Spectrum Efficiency Prediction for Real-World 5G NR Networks Based on Drive Testing Data

Zheng Xing[1†], Haoyun Li[1†], Wenjie Liu[1], Zixiang Ren[1], Junting Chen[1], Jie Xu[1], and Cai Qin[2]
[1] School of Science and Engineering, The Chinese University of Hong Kong(Shenzhen), 518172, China
[2] Huawei Technologies Co., Ltd, Shenzhen, China

***Abstract*—Therein, the problem of predicting the spectrum efficiency for 5G NR cellular networks based on the channel state information (CSI) statistics, such as reference signal received power (RSRP) and average channel quality information (CQI), are studied. This problem is challenging because there is no clear correspondence between the CSI statistics and the spectrum efficiency in a 5G NR network. Instead, there are many instantaneous factors that affect the efficiency, such as the adaptive beamforming strategy employed by the base station (BS) and the inter-cell interference due to the traffic at the neighboring cell. In this paper, a model-assisted data-driven approach based on deep neural networks (DNN) is developed. First, a mobility-aware approach is derived to assist the data preprocessing and feature engineering; second, an interference-aware DNN model is developed to jointly train the predictions on the interference with noise and the spectrum efficiency. In our experiments in a real world 5G NR network, the BSs are equipped with 64 dual-polarized antennas, and a commercial user device with 4 antennas travels along a fixed route to collect drive testing measurements over a period of several months. It is found that the model-assisted approach indeeds provides additional prediction accuracy of 1–2%over a purely data-driven approach, and the overall prediction error can be brought down to 15%.**

I. Introduction

Big picture and motivation: In 5G NR wireless communication networks, with the increasingly dense BSs deployed and the massive antennas operating at each BS, it becomes extremely challenging for the operator to optimize the network parameters, such as tuning the tilting angles of the antenna panels at the BSs. Spectrum Efficiency (SE), which is also called spectral efficiency or bandwidth efficiency, refers to the information rate that can be transmitted over a given bandwidth in a specific communication system. It is a measure of how efficiently a limited frequency spectrum is utilized by the physical layer protocol, and sometimes by the media access control address (MAC). One promising strategy that is under active research is to divide the network parameter optimization into two parts: The first part develops a model that maps the parameters of a BS to the channel quality, such as RSRPs, at each geo-location. The second part builds a model to predict the SE at each geo-location based on the RSRPs. As a result, the operator can optimize the network parameters by evaluating the SEs in the area of interest.

However, although there has been a lot of research effort on modeling the long-term statistics, e.g., RSRP, of the wireless channel, such as those based on ray-tracing techniques, it is still challenging to predict the SE based on the long-term channel statistics.

Big problem and importance: This paper aims at predicting the SE based on the long-term channel statistics. Specifically, we target at using RSRPs as the only input to the model, because it is well-known how to calculate the RSRP at every location based on the antenna configuration [xx]-[yy], and moreover, RSRP data is easy to collect for building a training dataset according to the current standard for 5G NR networks.

Challenge:

• (Partial information) In Rayleigh fading channels, the SE can be computed using Shannon's formula if the full channel statistics is available. However, in real 5G NR systems with massive MIMO antenna configurations, only partial channel statistics is available, e.g., only knowing the RSRPs of a few pre-defined *CSI beams*, and the channel distribution is rarely known.

• (Unknown precoding strategy) In practical system, the BS runs a complicated precoding mechanism for the transmission, including the two-layer precoding strategy and the hybrid automatic repeat request (HARQ) mechanism. Thus, it is almost impossible to find a precise model to compute the SE from RSRP.

• (Inter-cell interference) As the inter-cell interference also affects the SE, it is not clear how to characterize the SE degradation due to inter-cell interference in massive MIMO networks where only coarse RSRP information is available from the neighboring cell.

Classical approach and deficiency:

Specific problem and High-level (technical) idea (intuition) to solve specific problems: Our goal is to establish and train a deep learning model that predicts the SE based on coarse CSI information in a 5G NR massive MIMO network, where the training dataset is highly unbalanced and incomplete. The interference prediction neural network is established to predict SE by combining the interference information obtained with SS and CSI. Aiming at the problem of large error in low SE prediction, a new scheme is designed.

Key contributions and key findings:

• location-based data augmentation

- model the interference-plus-noise (IPN)
- An interference-assisted SE prediction scheme is designed
- A new SE prediction scheme is designed for low SE.

Paper Structure: The remainder of this article is organized as follows. Section II shows the collection system model.

† Equal first authorship

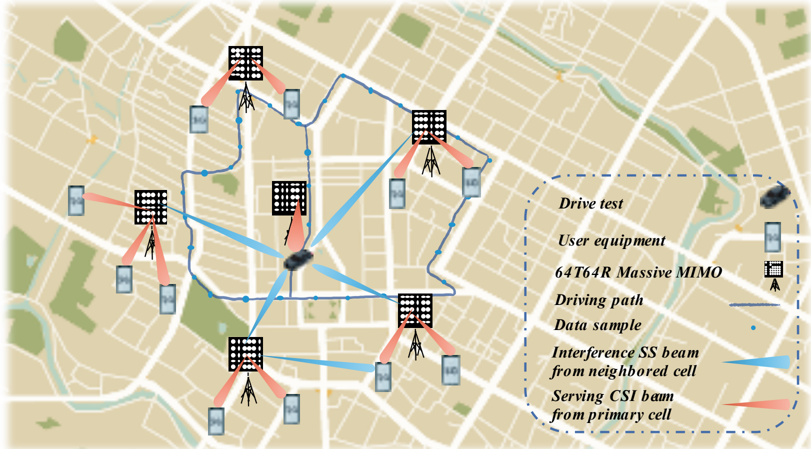

Figure 1: Massive MIMO Chengdu

Section III introduces our feature engineer techniques. Then we first design Interference Model Aided Prediction model in section IV. Then, we consider the prediction enhancement for sections with sparse training data in section V. Finally, we give experimental results in section VI.

