Ensemble-transfer-learning-based channel parameter prediction in asymmetric massive MIMO systems[*]

Zunwen HE[†1], Yue LI[1], Yan ZHANG[†‡1], Wancheng ZHANG[1], Kaien ZHANG[1],
Liu GUO[1], Haiming WANG[2]

[1]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
[2]State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China
[†]E-mail: hezunwen@bit.edu.cn; zhangy@bit.edu.cn

Abstract: Asymmetric massive multiple-input multiple-output (MIMO) systems have been proposed to reduce the burden of data processing and hardware cost in sixth-generation mobile networks (6G). However, in the asymmetric massive MIMO system, reciprocity between the uplink (UL) and downlink (DL) wireless channels is not valid. As a result, pilots are required to be sent by both the base station (BS) and user equipment (UE) to predict double-directional channels, which consumes more transmission and computational resources. In this paper we propose an ensemble-transfer-learning-based channel parameter prediction method for asymmetric massive MIMO systems. It can predict multiple DL channel parameters including path loss (PL), multipath number, delay spread (DS), and angular spread. Both the UL channel parameters and environment features are chosen to predict the DL parameters. Also, we propose a two-step feature selection algorithm based on the SHapley Additive exPlanations (SHAP) value and the minimum description length (MDL) criterion to reduce the computation complexity and negative impact on model accuracy caused by weakly correlated or uncorrelated features. In addition, the instance transfer method is introduced to support the prediction model in new propagation conditions, where it is difficult to collect enough training data in a short time. Simulation results show that the proposed method is more accurate than the back propagation neural network (BPNN) and the 3GPP TR 38.901 channel model. Additionally, the proposed instance-transfer-based method outperforms the method without transfer learning in predicting DL parameters when the beamwidth or the communication sector changes.

Key words: Asymmetric massive multiple-input multiple-output (MIMO) system; Channel model; Ensemble learning; Instance transfer; Parameter prediction

1 Introduction

To fulfill the requirements of high spectrum efficiency and high energy efficiency, massive multiple-input multiple-output (MIMO) arrays are widely used in fifth-generation mobile networks (5G). How-

ever, as the number of antenna elements increases continually to meet the increased demand in sixth-generation mobile networks (6G), the signal processing burden and hardware costs become higher and higher (Albreem et al., 2019, 2021; Radhakrishnan et al., 2021; Qiu et al., 2022). To solve this problem, asymmetric full-digital beamforming massive MIMO systems (Hong et al., 2020, 2021) have been put forward. Unlike conventional large-scale beamforming arrays, the asymmetric massive MIMO system employs a smaller number of receiving (Rx) radio frequency (RF) chains than transmitting (Tx)

chains, which can reduce the hardware cost and energy consumption. This asymmetric system with non-reciprocal Tx/Rx beam patterns has also been proven to realize wider coverage in the uplink (UL) and offer higher gain beams in the downlink (DL) (Hong et al., 2021). Thus, there is great potential for its application in future wireless communication networks.

However, an emerging problem in the asymmetric massive MIMO system is that the reciprocity between the UL and DL wireless channels is not valid. This means that even in a time division duplexing (TDD) system, it is impossible to directly use the UL channel estimation results to predict the DL channels. Thus, pilots are required by both base station (BS) and user equipment (UE), which occupy more transmission and computational resources (Wimalajeewa et al., 2017). This problem becomes severe in the DL because the number of Tx antenna elements on the BS side is very large.

Actually, the non-reciprocity between the UL and DL channels results from different Tx and Rx beam patterns; i.e., the UL/DL signals do not travel under the same propagation conditions. Analyzing the relationship between beam patterns and channel parameters can be useful in characterizing this non-reciprocity. Some research has been conducted to investigate the impacts of the beamwidth on channel parameters and can be divided into two categories. The first category is based on empirical models. In several studies (Wu et al., 2015; Kim et al., 2016; Li et al., 2017; Erden et al., 2020; Jiang et al., 2021), the channel parameter distributions of different Tx beamwidths were obtained by conducting signal acquisition experiments, to fit empirical models that can be used for channel parameter estimation. The estimation results can be quickly calculated by the empirical models. However, they could provide only statistical results and the accuracy was limited. In addition, they were applicable only to specific beamwidths and scenarios. The second category is to employ ray-tracing techniques, which are based on the principles of geometric optics (Tuan et al., 2016; Chen YS et al., 2019). The ray-tracing-based methods can provide accurate results, but may require a great amount of computation time.

