

Crowdsourcing or Network KPIs? A Twofold Perspective for QoE Prediction in Cellular Networks

Andrea Pimpinella, Andrea Marabita, Alessandro E. C. Redondi
Dip. Elettronica, Informazione e Bioingegneria, Politecnico di Milano
Email: name.surname@polimi.it

Abstract—Monitoring the Quality of Experience (QoE) of the customer base is a key task for Mobile Network Operators (MNOs), and it is generally performed by collecting users feedbacks through directed surveys. When such feedbacks are few in number, a MNO may predict the users QoE starting from objective network measurements, gathered directly from the users equipments through *crowdsourcing*. In this work, we compare such a traditional approach with a different one, where the data used for predicting the users QoE is gathered directly at the network access, using Key Performance Indicators (KPI) available on each base station. Although such KPIs are aggregated by design (i.e., they refer to the distribution of a population of users rather than to a single individual), we show through experiments with a country-wide dataset that their predictive power is comparable and in some cases superior than the one of crowdsourcing. Such a result is particularly attractive for MNOs, since network KPIs are generally much easily obtainable than crowdsourcing data.

Index Terms—QoE prediction, Mobile Cellular Network, Network Intelligence

I. INTRODUCTION

Mobile network operators (MNOs) constantly monitor the level of satisfaction of their customers in order to minimize the so called churn rate, i.e., the percentage of customers that stop their subscriptions and move to a different operator due to an unsatisfactory service. Indeed, knowing which users are not satisfied with the network service helps the operator in discovering potential issues in the Radio Access Network (RAN), driving investments to boost up specific parts of the network infrastructure or developing attractive marketing strategies.

Operatively, MNOs rely on the administration of surveys and questionnaires to monitor the level of satisfaction of their customers. Standard tools such as the Net Promoter Score (NPS) survey are generally used, asking users to indicate the likelihood of recommending the network operator to a friend or colleague on a scale from 0 to 10. In addition to such a generic survey, operators often ask customers to reply on very specific questions related to the Quality-of-Experience (QoE) of certain mobile network services (network coverage, voice and video quality, etc.), which can better highlight possible problems in the network. Unfortunately, performing such directed surveys is costly and cumbersome for operators, mainly due to the generic poor cooperative attitude of customers: most of the time, the surveys proposed to customers are in fact simply ignored.

A possible way to cope with this issue is to rely on artificial intelligence and predict the customers QoE rather than measuring it directly, through *crowdsourcing* [1], [2], [3]. With this approach, the operator gathers network measurements taken directly at the users terminals, such as signal quality and throughput, and correlates them with the QoE expressed in the surveys. This allows to train Machine Learning (ML) models which are then able to predict the QoE of a user given her network measurement. However, crowdsourcing has its own limitations. Indeed, to be able to gather measurements from users terminals, specific applications running in background and installed under the user consent are generally used. It follows that the fraction of users that participate to crowdsourcing campaigns is generally small. Moreover, several studies have confirmed that it is not a trivial task to gather reliable responses from crowds, especially when no reward systems are conceived [4], [5], [6].

At the same time operators have access to massive amount of measurements collected at the RAN and available for each eNodeB, such as KPIs and counters measuring cumulative uplink/downlink data volumes, distributions of transmission parameters, available radio resources and accessibility or handover requests/failures. Even though such measurements are aggregate by design and therefore unable to represent each user individually, it is of interest to study whether it is possible to use them as an alternative data source for performing QoE prediction: this is precisely the scope of this paper.

We leverage a country-wide dataset containing several thousands users including their QoE answers, crowdsourcing data and KPIs gathered from the visited eNodeBs to assess the performance of the two approaches for QoE prediction, i.e. crowdsourcing and network KPIs based. We use such data to train ML models and then predict users QoE relative to two specific network services, namely network coverage and video streaming. We comment on each approach design choices and we study the performance improvement obtained when fusing together the two available data sources.

The rest of this paper is structured as it follows: Section II summarises related works in the area of QoE prediction in mobile cellular networks, while Section III describes the datasets available for this work. Section IV comments the results obtained from predicting users QoE about network coverage and video streaming for both the crowdsourcing and the network KPIs based approaches. Finally Section V provides a discussion on the obtained results and on future

work directions.