II. System Model

A. The SE Model for 5G MIMO Networks

Consider a cellular network with $G+1$ BSs, where BS 0 only serves a drive test (DT) user and BSs $b=1,2,\ldots,G$ each serves multiple users. Each BS has N_t antennas[1] and the DT user has N_r antennas. The downlink channel from the bth BS to the DT user is given by $\mathbf{H}_b \in \mathbb{C}^{N_\mathrm{r}\times N_\mathrm{t}}$. Let $\mathbf{x}_b \in \mathbb{C}^{d_b}$ be the message transmitted at the bth BS, $b=0,1,\ldots,G$, where d_b are the numbers of data streams and $\mathbb{E}\{\mathbf{x}_b\mathbf{x}_b^\mathrm{H}\}=\mathbf{I}$. Let $\mathbf{V}_b \in \mathbb{C}^{N_\mathrm{t}\times d_b}$ be the transmit precoding matrix at the bth BS, where $\mathrm{tr}\{\mathbf{V}_b\mathbf{V}_b^\mathrm{H}\}\leq P$ with P denoting the total power constraint. The received signal at the DT user is given by

$$\mathbf{y}=\mathbf{H}_0\mathbf{V}_0\mathbf{x}_0+\sum_{b=1}^{G}\mathbf{H}_b\mathbf{V}_b\mathbf{x}_b+\mathbf{n}$$

where $\mathbf{n}\sim\mathcal{CN}(0,\sigma^2\mathbf{I}_{N_\mathrm{r}})$ is the additive complex Gaussian noise.

While the transmission mechanism is complicated in an actual 5G network, we only focus on the first-order behavior of the transmission system and employ a simple de-correlation model to assist the design of the neural network. Specifically, consider that the DT user employs a zero-forcing de-correlator and uses $\mathbf{u}_i$ to extract the ith data stream.[2] As a result, the SNR of the ith data stream is given by

$$\gamma_i=\frac{|\mathbf{u}_i^\mathrm{H}\mathbf{H}_0\mathbf{v}_{0,i}|^2}{\sigma^2+\sum_b\|\mathbf{u}_i^\mathrm{H}\mathbf{H}_b\mathbf{V}_b\|^2}$$

[1]In practice, N_t is a multiple of 4.

[2]In the low signal-to-noise ratio (SNR) regime, the BS may prefer to transmit only a single data stream , where zero-forcing can be identical to match filtering. At high SNR, zero-forcing is optimal.

where $\mathbf{v}_{0,i}$ is the ith column of the precoding matrix $\mathbf{V}_0$ for the DT user and $\|\cdot\|$ denotes the Euclidean norm. Thus, the instantaneous SE for the DT user can be computed as

$$r(\{\mathbf{H}_b\})=\sum_{i=1}^{d_0}\log_2\left(1+\gamma_i\right).$$

The BS is believed to use the "best effort" to design the precoding $\{\mathbf{V}_b\}$ on based on instantaneous information measured and reported by the DT user, such as, RSRP, CQI, and RI, as well as the transmission history, *e.g.,* due to the implementation of the HARQ scheme. The goal of the paper is to predict the average SE of the DT user $\bar{r}\triangleq\mathbb{E}\{r(\{\mathbf{H}_b\})\}$ based on the statistical information, *RSRP*, defined in the following subsection, where the expectation is taken over the small-scale fading of the MIMO channel at the order of hundreds of milliseconds.

B. Measurement Model for the RSRP

Denote $\mathbf{W}\in\mathbb{C}^{N_\mathrm{t}\times N_\mathrm{t}}$ as the DFT matrix with its ith column $\mathbf{w}_i$ being defined as the ith *CSI beam*. Consider to partition the index set $\{1,2,\ldots,N_\mathrm{t}\}$ into $M=N_\mathrm{t}/4$ subsets $\mathcal{C}_1,\mathcal{C}_2,\ldots,\mathcal{C}_M$. Define $\tilde{\mathbf{w}}_i \triangleq \frac{1}{4}\sum_{j\in\mathcal{C}_i}\mathbf{w}_j$ as the beamforming vector for the synchronization-signal reference-signal-received-power (SS-RSRP) signal. It is clear that the beam width of the SS signal is wider than that of the *CSI beam*.