Recently, methods based on artificial intelligence (AI) (Ye et al., 2018; Joo et al., 2019; Yang GS et al., 2019; Yang YW et al., 2019; Han et al., 2020; Lin B et al., 2021; Zhang S et al., 2021; Zhang SB et al., 2021) enabled fast and accurate channel statistical parameter prediction without relying on electromagnetic propagation maps. Although these works considered only a single link, i.e., UL or DL, they provided inspiration to solve the prediction problem in the non-reciprocal channels. In this paper, we propose a method based on ensemble learning and instance transfer to predict DL channel parameters, which are important for the link budget and adaptive transmission in asymmetric massive MIMO systems. Although the selected scenario is a typical urban one, our proposed method can also be applied in other outdoor scenarios, such as suburban or rural ones. It should be mentioned that the transfer learning module may be useful only when the new scenario is similar to the original one; e.g., they are different sectors in the same cell. The intrinsic physical mechanism is that although the Tx and Rx beam patterns are different, the BS and UE are located in the same area. Therefore, the UL and DL channels share a common part of the propagation environment. An AI-based model is built to characterize the relationship among the UL channels, DL channels, and propagation environments. Our work is the first one to apply ensemble transfer learning for channel parameter prediction in asymmetric massive MIMO systems. We investigate the correlation between the UL and DL propagation environments and design a two-stage feature selection method, which can reduce the negative impact of redundant features. An ensemble learning approach is used to combine multiple weak learners to improve the generalization of the channel parameter prediction model. In addition, to realize fast network deployment in new propagation conditions, we propose an instance transfer scheme that can use the knowledge within the original condition for the new model training. The proposed method is helpful for the design and optimization of asymmetric massive MIMO systems under non-reciprocal channel conditions.

In our proposed model, both the environment information and UL channel parameters are chosen as features to participate in the DL channel parameter prediction. However, too many training features or irrelevant features may increase the computational complexity of the model and even reduce the accuracy (Liu and Tang, 2014; Wang D et al., 2015; Lundberg and Lee, 2017; Yang GS et al., 2019; Wang ZG

et al., 2020). Thus, we design a two-step feature selection algorithm. Compared with classical feature selection methods which quantify the importance of multiple features and select the best number of features by ergodic search, our algorithm can automatically ascertain the number of most relevant features and therefore reduces the computation complexity of feature selection. Then, an ensemble-learning-based prediction model composed of multiple weakly supervised models is proposed to predict DL channel parameters.

A sufficient amount of training data are essential to ensure the prediction accuracy of AI-based methods. In practice, obtaining a large amount of training data under new propagation conditions is time-consuming and computationally intensive, thus limiting the generalization performance of AI-based methods. To settle this problem, we propose an instance transfer method to achieve prediction in new propagation conditions by applying the existing data from the original condition to the new model training.

Moreover, the aforementioned works concerning channel parameter prediction focus on path loss (PL) (Wu et al., 2015; Kim et al., 2016; Tuan et al., 2016; Li et al., 2017; Chen YS et al., 2019; Joo et al., 2019; Yang GS et al., 2019; Yang YW et al., 2019; Erden et al., 2020; Han et al., 2020) or delay spread (DS) (Kim et al., 2016; Li et al., 2017; Yang GS et al., 2019; Han et al., 2020). In this paper, we realize the prediction of many important parameters including PL, path number (PN), DS, azimuth angular spread of arrival (AASA), elevation angular spread of arrival (EASA), azimuth angular spread of departure (AASD), and elevation angular spread of departure (EASD). The prediction results are compared with the data generated by commercial ray-tracing software. It is verified that the proposed method can provide acceptable prediction accuracy with low complexity.

The major novelties and contributions of this paper are summarized as follows:

1. A prediction method based on ensemble learning and instance transfer is proposed to predict DL parameters in non-reciprocal DL channel parameters. The proposed method is able to predict multiple DL channel parameters including PL, PN, DS, and angular spreads. Simulation results show that this method can achieve rapid channel parameter prediction while ensuring their accuracy.

2. We propose a two-step feature selection algorithm that can determine a feature importance ranking and an optimistic feature combination. This algorithm reduces the computational complexity of feature selection and is beneficial for improving prediction accuracy.

3. The instance transfer method is introduced to assist the prediction model in new propagation conditions. By using this transfer-learning-based approach, it is possible to deploy our method in a new propagation condition where enough training data are difficult to collect within a short time.