II. RELATED WORK

Many authors during the last decade investigated and evaluated the feasibility of predicting both short-term [7], [3] and long-term [8], [2], [9] customers QoE relative to different network services. Short-term QoE concerns individual and time-limited sessions in which users are instructed to use a service (e.g., watching a video content on YouTube) under controlled network environments and are later asked about the quality of their experience. Instead, long-term QoE refers to the experience of users in the network during periods of several weeks or months. Regardless of the type of the prediction problem, common approaches for assessing the impact of network and services performance on customers experience is conducting either controlled laboratory experiments (which ensure full control of the evaluation process) or field trial experiments through crowd-sourced data (which capture end users experience in real network conditions). In [7] authors combine subjective controlled lab tests and passive end-device measurements collected through a field trial with QoE users feedbacks on five different applications. They show that down-link bandwidth fluctuations are crucial in determining the QoE of a service, especially for high-interactive ones. Similarly, Finley, et al. [8] combine network data, demographic data and QoE satisfaction surveys collected through the Netradar platform to study the significance of different predictors of user satisfaction. Leveraging a custom pop-up surveys and a five-point Likert scale for respectively users QoE data collection and quantification, the authors recognise the minimum download throughput as a performing QoE predictor. Also, in [3] authors use a rich QoE dataset taken from field trials in operational cellular networks to develop a decision-tree based model capable of predicting the quality of per-user overall experience and service acceptability with a success rate of 91% and 98% respectively. With similar aims, authors in [2] leveraged crowdsourcing-based LTE and WiFi network monitoring data to predict the quality of users' experience with respect to different applications, showing a 20% G-Mean improvement over baseline classifiers.

Finally, it is worth mentioning that QoE prediction is an efficient driver for optimising resources allocation and enforcing maintenance strategies to meet customer demands and expectations [10], [11]. With this view, in [10] the authors design the concept of a generic framework for ML-based QoE/KPI monitoring of HTTP Adaptive Streaming (HAS) services, showing its applicability in a concrete test case of YouTube usage via customers smart-phones. Also, in [11] the authors investigate the impact of QoE prediction errors in a crowdsourcing-based network monitoring system, obtaining insights which are generalizable and that provide interesting guidelines for network operators.

III. DATASETS

The datasets used in this work have been collected by one of the major European mobile operators over a period of five

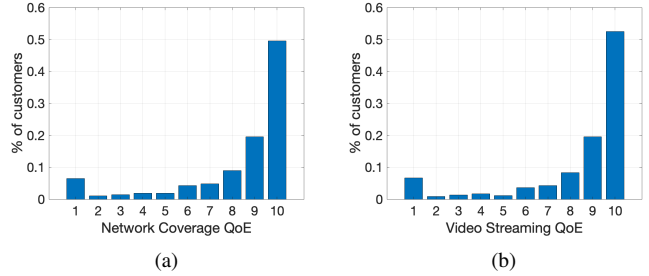


Fig. 1. Distribution of QoE feedbacks for the two scenarios.

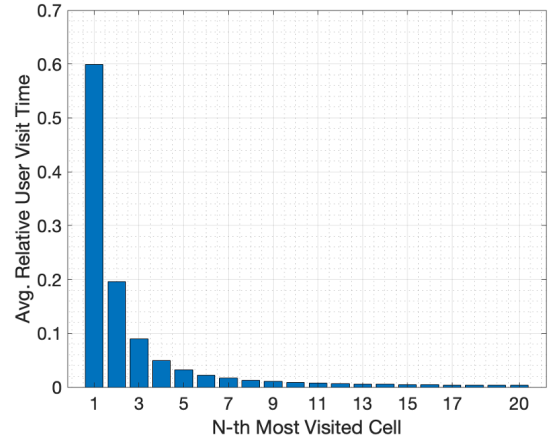


Fig. 2. Average distribution of users visit time versus the site's rank of importance. The first bar refers to the average proportion of time spent by users in the most visited site, i.e., in the home cell.

months between January and May 2020, and refer to the entire territory of Italy. Three different datasets have been used, which are detailed in the following sections. Note that, in order to protect users privacy, all identifiers have been anonymized through hashing.

