Towards Real-time User QoE assessment via Machine Learning on LTE network data

Umair Sajid Hashmi*†, Ashok Rudrapatna*, Zhengxue Zhao*, Marek Rozwadowski*, Joseph Kang*, Raj Wuppalapati* and Ali Imran†
*Bell Labs Consulting, Murray Hill, NJ, USA
†School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK, USA

***Abstract*—It is well known that current reactive network management would be unable to support the exponential increase in complexity and rapidity of change in future cellular networks. Keeping this in perspective, the goal of this paper is to investigate applicability of machine learning and predictive models to assess cell-level user quality of experience (QoE) in real-time. For this purpose, we leverage a 5-week LTE metrics data collected at cell level granularity for a national LTE network operator. Domain knowledge is applied to assess user QoE with network key performance indicators (KPIs), namely scheduled user throughput, inter-frequency handover success rate and intra-frequency handover success rate. Results indicate that applying boosted trees model on a subset of carefully selected non-collinear features allows high accuracy threshold-based estimation of user throughput and inter-frequency handover success rate. We also exploit the periodic nature of cell data characteristics and apply a recently developed time series prediction model known as PROPHET for future QoE estimation. By employing machine learning and data analytics on network data within an end-to-end framework, network operators can proactively identify low performance cell sites along with the influential factors that impact the cell performance. Based on the root cause analysis, appropriate corrective measures may then be taken for low performance cell sites.**

***Index Terms*—Machine learning, user quality-of-experience, PROPHET model, gradient boosted trees, mobile network data**

I. Introduction

As we approach the era of 5th generation wireless networks (5G), network operators are striving ever so hard to cope up with the exponentially growing data demands and diversity of devices connected to the network. Ensuring adequate user quality of experience (QoE) in ultra-dense 5G networks becomes ever so important for multiple reasons, for instance, reducing customer churn as well as supporting advance 5G AR (Augmented reality) / VR (Virtual Reality) use cases. Existing user QoE assessment techniques are rendered impractical in these scenarios due to their passive operation. Whether the assessment is done via mobile applications, probe measurements or MDT (Minimization of Drive Test) reports, there is an inherent delay in these methodologies which causes any remedial action to be sluggish. Since the network states in dense networks changes instantly, real-time evaluation of the user QoE is essential to enable proactive network management.

In this paper, we apply well established machine learning (ML) methodologies to investigate real-time user QoE measurement via cell level surrogate key performance indicators (KPIs). We employ real-time counters and metrics measured at different network elements of a country wide LTE operator. The data was collected over a total duration of 10 weeks with a time granularity of 1 hour. From the cell level performance data, we shortlist three user QoE assessment KPIs: i) downlink user throughput, ii) handover (HO) success percentage due to mobility from one cell to another, and iii) hand over success rate from one radio frequency (RF) band to another within the same cell. While the selected KPIs do not necessarily express the QoE at user level, they are suitable indicators for the average perceived QoE for an arbitrary user within a cell's coverage area. The three-fold goal of this work is summarized as follows:

- We validate our hypothesis that user QoE can be accurately estimated by using limited set of real-time LTE network metrics. By applying off-the-shelf regression and classification models, we investigate the performance of the algorithms in terms of their ability to accurately predict from a set of LTE counters whether a cell's performance in terms of three QoE KPIs is below the stated threshold.
- After establishing that real-time QoE performance level of a cell can be accurately estimated, we apply a recently developed time series based predictive model known as PROPHET on the LTE data to predict future cell level QoE performance. The QoE performance prediction is performed for near-future and further-future scenarios, to assess how well the model can capture the hourly variation for the next 48 hours in case of former, and capture the weekly trend in user QoE KPI for the latter scenario.
- From the correlation and feature importance analysis, we identify multi-collinearity amongst selected metrics. By training our ML models on non-collinear features, we obtain the relative influence of metrics on the user QoE KPIs. Consequently, the network operator can improve the overall user QoE

978-1-7281-1220-6/19/$31.00 ©2019 IEEE

by applying remedial actions to address the highly influential network metrics.