2 System model

As illustrated in Fig. 1, we consider an asymmetric massive MIMO system in an urban scenario, which is composed of a BS and n UEs. The array on the BS side is equipped with non-reciprocal Tx/Rx beam patterns.
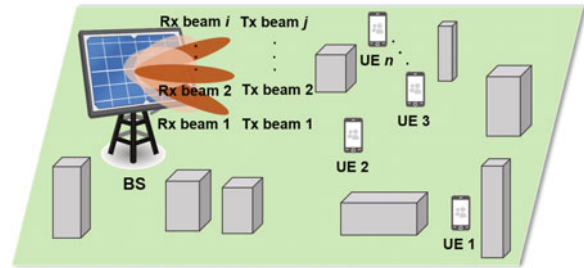
Fig. 1 A typical application scenario of the asymmetric massive MIMO system (BS: base station; MIMO: multiple-input multiple-output; UE: user equipment)

Consider the UL and DL between the BS and the ith UE, where $i = 1, 2, \ldots, n$. $\{\theta_{\text{UL}}^i\}$ and $\{\theta_{\text{DL}}^i\}$ represent the UL and DL channel parameter sets, respectively. Then, $\{\theta_{\text{UL}}^i\}$ can be expressed as

$$\{\theta_{\text{UL}}^i\} = \{\text{PL}_{\text{UL}}^i, \text{PN}_{\text{UL}}^i, \text{DS}_{\text{UL}}^i, \text{AASA}_{\text{UL}}^i, \text{EASA}_{\text{UL}}^i,$$
$$\text{AASD}_{\text{UL}}^i, \text{EASD}_{\text{UL}}^i\},$$
$$(1)$$

where PL_{UL}^i, PN_{UL}^i, DS_{UL}^i, $\text{AASA}_{\text{UL}}^i$, $\text{EASA}_{\text{UL}}^i$, $\text{AASD}_{\text{UL}}^i$, and $\text{EASD}_{\text{UL}}^i$ are the PL, PN, DS, AASA, EASA, AASD, and EASD of the ith UE's UL channel, respectively.

Similarly, $\{\theta_{\text{DL}}^i\}$ can be expressed as

$$\{\theta_{\text{DL}}^i\} = \{\text{PL}_{\text{DL}}^i, \text{PN}_{\text{DL}}^i, \text{DS}_{\text{DL}}^i, \text{AASA}_{\text{DL}}^i, \text{EASA}_{\text{DL}}^i,$$
$$\text{AASD}_{\text{DL}}^i, \text{EASD}_{\text{DL}}^i\},$$
$$(2)$$

where $\mathrm{PL}_{\mathrm{DL}}^i$, $\mathrm{PN}_{\mathrm{DL}}^i$, $\mathrm{DS}_{\mathrm{DL}}^i$, $\mathrm{AASA}_{\mathrm{DL}}^i$, $\mathrm{EASA}_{\mathrm{DL}}^i$, $\mathrm{AASD}_{\mathrm{DL}}^i$, and $\mathrm{EASD}_{\mathrm{DL}}^i$ are the PL, PN, DS, AASA, EASA, AASD, and EASD of the ith UE's DL channel, respectively.

Environment features are those features that reflect the characteristics of the communication environment. In addition to the commonly used UE coordinates and propagation distance (Luo et al., 2019; Yang GS et al., 2019), we further consider the properties of buildings between BS and the ith UE. The environment features $\{\theta_{\mathrm{EN}}^i\}$ are given by

$$\{\theta_{\mathrm{EN}}^i\} = \{x^i, y^i, d^i, \Omega^i, L^i, N_{\mathrm{building}}^i, H_{\mathrm{average}}^i, H_{\mathrm{first}}^i, \\ H_{\mathrm{last}}^i\}. \tag{3}$$

Detailed explanations of $\{\theta_{\mathrm{UL}}^i\}$, $\{\theta_{\mathrm{DL}}^i\}$, and $\{\theta_{\mathrm{EN}}^i\}$ are given in Table 1. An example is given in Fig. 2. In this example, $N_{\mathrm{building}}^i = 3$ and H_{average}^i is the average height of three buildings. Then, the problem is to build an appropriate model to predict $\{\theta_{\mathrm{UL}}^i\}$ by using $\{\theta_{\mathrm{DL}}^i\}$ and $\{\theta_{\mathrm{EN}}^i\}$.
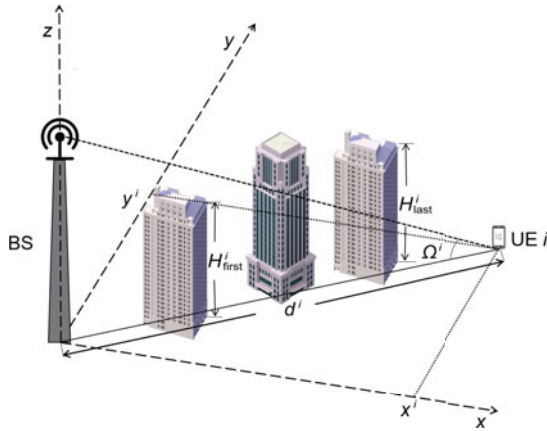
Fig. 2 The explanation of environment features (BS: base station; UE: user equipment)