A. QoE dataset

This dataset contains timestamped ground truth QoE feedbacks from users, expressed as grades from 0 to 10, obtained with directed surveys that the operator administers periodically to its customers. In particular, we restrict our attention to two scenarios, where QoE feedbacks are relative to the quality of network coverage (Q_c) or video streaming (Q_v). Over the five months period, Q_c contains about 6.2k feedbacks while Q_v contains about 5.6k feedbacks. As one can observe in Figure 1, which shows the distribution of satisfaction grades for the two considered services, both distributions are highly skewed, with the majority of users reporting positive feedbacks. It is possible to discretise the grades into two classes, with respect to a predefined satisfaction threshold \mathcal{T} : users whose vote is less or equal than \mathcal{T} are grouped together as *Unsatisfied*, while the opposite happens for *Satisfied* users. In this work we set $\mathcal{T} = 6$, to which corresponds a percentage of users unsatisfied with network coverage and video streaming of roughly 17% and 15%, respectively.

TABLE I
SUMMARY OF THE FEATURES USED FOR CROWDSOURCING-BASED QoE PREDICTION ACCORDING TO THE WORK IN [9]

Measurement	Description	Scenario
3G and 4G Full Service Time Ratio	Ratio between the time spent by a user under full network service and the total 3G/4G user time in the last d days	Coverage
3G and 4G Limited Service Time Ratio	Ratio between the time spent by a user under limited network service and the total 3G/4G user time in the last d days	Coverage
Min, Max and Average SNR	The minimum, maximum and average SNR observed by the user terminal in the last d days	Coverage
Min, Max and Average RSRQ	The minimum, maximum and average RSRQ observed by the user terminal in the last d days	Coverage
3G and 4G Video Download Time and Volume	The cumulative time spent and traffic volume downloaded by the user in the last d days under 3G or 4G connection, restricted to video applications (YouTube, Netflix, etc.)	Video
3G and 4G Video Throughput	The ratio between 3G/4G video download volume and time	Video
Overall Average Throughput	The ratio between the total video download volume and time, regardless of the radio technology used	Video
3G and 4G Throughput and Volume peaks	The maximum observed video throughput and Downloaded Volume	Video
4G video Download Time Ratio	The ratio between the video download time in 4G and the total video download time (3G and 4G)	Video

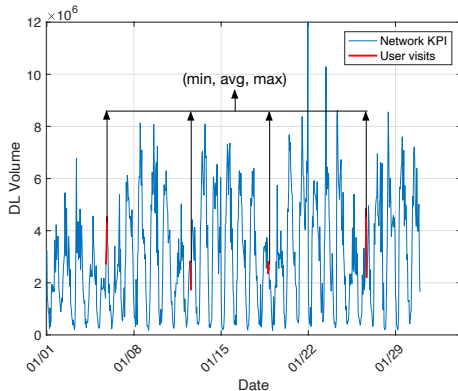


Fig. 3. Network KPIs features engineering process from a sample user's home cell in a period of $d = 30$ days before user's survey response date (here issued on 01/31). Blue line refers to the raw KPI data of the considered period while red lines refer to values of the KPI in the user's days of visit.

B. Crowdsourcing dataset

The second dataset contains network measurements gathered directly from users terminals through an ad-hoc mobile application, running in background under the users consent. The application periodically logs several active and passive network measurements relative to low-level network indicators (e.g., average cell signal strength and channel quality indicators, daily time spent by the user in full or limited service conditions, etc.) as well as application level indicators (e.g., session downlink/uplink data volume, duration and throughput) of different applications run in foreground by the user. Each measurement is timestamped and marked with the user anonymized identifier. We pre-process the dataset by considering only the users who replied to QoE feedbacks (i.e., whose identifiers are present in the Q_c or Q_v datasets), resulting in about 2k users for each scenario. According to our previous study [9], we restrict our attention to a subset of features (10 for network coverage and 14 for video streaming), summarized in Table I. Regardless of the scenario, features are

computed independently for each user, considering different periods of d days before the date of the survey. The variable d relates to the so-called *user memory*, i.e., for how long a disservice is able to impact the response to a survey. Similarly to what done in [9], we do not use a single value of d , but we set its value to 30, 60 and 90 days before the survey response, each time computing the corresponding subset of features.¹ Finally, we concatenate the three subsets, obtaining a total of 30 and 42 features to be used for each user in network coverage and video streaming scenarios, respectively.

