAE have the best values.

Figure 8 The estimation accuracy of four QOE estimation algorithms. (a) Pearson correlation (b) RMSE (c) RRSE (d) MAE

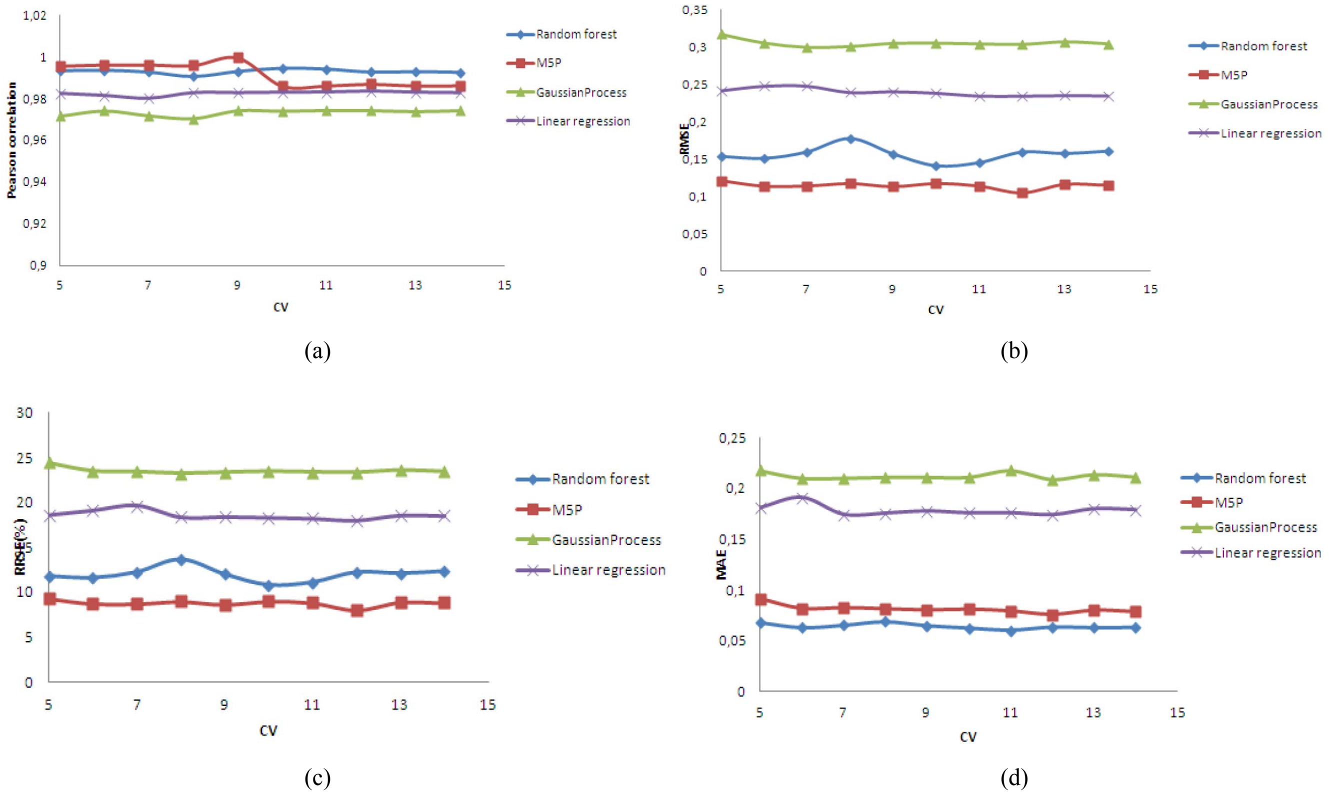


Table 5 The accuracy comparison of M5P and RF

	CV	*14*	*13*	*12*	*11*	*10*	*9*	*8*	*7*	*6*	*5*
Random Forest	r	0.9925	0.9928	0.9926	0.9939	0.9943	0.9929	0.9907	0.9926	0.9934	0.9932
	RMSE	0.1604	0.1572	0.159	0.1446	0.1405	0.1567	0.1776	0.1593	0.1506	0.1533
	RRSE	12.354	12.0901	12.2388	11.1286	10.8111	12.0058	13.64	12.2324	11.6041	11.8064
	MAE	0.063	0.0626	0.0633	0.0599	0.0622	0.0648	0.0691	0.0652	0.0626	0.0682
	time	0.065 sec									
M5P	R	0.9861	0.986	0.9868	0.986	0.9859	0.9962	0.9959	0.9962	0.9962	0.9956
	RMSE	0.1147	0.1158	0.1043	0.1136	0.1169	0.1128	0.1169	0.1136	0.1134	0.1208
	RRSE	8.8286	8.9069	8.0299	8.8349	8.9913	8.6427	8.9795	8.7243	8.7364	9.3065
	MAE	0.0785	0.0799	0.0758	0.0792	0.0809	0.0802	0.0813	0.0824	0.0817	0.0912
	time	0.02 sec									

