A Machine Learning Approach that meets Axiomatic Properties in Probabilistic Analysis of LTE Spectral Efficiency

Yoonjoo Lee, Yunbae Kim, Seungkeun Park
Radio Resource Research Section
ETRI
Daejeon, South Korea
yjl7436@gmail.com

***Abstract*—Recent maturity of Long Term Evolution (LTE) has raised interest in assessing how efficiently the cells are deployed. Since Cell Spectral Efficiency (CSE) is a principal indicator of such assessment, we calculate the Cumulative Distribution Function (CDF) of CSE using the measurement data obtained from Rohde & Schwarz TSME which can decode the control channel of the downlink signal. We adopt data processing methods from our previous work for deriving the CDF of CSE. In this paper, we propose a deep neural network model which is modified from the previous one to guarantee that the CDF generated from the model satisfies the probability axioms for any cell configurations. Predicted results from the measured data validate our analysis.**

***Index Terms*—LTE, cell spectral efficiency, cumulative distribution function, deep neural network**

I. Introduction

As Long Term Evolution (LTE) grows enough to cover the increasing traffic demands of mobile users, attention has been focused on evaluating how efficiently LTE cells are deployed from various perspectives. Since cell deployment directly affects the quality of transmissions, analyzing Cell Spectral Efficiency (CSE) can be a way of carrying out the evaluation. CSE is defined as the aggregate ThroughPut (TP) of the users normalized by the number of cells and the channel bandwidth. Under this definition, the analysis requires scheduling records on how the bits are transmitted through Resource Blocks (RBs) for all users at the same time. Although such information can be easily accessible in operators' side, it is usually not disclosed to the public, and several research works have been based on in-field measurement data [1]–[3]. In our previous work [1], noting the Physical Downlink Control CHannel (PDCCH) of an LTE downlink signal includes the necessary scheduling information, we utilized a popular commercial off-the-shelf scanner called Rohde & Schwarz (R&S) TSME to decode LTE PDCCH and analyzed the CSE performance using data obtained from R&S ROMES software.

Since the CSE value depends on the user's varying channel condition, performance analysis of CSE needs to be addressed from a stochastic point of view. In [1], novel transformation methods are devised so that the transformed results for cell TP and RB utilization approximately follow a joint Gaussian distribution, which is briefly reviewed in Section II. By exploiting the transformations, the Cumulative Distribution Function (CDF) of CSE can be computed by using Monte Carlo integration technique. Also, a Deep Neural Network (DNN) is applied to extend the analysis of all network configurations, not just actual measured cases. Since we focus on the CDF, it is necessary to ensure that the results of applying DNN satisfy the axiomatic properties of probability. In this work, we modify the DNN model given in [1] so that those properties are satisfied when computing the CDF of CSE for any cases.

The rest of the paper is structured as follows. Section II presents the review of data processing given in [1]. Section III explains our proposed model and the results of applying it. Finally, Section IV concludes this work.

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00109, Development of Frequency Analysis Technology for the Virtuous Circulation of Radio Resource).

II. Data Processing

A. Data description

The data was measured for four days at Gangnam Station, one of the most congested area of South Korea, and for sub-GHz frequency of 10 MHz bandwidth. The data of the PDCCH decoded by the TSME was provided via ROMES as a statistical result for the sampled subframes during the predefined Observation Internal (OI). To analyze the performance of CSE from a stochastic point of view, the following three variables defined as in [4] are used, all of which are measured during the OI. The 'Number of RNTIs' represented by N means the number of different Radio Network Temporary Identifiers (RNTIs) detected. The 'RB usage of Cell' represented by R is calculated as [Sum of all used RBs in measured subframes] / [Sum of all RBs in measured subframes]. The 'Average Scheduled TP of Cell' represented by T is calculated as [Sum of all transmitted bits] / [Number of measured subframes]. We use R as a value in $(0, 1)$ and T as the value in kbps units, and from now on consider N, R, and T as random variables. In addition, the goal of our paper is to compare how CSE performance differs as the number of RNTIs increases, so all data is collected and analyzed regardless of cell and measurement date.

Let $\mathcal{D}$ the set of all the data consisting of (r, t, n) triples of realized values of R, T, and N, and let $\mathcal{D}_n$ be the set of (r, t) pairs where $N = n$. For the description of the contents, we called $\mathcal{D}$ as a set of all data, which was composed of (r, t, n) of realized values of R, T, and N, and also called $\mathcal{D}_n$ as a set of (r, t) pairs when $N = n$. Since the actual data shows high time correlations having multiple runs of the same realizations, we randomly choose K pairs from $\mathcal{D}_n$ and regard it as a random sample from the conditional joint distribution $(R, T)|N = n$.