In practice, we can obtain partial data for training purposes by measurement or other means. The training set includes UL channel parameters, environment features, and DL channel parameters, which are denoted as $\{\theta_{\mathrm{UL,train}}^{i'}\}$, $\{\theta_{\mathrm{EN,train}}^{i'}\}$, and $\{\theta_{\mathrm{DL,train}}^{i'}\}$ respectively, where i' denotes the index of the UE whose DL parameters are known. To predict the DL parameters of other UEs in the same condition, the first step of the proposed prediction method is to select features from the UL parameters and environment features as inputs of the ensemble learning model. We design a two-step feature selection method, in which the SHapley Additive exPlanations (SHAP) algorithm is employed to obtain the feature importance, and then the minimum description length (MDL) algorithm is used to determine the number of selected features. Second, the ensemble learning model can be built with a training set to realize DL channel parameter prediction. The UL channel parameters and the environment features of the predictive channel are expressed as $\theta_{\mathrm{UL}}^{i''}$ and $\theta_{\mathrm{EN}}^{i''}$, respectively, where i'' denotes the index of the UE that is waiting for the prediction, and are inputted into the trained ensemble learning model to predict DL channel parameters.

The accuracy of the built model relies on the amount of training data. However, in practical applications, it is difficult to obtain sufficient training data in a short time for a new propagation condition. Thus, we introduce the instance transfer method to quickly build the DL channel parameter prediction model in a new condition. The new condition contains changes in the Tx beamwidth of the BS or changes in the sector where the UEs are located. Assume that the index of the UE in the training set within the new condition is j'. Similarly, it consists of the UL channel parameters $\theta_{\mathrm{UL,train}}^{j',\mathrm{new}}$, environment features $\theta_{\mathrm{EN,train}}^{j',\mathrm{new}}$, and DL channel parameters $\theta_{\mathrm{DL,train}}^{j',\mathrm{new}}$. Unlike the above prediction steps, we perform feature screening for the training set under the new channel propagation condition based on the feature selection results obtained in the first step. In the third step, the instance transfer model is trained together with $\theta_{\mathrm{UL,train}}^{j',\mathrm{new}}$, $\theta_{\mathrm{EN,train}}^{j',\mathrm{new}}$, and $\theta_{\mathrm{DL,train}}^{j',\mathrm{new}}$. The fourth step is to input the UL channel parameters and environment features under the new conditions into the trained instance transfer model, which will output the corresponding DL channel parameters under the new conditions after calculation.

3 Downlink prediction for the same propagation condition

In this section, we consider the prediction of DL parameters in the same propagation condition with the training set. As mentioned above, to achieve balance between the prediction accuracy and the time consumption, the first step of the prediction model training process is selecting the features from $\{\theta_{\mathrm{UL}}^i\}$ and $\{\theta_{\mathrm{EN}}^i\}$. The second step is to feed the selected features and DL parameters of the training samples