The rest of the paper is organized as follows. In section II, we present a detailed background of the use of ML techniques on mobile network generated data. Section III introduces the operator data and some exploratory analysis. We present a user QoE assessment methodology and framework for predictive user QoE improvement in LTE and future networks in section IV. In section V, we evaluate the performance of standard ML techniques for real-time cell level QoE assessment. Section VI presents the future prediction results generated from Facebook's PROPHET predictive model. Conclusion and future work are discussed in section VII.

II. Overview of ML applications for Cellular Data

Application of machine learning techniques for improving the operational efficiency of mobile networks gained a lot of traction within this decade. Imran et al. provided a comprehensive framework to leverage the huge amount of network data available to create end-to-end visibility resulting in an improvement in the network's response time to outages and failures [1]. We will review work done within the three core ML paradigms: supervised learning, unsupervised learning and reinforcement learning. Supervised learning schemes have been applied in multiple domain within cellular networks, for instance, for mobility prediction [2], resource allocation [3], load balancing [4], fault classification [5] and cell outage management [6]. As an example, outage detection is performed using k-nearest neighbor (k-NN) algorithm that classifies a sample based on the class distribution within the k closest points measured via some distance measurement, such as the Euclidean or the Manhattan distance [7]. Neural networks is another widely used supervised scheme which has been employed in the context of cellular networks for traffic prediction and handover management in LTE networks. Using their ability to maximize the multi-dimensional separation in classification problems, support vector machines (SVM) have been utilized for automatic outage detection in cellular networks [8]. Partitioning algorithms, such as decision trees, have also been applied to cellular network data for self-organizing networks (SON) coordination [9], and cell outage detection [10]. Predictions based on models that leverage a user's behavior in the recent past paves the way for efficient network management. Also known as recommender systems, this memory-based technique has been applied to cellular networks for cooperative cell outage management [11] as well as proactive caching in information centric networks [12].

In the case of unlabeled data, there are various unsupervised learning algorithms which aim to infer the underlying relationship between the input parameters and the outputs without the presence of ground truth data. Some examples of unsupervised learning techniques are clustering algorithms, self-organizing maps (SOM), density-based clustering and outlier detection. K-means, which is by far the most commonly used distance-based clustering algorithm has been used extensively in wireless communication, for instance in proactive caching [12] and cell outage management [13]. In [14], the authors perform clustering and anomaly detection on a broadband operator customer complaints data to identify the spatiotemporal signatures of the faults and their relationship to the geographical and time-based attributes along with the fault causes. Self-organizing maps which are used to visualize similarity relationships within large dimensional data sets provided higher classification accuracy as compared to K-means clustering for a variety of synthetic and real-world datasets in [15]. Cellular applications of SOM have been analyzed in handover optimization [16] and anomaly detection using faults data [17]. Reward based learning methodologies, also known as reinforcement learning have also been useful in policy updation by network nodes to recursively improve their rewards. There is the expectation vs exploitation trade-off where each entity decides whether they would like to explore new system states in search of higher award (exploration) or maximize their reward based on the current known actions (exploitation). Reinforcement learning has been used with Fuzzy Q-Learning process for radio parameter configuration [18] and achieving SON targets of coverage and capacity optimization [19]. Other stochastic models such as semi-markov model has been used to predict spatio-temporal patterns in mobile user mobility [20]. Based on the concept of participatory sensing, such mobility prediction techniques using live network data allows the operators to perform proactive network management. Yet another ML scheme utilized in mobile networks is transfer learning, which is based on application of a training model from one spatio-temporal region to another highly similar region based on some predefined similarity index [21]. Transfer learning is applicable on different kinds of ML objectives, including regression, classification and clustering.

Now that we have presented a detailed overview of the application of ML in cellular network optimization, it is pertinent that we highlight how this work differentiates with the existing literature. The novelty lies in defining multi-dimensional user QoE measures that can be accurately estimated from live network measures. Through some intuitive feature selection and engineering, we have demonstrated that from a large data set of live network measures, we may select as few as 6-8 LTE metrics for both real-time and future prediction of user QoE. Additionally, the results identify LTE metrics that highly influence the classification process and therefore provide insights for the operator to perform remedial actions to minimize the variation in those parameters that cause the network to operate in low user QoE state.