C. Network KPIs dataset

The third dataset is composed of network KPIs gathered directly from the eNodeBs visited by each user who replied to a QoE survey (about 3.5k users for the coverage scenario and 3.2k for the video one). To populate this dataset we leverage the knowledge of each user's network visit times, i.e., a list containing which eNodeBs were visited by each user in the d days before the survey response and for how long. Figure 2 shows the distribution of such visit times, considering the whole five months period: as one can see, users spend on average almost 60% of their time in just one eNodeB, which we refer to as *home cell*. We therefore restrict our analysis to each user's home cell, i.e., to roughly 3k different network cells, extracting for each of them a subset of the available KPIs (summarized in Table II). In particular, we considered three specific KPIs categories, which are the ones more likely impacting on users QoE [12], [13]:

- **Accessibility KPIs:** provide information about whether network services can be successfully accessed by users or not. These include success rates for both Radio Resource Control (RRC) connection request/re-establishment and Radio Access Bearer (E-RAB) assignment;
- **Mobility KPIs:** describe the capability of the network to guarantee continuous and ubiquitous service to users. This

¹In [9], the maximum user memory was set to 30 days due to data limitation reasons. Following experiments showed a small performance improvement when considering even larger periods such as 60 and 90 days.

TABLE II
SUMMARY OF THE FEATURES USED FOR NETWORK KPI-BASED QoE PREDICTION

Measurement	Description	Type
E-RAB, Call and RRC Setup SR	Success rate for Radio Access Bearer, Call and RRC connection setup.	Accessibility
RRC CR SR	RRC Connection Re-establishment Success Rate	Accessibility
RAN Unavailability Rate	Radio Access Network unavailability rate (4G only)	Accessibility
Intra/Inter-frequency HO-out SR	Success Rate for both intra- and inter-frequency outgoing handover	Mobility
Intra/Inter-frequency HO-in SR	Success Rate for both intra- and inter-frequency ingoing handover	Mobility
Inter-RAT handovers SR	Success Rate for LTE to WCDMA and LTE to GSM handover	Mobility
Downlink/Uplink Volume	Cumulative traffic volume in downlink/uplink	Usage
Downlink/Uplink Throughput	Average throughput observed in downlink/uplink	Usage
Avg/Max Connected Users	Average and peak number of RRC-connected users	Usage

category includes success rates of inter/intra frequency handovers for both outgoing and ingoing users mobility;

- Usage KPIs: detail the activity of the users in the network and the congestion level at the cell site. In this work, the former aspect is included by considering cumulative traffic volumes and average throughput observed in downlink and uplink whereas the latter is represented by the average and maximum number of RRC-connected users.

Each KPI is available in form of a hourly sampled time series, which we pre-process in order to be fed as input for prediction. To do so, we leverage the knowledge of users visit times and retain from each time series only those samples which correspond to each user’s visit. Since users generally visit their home cell different times during a given d -day period, we aggregate all the gathered samples by computing the minimum, maximum and average value, for each KPI. Figure 3 illustrates such a process, considering as example the data volume downloaded by a user within a 30-days period before her survey response date. At the end of this process, we obtain 47 features to be used for both network coverage and video streaming QoE prediction. Similarly to what done in [9], we apply to the computed features a log-like transformation, to make their statistical distributions more similar to Gaussian. Finally, features are standardized to mean and variance, a pre-process which benefits ML methods working with normally distributed inputs.

IV. EXPERIMENTS

A. Training the classifiers

We aim to compare the performance of the two different approaches, crowdsourcing or network KPIs, in predicting the users QoE. The problem we consider has the form of a binary classification problem, such that each user can be classified as either satisfied (class 0, negative case) or unsatisfied (class 1, positive case or alarm). We take therefore a supervised learning approach: among the several available ML classifiers to be trained we select the Random Forest algorithm, which is widely known to perform well in general and has been successfully applied in the past for similar problems [3], [9]. To compare the two approaches to QoE prediction we proceed as it follows:

- 1) First, we select the subset of users who appear in both datasets, i.e. for whom both crowdsourcing and network