Table 1 Definitions of UL/DL parameters and environment features

Parameter/Feature	Description
$\mathrm{PL}_{\mathrm{UL}}^i/\mathrm{PL}_{\mathrm{DL}}^i$	Path loss of the ith UE's UL/DL channel
$\mathrm{PN}_{\mathrm{UL}}^i/\mathrm{PN}_{\mathrm{DL}}^i$	Path number of the ith UE's UL/DL channel
$\mathrm{DS}_{\mathrm{UL}}^i/\mathrm{DS}_{\mathrm{DL}}^i$	Root-mean-square delay spread of arrival of the ith UE's UL/DL channel
$\mathrm{AASA}_{\mathrm{UL}}^i/\mathrm{AASA}_{\mathrm{DL}}^i$	Azimuth angular spread of arrival of the ith UE's UL/DL channel
$\mathrm{EASA}_{\mathrm{UL}}^i/\mathrm{EASA}_{\mathrm{DL}}^i$	Elevation angular spread of arrival of the ith UE's UL/DL channel
$\mathrm{AASD}_{\mathrm{UL}}^i/\mathrm{AASD}_{\mathrm{DL}}^i$	Azimuth angular spread of departure of the ith UE's UL/DL channel
$\mathrm{EASD}_{\mathrm{UL}}^i/\mathrm{EASD}_{\mathrm{DL}}^i$	Elevation angular spread of departure of the ith UE's UL/DL channel
x^i, y^i	Location coordinates of the ith UE
d^i	Two-dimensional distance from BS to the ith UE
Ω^i	Angle between the horizontal plane and the connection between BS and the ith UE
$L^i=0$ or 1	Propagation condition between BS and the ith UE, line-of-sight (1) or non-line-of-sight (0)
N_{building}^i	Number of buildings located in the propagation path between BS and the ith UE
H_{average}^i	Average height of buildings located in the propagation path between BS and the ith UE
H_{first}^i	Height of the first building located in the propagation path between BS and the ith UE
H_{last}^i	Height of the last building located in the propagation path between BS and the ith UE
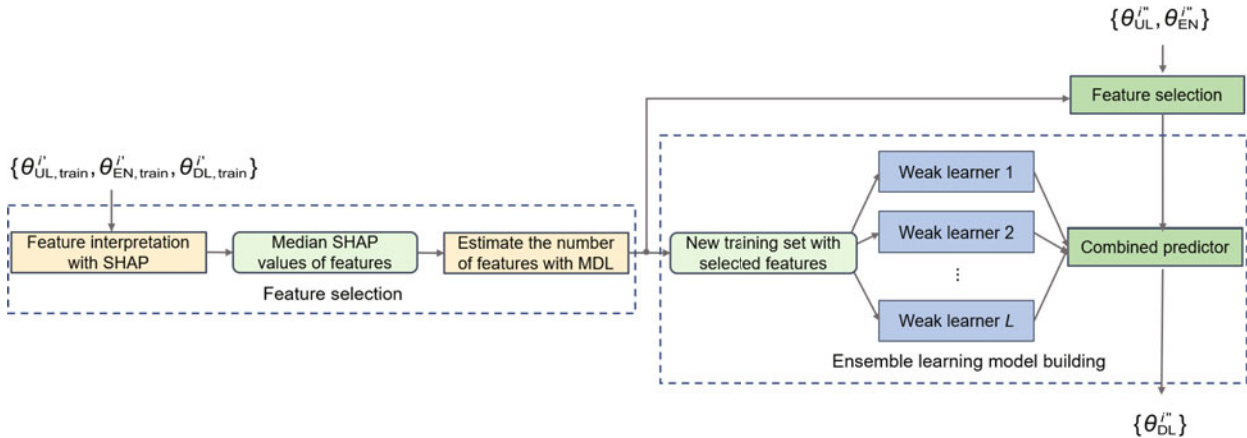
Fig. 3 Framework of the proposed downlink parameter prediction in the same condition (SHAP: SHapley Additive exPlanations; MDL: minimum description length)

into the weak learners of the ensemble learning model for training. The overall process is shown in Fig. 3.

3.1 Two-step feature selection

In Section 2, we have listed many features including the UL parameters and environment features. However, some features may have little relevance to the predicted targets. Using these features for training will increase the burden on the model, and even cause a decline in the accuracy of predictions. Moreover, for different DL channel parameters and different propagation conditions, the relationship between the features and the predicted parameters may differ. We propose a two-step feature selection method. The features are interpreted with the SHAP algorithm. The optimal number of features is then decided with the MDL algorithm.

3.1.1 Feature interpretation

The SHAP algorithm measures how much each feature contributes to each training sample. The median SHAP values reflect the importance of features for prediction and are used as the measurement of feature importance. Consider the i'^{th} training

sample; if we denote the basic prediction model used to calculate the SHAP values as G, the feature interpretation process can be expressed as

$$G(\theta_{\mathrm{UL}}^{i'}, \theta_{\mathrm{EN}}^{i'}) = \phi_0 + \sum_{k=1}^{K} \phi_k^{i'} z_k^{i'}, \qquad (4)$$

where $\phi_k^{i'}$ is the SHAP value of the kth feature for the i'^{th} training sample. K is the total number of features. $z_k^{i'} \in \{0,1\}^K$ are the sample coalitions. $z_k^{i'} = 1$ means that the kth feature is presented in the coalition and $z_k^{i'} = 0$ means that it is absent. ϕ_0 is the mean predicted value of the training samples and can be calculated as

$$\phi_0 = \mathbb{E}(\{G(\theta_{\mathrm{UL}}^{i'}, \theta_{\mathrm{EN}}^{i'})\}). \qquad (5)$$