Figure 9 M5P tree of our model

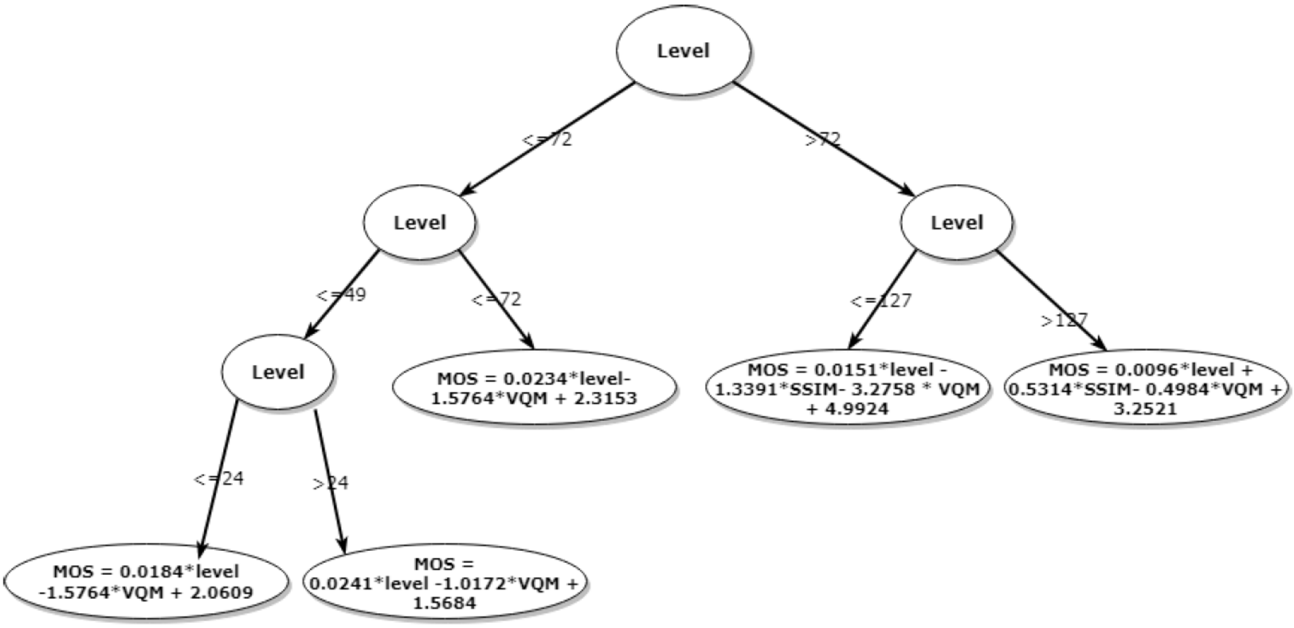


Figure 9 presents that the distribution of our tree is according to the QoS parameters of SDN networks and application parameters (levels), we obtain five linear equations in function of VQM, SSIM and levels also. If the level is lower than 72 (packet loss <1%, high resolution) the estimated MOS depends of the number of levels, VQM and SSIM, on the other side (level = <72), we have a standard quality, the SSIM has no effect on the predicted MOS. In both sides, M5P does not use blocking metric to calculate MOS.

9 Conclusion

The evaluation of the quality of experience of video streaming is a major problem in the context of the internet of the future and SDN networks. There are several methods in the literature that use either network parameters or characteristics of the voice or video signal. These methods have several limitations, such as the impossibility of using them in a real-time context, the complexity of their algorithms, or even their precision compared to subjective tests. These limitations have prompted us to introduce new techniques to estimate the quality of experience. In this paper, we proposed several contributions in the field of quality evaluation of video streaming. We used the Degradation Category Rating (DCR) methodology to better study the variation in the quality of the multimedia streams transmitted on an SDN and to estimate in real time this quality. We test in this paper three famous machine learning types: decision tree, meta learning and functions learning with different *k*-fold cross validation then we calculate RMSE, *r*, MAE, RRSE to measure the performance of each algorithm. After the analysis we find that M5P has the best performances, also the lowest complexity time. So our model will be based on M5P algorithm.

References

Abar, T., Letaifa, A.B. and El Asmi, S. (2017) 'Objective and subjective measurement QOE in SDN networks', *Proceedings of the 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, pp.1401–1406.

Boutaba, R. and Da Fonseca, N.L.S. (2015) 'Cloud architectures, networks, services, and management', *Cloud Services, Networking, and Management*, pp.1–22.