III. Operator Data Set

The available cell level data set for this analysis comes from an LTE operator operating at 800 and 1800 MHz bands. The network performance data is collected for 7,000 cell sites over a duration of 5 weeks with a time granularity of 1 hour between subsequent measurements. This culminates to 17 million data points in total. Each data point includes more than 80 LTE performance counters and metrics, some of which are obtained in real-time at different network nodes while others are calculated by using multiple counters with a delay of a few hours.

The available data set can be divided into following sub categories:

i) *Cell throughput counters:* General cell counters include measures such as available physical resource blocks (PRBs) per hour, data volume transmitted at medium access control (MAC) and packet data convergence protocol (PDCP) layers during DL sessions, and number of scheduled user equipments (UEs) per transmission time interval (TTI) for every cell.

ii) *Cell Level KPIs:* The aforementioned real-time cell counters are used to calculate different cell level key performance indicators. Examples include downlink traffic volume, downlink PRB utilization, average downlink user throughput and downlink spectrum efficiency.

iii) *Hybrid automatic repeat request (HARQ) KPIs:* Real-time HARQ counters yield the hourly number of HARQ requests in both uplink and downlink at different LTE modulation and coding schemes (MCS), i.e. quadrature pulse shift keying (QPSK), 16-quadrature amplitude modulation (QAM) and 64-QAM. These counters are then utilized to assess the channel propagation conditions for each cell in terms of the calculated QPSK transmission ratio.

iv) *Accessibility KPIs:* These include success rate for the radio resource control (RRC) connection requests from the UE to the network. Additionally, we have the EPS radio access bearer (ERAB) success rate which reflects successful bearer assignment to the UEs.

v) *Retainability KPIs:* We analyze handover (HO) success rates for inter-frequency (inter-freq) and intra-frequency (intra-freq) HO attempts. Additionally, the data provides UE context drop rate due to radio failures and transport block failure rate values per hour for each cell within the analyzed network region.

The first challenge in this work is identification of the surrogate user QoE KPIs that are as close to an accurate representation of the user experience as possible. In the absence of user-level measurements, these KPIs should give us a fair picture of the average user QoE within a cell's coverage area. With the help of domain experts, we shortlist three user QoE KPIs for analysis: i) scheduled user throughput, ii) inter-frequency HO success rate, and iii) intra-frequency handover success rate. The user QoE KPIs are determined by taking the mean over all users in a cell for each hourly time frame. The distribution of the user QoE KPIs for the 800 MHz band is presented as Fig. 1. To analyze any possible correlation or dependence between the inter-freq HO success rate and intra-freq HO success rate KPIs, we investigate the KPIs' trends with respect to time of the day. For this purpose, we look at the total number of inter-freq and intra-freq HO failures for each hour. We observe contrasting trends for the inter and intra-frequency failures (when success rate $<$ 98%) at different hour intervals. Fig. 2 reveals that while the inter-frequency HO failures occur in high numbers during peak data load hours, the trend does not hold for intra-freq HO failures. This may be a direct effect of high average user mobility in low traffic hours. Additionally, during busy hours, network is more likely to be in high utilization, and hence more preponderance of inter-freq HO, hence more such HO failures.

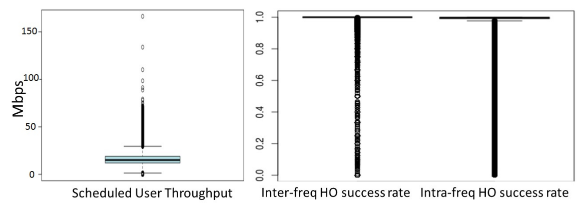

Fig. 1. Box Plot distributions for the user QoE KPIs.

