World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:14, No:6, 2020

Uplink Throughput Prediction in Cellular Mobile Networks

Engin Eyceyurt, Josko Zec

***Abstract*—**The current and future cellular mobile communication networks generate enormous amounts of data. Networks have become extremely complex with extensive space of parameters, features and counters. These networks are unmanageable with legacy methods and an enhanced design and optimization approach is necessary that is increasingly reliant on machine learning. This paper proposes that machine learning as a viable approach for uplink throughput prediction. LTE radio metric, such as Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and Signal to Noise Ratio (SNR) are used to train models to estimate expected uplink throughput. The prediction accuracy with high determination coefficient of 91.2% is obtained from measurements collected with a simple smartphone application.

***Keywords*—**Drive test, LTE, machine learning, uplink throughput prediction.

I. Introduction

CELLULAR broadband data usage is growing steeply, facilitated by mature fourth generation (4G) and emerging fifth generation (5G) of the wireless cellular technology. Therefore, it is essential to maintain the Quality of Service (QoS) ensuring stable networks and high-speed data rates. Data rates in wireless communication systems have increased significantly from modest beginnings in the second generation (2G) to current faster-than-wireline wireless broadband data speeds in the 5G of cellular standards. Web content, video streaming and social networking providers now offer high-quality audio and video content to mobile users. Improved user experience is driving demand for mobile data to approximately double every two years and the total monthly volume is expected to reach 77.5 exabytes in 2022 [1], [2]. This growth will be fueled by numerous new applications under the umbrella of the Internet-of-Things (IoT) concept, including exploding autonomous vehicle (AV) industry. The current worldwide Covid-19 pandemic is straightforward evidence of the urgent need for such innovative technologies.

The data volume in downlink direction from the base station to the mobile user equipment (UE) is typically higher than uplink volume from the UEs to the base stations. Therefore, the cellular operators sometimes concentrate more on downlink and deploy asymmetry between uplink and downlink channels. This asymmetry causes lower upload data speed than download speeds. Moreover, cellular standards generally accommodate more metric and observables in the downlink than uplink. However, some applications, such as AV control, rely mostly on uplink to assist in emergency remote intervention that is legally required in the AV industry. High uplink throughput is necessary in AV control to carry video and sensor data and thus enable remote vehicle control. Numerous other applications require comparable uplink and downlink data rates, such as video conferencing, file sharing and VoIP. Uplink volume constituted approximately 13% percent of the total mobile data traffic in 2014. This ratio is expected to reach 30% in 2020 [3], [4]. It is estimated that today approximately 20% of users receive unsatisfactory uplink data rates [5].

Number of parameters, vendor features and counters in cellular networks has grown to levels unmanageable by humans using traditional rule-based methods. Even most experienced engineers cannot feasibly process all information generated within a network. Machine learning (ML) algorithms, in combination with big data technology delivered on powerful computing platforms, are being increasingly used in design and optimization of cellular networks [6], [7]. Modeling downlink and uplink UE throughput is one obvious ML use case because data throughput is the primary metric in today's exclusively packet-switched cellular networks, replacing traditional call drop rates in earlier years of cellular communications, dominated by circuit-switched voice service.

Current dominant cellular technology globally is the Long-Term Evolution (LTE), commonly referred to as 4G, with 5G being rolled out and expected to become leading radio access mechanism in the near future. This paper addresses modeling and predicting uplink throughput from a variety of radio measurements in 4G. These predictions, for example, can be the key input for AV operators to route their robotic vehicles along routes that will ensure highest uplink throughput. High uplink throughput may be critical in case of emergencies that require remote vehicle operation instead of vehicle's complete autonomy.

The scope of this paper is to introduce the topic of uplink throughput prediction via a simple measurement collection platform available in the form of smartphone application. Data collected via this platform will be used to initiate model development and create a framework for future modeling from more comprehensive data sets, such as wide-area drive tests with professional-rate scanners, crowdsourcing data, network measurements and cell traces. Thus, this paper can be considered as the first in a series.

The paper continues with recognizing related work, describing collection of data used in modeling, defining

Engin Eyceyurt is with the Florida Institute of Technology, Melbourne, FL 32901 USA (corresponding author, phone: 321-735-2161; e-mail: eeyceyurt2013@my.fit.edu).

Josko Zec is with the Electrical Engineering Department, Florida Institute of Technology, Melbourne, FL 32901 USA

methodology behind uplink throughput prediction and concludes with target future work.