B. Data transformation

As given in [1], a Quantile-Quantile (Q-Q) plot of $R|N = n$ against theoretical normal quantiles shows an 'S'-shaped curves in general, and it inspires the application of the transformation family as follows:

$$U(r; u_1, u_2, u_3) = u_1 \log(r) - u_2 \log(1 - r) + u_3,$$

where $u_1, u_2 \geq 0$, and $u_3 \in \mathbb{R}$. Similarly, observing that $T|N = n$ usually shows a Q-Q plot which is downward arced shape, the following family of transformation is applied.

$$V(t; v_1, v_2, v_3) = v_1 \log(t) + v_2 t + v_3,$$

where $v_1, v_2 \geq 0$, and $v_3 \in \mathbb{R}$. The optimal parameters (u_1, u_2, u_3) and (v_1, v_2, v_3) are obtained so that mean squared errors against the theoretical quantiles of the standard normal are minimized. In other words, the optimal parameters minimize $J_R(u_1, u_2, u_3) = \sum_{k=1}^{K} \{q_k - U(r_k; u_1, u_2, u_3)\}^2$ and $J_T(v_1, v_2, v_3) = \sum_{k=1}^{K} \{q_k - V(t_k; v_1, v_2, v_3)\}^2$ where $q_k = \Phi^{-1}((k - 0.375)/(k + 0.25))$ for $1 \leq k \leq K$ and $\Phi(\cdot)$ is the CDF of the standard normal distribution.

C. Processing for SE analysis

As described in [1], normality tests are applied to configure the data for SE analysis. We use Shapiro-Wilk and Royston tests to check the marginal and joint normalities, respectively, in the sense that the former does not mean the latter. For a data point to be organized, we begin with choosing a sample n for N with probabilities $P(N = n) = |\mathcal{D}_n|/|\mathcal{D}|$. Then, a set of pairs $\{(r_k, t_k)\}_{k=1}^{K}$ of size K is chosen from $\mathcal{D}_n$. With the optimal parameters (u_1, u_2, u_3) and (v_1, v_2, v_3) for this set of pairs, Shapiro-Wilk test is applied to each of $\{U(r_k; u_1, u_2, u_3)\}_{k=1}^{K}$ and $\{V(t_k; v_1, v_2, v_3)\}_{k=1}^{K}$. If the normality assumption is accepted for both sets, Royston test is applied to the transformed pairs $\{(U(r_k; u_1, u_2, u_3), V(t_k; v_1, v_2, v_3))\}_{k=1}^{K}$. When the p-value of the set is greater than the chosen significance level, we can see that the set of transformed results is a random sample of the joint normal distribution $\mathcal{N}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}1&\rho\\\rho&1\end{smallmatrix}\right]\right)$ where ρ is the sample covariance. Then we structure a data point as $(n, \boldsymbol{\theta})$ with $\boldsymbol{\theta} = (u_1, u_2, u_3, v_1, v_2, v_3, \rho)$. A summary for this process is provided in Fig. 1 (which is Fig. 6 in [1]).

III. Cell Spectral Efficiency Analysis

Let CSE be denoed by S, then S can be calculated as $T/(wR)$, where w be the bandwidth in kHz unit, based on the

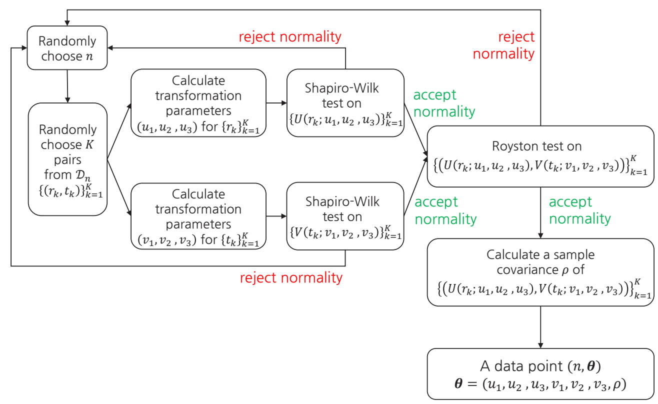


Fig. 1. A summary of processing the data.