- KPIs data are available. This results in about 1.5k users for the coverage scenario and 1.3k users for the video one;
- 2) Then, for each scenario (coverage or video) and each approach (crowdsourcing or network KPIs) independently, we divide the set of common users in train, test and validation sets according to a 80-10-10 splitting ratios and 10-fold cross validation. Similarly to what done in [9], we first use the train and test set to select the best set of hyper-parameters (number of decision trees, number of features to be considered in each tree, etc.) of the random forest. Secondly, we train the classifier with the best hyper-parameters found and finally we test its performance on the validation set. Note that for each observation of the validation set, the classifiers output the probability that it belongs to the class of unsatisfied users: by thresholding such a probability, the classifier effectively takes a decision and outputs the prediction outcome;
- 3) Since the two classes (satisfied and unsatisfied users) are highly unbalanced (as shown in Figure 1, we show the performance in terms of Receiver Operating Characteristic (ROC) curves, obtained observing the True Positive Rate (TPR) and False Positive Rate (FPR) of the classifier at different prediction thresholds. We recall that the TPR, or *recall* r , equals the percentage of unsatisfied users correctly detected by the classifier in the validation set, while the FPR is the rate of false alarms. Also, to summarize the ROC curve in a single value, we compute the Area Under the Curve (AUC) value. For each working point (FPR,TPR) of the classifier, we also compute the *precision* p , i.e., the ratio between the number of correctly predicted unsatisfied users and the number of actually unsatisfied users. Finally, we select on the ROC curve a specific working point, the one maximizing the F1-score of the classifier, defined as:

$$\text{F1-score} = 2 \frac{p \cdot r}{p + r} \quad (1)$$

- 4) As a last step, we also compute the performance of a classifier working with both the sets of features (crowdsourcing and network KPIs) for each user, and we refer to this last approach as *joint* approach.

B. Crowdsourcing VS Network KPI

Figure 4 and 5 show the results obtained for coverage and video scenarios, respectively. Also, we summarise in Table III

TABLE III
QOE PREDICTION PERFORMANCE OBTAINED WHEN USING CROWDSOURCING, NETPERFORM KPIS AND JOINT APPROACHES.

Scenario	Video				Coverage			
	AUC	FPR _{F1}	TPR _{F1}	F1-Score	AUC	FPR _{F1}	TPR _{F1}	F1-Score
Crowdsourcing	0.59	0.45	0.57	0.36	0.58	0.37	0.52	0.36
Net. KPIS	0.60	0.34	0.51	0.37	0.63	0.36	0.55	0.39
Joint	0.62	0.41	0.62	0.4	0.63	0.45	0.63	0.39

the working points (FPR_{F1}, TPR_{F1}) maximizing the classifier’s F1-score and the corresponding F1-score value. Two main observations can be made from the inspection of the results:

- For what concerns the coverage case, we observe that using network KPIs as features for predicting users QoE works better than using crowdsourcing data, with an AUC of 0.63 against 0.58 and a F1-score about 8% better. The same holds also for the video case, where AUC values equal 0.60 and 0.59 when network KPIs and crowdsourcing data are used respectively, although with a smaller F1-score gap (about 3%). This can be explained considering that users QoE on network coverage is by nature more "objective" than their QoE on video streaming, where the impact of each user experience is greater in determining the satisfaction level;
- The *joint* approach (which uses both sets of features) outperforms the other methods in both scenarios. This is somehow expected, as the classifiers can learn from the point of views of both end users and network access. Considering the F1-score, in video streaming scenario the joint model outperforms the crowdsourcing approach by more than 11% and the network KPIs approach by about 8%. Instead, for the case of network coverage, the joint model improves the F1-score by 8% compared to crowdsourcing-based model while the improvement is negligible compared to using only network KPIs.

We also train the classifiers considering all possible users available for each of the two data sources, that is, regardless of the scenario, about 2k users for crowdsourcing and 3.5k for network KPIs one. Results are shown in Table IV: although the comparison is not exactly fair in this case, as the classifiers are trained with a different number of training samples, we highlight the remarkable improvement in performance of the network KPIs approach. Summarizing, our results show that the predictive power of network KPIs data is comparable to and in some cases even superior than using crowdsourcing data. This could be of great interest for mobile network operators, which can generally obtain such KPIs much more easily compared to crowdsourcing measurements. To conclude, we recall that one of the primary use of QoE prediction is to identify network sites with possible issues. Since each site is visited by many users, it can be shown that individual misclassification errors are somehow alleviated when grouped on a single network element/area [11]. This means that, even with modest prediction performance such as the one obtained in this work, it is possible to exploit users QoE prediction as a way to reinforce network monitoring systems.