II. Related Work

Cellular operators rank throughput as the most important metric when evaluating network quality. Therefore, modeling throughput is a major step in network design and optimization. Cellular mobile data throughput models may be grouped into history-based, calculation-based, and ML -based predictions. In history-based predictions, the throughput is predicted extrapolating past behavior. Throughput data are collected and archived, and seasonal predictions are made accommodating various parameters, such as weekend/weekdays, leaf coverage, traffic patterns, etc. Calculation-based methods attempt to model throughput as a function of various variables in an analytical form. However, neither of these two approaches resulted in satisfactory prediction accuracy, particularly on uplink.

Previous studies related to` throughput predictions are mainly focused on downlink direction. Konishi et al. evaluate the downlink throughput while video streaming and achieve 81% prediction accuracy in the daytime, with the accuracy dropping to 72% at night [8]. Yue et al. use LTE radio measurements from stationary, local routes, highways, and pedestrian routes to predict downlink throughput. In that study, downlink from two US cellular operators are compared and the link forecast is made via ML with the prediction rate of between 83% and 96% [9]. Liu and Lee use ML algorithms on data taken during an eight-day collection period and make the downlink traffic and throughput estimations in 3G systems [10]. Lee et al. also work on time-series and ML algorithms deploying neural networks and multiple linear regression methods to predict hourly downlink throughput [11]. Wei et al. utilize the method of recurrent neural networks to make estimation on the downlink throughput stating that this method decreases the prediction errors by 29% percent compared to traditional prediction algorithms [12]. Oussakel et al. work on uplink throughput prediction and throughput sensitivities to noise levels and training size; however, information about the prediction rate is not shared [13].

III. Data Collection

Data used in throughput modeling have been collected in Melbourne, FL area, where suburban morphology dominates. *RantCell Test Analytics* measurement application is installed on a Samsung Note 10 smartphone and all tests are made with the same device. Data collection took five weekdays (from March 30 to April 3, 2020) under same weather conditions and stable, although atypical traffic driven by partial lockdown due to Covid-19 pandemic. The measurements are taken along different routes each day to observe a potentially diverse range of LTE radio conditions and associated uplink throughput rates. The average car speed was 40 mph. The phone is set with the same configurations for each drive test. Some of the unnecessary application features are disabled to avoid additional battery consumption. The following metric is generated within a drive test:

- Timestamp: timestamp of the data sample
- Bit rate: Uplink throughput value in Mbps
- Coordinate: Latitude and longitude from the embedded phone GPS
- Network type: The network technology (5G, LTE, 3G or Wi-Fi)
- RSRP: The average power of cell-specific signals. RSRP is used in cell selection and coverage estimation. The range of RSRP values is between -140 dBm and -44 dBm.
- RSRQ: The quality of signal and ranging between -20 dB and -3 dB
- SNR: Signal to noise ratio in dB [14]

IV. Uplink Throughput Prediction

In this section, an attempt to predict the uplink throughput via ML modeling will be described. The process contains the data collection, feature extraction, training, and estimation. Similar process is applicable regardless of input data and measurements used in modeling. The modeling process used in the uplink throughput prediction is illustrated in Fig. 1. Measurements contain both features/parameters (radio measurement matrix x) and targets/labels (uplink throughput vector y). Features and labels are extracted from the measurements and divided into training and estimation (testing) sets, a common procedure in ML. Training set is used in model generation and testing set is used to assess model accuracy.

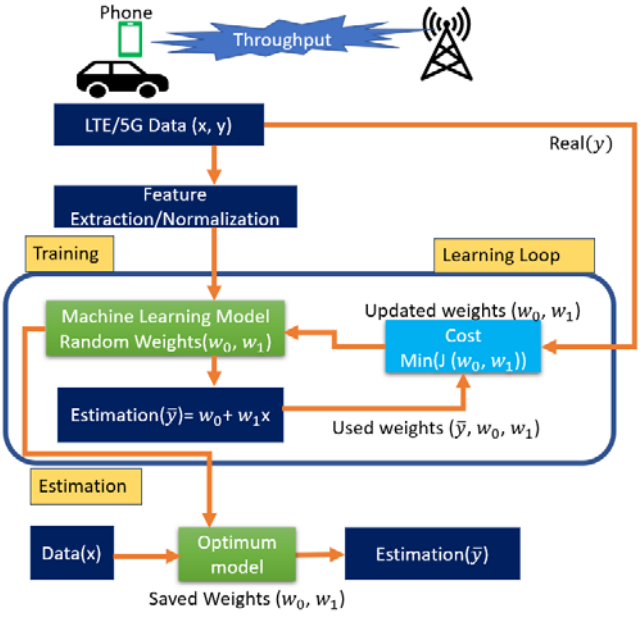

Fig. 1 Modeling process block diagram

Prior to training/prediction stages, correlation analysis is deployed to estimate correlation between each feature and the target (uplink throughput). Correlation is followed by the training and the prediction stages. Following the prediction, feature importance and prediction accuracy are analyzed. The first correlated feature is the downlink RSRP. RSRP is the