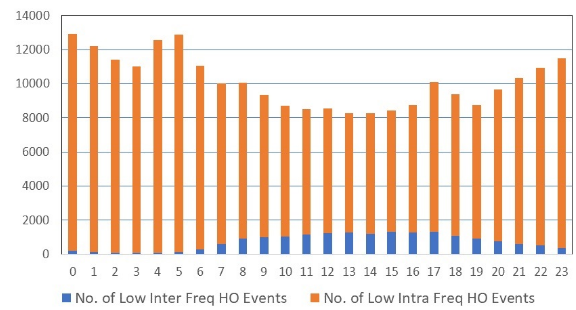

Fig. 2. Frequency Chart of the Low Inter- and Intra-Frequency events v/s Hour of the day.

The next task is to filter out features (or predictors) for the user QoE analysis that cover all the available counters and performance measures while avoiding redundancy in the selected feature set. The Spearman rank correlation between the user QoE KPIs and the selected feature set is given in Fig. 3. For the scheduled user throughput, we observe that higher percentage of hybrid ARQ transmissions with QPSK modulation in a cell has the strongest negative impact on the average throughput. Transport block failure rate, which occurs when a user is out of coverage, understandably also has a significant negative correlation with the user throughput. For the inter and intra-frequency HO success KPIs, we

Fig. 3. Spearman's rank correlation coefficient between user QoE KPIs and selected features.

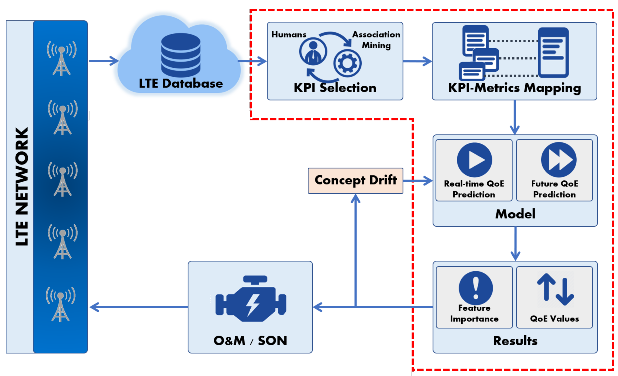

Fig. 4. UE QoE Assessment and Network Management Framework.

have their relevant load metrics, which are the number of corresponding HO requests as the major correlated metric. For the mobility-based HO success measure (i.e. the intra-frequency HO success rate), high cell load (PRB utilization) also impacts the HO success rate.

IV. QoE Assessment Framework

Fig. 4 illustrates our proposed user QoE assessment framework based on the selected user QoE KPIs and the predictor set with the dotted area depicting the scope of the analysis in this paper within the larger framework. The main idea behind the framework is to present end-to-end visibility from real-time data collection to user QoE prediction and proactive mechanisms to address high influence predictors that push cell level performance to unsatisfactory levels. As seen from the framework, live LTE metrics are fed in a cloud database. KPI selection and KPI-metric mapping process can be initially performed by a domain expert and later simplified through association mining to remove redundant KPI-metric relationships from the data being fed to the model. We employ two models: one predicts real-time cell QoE state and the other assesses time series trends to predict the cell QoE state in both near and long-term future time frames. The results that include QoE states and predictor importance for each cell are fed in the SON engine and concept drift block. The SON engine performs remedial actions to uplift a cell from Low QoE state to High QoE state. Based on the root cause, these actions may involve antenna tilt optimization, load balancing through cell individual offset (CIO) bias and offloading users to Wi-Fi or other access networks. Concept drift involves time-evolving stream classifiers that employ a drift detection mechanism using sliding window approaches and process a limited amount of incoming data to detect changes and correspondingly react in real time. Using techniques such as ensemble classifiers, this block performs drift detection and updates the classifiers in the models without having to re-train on the entire data set.