The SHAP value $\phi_k^{i'}$ is the contribution of the kth feature in the prediction of the i'^{th} sample. When $\phi_k^{i'} > 0$, the kth feature increases the prediction value and has a positive effect on the prediction. Conversely, when $\phi_k^{i'} < 0$, the kth feature decreases the prediction value and has a negative effect on the prediction. Unlike other feature interpretation approaches, which yield the feature importance without differentiating samples, the SHAP value reflects the influence of the feature on the prediction of each sample. $\phi_k^{i'}$ of the kth feature for the i'^{th} training sample is expressed as

$$\begin{aligned}
\phi_k^{i'} = &\sum_{S \subseteq \{\theta_{\mathrm{UL}}^{i'}\} \cup \{\theta_{\mathrm{EN}}^{i'}\} \backslash \{f_k^{i'}\}} \frac{|S|!(K-|S|-1)!}{K!} \\
&\cdot [G_{S \cup \{f_k^{i'}\}}(\theta_{\mathrm{UL}}^{i'}, \theta_{\mathrm{EN}}^{i'}) - G_S(\theta_{\mathrm{UL}}^{i'}, \theta_{\mathrm{EN}}^{i'})], \\
&k = 1, 2, \ldots, K.
\end{aligned} \qquad (6)$$

Here, $\{\theta_{\mathrm{UL}}^{i'}\} \cup \{\theta_{\mathrm{EN}}^{i'}\} \backslash \{f_k^{i'}\}$ means removing $f_k^{i'}$ from $\{\theta_{\mathrm{UL}}^{i'}\} \cup \{\theta_{\mathrm{EN}}^{i'}\}$. S is all subsets of $\{\theta_{\mathrm{UL}}^{i'}\} \cup \{\theta_{\mathrm{EN}}^{i'}\} \backslash \{f_k^{i'}\}$. To compute the contribution of the kth feature to the prediction result, the model $G_{S \cup \{f_k^{i'}\}}$ is trained with $\{f_k^{i'}\}$ and S. G_S is the model trained only with S. Each sample in the training set has its own SHAP value. To eliminate the influence of a few extreme samples on the results, the median of the samples' SHAP values is selected to represent the overall importance of the feature. The median SHAP values of features after normalization are illustrated as $s_1, s_2, \ldots, s_K$, which represent the importance value of each feature in the model training.

3.1.2 Estimation of the number of features

After calculating the median SHAP values of all features, we introduce the MDL-based threshold calculation algorithm to compute the number of selected features. The MDL criterion is commonly used for accurate source enumeration (Wax and Ziskind, 1989; Huang et al., 2009; Bazzi et al., 2016; Lin CH et al., 2018). In this paper, the algorithm is used to determine the number of features involved in the prediction. The number of selected features computed by MDL is $\tilde{K}$ and the selected features are stated as $f_1', f_2', \ldots, f_{\tilde{K}}'$. The objective function of the MDL criterion used in calculating the feature selection threshold is defined as

$$\mathrm{MDL}(k) = -(K-k)\log_2 \left(\frac{\prod\limits_{a=k+1}^{K} s_a^{\frac{1}{K-k}}}{\frac{1}{K-j} \sum\limits_{a=k+1}^{K} s_a} \right). \qquad (7)$$

The number of selected features $\tilde{K}$ is estimated as

$$\tilde{K} = \arg \min_j \mathrm{MDL}(j), j \in \{0, 1, \ldots, K-1\}. \qquad (8)$$

Then, we select the features with the top $\tilde{K}$ importance values. These features compose the new training set and are inputted into the ensemble learning model to realize the prediction of DL channel parameters.

3.2 Construction of the ensemble learning model

The idea of ensemble learning is to improve the performance of prediction by merging weak learners. The weak learners are decision trees and are trained with $\theta_{\mathrm{UL}}^{i'}$, $\theta_{\mathrm{EN}}^{i'}$, and label $\theta_{\mathrm{DL}}^{i'}$. After training, each weak learner is given a weight w_l, $l = 1, 2, \ldots, L$. L is the total number of the weak learners. In the prediction process of the i''^{th} prediction sample, each weak learner can obtain a prediction result $p_l^{i''}$. The final prediction result of the ensemble learning model is the sum of the prediction results of the weak learners multiplied by the weights, expressed as

$$\hat{\theta}_{\mathrm{DL}}^{i''} = \sum_{l=1}^{L} p_l^{i''} w_l. \qquad (9)$$

Boosting is the most popular branch of ensemble learning. The focal point of boosting is reducing

bias. Therefore, boosting methods can build a strong integration learner from several base learners, thus enabling powerful generalization. Three boosting methods, such as adaptive boosting (AdaBoost) (Hu et al., 2008), extreme gradient boosting (XGBoost) (Chen TQ and Guestrin, 2016), and light gradient boosting (LightGBM) (Wang DH et al., 2017), are selected for DL parameter prediction.