Cai, X., Wang, H. and Cui, Z. et al. (2018) 'Bat algorithm with triangle-flipping strategy for numerical optimization', *International Journal of Machine Learning and Cybernetics*, Vol. 9, No. 2, pp.199–215.

Dorigo, M. and Stützle, T. (2019) 'Ant colony optimization: overview and recent advances', *Handbook of Metaheuristics*, Springer, Cham, pp.311–351.

ITU-R Recommendation BT.500-11 (2002) *Methodology for the Subjective Assessment of the Quality of Television Pictures*.

ITU-T Recommendation P.910 (2008) *Subjective Video Quality Assessment Methods for Multimedia Applications*.

ITU-T SG12 (2007) 'Definition of quality of experience', *TD 109rev2 (PLEN/12)*, Geneva, Switzerland.

Joskowicz, J. and Ardao, J.C.L. (2011) 'Combining the effects of frame rate, bit rate, display size and video content in a parametric video quality model', *Proceedings of the 6th Latin America Networking Conference*, ACM, New York, USA, pp.4–11.

Lemke, C., Budka, M. and Gabrys, B. (2015) 'Metalearning: a survey of trends and technologies', *Artificial Intelligence Review*, Vol. 44, No. 1, pp.117–30.

Machado, V.A., Silva, C.N. and Oliveira, R.S. et al. (2011) 'A new proposal to provide estimation of QoS and QoE over WiMAX networks: an approach based on computational intelligence and discrete-event simulation', *Proceedings of the IEEE 3rd Latin-American Conference on Communications*, IEEE, Brazil, pp.1–6.

Machado, V.A., Silva, C.N. and Oliveira, R.S. et al. (2011) 'A new proposal to provide estimation of QoS and QoE over WiMAX networks: An approach based on computational intelligence and discrete-event simulation', *Proceedings of the IEEE 3rd Latin-American Conference on Communications*, IEEE, pp.1–6.

Martinez-Yelmo, I.S.I. and Guerrero, C. (2010) 'Fair quality of experience (qoe) measurements related with networking technologies', *Proceedings of the International Conference on Wired/Wireless Internet Communications*, Springer, Berlin, Heidelberg, pp.228–239.

Menkovski, V., Oredope, A., Liotta, A. et al. (2009) 'Predicting quality of experience in multimedia streaming', *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia*, ACM, pp.52–59.

Mininet Simulation (2016) *An Instant Virtual Network on your Laptop (or other PC).* Available online at www.mininet.org.

MSU Codes (2016) *MSU Graphics and Media Lab*, *Video Group, MSU codecs*. Available online at: www.compression.ru/video/

MSU Documentation (2016) *CS MSU Graphics and Media Lab Video Group*.

NTIA VQM (2016) Available online at: http://www.its.bldrdoc.gov/vqm

Pande, G. (2013) 'Metrics for video quality assessment in mobile scenarios', *International Journal of Digital Multimedia Broadcasting*, pp.1–12.

Pokhrel, J. (2014) *Intelligent quality of experience (QoE) analysis of network served multimedia and web contents*, Thèse de Doctorate, Institut National des Télécommunications.

Rodriguez, D.Z., Rosa, R.L. and Bressan, G. (2013) 'Predicting the quality level of a VoIP Communication through intelligent learning techniques', *Proceedings of the 7th International Conference on Digital Society (ICDS)*, pp.42–47.

SDN (2012) *Software-Defined Networking: The New Norm for Networks*, Open Networking Foundation, Menlo Park, CA, USA.

Sheikh, H.R., Sabir, M.F. and Bovik, A.C. (2006) 'A statistical evaluation of recent full reference image quality assessment algorithms', *IEEE Transactions on Image Processing*, Vol. 15, No. 11, pp.3440–3451.

Tan, X., Gustafsson, J. and Heikkilä, G. (2006) 'Perceived video streaming quality under initial buffering and rebuffering degradations', *MESAQIN Conference*, Vol. 90.

Vilalta, R. and Drissi, Y. (2002) 'A perspective view and survey of meta-learning', *Artificial Intelligence Review*, Vol. 18, No. 2, pp.77–95.

Vranješ, M., Rimac-Drlje, S. and Žagar, D. (2007) 'Objective video quality metrics', *ELMAR*, IEEE, Zadar, pp.45-49.

WEKA Tool (2017) *Machine Learning Software in Java*. Available online at: http://www.cs.waikato.ac.nz/ml/weka/

Wu, H.R. and Yuen, M. (1997) 'A generalized block-edge impairment metric for video coding', *IEEE Signal Processing Letters*, Vol. 4, No. 11, pp.317–320.