AI+ Wireless Competition Task and Transformer-Based Competition Solution

Competition Introduction

The cloud-native network architecture, with its advantages of robustness, agility, low cost, predictability, and ease of management, has become a trend in the future development of wireless communication networks. Open and cloud-based wireless communication networks are moving away from traditional tightly coupled models. Implementing decoupling at the base station end (such as uplink-downlink decoupling and control plane-data plane decoupling) can further enhance the flexibility of future networks and leverage the prominent advantages of cloud-native networks in flexible resource allocation and improving spectrum efficiency. Feedback-free communication systems, through feedback of auxiliary information (such as geographic coordinates) at the application layer, can reduce reliance on physical layer channel measurements, thereby reducing air interface resource overhead.

In recent years, Artificial Intelligence (AI) technology has shown tremendous potential in the field of wireless communication, and the use of AI algorithms to reduce communication air interface overhead has gradually become a consensus in academia and industry. In feedback-free communication, AI technology can help communication systems discover hidden patterns between auxiliary information and physical channels within a large amount of data, thus achieving feedback-free communication. However, how to optimally utilize auxiliary information to assist communication is currently a technological focus and challenge.
https://www.datafountain.cn/competitions/645

Challenge Objectives

The competition will provide tuples of <geographic coordinates, wireless channel response> as data, as illustrated in Figure 1. The competition encourages participants to adopt a data-driven approach and use deep learning methods to design a downlink transmission scheme suitable for feedback-free communication systems.

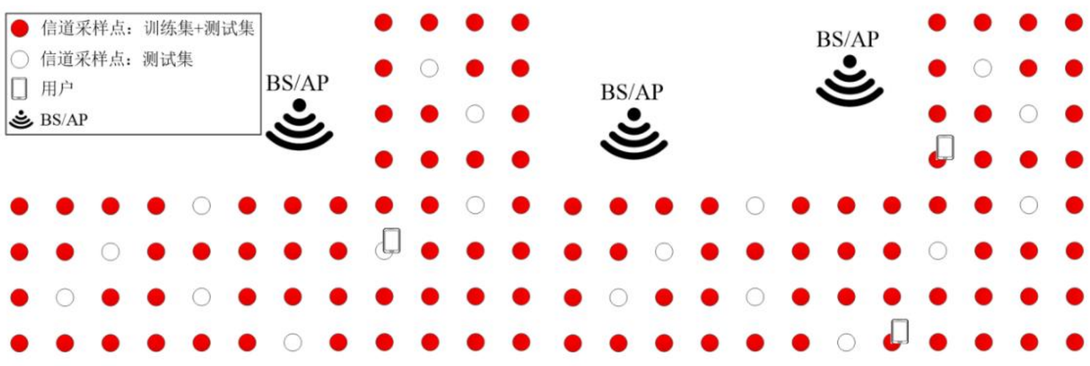


Figure 1

The scenario for the semi-finals is a "dual base station single user" setup, where two base stations transmit the same information to the user using a single-stream approach. Participants are required to design a neural network based on the user's geographical location coordinates and wireless communication knowledge. Using the designed neural network, they should obtain the transmit precoding vectors for the two base stations and perform beamforming at the base stations.

Data Overview

The channel data for the competition is generated through professional channel detectors at a center frequency of 5.4GHz. The transmitter uses a 4×4 dual-polarized planar antenna array with a height of 2.8m, while the receiver employs a 4×1 dual-polarized linear antenna array with a height of 1.38m. Both ends of the antennas consist of ±45° dual-polarized microstrip antennas, each with a specific radiation direction and beamwidth, arranged uniformly with fixed spacing of half a wavelength. The layout of the antennas is as shown in the figure below. The measured channel impulse response is transformed into the channel frequency response of 120 subcarriers, with a subcarrier spacing of 60KHz.

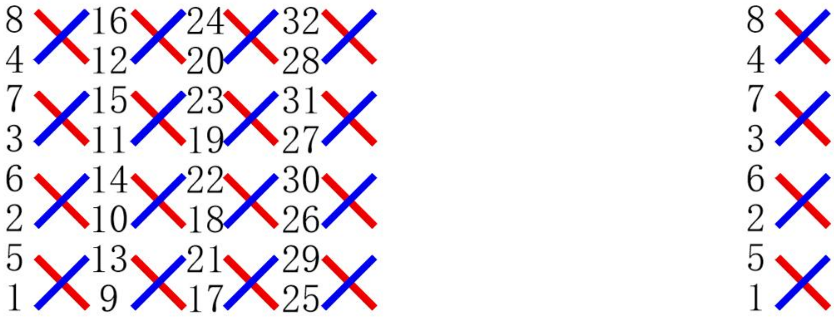


Figure 2: Antenna Numbering at the Transmitter and Receiver

For the semi-finals, two MAT files are provided, one containing data for <user geographical coordinates, wireless channel responses from Base Station 1 to the user>, and the other for <user geographical coordinates, wireless channel responses from Base Station 2 to the user>. The provided geographical coordinates in the dataset are in three-dimensional Cartesian coordinates, and the format of the wireless channel response is (Nr, Nt, K, 2), where Nr=8 represents the number of receiving (user) antennas, Nt=32 represents the number of transmitting (base station) antennas, and K=120 represents the number of subcarriers. The last dimension represents the real and imaginary parts of the channel response, respectively.