definition. Let some transformations $U(\cdot)$ and $V(\cdot)$ that make $(U(R), V(T))|N = n$ approximately follow the joint normal $\mathcal{N}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}1&\rho\\\rho&1\end{smallmatrix}\right]\right)$ be given. Then, the CDF $P(S \leq \eta|N = n)$ for $S|N = n$ can be derived as follows:

$$\begin{aligned} P(S \leq \eta|N = n) &= P\left(\frac{T}{wR} \leq \eta \middle| N = n\right) \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{V\left(U^{-1}(x)w\eta\right) - \rho x}{\sqrt{1-\rho^2}}\right) f_{U(R)|N=n}(x)dx. \\ &(\because\ V(T)|U(R) = x, N = n \sim \mathcal{N}\left(\rho x, 1-\rho^2\right)) \end{aligned} \tag{1}$$

By Monte Carlo integration technique, the integral (1) can be computed as

$$\frac{1}{\Gamma}\sum_{i=1}^{\Gamma} \Phi\left(\frac{V\left(U^{-1}(x_i)w\eta\right) - \rho x_i}{\sqrt{1-\rho^2}}\right), \tag{2}$$

where $\{x_i\}_{i=1}^{\Gamma}$ is a random sample of size Γ from $\mathcal{N}(0, 1)$.

A. Proposed model

In our previous work [1], from a data point $(n, \boldsymbol{\theta})$, input is set as (η, n, x) where η and x are random samples from the uniform distribution on a range $(0, 10)$ and the standard normal $\mathcal{N}(0, 1)$, respectively. With this input, an output is set to the argument of $\Phi(\cdot)$ in (1) with $\boldsymbol{\theta}$ as

$$y_o(\eta, n, x; \boldsymbol{\theta}) = \frac{V\left(U^{-1}(x; \boldsymbol{\theta})w\eta; \boldsymbol{\theta}\right) - \rho x}{\sqrt{1-\rho^2}}. \tag{3}$$

Then such an input-output pair is applied to a DNN as a data point.

When $\eta \geq 0$, the CDF of CSE should have a value of 0 at $\eta = 0$ and increase to 1 as η increases. Since the argument of $\Phi(\cdot)$ in (1) goes to $-\infty$ when η goes to zero and increases when so does η, the desired properties are satisfied. As in [5], instead of using the argument of $\Phi(\cdot)$ in (1) as an output of a DNN as (3), we approximate it by a parametrized function satisfying intended properties. Our proposed function is given as follows.

$$\begin{aligned} y_m(\eta, n, x; \boldsymbol{\theta}) &= c_0 + c_1 \log \eta + c_3 \eta^{c_2} \\ &= c_0 + c_1 \log \eta + c_3 e^{c_2 \log \eta}, \end{aligned} \tag{4}$$

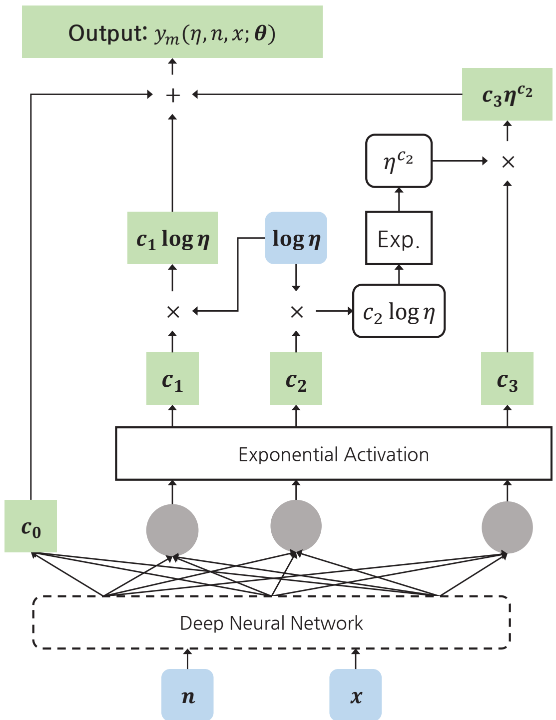


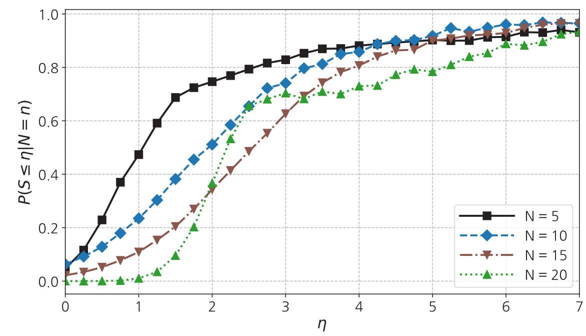
Fig. 2. The DNN structure of our proposed model. 'Exp.' means an exponential activation function.