V. Real-time User QoE Prediction

In this section, we provide results and analysis to evaluate the following hypothesis: *"Can we employ machine learning algorithms to accurately predict the real-time user QoE state of a given cell from selected LTE metrics?"*. For this purpose, all the 17 million data points are distributed in a user QoE class that is dependent on a threshold for each of the user QoE KPI. Once again, we utilize the domain experts' knowledge to define binary user QoE states ("High" and "Low") with the following demarcation levels: mean user scheduled throughput considered High if greater or equal to 10 Mbps and Low otherwise; mean inter- and intra-frequency handover success rates considered High if greater or equal than 98% and Low otherwise. This corresponds to a total of $2^3 = 8$ user QoE classes. The class data distribution shown in Fig. 5 reveals that the available data is highly skewed (about 83%) towards the HHH user QoE class which corresponds to the cell sites satisfying the QoE criteria for all the three KPIs in the hourly time frame. The QoE states are represented as $\text{QoE}_1\text{QoE}_2\text{QoE}_3$, such that $\text{QoE}_i = \{\text{H,L}\}$, where "H" and "L" represent High and Low user QoE states respectively. Similarly, the ordered subscripts 1,2,3 denote the mean hourly scheduled user throughput, inter-frequency HO success rate and intra-frequency HO success rate respectively.

Now, to simplify the multi-class classification and enable better understanding of the influential predictors for each user QoE KPI, we divide the problem into three binary classification sub-problems, one per QoE KPI. To observe peculiar frequency dependent behaviors, we develop models separately for 800 and 1800 MHz

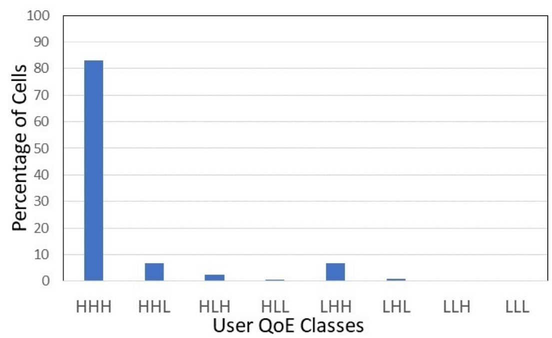

Fig. 5. Box Plot distributions for the user QoE KPIs.

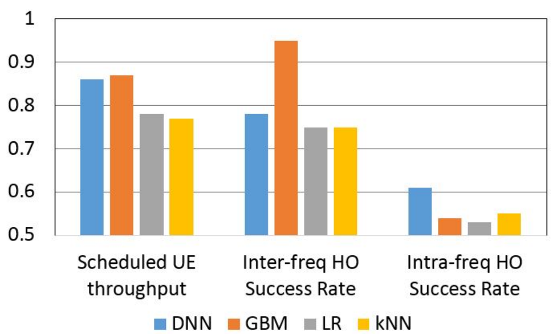

Fig. 6. AUROC Scores of Real-Time User QoE Prediction Models.

data. Stratified sampling is performed on the training data set for each of the six models (3 user QoE KPIs x 2 frequency bands) to enable balanced representation of High and Low classes in the training data. Real-time user QoE class prediction is performed using four standard techniques, namely: i) deep neural networks (DNN), ii) generalized boosted trees (GBM), iii) k nearest neighbors (kNN), and iv) logistic regression (LR).

Fig. 6 displays the model prediction results when unseen data (test data set) is passed through the trained models for the user QoE KPIs at 800 GHz. Due to the skewness in test data class distribution, we use the area under the receiver operating characteristic curve (AUROC) as the measure for model evaluation. The boosted trees (GBM) method outperforms other models in prediction of the QoE state for scheduled throughput and inter-freq HO success rate. While these two KPIs are predicted quite accurately from the selected LTE metrics, we observe that the intra-freq HO success rate KPI is not modeled efficiently with the selected attributes. Since this KPI is related to how well mobile users transition from one cell to another, GPS based user-level statistics such as individual speed would be a more relevant measure for predicting intra-freq HO success rate.

Fig. 7 depicts the relative influence of the predictors for the training models generated via gradient boosted trees. The measure is an indication of the relative frequency of the usage of a predictor in the tree splitting process. From the predictors' relative influence, we see that the primary distinguishing features between a High and Low throughput cell are percentage of QPSK transmissions in hybrid ARQ and cell load expressed as downlink PRB utilization. To increase the user QoE through a higher average scheduled throughput, an LTE operator may focus on small cell deployment / SON strategies that improve the channel conditions especially for cell edge users. Moreover, intelligent load balancing schemes may alleviate highly loaded cells during peak hours thereby reducing cell congestion and improving overall user QoE. For the inter-freq HO success, the number of inter-freq HO requests is the primary distinguishing feature between High and Low performance cell. We observe multiple factors having significance feature importance for intra-freq HO success rate in the trained GBM model. The dependence on a large number of features contributes to the inability of the boosted trees model to differentiate between High and Low performance. Additionally, for intra-freq HO, the model performance is expected to improve if average mobility statistics are incorporated while training the models.

