QoE Estimation in mobile networks using Machine Learning[1]

Jorge Mesquita L.[a], Guillermo Osorio M.[a], María E. García-Díaz [a], Diego P. Pinto-Roa [a2], Carlos Núñez-Castillo [a] and Miguel García-Torres.[b]

[a]Facultad Politécnica - Universidad Nacional de Asunción

[b]Division of Computer Science - Universidad Pablo de Olavide

Abstract

Quality of experience (QoE) can be defined as the overall level of acceptability of an application or service, as perceived by the end-user. The perceived QoE of mobile user plays a key role in the business of the telecom carriers. This work has focused in design a model capable of predicting the QoE of the end-user using a number of Machine Learning approaches, based on quality of service (QoS) metrics from different sources like the mobile device, the mobile network and also subjective metrics given by the user (QoE and Mood surveys) in a real life setup. An android app, a metric collection platform, a system for data processing and semi-automatic analysis of metrics has been developed as a part of this work. The experimental results show that by assembling a combined model of the algorithms with best observed individual performance, improvements in the overall performance of the prediction can be achieved.

Keywords: QoE, QoS, Machine Learning, Enssembled Algorithms, Android App.

1. Introduction

This work proposes a platform of estimation of the Quality of Experience (QoE) in a mobile network environment using agents installed in mobile devices to collect information on Quality of Service metrics (QoS) and user insights. The project is divided into two phases: in a preliminary phase the data is collected, and then submitted in the second phase, applying machine learning techniques to generate classification models [22]. In the first phase, the following inputs are taken: subjective valuations of QoE provided by the user, mobile device metrics and network parameters. This data is processed and converted to a standard format[3] for data analysis. In the second phase, the input data is subjected to the process of generating correlation models. At this point the performance statistics of the classifiers are processed and a assembly classification model proposed from the best performing classifiers.

[1]This work was partially supported by grant PINV15-257 CONACYT-Paraguay

[2]E-mail Corresponding Author: dpinto@pol.una.py

[3]*ARFF* of *WEKA* (*Attribute-Relation File Format*)

2. Preliminary Notions

We will briefly define some preliminary terms that are relevant to this study [1].

1. Quality of Service (QoS).

   The *UIT / ETSI* presents a definition of perceived service quality that reflects the customer's experience when using a particular service. In Rec. UIT-T G.1000 [10] four definitions are distinguished, reflecting different views on the subject *QoS*:

   (a) Requirements of *QoS* of the client: generally expressed in non-technical language, influenced by the customer's experience with similar telecommunications services or opinions of other customers.

   (b) *QoS* offered by the provider: expressed as the Service Level Agreement (*SLA*) from the supplier to the customer.

   (c) *QoS* achieved by the provider: It is the effect observed by the techniques and configurations applied in the network by the provider.

   (d) *QoS* perceived by the client: It is the one experienced by the client.

2. Quality of Experience (*QoE*)

   The term *QoE* has gain mayor popularity recently. The definition can be found in Rec. UIT-T P.10 [11]: "the general acceptability of an application or service, as perceived subjectively by the customer". The overall assessment of *QoE* is affected by environmental, psychological and sociological factors, including user expectations and experience with similar services, other opinions, pricing policies, characteristics of the particular location where the service is received, etc.

   The evaluation methods of *QoE* could be classified into:

   (a) Qualitative (Subjective): Qualitative Methods are constructed with the participation of people, a representative sample of the population, who used a particular service. In these methods, the service is evaluated in a controlled environment and people complete a survey with numerical value ratings.

   (b) Quantitative (Objectives): Quantitative Methods provide an assessment of *QoE* based on the measurement of various parameters related to quality of service indicators in the network signal at the output of the transmission channel.

   (c) Hybrids (*PSQA*[4]): Takes the best of subjective and objective models, the results are in terms of Mean Opinion Score (*MOS*[5]), this is a non-intrusive method and the data is obtained in real time. The parameters that cause the distortion are related to the perceived quality.