The *CSI-RSRP* measured at the kth receive antenna for the ith *CSI beam* transmitted by the bth BS is defined as the average of the received signal power $\bar{g}_{b,i}^{[k]}=\mathbb{E}\{|\mathbf{e}_k^\mathrm{T}\mathbf{H}_b\mathbf{w}_i|^2\}$, where $\mathbf{e}_k$ is a vector of zeros except for the kth entry being 1, and the expectation $\mathbb{E}\{\cdot\}$ is taken over the small-scale fading. Similarly, the *SS-RSRP* measured at the kth receive antenna for the ith *CSI beam* transmitted by the bth BS is defined as $\tilde{g}_{b,i}^{[k]}=\mathbb{E}\{|\mathbf{e}_k^\mathrm{T}\mathbf{H}_b\tilde{\mathbf{w}}_i|^2\}$.

Let $\{\bar{g}_{b,(i)}^{[k]}\}$ be the *ordered* CSI-RSRP, such that $\bar{g}_{b,(1)}^{[k]}\geq\bar{g}_{b,(2)}^{[k]}\geq\cdots$. The ordered SS-RSRP $\{\tilde{g}_{b,(i)}^{[k]}\}$ is defined in the same way. However, due to practical constraints, only the CSI-RSRP for the serving cell $b = 0$ is available in our system, and only the $N = 8$ strongest CSI-RSRP $\bar{g}_{b,(1)}^{[k]}, \bar{g}_{b,(2)}^{[k]}, \ldots, \bar{g}_{b,(N)}^{[k]}$ over a time frame of hundreds of million seconds are recorded. For SS-RSRP from the neighbor cells, only the strongest one $\tilde{g}_{b,(1)}^{[k]}$ is recoded. Furthermore, only two receive antennas $k = 1, 2$ record the RSRP.

C. Data Collection

With the goal of developing a deep learning model to predict SE from RSRP, we collect a massive amount of auxiliary 5G NR data to assist the training for the model. In LTE, the UE measurements are performed for RSRP and SNR associated with CRS (Cell-Specific Reference Signal). 5G NR uses SS (Synchronization Signal) and CSI (Channel State Information) instead of CRS. Without the measurement of primary cell (PC) beam antenna channel-state-information reference-signal-received-power (CSI-RSRP) (dBm), which is mentioned in Section II-B, we also measured date and time of collection, geographic location of collection, rank indicator (RI), downlink MAC count, downlink initial block error rate (BLER), physical downlink shared channel (PDSCH) resource block (RB) number per slot, downlink MAC throughput (Mbit/s), average CQI, downlink average modulation and coding scheme (MCS), Neighbored cell (NC) SS-RSRP in detected cells (dBm).

The data collection campaign was conducted in both Chengdu and Shenzhen, two major cities in China. Three datasets are formed. Dataset 1: 86.88 hours 142.3 km driving data from April to October 2020 in Chengdu; Dataset 2: 18.95 hours 26.5 km driving data in December 2020 in Chengdu, with different network parameters configuration; Dataset 3: 12.08 hours 18.5 km driving data from August to September 2020 in Shenzhen. Because the parameters of the Massive MIMO antenna are changed after December 8, 2020, dataset 2 is further divided into a validation dataset and a parameter-changed test dataset with December 8 as the segmentation boundary. We use Dataset 1 to be the training dataset and dataset 3 to be the area-changed test dataset. Data is collected in a non-uniform time sequence, with 10 to 60 samples per second.

III. Design of a Preliminary NN Model

A. Data preprocessing

Recall that the DT user for measurement data collection is a vehicle moving at variable speed, with occasional stops. As a result, the original dataset is noisy with a large amount of missing values. In addition, there is a significant amount of "outliers" due to transmitting small packets for hand-shaking from higher layers. Yet, the goal of this paper is not to predict the instantaneous SE, but an average one under the "best effort" transmission. Therefore, we perform outlier removal and normalization as follows.

- **Outlier removal:** In statistics, an outlier is a data point that differs significantly from other observations. An outlier may be due to variability in the measurement, and it can cause serious problems in statistical analyses. We wish to discard them according to the cumulative distribution function (CDF) in several features, including CSI-RSRP, SS-RSRP, MAC throughput, and initial BLER. Their CDFs are shown in Fig 2. The main value range of CSI-RSRP is -100dBm to -60dBm, the main value range of SS-RSRP is -110dBm to -70dBm, the main value range of MAC throughput is 200Mbits/s to 1400Mbits/s, and the main value range of initial transmission BLER is 1% to 13%. The proportion of values greater than -60dBm in CSI-RSRP is about 3.084%, the proportion of values less than 5% in initial BLER is about 4.267%, the proportion of values less than 100Mbits/s in MAC throughput is about 10.479%, and the proportion of values greater than -70dBm in SS-RSRP is about 2.041%. We set the threshold according to these proportions and removed 0.378% data sample finally.

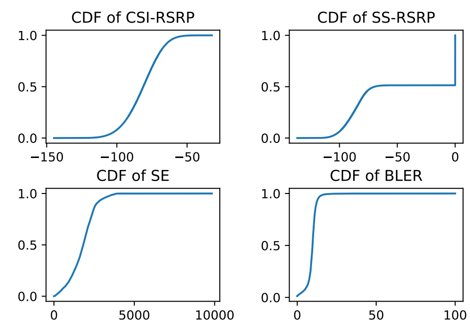

Figure 2: The CDF of the feature CSI-RSRP, SS-RSRP, MAC, and BLER

- **Normalization:** Denote $\mathbf{X}$ as the data matrix with columns representing the features and rows representing the samples. Let x_{ij} be the jth feature of the ith sample, $\mathbf{x}_{i\cdot}$ be the ith sample, and $\mathbf{x}_{\cdot j}$ be the jth feature vector. We normalize each feature to a number in $[0, 1]$. It is normalized as $x'_{ij} = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})}$, where $\min_j(x_{ij})$ is the minimum value of the jth feature, $\max_j(x_{ij})$ is the maximum value of the jth feature.