VI. Future User QoE Prediction

In this section, we discuss the results from a time series prediction model with the aim to estimate the future user QoE states from current LTE metrics and QoE KPI values. The analysis was performed both for short-term and long-term time series predictions. We employ a recently proposed open source time prediction model by Facebook, known as PROPHET [22]. PROPHET is a modular time series regression model with three components: trend, seasonality and holidays (eq. 1 in [22]). It outperforms the prior state-of-the-art automatic prediction models, such as ARIMA (Auto-Regressive Integrated Moving Average), by virtue of incorporating weekly and seasonal trends, and reducing errors arising from trend variations. The model is flexible in accommodating seasonality effects and yielding interpretable insights for future parameter estimation.

As a case study, we evaluate the PROPET model on four diverse cells within the LTE network. The four cells capture a diverse variety of propagation environments, user density and mobility scenarios. The cells labelled from 1 to 4 hereon have the following propagation conditions: Cell 1 - suburban residential, Cell 2 - rural with a close by highway, Cell 3 - dense urban near city center, and Cell 4 - suburban with a close by highway. Future forecast modeling via PROPHET is performed both for immediate short-term as well as distant long-term QoE prediction. For the short-term prediction, we train the PROPHET model on a five-week LTE data set and predict the three user QoE KPIs for the coming two days. The same data is used for long-term prediction, the only difference being that instead of predicting the

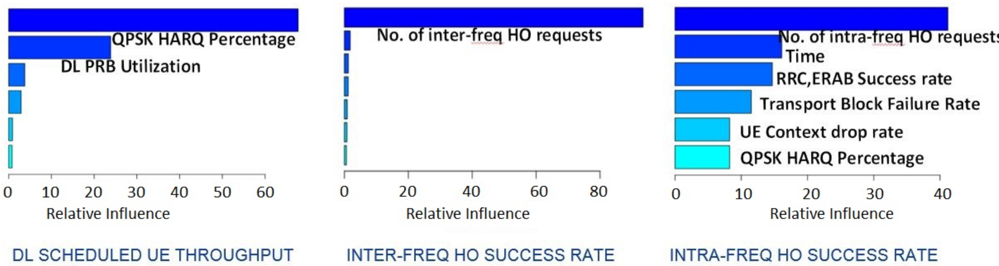

Fig. 7. Relative Influence measure of the predictors for the QoE KPIs using gradient boosted trees.

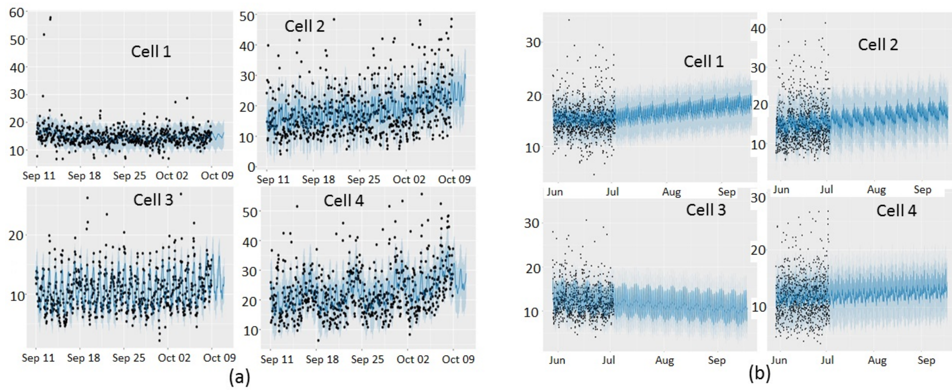

Fig. 8. PROPHET model prediction on scheduled user throughput (Mbps) in (a) short-term and (b) long-term future time frames.

immediate future, we predict the QoE performance for two days which are ten weeks in future from the last sequential training data point.