where $c_0 \in \mathbb{R}$ and $c_1, c_2, c_3 \geq 0$. The parameters c_0, c_1, c_2, and c_3 are estimated by applying a DNN in the way of approximating the value $y_o(\eta, n, x; \boldsymbol{\theta})$ in (3) by $y_m(\eta, n, x; \boldsymbol{\theta})$ in (4). The structure of a DNN model for the proposed model can be configured as given in Fig. 2.

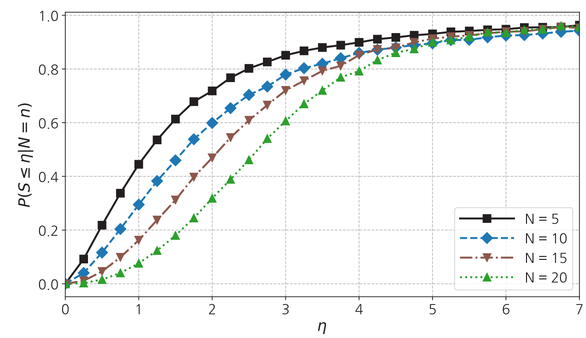
B. Results

For DNN structure, we use standard multilayer perceptron with Rectified Linear Unit activation function, which is widely used in modern NNs [6]. We consider two cases: the 'Previous' model based on (3) and the 'Proposed' model based on (4). DNN learning is performed with a data set of size 10000, 60% of which is used as a train set and the rest used as a validation set. The validation set is used for applying early stopping technique to prevent overfitting. In addition, an $l2$ penalty with a coefficient 0.01 is applied for both models. The DNN design parameters are determined to minimize validation loss. After model being learned through inputs (η, n, x) and respective outputs for considered models, we calculate the CDF of CSE for any input (η, n, x) by putting the output of the trained model into $\Phi(\cdot)$ and calculating (2) with $\Gamma = 1000$.

The results of the CDF of CSE for the considered models are given in Fig. 3. Two layers with 32 hidden units for the 'Previous' model and five layers with 8 hidden units for the 'Proposed' model are selected as DNN design parameters. As can be seen in Fig.3-(a), there are some cases that the previous DNN model generates the CDFs that do not satisfy the monotonicity and break the tendency. Such phenomena especially occur when there is just small number of data ($n = 10$) or not measured ($n = 15$ and 20). On the other hand,



(a) Previous model



(b) Proposed model

Fig. 3. Predicted results for the CDF of CSE from previous and proposed models.

we can check from Fig.3 (b) that the 'Proposed' model gives the CDFs that satisfy the axiomatic properties of probabilities and show a stable tendency even for unmeasured cases.

IV. Conclusion

In this work, we proposed a new model that improves our previous DNN model to enable a stable analysis of the CSE of LTE from a probabilistic point of view. By applying a parametrized function to the output of DNN learning, we have ensured that the axiomatic properties of probability are met in any case in deriving the CDF of CSE. The analysis was based on in-field measurement data provided by R&S TSME, and the measured data was used to check the validity of our proposed model through a comparison with predicted results.

References

[1] Y. Lee, Y. Kim, Y. Park and S. Park, "Probabilistic Analysis of Spectral Efficiency for LTE based on PDCCH Measurement Data," accepted for publication in *IEEE Commun. Lett.*, Jun. 2019.

[2] S. Kumar, E. Hamed, D. Katabi, and L. E. Li, "LTE Radio Analytics Made Easy and Accessible," in *Proc. ACM SIGCOMM*, pp. 211-222, Aug. 2014.

[3] R. Falkenberg, K. Heimann, and C. Wietfeld "Discover Your Competition in LTE: Client-Based Passive Data Rate Prediction by Machine Learning," in *Proc. IEEE GLOBECOM*, Dec. 2017.

[4] J. Schilbach, "R&S©TSMW, TSME, TSMA LTE Downlink Allocation Analysis," Application Note Rohde & Schwarz, 2016.

[5] H. E. Hammouti, M. Ghogho, and S. A. R. Zaidi, "A Machine Learning Approach to Predicting Coverage in Random Wireless Networks," in *Proc. IEEE GLOBECOM Workshops*, Dec. 2018.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016.