B. Location-based data augmentation

Conventionally, missing values can be filled using linear interpolation over time. Our measurement campaign also revealed that user performance in the cellular network has significant correlation also with the location, since the signal propagation strongly depends on the local environment. Recall that the vehicle moves in a variable speed in practice. Therefore, a better data completion scheme is to interpolate and filter the data according to both time and spatial adjacency.

We design a 1D filter to interpolate the data and smooth the features. Let l_i be the cumulated travel distance of the DT user when data sample i is collected. Note that the time adjacency is also captured in the sequence l_i, since the vehicle speed is bounded. Given the ith data sample, the kth sample's index which satisfy $|l_i - l_k| \leq W$ creat the neighborhood index set

Ω_i. let $w_{i,k} = e^{-\frac{|l_i - l_k|}{W/2}}$ be the weight associated with the observed data sample index in the neighborhood index set Ω_i of the ith data sample, where $w_{i,k}$ captures the correlation both in space and time, the further it travels, the less weight. The parameter W decides how much sample data we use in the set Ω_i. The estimated value of the $i$the sample jth feature x_{ij} is given by

$$x'_{ij} = \frac{\sum_{k \in \Omega_i} w_{i,k} x_{kj}}{\sum_{k \in \Omega_i} w_{i,k}}$$

The choice of the kernel bandwidth W is crucial for our experimental result. With the increasing of the value of W, the mean absolute percentage error (MAPE) increase first and then decrease. **There will be a supplement later about the optimization of W.**

The location-based data augmentation method is only applied to the training dataset and the validation dataset. We don't apply the smooth step to the test dataset, and we use the nearest value to interpolate it. Let x_{ij} be the jth feature of the ith sample. For the jth feature, the nearest ith sample non-missing value $x_{i,j}$ in time from the current kth sample missing value $x_{k,j}$ is used to interpolate the current missing sample value, that is $x_{k,j} = x_{i,j}$.

C. Feature selection

In practical applications, we can only measure the RSRP data. As a result, our goal becomes to use Beam-Antenna CSI-RSRP (dBm) and SS-RSRP in Detected Cells(dBm) data to predict SE. Downlink MAC throughput and PDSCH RB number are used to calculate the SE according to analysis in section II.

IPN is a factor that cannot be ignored when we analyze the communication problem. Analysis revealed RI, MCS, CSI RSRP, and SS-RSRP have the strongest correlation with IPN. So, we will use them to build the mathematical model of IPN.

CQI, RI, and MCS usually represent the interference strength. The correlation between SE and them is strong which is shown in Fig3. We use Pearson's rank correlation coefficient to measure the correlation between them. Intuitively, the Pearson correlation between two variables will be high when observations have a similar rank between the two variables, and low when observations have a dissimilar rank between the two variables. Pearson's correlation assesses linear relationships. It is the ratio between the covariance of two variables and the product of their standard deviations; thus it is essentially a normalised measurement of the covariance, such that the result always has a value between -1 and 1. $\mathbf{x}_{\cdot j}$ is the jth feature vector and $\mathbf{x}_{\cdot m}$ is the mth feature vector. The formula for ρ of the mth feature and the jth feature is:

$$\rho_{\mathbf{x}_{\cdot m}, \mathbf{x}_{\cdot j}} = \frac{\text{cov}(\mathbf{x}_{\cdot m}, \mathbf{x}_{\cdot j})}{\sigma_{\mathbf{x}_{\cdot m}} \sigma_{\mathbf{x}_{\cdot j}}}$$

where $\text{cov}(\mathbf{x}_{\cdot m}, \mathbf{x}_{\cdot j})$ is the covariance of the variables $\mathbf{x}_{\cdot m}$ and $\mathbf{x}_{\cdot j}$, $\sigma_{\mathbf{x}_{\cdot m}}$ and $\sigma_{\mathbf{x}_{\cdot j}}$ are the standard deviations of the variables $\mathbf{x}_{\cdot m}$ and $\mathbf{x}_{\cdot j}$.

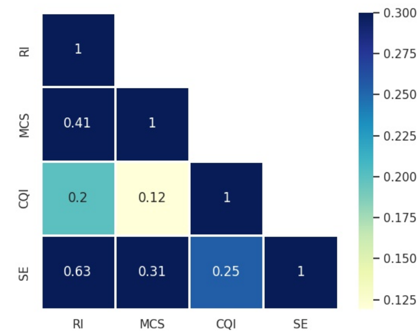

Figure 3: Pearson's rank correlation coefficient

The correlation of mth feature and the jth feature is none if $\rho_{\mathbf{x}_{\cdot m}, \mathbf{x}_{\cdot j}} \in [0.0, 0.09]$, weak if $\rho_{\mathbf{x}_{\cdot m}, \mathbf{x}_{\cdot j}} \in 0.1, 0.3$, medium if $\rho_{\mathbf{x}_{\cdot m}, \mathbf{x}_{\cdot j}} \in [0.3, 0.5]$, and high if $\rho_{\mathbf{x}_{\cdot m}, \mathbf{x}_{\cdot j}} \in [0.5, 1.0]$. As shown in the Fig 3, RI, MCS, and CQI are relevant to SE, MCS and RI are relevant.