[4] Pseudo-Subjective Quality Assessment

[5] Mean Opinion Score

3. **Machine Learning**

    (a) Types of Learning: Depending on the data or feedback available for the learning system, types of learning can be classified into [16]: Supervised Learning and Unsupervised Learning. This work falls under the category of Supervised Learning, in which item labels are provided within the training dataset, so the base knowledge of the system is formed by labeled examples.

    (b) Classification, Regression and Clustering: In the Classification, a category is assigned to each item. For instance, the classification of QoE is based on the parameters of *QoS*. In Regression, a numeric real value is predefined for each item. Examples of regression include predicting variations in economic variables. The Clustering seeks to divide items into homogeneous regions, and is often used to analyze a large dataset. For instance, in the context of social network analysis, clustering algorithms attempt to identify "communities" within large groups of people.

    (c) Assembled algorithms: There are conditions under which using a system based on the assembly of classifying algorithms can be more beneficial than using the individual algorithms, this is because the strengths of individual classifiers that make up the assembly [15] are taken in advantage.

4. Related work

Currently there are several surveys that were carried out around *QoE* and *Machine Learning*.

In [2] a general description of the correlation models is presented *QoE-QoS* based on automatic learning techniques. According to the type of learning, they propose a categorization of the correlation models. For each category, they review the main existing works citing the learning methods implemented and the parameters of the model, measuring *QoE*, parameters of *QoS* and type of service.

In [3] a state of the art survey is provided on the application of a data-driven approach in the evaluation of *QoE*. First, they describe how to choose the factors that influence the *QoE*. They investigate and discuss the strengths and deficiencies of existing automatic learning algorithms for modeling and forecasting the *QoE* of the users. Finally, they describe their research work on how to evaluate the *QoE* in unbalanced data sets.

In [4] provides an exhaustive survey of the literature related to automatic learning algorithms applied to *SDN*[6], presenting an overview of the automatic learning algorithms. In addition, it reviews how the algorithms are applied in the field of *SDN*, from the perspective of traffic classification, routing optimization, *QoS* prediction / *QoE*, management of sources and security.

[6]Software Defined Network

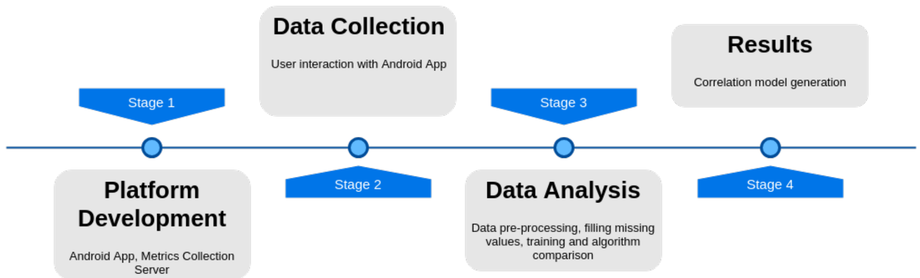

Figure 1: Overview of the phases involved in this work

In [5] the application of different *Machine Learning* techniques are jointly presented in several key network areas through different network technologies.

In [7] proposed to use a hybrid model for the generation of a model related to *QoS* and *QoE*, for the context of a specific video streaming and web browsing application. A conclusion of this work is that the network parameters that most affect the *QoE* are bandwidth and packet loss.

5. Proposed Methodology

This work can be summarized in a series of phases, from the development of the platform to the generation of the correlation model. The figure 1 summarizes the phases mentioned above.

The objectives of this work were the design, development and the implementation of:

1. A mobile application, for the Android platform, for the collection of network parameters and user device, and the performance of subjective tests for the evaluation of the *QoE*.

2. A platform for collecting, analyzing and classifying data traffic on a mobile network.

3. On the basis of the above, and using automatic learning techniques, estimate the *QoE* by correlation models of the various parameters influencing the network and device level.