fundamental measurement for the LTE coverage because it captures received signal (in dBm) from a constant-power reference signal. This measurement is not affected by loading nor interference. Thus, RSRP serves to estimate coverage of each radio cell and is the basis for reselection and handover decisions. Since uplink and downlink are transmitted at different frequencies, signal levels fade independently and strong instantaneous RSRP does not necessarily mean strong uplink signal. However, on average, RSRP captures coverage conditions that are partially reciprocal between uplink and downlink. For example, high RSRP may indicate outdoor UE close to the cell and these are favorable conditions on both links. Thus, in absence of equivalent uplink metric, RSRP is corelated with uplink throughput and scatter plot of uplink throughput as a function of RSRP is presented in Fig. 2.

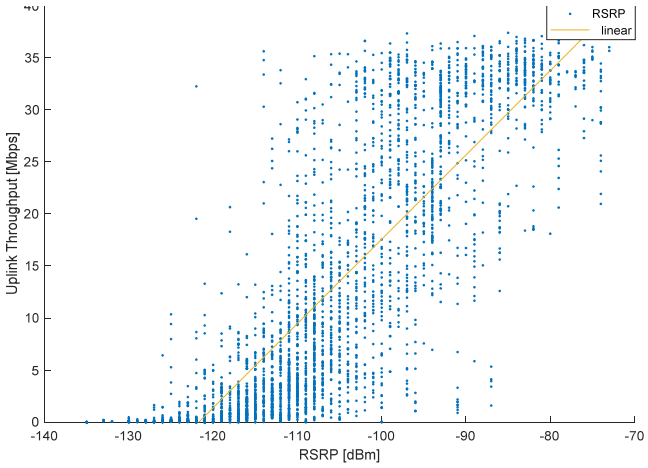

Fig. 2 Scatter plot of RSRP and uplink throughput

Linear trend line is included in the scatter plot and it shows expected increase of uplink throughput with growing RSRP. Cluster of RSRP measurements below -110 dBm is associated with low uplink throughputs and the highest throughputs are concentrated in the highest RSRP regions. This clear trend confirms that RSRP is useful in predicting associated uplink throughput with sufficient correlation.

RSRQ is another fundamental LTE radio metric. It captures ratio between received downlink power from reference signal resource blocks of serving cell and the total power in downlink direction, including all UEs in serving and adjacent cells. Therefore, such ratio is always less than 1 (negative in dB). This ratio depends on load in serving and adjacent cells and cannot be used as direct measure of pure cell coverage.

RSRQ ranges are between -20 dB and -3 dB, where -20 dB indicate high interference and -3 dB indicate clear conditions with little interference. However, direct interpretation of RSRQ as the measure of interference is affected by the nature of the power contribution in the RSRQ denominator. If most wideband power in the RSRQ denominator comes from the in-cell traffic, that is not interference because same cell UEs are orthogonal and do not mutually interfere. If most power in the RSRQ denominator is contributed from adjacent cells, orthogonality is not maintained among cells and that power constitutes the real downlink interference.

Scatter plot showing uplink throughput as a function of the RSRQ is shown in Fig. 3, together with the linear trend line. While trend line shows expected increasing trend with the RSRQ, density of the points is not as clearly concentrated in regions, as in RSRP comparison with wider spread at each RSRQ value. Therefore, correlation between the uplink throughput and the RSRQ is expected to be lower than the correlation between the uplink throughput and the RSRP.