D. SE prediction based on RSRP(Model A)

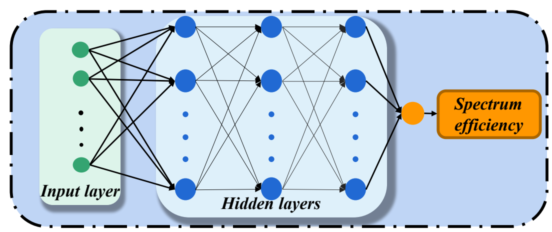

Figure 4: SE prediction based on RSRP

In this paper, the SE of a digital communication system is measured in $Mbit/block$. The calculation formula of SE is

$$SE = \frac{d}{p}$$

where d is downlink MAC throughput, which is measured in $Mbit/s/RB$, and p is PDSCH RB number per slot.

We use the CSI-RSRP from primary cell and SS-RSRP from neighbored cell as the input to predict SE directly. The hidden layers include three FC(fully connect) layers which is used to fit the non-linear correlation between SE and RSRP. The activation function of the three layers is rectified linear unit (ReLU) function, and the output activation function is the linear function. We use Adam optimizer.

The initial learning rate is set to 0.001, and when the loss function of 10 cycles does not drop by more than 0.0001, the learning rate is adjusted to 0.9 times the previous one. In addition, we have added a cooling time parameter that adjusts the learning rate no less than 10 cycles. The initial value of the weight determines the starting point of model training. Good initialization can speed up the training process and prevent the model from converging to an unreasonable local minimum. To avoid weights changing in the same direction during training, we generally do not re-initialize ownership to the same value, such as a full 0 matrix or a full 1 matrix. Weight initialization in this network uses Glorot Uniform, the weights are subject to the uniform distribution of $[-\sqrt{\frac{6}{m+n}}, \sqrt{\frac{6}{m+n}}]$, where n is the dimension of the input layer and m is the dimension of the output layer.

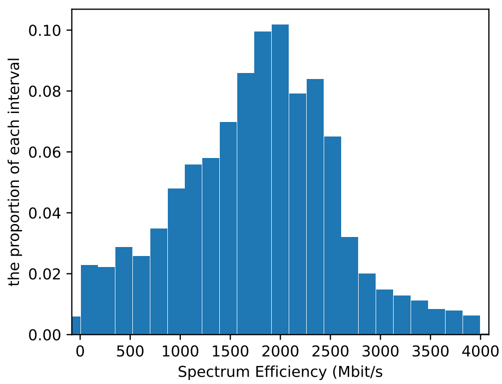

Figure 5: The PDF of SE

The spectrum efficiency range of the training dataset is divided into 1000 intervals, and the proportion of spectral effect labels within each interval is counted as probability density. Denote the probability density of the SE value y_i is $f(y_i)$. The MAPE function is used to calculate the loss of the multi-layer perceptron (MLP) neural network.

$$Loss_{Model\,A} = \frac{1}{n}\sum_{i=1}^{n} w_i |\frac{\widetilde{y_i} - y_i}{y_i}|$$

where the weight $\omega_i = \frac{1}{f(y_i)}$, n is the batch size, $\widetilde{y}$ is the predicted SE, y_i is the true SE.

IV. Interference Model Aided Prediction

A. Modeling the interference and noise

The signal-to-interference-and-noise ratio (SINR) is commonly used in wireless communication as a way to measure the quality of wireless connections. Typically, the energy of a signal fades with distance, which is referred to as a path loss in wireless networks. In a wireless network one has to take other factors into account (e.g. the background noise, interfering strength of other simultaneous transmission). The concept of SINR attempts to create a representation of this aspect. the SINR is defined as the power of a certain signal of interest divided by the sum of the interference power (from all the other interfering signals) and the power of some background noise.

$$SINR = \frac{P}{I+N}$$

where P is the power of the incoming signal of interest, I is the interference power of the other (interfering) signals in the network, and N is some noise term, which may be a constant or random. Like other ratios in electronic engineering and related fields, the SINR is expressed in dB.

The interference analysis is important when we solve the communication problem. Here we denote interference and noise as IPN. We first build the mathmatical model of P according to the CSI-RSRP from primary cell.

$$P_{r_{mW}} = \frac{1}{RI^2}\sum_{i=1}^{RI} max\left[Primary\,cell\,CSI\;RSRP_{mW}\,(i)\right]$$

where RI is the rank indicator we measure on the user equipment (UE) side. Depending on instruction from the network, UE may periodically or aperiodically measure RI and report it to Network. RI is an indicator showing how well multiple Antenna work. We usually say "Each of the multiple antenna (e.g, each antenna in MIMO configuration) works well if the signal from each antenna has NO correlation to each other". "No correlation" implies "no interference to each other". Maxmum RI value is very closely related to the number of Antenna. Maximum RI is same as number of antenna on each side if the number of Tx antenna and Rx antenna is same. If the number of Tx and Rx are different, the one with less antenna is the same as Max achievable RI. Max RI means "No Correlation between the antenna", "No interference to each other", "Best Performance". For example, in case of 2x2 MIMO, the RI value can be 1 or 2. When the value 2 in this case means "No Correlation between the antenna", "No interference to each other", "Best Performance". If the value is 1, it implies that the signal from the two Tx antenna is percieved by UE to be like single signal from single Antenna, which means the worst performance. Here we only add up the strongest RI beam values. Then we multiply $\frac{1}{RI}$ to calculate the average power of the strongest RI beam values and multiply $\frac{1}{RI}$ again to calculate the total power evenly allocated to each beam. Then we can calculate SINR according to MCS.