Design of the proposed model

Figure 2 exposes the design of the proposed model. In this model, a mobile agent (Android Application) is installed on the user's mobile device. This agent is responsible for obtaining the metrics corresponding to the user (*MOS*), the network and the mobile device. This data is stored in a local repository on the mobile device, and then, through a scheduled task, synchronize this data with a central metric repository. After the data is synchronized to the central repository, it is pre-processed before being subjected to training with *Machine*

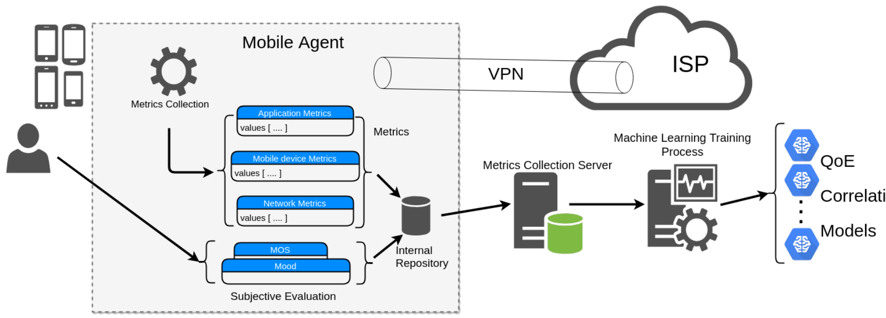

Figure 2: Proposed architecture for data collection and analysis

Learning. Once generated the correlation models of *QoE* using different algorithms, performance metrics are evaluated to generate an assembled model from them.

Some elements related to implementation are described below.

1. VPN Server: The analysis process of this work required access to information from the IP packet headers of the underlying traffic. A solution based on a VPN connection does not require administrator privileges to access information from the IP traffic headers. An OpenVPN server application was used [12] to collect network traffic information in a local database of the mobile device. The server is installed within the operator's network and the client is deployed from the mobile application.

2. Metrics Collection Server: The metrics collection server is a Django [13] based implementation which is in charge of making available an *API* to standardize communication between and the Mobile Application and the server, which has the task of collecting the data sent from the different mobile terminals that participate in the data collection program for this work.

3. Mobile Application *QoE Analyzer*: It is the implementation of the mobile application that incorporates the client of the *VPN* server described above, in addition of controlling the logic of data collection, the periodicity in which the data is requested, the *MOS* of the user and the synchronization of the local repository with the remote Metric Collection Server.

4. Statistical Processing: The performance statistics of the models generated from the analysis process are averaged, obtaining an average deviation as well. For each metric taken, the mean and mean deviation values represent the final performance of the algorithm. As the *dataset* analyzed had missing values, these were filled in using missing data filling strategies [14].

Experiment configuration

In this work, the Hybrid model was used as an evaluation method, which allows the quantification of the *QoE* close to the assessment that human observers could make. This methodology, therefore, will allow us to implement some tests to generate the correlation model between the user experience and some service parameters at the network, device and application levels. For the training and generation of Machine Learning correlation models, the *WEKA* software was used [22]. The list of algorithms submitted in the training process were based on those used in [7]: `A2DE`, `A1DE`, `Random Forest`, `Simple Logistic`, `Bayes Net`, `DTNB`, `Naive Bayes`, `RBF Network`, `J48`, `REPTree`, `Multi Layer Perceptron`, `SMO`, `Decision Table`, `Random Tree`, `RBF Classifier`, `IB1`, `IN5` and `IB10`. For details about these algorithms please see in [15, 23, 24]. For all algorithms, *10-fold cross validation* were used. The approach of this work is based on Supervised Learning, Classification and Ensemble algorithm with combination of experts in a weighted majority votes [15]. To perform data collection, the user must install the Mobile Application on their Android device. Once the application is installed, the user must start a test that will ask for up to 10 ratings of quality of experience (*MOS*) in the use of the network, with a periodicity of 1 minute for each rating. One *test* is composed of: An interval of one minute of use of some mobile application of user preference (*YouTube*, *Facebook*, etc.), and an express rating of the user (in number of stars) about their experience, where one, two, three, four, and five indicate poor, uncertain, fair, good and excellent, respectively.