TABLE I
PROPHET MODEL ACCURACY ON QOE KPIS AND KEY PREDICTORS

QoE KPI/Predictor	Timeframe	Cell	RMSE
Scheduled UE throughput	Short-Term	1	2.7 Mbps
		2	8.7 Mbps
		3	3.7 Mbps
		4	7.7 Mbps
QPSK Percentage	Short-Term	1	0.09
		2	0.14
		3	0.11
		4	0.09
PRB Utilization	Short-Term	1	0.03
		2	0.02
		3	0.05
		4	0.04
Intra-freq HO Success Rate	Short-Term	1	4.6 %
		2	1.8 %
		3	1.1 %
		4	2.4 %
Scheduled UE throughput	Long-Term	1	6.5 Mbps
		2	6.8 Mbps
		3	2.6 Mbps
		4	8.2 Mbps

The PROPHET modeling for scheduled user throughput is illustrated for the four selected cells in Fig. 8. We can notice that while the short-term prediction (Fig. 8a) captures hourly and weekly throughput trends, the long-term future QoE prediction (Fig. 8b) demonstrates the average increasing / decreasing trend and is less sensitive to hourly throughput fluctuations. The root mean square error (RMSE) results of the estimation accuracy for a few combinations of KPI and time frames are given in Table I. As evident from the presented RMSE values, the PROPHET model can capture the future trends for both the user QoE KPIs as well as the influential predictors. We do not include inter-freq HO prediction results because of the sparsity in the training data vector caused by a large proportion of time series data having nil inter-freq HO requests. The estimation errors are high for rural and suburban cells with fast user traffic due to presence of highways. Long-term prediction yields lower RMSE for Cells 2 and 3 because a larger time period allows better fitting of the trend component with a lower uncertainty interval. However, in case of Cell 1, the long-term prediction errors are high due to a large increasing trend slope. The five-week data throughput measurements for Cell 1 overestimated the throughput gradient for the next 10 weeks. The actual throughput, on the other hand, could not maintain the positive trend suggested by PROPHET, which eventually resulted in larger RMSE values.

VII. CONCLUSION

This paper has presented an approach for real-time user quality of experience estimation through surrogate LTE performance KPIs predicted through machine learning techniques. Current network management is performed in reactive mode, for instance in response to MDT reports or outage alarms in the network's operations and management (O&M) dashboards. Our work addresses this issue by utilizing ML techniques that enable network operators to detect low performing cell sites in a timely manner. In the proposed end-to-end network management framework, live LTE metrics collected at cell sites are stored at a centralized server from where they are fetched and filtered to prepare database for QoE model training. To test our hypothesis, we develop separate models on an LTE operator's data for the three user QoE KPIs, namely, scheduled user throughput, inter-frequency handover success rate and intra-frequency handover success rate. For real-time QoE prediction, we compare four machine learning techniques and observe that the gradient boosted trees method yields highest prediction accuracy with AUC scores of 0.87 and 0.95 for scheduled throughput and inter-frequency handover success rate respectively. The model training for intra-frequency handover success rate lacked reliability because mobility related information was not available in the data set. The trained models also yield relative influence measure of the predictors for the QoE KPIs. This provides insights to the operator for re-aligning the proactive measures in order to curb the effects from these highly influential predictors. For future QoE KPI and predictor estimation, we use Facebook's PROPHET model which captures the hourly fluctuations and weekly trends with high accuracy, particularly for cells operating in dense urban regions. In summary, this case study presents a strong case for leveraging well established and powerful tools from the domain of machine learning and data analytics for proactive network management and user QoE enhancement.

ACKNOWLEDGEMENT

This work was carried out during Umair Hashmi's summer internship at Nokia Bell Labs in New Jersey, USA. The QoE management framework is inspired from Dr. Ali Imran's work. For more information about Dr. Imran's projects, please visit www.ai4networks.com.

REFERENCES

[1] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: how to empower SON with big data for enabling 5G," *IEEE Network*, vol. 28, no. 6, pp. 27–33, Nov 2014.