$$SINR_{dB} = m(MCS)$$

where m is the function between MCS and SINR. For any communication technology, Modulation and Coding Scheme (MCS) defines the numbers of useful bits which can carried by one symbol. In contrast with 5G or 4G, a symbol is defined as Resource Element (RE) and MCS defined as how many useful bits can be transmitted per Resource Element (RE) . MCS depends on radio signal quality in wireless link, better quality the higher MCS and the more useful bits can be transmitted with in a symbol and bad signal quality result in lower MCS means less useful data can be transmitted with in a symbol. Modulation defines how many bits can be carried by a single RE irrespective of whether it's useful bit or parity bits. 5G NR supports QPSK,16 QAM, 64 QAM and 256 QAM modulation for PDSCH. Code rate can be defined as the ratio between useful bit and total transmitted bit (Useful + Redundant Bits). These Redundant bits are added for Forward Error Correction (FEC). In other words we can it is the ratio between the number of information bits at the top of the Physical layer and the number of bits which are mapped to PDSCH at the bottom of the Physical layer. We can also say, it a measure of the redundancy which is added by the Physical layer. A low coding rate corresponds to increased redundancy. There are about 32 MCS Indexes (0-31) are defined and MCS Index 29,30 and 31 are reserved and used for re-transmission. 3GPP Specification 38.214 has given three tables for PDSCH MCS namely 64 QAM Table, 256 QAM Table and Low Spectral Efficiency 64 QAM Table. The selection of the table is according to the application scenario. As noted earlier, we can get the MCS index from 0 to 28 and its corresponding SINR value. In practice, we can measure the MCS float value and calculate the corresponding SINR according to the funciton m. How to get the function according to the measurement table is our problem. Consider the bivariate data $(X_1, Y_1), \cdots, (X_n, Y_n)$,which form an independent and identically distributed sample from a population (X, Y). Here X represent MCS and Y represent SINR. Of interest is to estimate the regression function $m(x_0) = E(Y = x_0)$ and its derivatives $m'(x_0), m''(x_0), \cdots, m^{(p)}(x_0)$. To help us understand the estimation methodology, we can regard the data as being generated from the model

$$Y = m(X) + \sigma(X)\varepsilon,$$

where $E(\varepsilon) = 0$, $Var(\varepsilon) = 1$, and X and ε are independent. Suppose that the $(p+1)^{th}$ derivative of $m(x)$ at the point x_0 exits. We then approximate the unknown regression function $m(x)$ lopcally by a polynomial of order p. A Taylor expansion gives, for x in a neighborhood of x_0,

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2$$

$$+ \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^2.$$

This polynormial is fitted locally by a weighted least squares regression problem:

$$min \sum_{i=1}^{n} \{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\}^2 K_h(X_i - x_0),$$

where h is a bandwidth controlling the size of the local neighborhood, and $K_h(\cdot) = K_h(\cdot/h)/h$ with K a kernel function assigning wights to each datum point. K is a symmetric probability density function with bounded support.

Denote by $\hat{\beta}_j$, $j = 0, \cdots, p$, the solution to the least squares problem. It is clear from Taylor expansion in ss that $\widehat{m_v(x_0)} = v!\hat{\beta}_v$ is an sitimator for $m^{(v)}(x_0)$, $v = 0, 1, \cdots, p$. To estimate the entire function $m^{(v)}(\cdot)$ we solve the above weighted least squares problem for all points x_0 in the domain of interest. It is more convenient to work with matrix notaion. Denote:

$$X = \begin{bmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{bmatrix},$$

$$y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}.$$

Futher, let W be the $n \times n$ diagonal matrix of weights:

$$W = diag\{K_h(X_i - x_0)\}.$$

Then the weighted least squares problem can be written as

$$\min_{\beta}(y - X\beta)^T W (y - X\beta),$$

with $\beta = (\beta_0, \cdots, \beta_p)^T$. The solution vector is provided by weighted least squares theory ans is given by

$$\hat{\beta} = (X^T W X)^{-1} X^T W y.$$

There are several important issues wich have to be discussed. First of all there is the choice of the bandwidth parameter h, which plays a rather crucial role. A large bandwidth under-parametrizes the regression function, cause a large modelling bias, while a too small bandwidth over-parametrizes the unknown funciton and results in noisy estimates. **There will be a supplement about the choose of h.**

The second issue in local polynomial fitting is the choice of the order p of the local polynormial. Since the modeling bias is primarily controlled by the bandwidth, this issue is less crucial however. For a given bandwidth h, a large value of p would expectedly reduce the modelling bias, but would cause a large variance and a considerable computational cost. There is a general pattern of increasing variability: for estimating $m^{(v)}(x_0)$, there is no increase in variability when passing from an even $p = v + 2q$ order fit to an odd $p = v + 2q + 1$ order fit, but when passing from an odd $p = v + 2q + 1$ order fit to the consecutive even $p = v + 2q + 2$ order fit there is a price to be paid in terms of increased variability. Therefore, even order fits $p = v + 2q$ are not recommended. Since the bandwidth is used to control the modelling complexity, we decide to use the lowest odd order $p = v + 3$.**There will be a supplement about the choose of p.**

The last issue concerns the choice of the kernel functino K. SInce the estimate is based on the local regression no negative weight K should be used. For all choices of p and v the optimal weight function is $K(z) = \frac{3}{4}(1 - z^2)_+$, the Epanechnikov kernel, which minimizes the asymptotic MSE of the resulting local polynomial estimators. Local polynomial fitting is nearly optimal in an asymptotic minimax sense.