5. Results

In the data-set evaluated, the most used applications, in decreasing order, were: YouTube, WhatsApp, Chrome and Facebook. Taking into account the statistical fashion of the data-set of the 930 instances, 419 correspond to the *MOS* 5. This represents a percentage of 45.0538%, and therefore we could consider this value as the minimum acceptable hit percentage for the classifying algorithms that will generate correlation models on the data-set.

Statistics of Collected Data. The class distribution can be seen in the Table 1. Moods were collected in a total of 449 instances (48% of the samples). The distribution can be seen in the following table 2. In the table 3 the attribute dictionary is presented.

Compiled Model Statistics. The assembled model was built from the best performing algorithms in each metric. These algorithms were `RandomForest`, `IBk5` e `IBk10`. The performance of assembled model can be seen in the Table 4.

Table 1: Class distribution of the original dataset

MOS	Sample Quantity
1	93
2	84
3	135
4	199
5	419

Table 2: Distribution of moods from the original dataset

Moods	Sample Quantity
1	4
2	0
3	104
4	179
5	162

Table 3: Features used

Abbreviation	Description
dev	ID of the device participating in the data collection tests
app	Foreground application
bl	Battery Level
bt	Battery Temperature
ii	Absolute MOS index of all tests performed by the user
ip	Test index in the current test
tn	Current test index
d	Distance of current MOS from the beginning of a set of tests
bw	Bandwidth
df	Delay to the first IP hop that responds to PING from the Telephone to the Cellular Network
sf	Number of Hops to the first IP Hop that responds to PING
dv	Delay to VPN server from Smart Phone
sv	Number of Hops to VPN Server
ds	Device Qualification (values 1 to 10)
mood	Mood of the User in the current instance (values 1 to 5)
ns	Signal Type Qualification (values between 2 and 4 based on the technology used)
st	Signal strength (values 1 to 4)
ram	Free RAM
sdk	Android Platform Indicator (Values 21 to 28)
cpu	CPU Temperature
mos	User Experience

Table 4: Performance of Assembled Model

	Average	Deviation
Hit rate %	59.3655913978495	0.655913978494623
Kappa	0.413326644173476	0.00972773255237
F Measure	0.580813229978095	0.006963645888596
PRC	0.64976287043897	0.002309985869439
ROC	0.834654664271629	0.001336444984176

The statistical metrics used are: Percentage of hits [17], Kappa [18], F Measure [19], ROC Curve [20] and PRC Curve [21]. Taking into account the assembled model, it presents a proportional improvement of 1% in relation to the average performance of the selected algorithms.

5. Conclusions and Future Work

From the exposed results it can be observed that the bandwidth is the least related to the user's valuations in the interval of one minute corresponding to the *test*. This could be due to the fact that the applications most used during the tests of this work have mechanisms of self-regulating video quality, use *buffers* or non-interactive content, to counteract an unfavourable environment of *QoS* during the last minute evaluated. The relatively low bandwidth required for instant messaging could also favor the low correlation of this metric with the *QoE*, whereas the multimedia content associated with it is not linked to the requirements of typical applications *streaming* or in real time. On the other hand, taking into account the experiment carried out in this work, most users rated the quality of experience as Excellent and Very Good (57.84%). This could give rise to the idea that the network infrastructure used by the mobile operator satisfies, to some extent, the requirements of *QoS* that allow the user to have a good quality of experience in the use of the mobile network. While improving the performance of the model assembled for the prediction of the *QoE* is barely significant in relation to the average of the selected individual algorithms, these values could be increased by introducing more metrics of *QoS* in the training of the predictive model that the mobile operator normally relies on, such as, for example, the loss of packets within the network. Finally, an assembled model presents a more general classifying model independent of the training data-set.

Acknowledgments

This work is an extension of the international project MOSAIC granted by Stic-AmSud 2013 (`http://www-public.it-sudparis.eu/~maag/MOSAIC/index.html`)

References

[1] R. Stankiewicz, P. Cholda and A. Jajszczyk,"QoX: What is it really?," in IEEE Communications Magazine, vol. 49, no. 4, pp. 148-158, April 2011. doi: 10.1109/MCOM.2011.5741159

[2] S. Aroussi and A. Mellouk, "Survey on machine learning-based QoE-QoS correlation models," 2014 International Conference on Computing, Management and Telecommunications (ComManTel), Da Nang, 2014. DOI: 10.1109/ComManTel.2014.6825604.