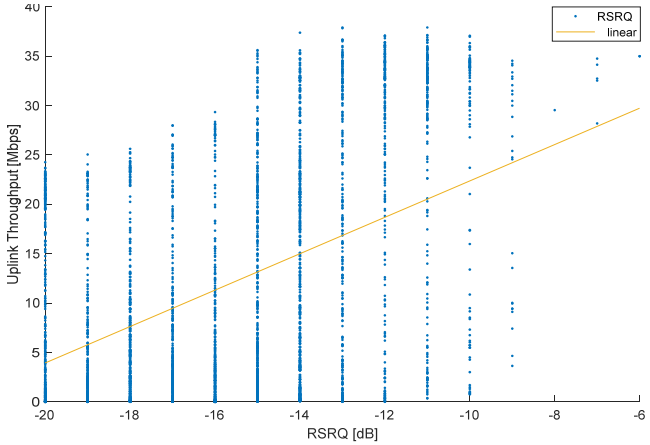

Fig. 3 Scatter plot of RSRQ and uplink throughput

The third investigated feature for correlation with the uplink throughput has been the SNR estimated at the UE receiver. This metric applies to the baseband downlink signal after down-conversion from the high frequency wideband carrier. This measurement is expected to correlate closely with downlink throughput, but also maintain correlation with the uplink throughput, although not as directly as in the downlink direction. That is confirmed with measured data presented on the scatter plot in Fig. 4 together with linear regression line. Measurements with low uplink throughput are concentrated in the low SNR region and the measurements with high uplink throughput are grouped around higher SNR values. Therefore, SNR is, similarly to RSRP, well correlated and will help in predicting the uplink throughput. The slope of the uplink throughput versus SNR is lower than the slope of the uplink throughput versus RSRP.

Correlations behind scatter plots in Figs. 2-4 have been calculated and presented in Table I. The correlation coefficients are calculated using Spearman's rank-order correlation. It is more robust to outliers than common Pearson's linear correlation and is preferred for describing dependency among parameters that are not normally distributed [15]. Spearman's correlation factor ranges between -1 and +1. Values close to +1 indicate two parameters ranked closely together, near-zero values indicated weak correlation and those near -1 indicate ranking in the opposite direction.

$$\rho = \frac{6 \sum_{i=1}^{n} (Rank\,(x_i) - Rank\,(y_i))^2}{n^3 - n} \quad (1)$$

where ρ = spearman's rank-order correlation factor; n = number of data pairs.

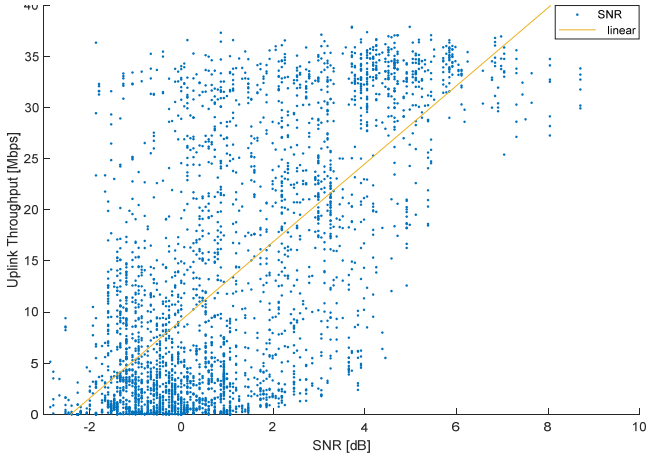

Fig. 4 Scatter plot of SNR and uplink throughput

As visually observed in the scatter plots, RSRP is the most-correlated feature with the uplink throughput (85%), followed by SNR (62%) and RSRQ (29%). These three metrics, as features, and uplink data rates, as labels, are inserted into stochastic descent-based ML algorithm after the feature normalization. Algorithm converges to a model that estimates uplink throughput from these three features. Mean square errors are calculated between measured and estimated uplink throughputs, and weights are updated with the direction of gradient in the algorithm loop until the process converges. The algorithm is implemented within the Python's Scikit-learn environment, one of the leading platforms for ML development. Scikit facilitates data preprocessing, such as data cleanup and feature normalization.

TABLE I
CORRELATION BETWEEN UL THROUGHPUT AND RSL VALUES

	LTE Parameters		
	RSRP	**RSRQ**	**SNR**
Correlation with Uplink Throughput	0.85	0.29	0.62

Measurements collected over a 5-day period are divided into training and test sets where 80% of data are randomly assigned as training set and fed into gradient boosting regressor model. The remaining data are used for the test set, which is used to assess model prediction accuracy. Gradient boosting regressor is chosen among a rich set of Scikit regression modules to train the model. The trained model is deployed on the test measurement subset to predict the uplink throughput. Predicted uplink throughputs are compared with test labels via the coefficient of determination R^2 defined as:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (2)$$

where, y_i is the individual throughput measurement, $\hat{y}_i$ is the predicted throughput in i^{th} measurement and $\bar{y}$ is the mean uplink throughput. The value calculated from collected data is 0.91, which indicates good prediction capabilities using simple metric.