After get the regression function $m(\cdot)$, we can use the ratio relationship to calculate IPN and the mathmatical model of IPN is

$$IPN_{dBm} = P_{r_{dBm}} - m(s)$$

where s is the MCS value collected in real time.

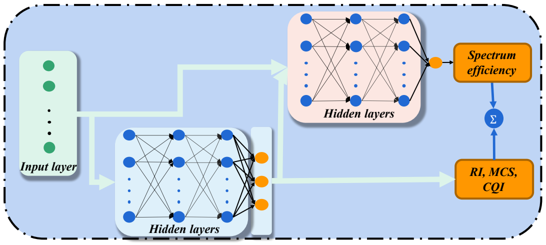

Figure 6: MCS and RI aided SE prediction

B. MCS ,CQI and RI aided SE prediction(Model B)

As shown in the Fig3, the correlation between MCS ,CQI , RI and SE is high. We designed the MCS ,CQI and RI aided SE prediction model to get better prediction result.

C. Joint interference and SE prediction(Model C)

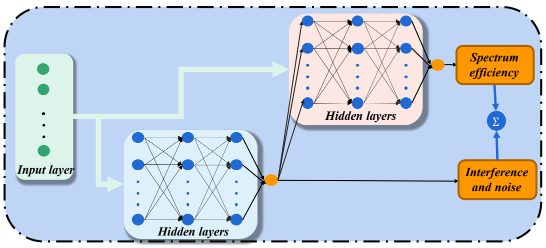

Figure 7: Joint interference and SE prediction

SE is closely related to IPN If we can model IPN and input the interference information into the SE prediction network, the accuracy of SE prediction is very likely to be improved. We design joint interference and SE prediction neural network which is called Model B. We use three FC layers to fit the non-linear connection between the RSRP signal and IPN. We still use three FC layers to fit the non-linear connection between the RSRP signal and SE, but we input the predicted IPN to the SE prediction network. The IPN can give the information of the interference and noise to the SE prediction network. The activation function of the three FC layers is the ReLU function, and the output activation function is the linear function. The loss function is shown as followed:

$$Loss_{Model\ C} = \frac{1}{n}(\alpha \sum_{i=1}^{n} w_i |\frac{\widetilde{y_i} - y_i}{y_i}| + (1-\alpha) \sum_{i=1}^{n} |\frac{\widetilde{g_i} - g_i}{g_i}|)$$

where α is the proportional control factor.$0 \leq \alpha \leq 0.5$ means that the loss of IPN is more important than the loss of SE. $0.5 \leq \alpha \leq 1$ means the opposite.

V. Prediction Enhancement for Sections with Sparse Training Data

A. Design of SE classification network

Because of the low performance of the low SE, we want to decrease the prediction error of SE in the low SE section. We find that, if we tell the SE prediction neural network that the input sample is low SE and fit it particularly, the SE prediction

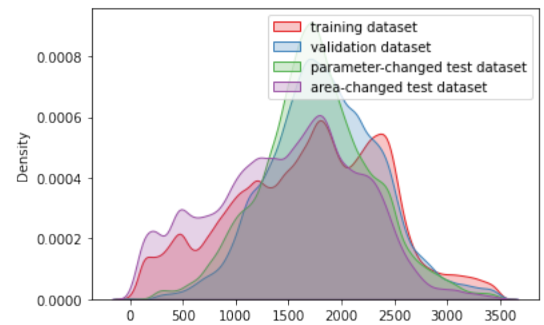

Figure 8: The distribution of SE

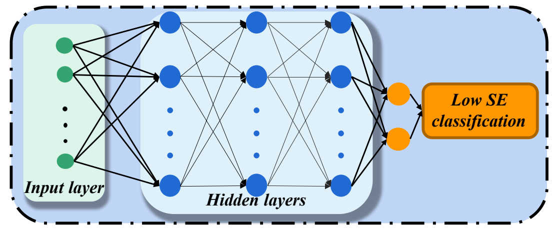

Figure 9: SE classification network

neural network will use this information to fit the sample in the low SE section better. So here we design a three FC layers network to achieve the classification goal. The low SE classification result is also connected with IPN, so we input the predicted IPN to the SE classification network. The activation function of the FC layers is the ReLU function, and the output activation function of SE is the softmax function. The cross-entropy loss function is calculated between the predicted category probability output and the one HOT form of the real category. The output activation function of IPN is still the linear function.