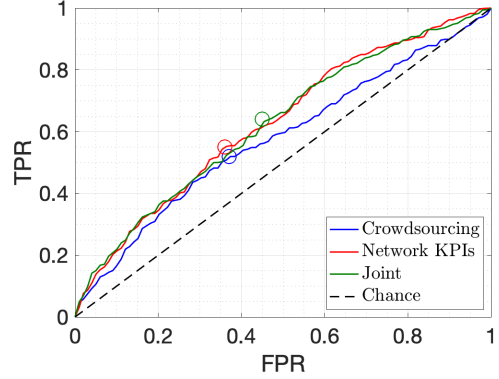


Fig. 4. ROC curve for network coverage QoE prediction. Blue, red and green circles refer to the working points (FPR_{F1}, TPR_{F1}) that maximize the F1-score of the classifier when crowdsourcing, network KPIs and joint approach is used, respectively.

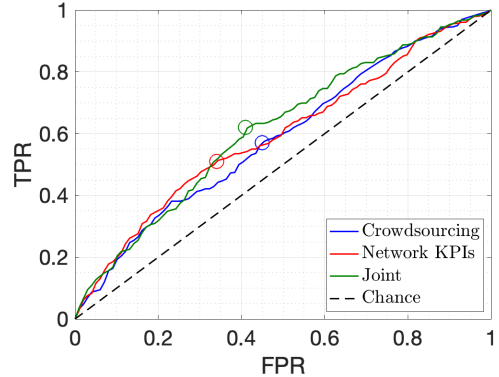


Fig. 5. ROC curve for video streaming QoE prediction. Blue, red and green circles refer to the working points (FPR_{F1}, TPR_{F1}) that maximize the F1-score of the classifier when crowdsourcing, network KPIs and joint approach is used, respectively.

C. Impact of user memory

As a last experiment, we evaluate what is the impact on prediction performance of the parameter d , which controls the time period over which features are computed and is therefore related to the user memory. On the one hand, crowdsourcing based input features are already computed for different user memories d , such that the task of selecting the most appropriate period is left to the classifier training step. On the other hand, the same approach would not be applicable with the network KPIs approach, due to the excessively high number of features to be computed (47 features are computed

TABLE IV
QoE PREDICTION PERFORMANCE WHEN 2K (CROWDSOURCING) AND 3.5K (NETWORK KPIS) USERS ARE USED FOR TRAINING.

Metric	AUC		F1-Score	
	Crowdsourcing	Net. KPIS	Crowdsourcing	Net. KPIS
Video	0.56	0.59	0.34	0.41
Coverage	0.58	0.6	0.38	0.39

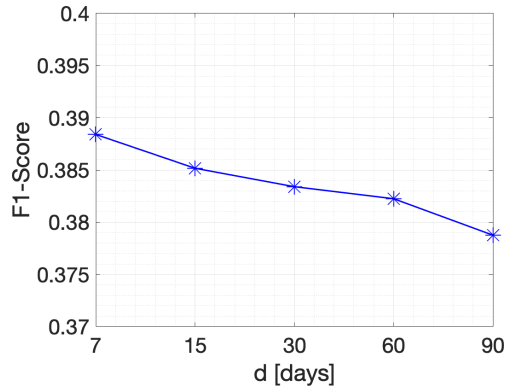


Fig. 6. F1-score of the classifier versus user memory length for network coverage QoE prediction.

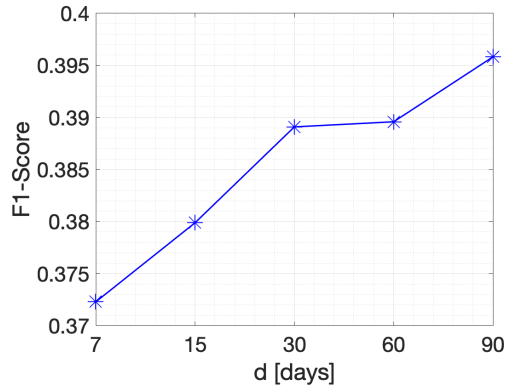


Fig. 7. F1-score of the classifier versus user memory length for video streaming QoE prediction.

for each value of d). Thus, we aggregate network KPIS for $d \in [7, 15, 30, 60, 90]$ days, training a different classifier each time and evaluating the results. Figures 6 and 7 show the best F1-score obtained for the coverage and video scenarios, respectively. For what concerns coverage, the F1-score has an increasing trend with shorter user memories (e.g., 7 days), while for the video scenario the opposite happens, with longer periods showing higher scores. The latter results is also indirectly confirmed by experiments on the crowdsourcing dataset, where enlarging the user memory d from 30 days (as in [9]) to 90 days produced a small performance improvement.