Following the ML training and prediction, the importance of each featured variable is investigated to rank features according to the impact on prediction capability. Calculated importance is illustrated in the bar chart in Fig. 5. Table I confirms that the RSRP metric is the main contributor to the prediction algorithm, followed by the baseband SNR, and finally, RSRQ.

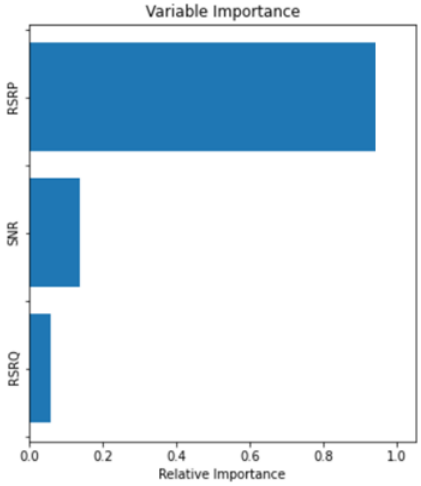

Fig. 5 Variable importance chart for the prediction

V. CONCLUSION

In this paper, we introduced a ML-based uplink throughput prediction applicable to cellular networks of 4^{th} and 5^{th} generations of standard. Although based on a very limited feature set, RSRP, RSRQ, and SNR, the model achieved somewhat surprising accuracy. The determination coefficient reached 91.2% indicating good predictability of uplink throughput, using a simple and common LTE metric. Our future work will include deploying more comprehensive measurement sources and expanding the feature list. Another future target is focusing on low-throughput spots where critical applications, such as AV, might experience outages.

ACKNOWLEDGMENT

The authors acknowledge the supports of the Open Access Subvention Fund by the Florida Tech Libraries.

REFERENCES

[1] J. Thomas Barnett, "Cisco Visual Networking Index (VNI) Global and Americas/EMEAR Mobile Data Traffic Forecast, 2017–2022," Cisco, 2019.
[2] Y. Egi, E. Eyceyurt, I. Kostanic, C. E. Otero, "An Efficient Approach for Evaluating Performance in LTE Wireless Networks," Las Vegas, 2017.
[3] Union, International telecommunication, "IMT traffic estimates for the years 2020 to 2030," Geneva, 2015.
[4] P. &. i. Inc, "Global Mobile Data Traffic Forecast, 2012 – 2017," Austin TX, 2013.
[5] Ericsson Mobility Report, Managing User Experience, Ericsson, 2016.
[6] Otero, Y. Egi C., "Machine-Learning and 3D Point-Cloud Based Signal Power Path Loss Model for the Deployment of Wireless Communication Systems," *IEEE Access*, vol. 7, pp. 42507-42517, 2019.

[7] Y. Egi, C. Otero, M. Ridley and E. Eyceyurt, "An Efficient Architecture for Modeling Path Loss on Forest Canopy Using LiDAR and Wireless Sensor Networks Fusion," in *23rd European Wireless Conference*, Dresden, Germany, 2017.
[8] H. Konoshi, K. Kanai, J Katto, "Improvement of Throughput prediction Accuracy for Video Streaming in Mobile Environment," in *IEEE 3rd Global Conference on Consumer Electronics*, Tokyo, 2014.
[9] C. Yue, R. Jin, K. Suh, Y. Qin, B. Wang and W. Wei;, "LinkForecast: Cellular Link Bandwidth Prediction in LTE Networks," *IEEE Transactions on Mobile Computing,* vol. 17, pp. 1582-1594, 2018.
[10] Y. Liu and J.Y.B. Lee, "An Empirical Study of Throughput Prediction in Mobile Data Networks," in *IEEE Global Communications Conference*, San Diego, 2015.
[11] D. Lee, D. Lee, M. Choi and J. Lee, "Prediction of Network Throughput using ARIMA," in *International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Fukuoka, Japan, 2020.
[12] B. Wei, M. Okano, K. Kanai, W. Kawakami and J. Katto, "Throughput Prediction Using Recurrent Neural Network Model," in *IEEE 7th Global Conference on Consumer Electronics*, Nara, 2018.
[13] I. Oussakel, P. Owezarski, P. Berthou, "Cellular Uplink Bandwidth Prediction Based on Radio Measurements," in *MobiWac*, Miami, 2019.
[14] E. T. 1. 2. V. (2018-07), "Technical Specification Physical layer measurements, 5G; NR," 2018.
[15] D. S. Mehta and S. Chen, "A spearman correlation based star pattern recognition," in *IEEE International Conference on Image Processing (ICIP)*, Beijing, 2017.