4 Downlink prediction for the new propagation condition

Based on the method depicted in Section 3, the DL parameters of the channel under the same propagation conditions can be predicted. When we consider the development in a new propagation condition, the Tx beamwidth may be set with more optional values, and the users may be located in different sectors. It is difficult to obtain enough training samples under the new condition in a short time, which affects the prediction accuracy of the data-driven ensemble learning model. In this paper, we further propose a method for the prediction of DL channel parameters with a small number of samples under new propagation conditions based on instance transfer.

We employ the two-stage TrAdaBoost.R2 (Pardoe and Stone, 2010) algorithm, which divides the weight adjustment of source domain instances and target domain instances into two stages. In the first stage, only the weights of source instances are adjusted to a certain point and the weights of target instances remain unchanged. In the second stage, the weights of all source instances are frozen and the weights of the target instances are updated. The process of the proposed prediction method based on the two-stage TrAdaBoost.R2 algorithm is described as follows.

Assume a source domain training set named $\boldsymbol{\Theta}'^{\text{ori}}_{\text{train}}$, which includes n samples, and a target domain training set named $\boldsymbol{\Theta}'^{\text{new}}_{\text{train}}$, which includes m samples. $\boldsymbol{\Theta}'^{\text{total}}_{\text{train}} = \boldsymbol{\Theta}'^{\text{ori}}_{\text{train}} \cup \boldsymbol{\Theta}'^{\text{new}}_{\text{train}}$ is the total training set. The UL channel parameters of the ith sample in $\boldsymbol{\Theta}'^{\text{total}}_{\text{train}}$ are indicated as $\{\theta'^i_{\text{UL}}\}$, and the environment features of the ith sample are named $\{\theta'^i_{\text{EN}}\}$. The predictive DL channel parameters of the ith sample are named $\{\theta'^i_{\text{DL}}\}$. $i = 1, 2, \ldots, n$ represents the index of the samples from the original training set and $i = n+1, n+2, \ldots, n+m$ represents the in-

dex of the samples from the new training set. Use w to represent the instance weight and e to represent the prediction error rate. The DL channel parameter prediction model in the new propagation condition is named M_{Trans}. At first, initialize the weight distribution W_1 of $\boldsymbol{\Theta}'^{\text{ori}}_{\text{train}}$ and $\boldsymbol{\Theta}'^{\text{new}}_{\text{train}}$. t represents the number of training iterations and is set to 1 at initialization.

$$W_1 = \{w_{1,1}, \ w_{1,2}, \ ..., \ w_{1,n+m}\}, \qquad (10)$$

where $w_{1,i}$ represents the weight of the ith sample in the first iteration and is given by

$$w_{1,i} = \frac{1}{n+m}. \qquad (11)$$

For each iteration $t = 2, 3, \ldots, T$, construct a regression estimator M^t_{Trans} by training it with $\boldsymbol{\Theta}'^{\text{ori}}_{\text{train}}$ and calculate the average error e_t only on $\boldsymbol{\Theta}'^{\text{new}}_{\text{train}}$ as

$$e_t = \sum_{i=n+1}^{n+m} \frac{w_{t,i} \left| \theta'^i_{\text{DL}} - M^t_{\text{Trans}}(\theta'^i_{\text{UL}}, \theta'^i_{\text{EN}}) \right|}{\sum\limits_{i=n+1}^{n+m} w_{t,i}}. \qquad (12)$$

Then, update the weights of $\boldsymbol{\Theta}'^{\text{new}}_{\text{train}}$ as

$$w_{t+1,i} = \frac{w_{t,i}\beta_t^{-\left| \theta'^i_{\text{DL}} - M^t_{\text{Trans}}(\theta'^i_{\text{UL}}, \theta'^i_{\text{EN}}) \right|}}{Z_t}, \qquad (13)$$
$$i = n+1, \ n+2, \ldots, \ n+m,$$

where $w_{t,i}$ represents the weight of the ith sample in the tth iteration. Z_t is a normalized constant. β_t is provided by

$$\beta_t = \frac{e_t}{1 - e_t}. \qquad (14)$$

The weights of $\boldsymbol{\Theta}'^{\text{ori}}_{\text{train}}$ remain unchanged. This process is the first stage of DL parameter prediction for a new propagation condition. In the second stage of DL parameter prediction for a new propagation condition, the error rate ε_t of $\boldsymbol{\Theta}'^{\text{ori}}_{\text{train}}$ is calculated as

$$\varepsilon_t = \sum_{i=1}^{n} \left| \theta'^i_{\text{DL}} - M^t_{\text{Trans}}(\theta'^i_{\text{UL}}, \theta'^i_{\text{EN}}) \right| w_{t,i}. \qquad (15)$$