Additionally, the position of Base Station 1 is (5.7, 9.2, 2.8), and the position of Base Station 2 is (22.21, 9.2, 2.8).

Evaluation Criteria

The competition score is determined by the spectral efficiency of the participants' transmission scheme, calculated as follows:

$$R = \frac{1}{KL} \sum_{l=1}^{L} \sum_{k=1}^{K} \log_2 \left(1 + \frac{\left\| H_{l,k}^{(1)} f_{l,k}^{(1)} + H_{l,k}^{(2)} f_{l,k}^{(2)} \right\|_2^2}{\sigma_n^2} \right)$$

Network Approach Overview

In the semi-final stage, the scenario used is "dual base station single user," where two base stations transmit the same information to the user using a single-stream approach. Consequently, in the semi-finals, two models are required for prediction. To fully utilize the fitting capacity of neural networks and use the official evaluation metric as the loss function, we train both models as subnetworks simultaneously. Similar to the preliminary round, we directly output the precoding vectors. In the transformer algorithm we employ, we expand the input by increasing the dimensions, with each of the 120 subcarriers treated as a token in the transformer model.

We adopted a combined indexing approach, merging the respective 2400 data entries from data_train_1.mat and data_train_2.mat to create a dataset of 45,000 entries with columns named ["loc", "CSI_1", "CSI_2"] for joint model training. Compared to training two models separately using data_train_1 and data_train_2, this joint training approach offers certain advantages. The transformer serves as the fundamental network framework, with a 3D input comprising user geographical coordinates. Due to the periodicity observed between subcarriers, which have a 60KHz spacing, the user's 3D location is duplicated 120 times to create a 120x3 input for the transformer module. This approach takes advantage of the inherent periodicity in the data and allows the transformer model to process the information effectively.

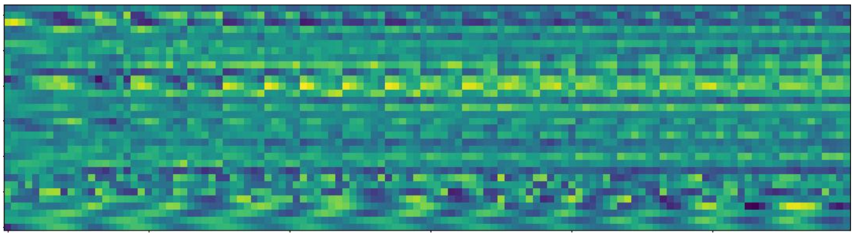


Figure 3: Sample Slicing

In the transformer module model, the conventional Feed-Forward Network (FFN) module typically uses two dense layers. However, we replace the dense layers with a Conv1D layer with kernel size 1. Due to the differences in initialization and computation precision in the NPU (Neural Processing Unit), the results obtained with Conv1D are slightly better than those with dense layers.

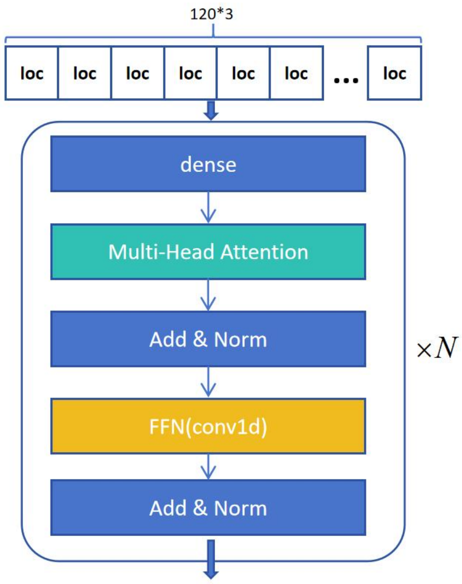


Figure 4: Replicated Geographical Locations & Network

Training Method

During the model training process, the learning rate is not constant. When the learning rate is higher, the model converges faster, while a lower learning rate leads to slower convergence. As the number of epochs increases, a stepwise learning rate reduction can help prevent overfitting and underfitting. In practical model training, batch training is used, with the model being saved after each training iteration. The learning rate is adjusted, and training continues until the model's performance reaches its optimum. The learning rate changes as shown in the figure. Using a stepwise learning rate, for the first 100 epochs, the learning rate is set to 1e-4, from epoch 100 to 107, the learning rate is 1e-5, and finally, a learning rate of 7e-7 is used.