[2] X. Chen, F. Mariaux, and S. Valentin, "Predicting a user's next cell with supervised learning based on channel states," in *2013 IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2013, pp. 36–40.

[3] A. Adeel, H. Larijani, A. Javed, and A. Ahmadinia, "Critical Analysis of Learning Algorithms in Random Neural Network Based Cognitive Engine for LTE Systems," in *2015 IEEE 81st Vehicular Technology Conference*, May 2015, pp. 1–5.

[4] C. A. S. Franco and J. R. B. de Marca, "Load balancing in self-organized heterogeneous LTE networks: A statistical learning approach," in *2015 7th IEEE Latin-American Conference on Communications (LATINCOM)*, Nov 2015, pp. 1–5.

[5] R. M. Khanafer, B. Solana, J. Triola, R. Barco, L. Moltsen, Z. Altman, and P. Lazaro, "Automated Diagnosis for UMTS Networks Using Bayesian Network Approach," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2451–2461, July 2008.

[6] S. Chernov, F. Chernogorov, D. Petrov, and T. Ristaniemi, "Data mining framework for random access failure detection in LTE networks," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, Sept 2014, pp. 1321–1326.

[7] W. Xue, M. Peng, Y. Ma, and H. Zhang, "Classification-based approach for cell outage detection in self-healing heterogeneous networks," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2014, pp. 2822–2826.

[8] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "A learning-based approach for autonomous outage detection and coverage optimization," *Trans. Emerging Telecommunications Technologies*, vol. 27, no. 3, pp. 439–450, 2016. [Online]. Available: https://doi.org/10.1002/ett.2971

[9] H. Y. Lateef, A. Imran, and A. Abu-dayya, "A framework for classification of Self-Organising network conflicts and coordination algorithms," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 2898–2903.

[10] C. M. Mueller, M. Kaschub, C. Blankenhorn, and S. Wanke, "A Cell Outage Detection Algorithm Using Neighbor Cell List Reports," in *Self-Organizing Systems*, K. A. Hummel and J. P. G. Sterbenz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 218–229.

[11] W. Wang and Q. Zhang, "Local cooperation architecture for self-healing femtocell networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 42–49, April 2014.

[12] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, Aug 2014.

[13] S. Chernov, D. Petrov, and T. Ristaniemi, "Location accuracy impact on cell outage detection in LTE-A networks," in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug 2015, pp. 1162–1167.

[14] U. S. Hashmi, A. Darbandi, and A. Imran, "Enabling proactive self-healing by data mining network failure logs," in *2017 International Conference on Computing, Networking and Communications (ICNC)*, Jan 2017, pp. 511–517.

[15] F. Bação, V. Lobo, and M. Painho, "Self-organizing Maps as Substitutes for K-Means Clustering," in *Computational Science – ICCS 2005*, V. S. Sunderam, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 476–483.

[16] N. Sinclair, D. Harle, I. A. Glover, J. Irvine, and R. C. Atkinson, "An Advanced SOM Algorithm Applied to Handover Management Within LTE," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1883–1894, Jun 2013.

[17] P. Sukkhawatchani and W. Usaha, "Performance evaluation of anomaly detection in cellular core networks using self-organizing map," in *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, vol. 1, May 2008, pp. 361–364.

[18] M. N. ul Islam and A. Mitschele-Thiel, "Reinforcement learning strategies for self-organized coverage and capacity optimization," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2012, pp. 2818–2823.

[19] R. Razavi, S. Klein, and H. Claussen, "Self-optimization of capacity and coverage in LTE networks using a fuzzy reinforcement learning approach," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2010, pp. 1865–1870.

[20] H. Farooq and A. Imran, "Spatiotemporal Mobility Prediction in Proactive Self-Organizing Cellular Networks," *IEEE Communications Letters*, vol. 21, no. 2, pp. 370–373, Feb 2017.

[21] E. Bastug, M. Bennis, and M. Debbah, "A transfer learning approach for cache-enabled wireless networks," in *2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2015, pp. 161–166.

[22] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, 09 2017.