B. Classification assisted SE prediction(Model D)

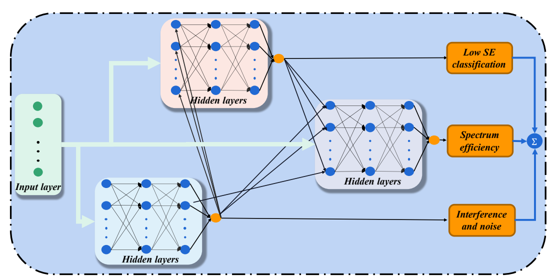

Figure 10: Classification assisted SE prediction

We can use the classification information as a new input of Model B. The total loss function is shown as follows:

$$Loss_{Model\ D} = \frac{1}{n}(\alpha \sum_{i=1}^{n} w_i |\frac{\widetilde{y_i} - y_i}{y_i}| + (1-\alpha) \sum_{i=1}^{n} |\frac{\widetilde{g_i} - g_i}{g_i}|)$$

$$-\frac{1}{n} \sum_{i=1}^{n} [f_i \times log(p_i) + (1-f_i) \times log(1-p_i)]$$

where f_i is the true label of the low SE classification, if the sample is in the low SE section, $f_i = 1$, otherwise $f_i = 0$. p_i comes from the output of the SE classification network. $0.5 < p_i \leq 1$ means the sample is predicted to be the low SE.

VI. Experiment Results

The four models are applied to the two datasets. We use the data collected in Chengdu to train and validate the model. Then we use the test dataset one which is collected after the Massive MIMO antennas are adjusted in Chengdu to test the trained model. The performance of the test dataset one can evaluate the robustness performance of the trained model when we change the configuration of the antenna. we also use test dataset two which is collected in Shenzhen to test the trained model. The performance of the test dataset two can evaluate the portability performance of the trained model when we use the trained model in different areas. The prediction error of Model B(0.149) is low than the prediction error of Model A(0.170). It proves that the error of spectrum efficiency prediction can be reduced by using IPN as part of the input of the spectrum efficiency prediction network. The low SE prediction error of Model A and Model B is high compared with other SE sections. Although the test dataset prediction error of the proposed Model C(0.193) is larger than Model B(0.149), the low SE prediction error(0.28) is lower than Model B's(0.74). The IPN prediction performance of the three models is roughly the same. The performance of test dataset two is worse than test dataset one, which means the portability performance of our model can be improved again in the future. We also calculate the standard deviation of the absolute percentage error (SDAPE) and standard deviation of the absolute error (SDAE) of the predicted result.

VII. Conclusion

In this paper, we adopted a model-assisted data-driven approach to build a DNN model to predict the spectrum efficiency in a 5G NR network based on CSI statistics, including CSI-RSRP, average CQI, and average RI. The challenge to address is that the spectrum efficiency depends on a lot of hidden factors that are not directly captured by the CSI data available. To circumvent the difficulty, we first developed a location-weighted averaging approach to de-noise the measurement data, and then designed a DNN architecture that is trained to jointly predict the interference with noise and the spectrum efficiency. It is found that, by first predicting the interference, the ultimate prediction accuracy on spectrum efficiency can be improved. This is verified by the massive amount of drive testing data collected in a real world 5G NR network at Chengdu, China, where the training and testing data are respectively collected at different periods with different network parameter configurations. An overall prediction error of 15% on the spectrum efficiency is achieved. [1]

In this work, we have studied the cost-effective design for FL. We analyzed how to optimally choose the number of participating clients (K) and the number of local iterations (E), which are two essential control variables in FL, to minimize the total cost while ensuring convergence. We proposed a sampling-based control algorithm which efficiently solves the optimization problem with marginal overhead. We also derived insightful solution properties which helps identify the design principles for different optimization goals, e.g., reducing learning time or saving energy. Extensive experimentation results validated our theoretical analysis and demonstrated the effectiveness and efficiency of our control algorithm. Our optimization design is orthogonal to most works on resource allocation for FL systems, and can be used together with those techniques to further reduce the cost.

References

[1] M. B. Afuosi and M. R. Zoghi, "Indoor positioning based on improved weighted KNN for energy management in smart buildings," *Energy and Buildings*, vol. 212, p. 109754, 2020.

Table I: The prediction result of the three proposed models

Output	Dataset	Error	Model A	Model B	Model C	Model D
SE	**training**	**MAPE(%)**	29.1		29.4	30.5
	validation	**MAPE(%)**	19.8		19.8	19.9
	parameter-changed test	**MAPE(%)**	17.0		14.9	19.3
		SDAPE(%)	15.0		16.7	16.3
		MAPE by sections(%)	[71.0 18.0 12.0 13.0]		[74.0 22.0 13.0 14.0]	[28.0 25.0 17.0 16.0]
	area-changed test	**MAPE(%)**	30.7		28.8	31.1
		SDAPE(%)	20.9		21.4	21.0
		MAPE by sections(%)	[72.0 24.0 28.0 32.0]		[78.0 26.0 20.0 19.0]	[32.0 31.0 36.0 38.0]
IPN	**training**	**MAE(dBm)**	3.8151		3.774	3.642
	validation	**MAE(dBm)**	3.501		3.431	2.547
	parameter-changed test	**MAE(dBm)**	3.541		3.56	3.517
		SDAE(dBm)	2.886		2.845	2.732
	area-changed test	**MAE(dBm)**	4.106		3.93	3.993
		SDAE(dBm)	3.182		3.049	3.044