[3] H. Ruochen, W. Xin, Z. Liang, L. Chaoping, M. Hao, Jin. Jiefeng, "A survey of data-driven approach on multimedia QoE evaluation". Frontiers of Computer Science,2018. DOI: 10.1007/s11704-018-6342-7.

[4] J. Xie, F. R. Yu, T. Huang, R. Xie, J. Liu and Y. Liu, "A Survey of Machine Learning Techniques Applied to Software Defined Networking (SDN): Research Issues and Challenges" in IEEE Communications Surveys & Tutorials. DOI: 10.1109/COMST.2018.2866942.

[5] R.Boutaba, M. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano and O. Caicedo. "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities". Journal of Internet Services and Applications, 2018. DOI="10.1186/s13174-018-0087-2".

[6] M. Li, H. Wei and H. Liao, "Mobile terminal quality of experience analysis based on big data," 2016 16th International Symposium on Communications and Information Technologies (ISCIT), Qingdao, 2016, pp. 241-245. doi: 10.1109/ISCIT.2016.7751629

[7] B. Quiñónez, L.F. Fernández, D.P. Pinto-Roa, C. Nuñez and M.E. García, "Evaluation of the quality of experience for mobile data services". XXI Argentine Congress of Computer Science 2015 (CACIC'2015), October 2015, Junin, Argentina.

[8] S. Khirman and P. Henriksen, "Relationship between Quality-of-Service and Quality-of-Experience for Public Internet Service". 2002. Passive and Active Measurement Conference. Fort Collins, CO, USA.

[9] "ITU-T Recommendation E.800. Quality of Telecommunication services: Concepts, Models, Objectives and Dependability Planning. Terms and definitions related to the quality of telecommunication services" Tech. Rep. [Online]. Available: http://www.itu.int/rec/T-REC-E.800-200809-I

[10] "ITU-T Recommendation G.1000. Communications quality of service: A framework and definitions" Tech. Rep. [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=s&id=T-REC-G.1000-200111-I!!PDF-E&type=items

[11] "ITU-T Recommendation P.10/G.100 (11/17). Vocabulary for performance and quality of service". Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.10-201711-I!!PDF-E&type=items

[12] "OpenVPN for Android". Available:https://github.com/schwabe/ics-openvpn

[13] "High-level Python Web framework". Available: https://www.djangoproject.com/

[14] "Multivariate Imputation by Chained Equations". Available: https://cran.r-project.org/web/packages/mice/mice.pdf

[15] R. Polikar, "Ensemble based systems in decision making". IEEE Circuits and Systems Magazine (Volume: 6, Issue: 3, Third Quarter 2006). DOI: 10.1109/MCAS.2006.1688199. pp. 33. Available: https://ieeexplore.ieee.org/document/1688199

[16] Richard O Duda, Peter E Hart, and David G Stork. "Pattern classification". John Wiley & Sons. 1999.

[17] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy". Remote sensing of Environment, vol. 62, no. 1, pp. 77–89, 1997.

[18] J. L. Fleiss, "Measuring nominal scale agreement among many raters". Psychological bulletin, vol. 76, no. 5, p. 378, 1971.

[19] N. Japkowicz and M. Shah, "Evaluating learning algorithms: a classification perspective'. Cambridge University Press, 2011.

[20] "Using ROC curves in clinical investigation. Theoretical and practical issues". Available: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0716-10182012000200003

[21] "Example of Precision-Recall metric to evaluate classifier output quality". Available: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

[22] Machine Learning at Waikato University. Weka 3: Data Mining Software in Java. URL: https://www.cs.waikato.ac.nz/ml/weka/

[23] L. Breiman, "Random Forests". Machine Learning Journal, vol. 45, no. 1, p. 5-32, 2001.

[24] D. Aha, and D. Kibler, "Instance-based learning algorithms". Machine Learning Journal, vol. 6, p.37-66, 1991.