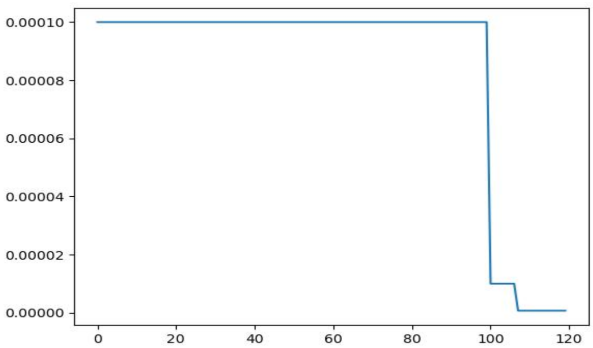


Figure 5: Learning rate

To write the loss function for spectral efficiency using MindSpore, due to certain functions in MindSpore 1.10 not supporting complex number calculations on NPU, in the code for computing spectral efficiency, real and imaginary parts are computed separately, and then combined to calculate the norm.

$$(a+bi)\times(c+di)=(ac-bd)+(ad+bc)i$$

Training Efficiency: When training with MindSpore on NPU, the efficiency is initially low, taking 100 minutes to complete one epoch. However, after implementing mixed-precision training using

Automatic Mixed Precision (AMP), the training time is reduced to 7 to 8 minutes per epoch. Using mixed-precision training can increase the training speed by more than tenfold.

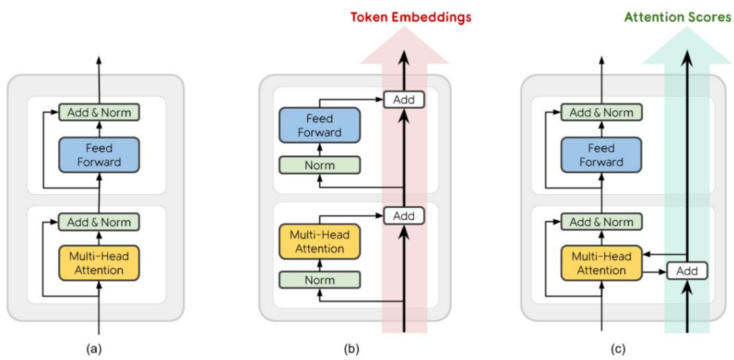
Conclusions

In the semi-finals, we continued to utilize a transformer-based algorithm to address the challenge of low input dimensionality and high output dimensionality, building upon the approach used in the preliminary round. Through deep learning models, our aim was to capture the intricate relationship between input and output, enabling us to discover the correspondence between coordinates and channels.

During the model training and debugging phase, we strictly adhered to the expectations of the organizers, using the MindSpore framework throughout the entire process. We made substantial efforts in various aspects, including feature extraction, model selection, model optimization, hyperparameter tuning, loss function design, and training strategy design. This experience allowed us to accumulate knowledge and expertise in using MindSpore and serves as a foundation for future research and work in the field of AI.

Other Tricks

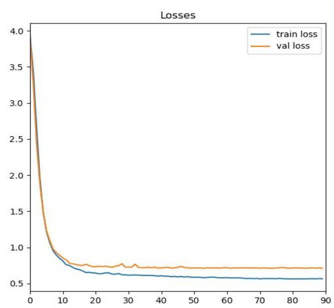
(1)RealFormer: it uses a Post-LN style Transformer as backbone and adds skip edges to connect Multi-Head Attention modules in adjacent layers. More formally, it adds Prev, the pre-softmax attention scores from the previous layer with shape.

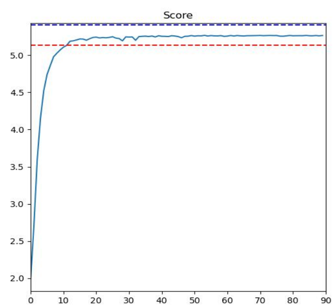


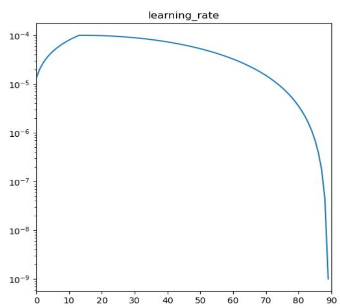
$$Attention(\boldsymbol{Q}_n, \boldsymbol{K}_n, \boldsymbol{V}_n) = softmax\left(\boldsymbol{A}_n\right)\boldsymbol{V}_n, \quad \boldsymbol{A}_n = \frac{\boldsymbol{Q}_n \boldsymbol{K}_n^{\top}}{\sqrt{d_k}}$$

$$Attention(\boldsymbol{Q}_n, \boldsymbol{K}_n, \boldsymbol{V}_n) = softmax\left(\boldsymbol{A}_n\right)\boldsymbol{V}_n, \quad \boldsymbol{A}_n = \frac{\boldsymbol{Q}_n \boldsymbol{K}_n^{\top}}{\sqrt{d_k}} + \boldsymbol{A}_{n-1}$$

(2) Function learning rate:







References

[1] Ruining He, Anirudh Ravula, Bhargav Kanagal, Joshua Ainslie RealFormer: Transformer Likes Residual Attention https://arxiv.org/abs/2012.11747

[2] Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. 2019. Generating long sequences with sparse transformers. preprint arXiv:1904.10509.