V. CONCLUDING REMARKS

QoE prediction is an important task for mobile network operators, as it helps to reduce the churn rate by tracking down possible issues in the network responsible of user dissatisfaction. In this paper, we have compared two approaches

to perform QoE prediction: a crowdsourcing based approach, which leverages measurements coming directly from users devices and a network KPIS based approach, which uses features computed from the eNodeBs visited by each user. Our experiments show that network KPIS, although aggregate in nature and representing a population of customers rather than an individual user, have a predictive power comparable to or even superior than crowdsourcing data and constitute therefore an interesting alternatives for MNOs. To conclude, we observe that the data used in this study refers to the period January-May 2020 in the country of Italy. Considering that since March the 9th Italy was in lock-down due to Covid-19 pandemic, part of the data used in this study refers to an extraordinary behaviours of the mobile network. In particular, the visit times distribution may be greatly affected due to the reduced mobility of users in such period: we plan to perform new experiments once the pandemic is over, including Network KPIS from other eNodeBs visited by each user (e.g., up to a certain cumulative percentage of visit time).

REFERENCES

- [1] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, "Quantification of youtube qoe via crowdsourcing," in *2011 IEEE Intl. Symp. on Multimedia*. IEEE, 2011, pp. 494–499.
- [2] E. Boz, B. Finley, A. Oulasvirta, K. Kilkki, and J. Manner, "Mobile qoe prediction in the field," *Pervasive and Mobile Computing*, vol. 59, p. 101039, 2019.
- [3] P. Casas, A. D'Alconzo, F. Wamser, M. Seufert, B. Gardlo, A. Schwind, P. Tran-Gia, and R. Schatz, "Predicting qoe in cellular networks using machine learning and in-smartphone measurements," in *2017 Ninth Intl. Conf. on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017, pp. 1–6.
- [4] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds," *Journal of Machine Learning Research*, vol. 11, no. Apr, pp. 1297–1322, 2010.
- [5] T. Hossfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing," *IEEE Trans. on Multimedia*, vol. 16, no. 2, pp. 541–558, 2013.
- [6] N. Shah, D. Zhou, and Y. Peres, "Approval voting and incentives in crowdsourcing," in *Intl. Conf. on machine learning*, 2015, pp. 10–19.
- [7] P. Casas, M. Seufert, F. Wamser, B. Gardlo, A. Sackl, and R. Schatz, "Next to you: Monitoring quality of experience in cellular networks from the end-devices," *IEEE Trans. on Network and Service Management*, vol. 13, no. 2, pp. 181–196, 2016.
- [8] B. Finley, E. Boz, K. Kilkki, J. Manner, A. Oulasvirta, and H. Hämmäinen, "Does network quality matter? a field study of mobile user satisfaction," *Pervasive and Mobile Computing*, vol. 39, pp. 80–99, 2017.
- [9] A. Pimpinella, A. E. Redondi, I. Galimberti, F. Foglia, and L. Venturini, "Towards long-term coverage and video users satisfaction prediction in cellular networks," in *2019 12th IFIP Wireless and Mobile Networking Conf. (WMNC)*. IEEE, 2019, pp. 146–153.
- [10] I. Orsolich and L. Skorin-Kapov, "A framework for in-network qoe monitoring of encrypted video streaming," *IEEE Access*, vol. 8, pp. 74 691–74 706, 2020.
- [11] A. Pimpinella, M. Repposi, and A. E. C. Redondi, "Unsatisfied today, satisfied tomorrow: a simulation framework for performance evaluation of crowdsourcing-based network monitoring," 2020.
- [12] M. Vaser and S. Forconi, "Qos kpi and qoe kqi relationship for lte video streaming and volte services," in *2015 9th Intl. Conf. on Next Generation Mobile Applications, Services and Technologies*, 2015, pp. 318–323.
- [13] U. S. Hashmi, A. Rudrapatna, Z. Zhao, M. Rozwadowski, J. Kang, R. Wuppapapati, and A. Imran, "Towards real-time user qoe assessment via machine learning on lte network data," in *2019 IEEE 90th Vehicular Technology Conf. (VTC2019-Fall)*, 2019, pp. 1–7.