Then, the weights of $\boldsymbol{\Theta}'^{\text{ori}}_{\text{train}}$ are updated as

$$w_{t+1,i} = \frac{w_{t,i}\gamma_t^{\varepsilon_t}}{Z_t}, i = 1, \ 2, \ \ldots, \ n, \qquad (16)$$

and the weights of $\boldsymbol{\Theta}'^{\text{new}}_{\text{train}}$ stay unchanged. γ_t is given by

$$\gamma_t = \frac{m}{n+m} + \frac{t}{T-1}\left(1 - \frac{m}{n+m}\right). \qquad (17)$$

Finally, we calculate $\tilde{t}$ and choose $M_{\mathrm{Trans}}^{\tilde{t}}$ as the DL parameter prediction model for the new propagation condition. The formula for $\tilde{t}$ is

$$\tilde{t} = \arg\min_{t} \varepsilon_t. \qquad (18)$$

The predicted values of the DL channel parameters can be acquired by inputting the selected features of the new propagation condition in the trained DL channel parameter prediction model.

5 Performance analysis

The simulation data are generated by commercial ray-tracing software, Wireless Insite (Međedović et al., 2012), which is an accurate and reliable urban wave propagation calculation tool for planning and designing mobile wireless networks. A typical urban propagation environment in Ottawa, Canada is considered (Fig. 4). It is an urban scene consisting of buildings and streets. The height distribution of buildings is from 9 to 51 m. The average height of buildings is 19.38 m. The area of the considered scenario is 1000 m × 630 m. The BS is fixed in the center, 30 m above the ground. The propagation environment is divided into three sectors centered on the location of the BS. The Rx beamwidth of the BS is 60°/120° and the Tx beamwidth is 10°/30°. All UEs employ omnidirectional antennas for transmission and reception and they are distributed along the roads. A total of 14 routes, namely routes 1–14, are considered. The height of each UE is 1.5 m. We consider 5635 different UE locations with 1 m spacing on each road. The channel parameters including PL, PN, DS, AASA, EASA, AASD, and EASD are calculated at each position. The parameter configurations are listed in Table 2.

The weak learners of the ensemble learning model are decision trees. The maximum tree depth is 16, and the number of ensemble members is 500. The performance of the proposed prediction method based on ensemble learning and instance transfer is evaluated. In terms of prediction accuracy and calculation time, the prediction results of the proposed method are compared with those of the 3GPP TR 38.901 (3GPP, 2020) channel model and the back propagation neural network (BPNN) (Gao et al., 2021).
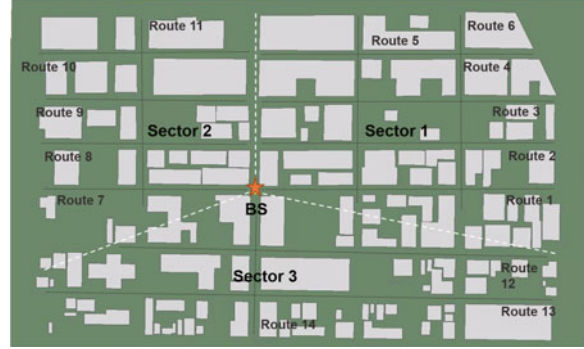

Fig. 4 Propagation environment and sector partition in an Ottawa urban area (BS: base station)

5.1 Evaluation metric and models for comparison

The mean absolute error (MAE) is used as a metric to measure the accuracy of the prediction results (Sterba and Kocur, 2009; Pan et al., 2019). It can be calculated as

$$\mathrm{MAE} = \frac{1}{B} \sum_{b=1}^{B} \left| \theta_{\mathrm{DL}}^{b} - \hat{\theta}_{\mathrm{DL}}^{b} \right|, \qquad (19)$$

where B is the total number of samples in the test data set, θ_{DL}^{b} is the actual DL parameter value of the bth sample, and $\hat{\theta}_{\mathrm{DL}}^{b}$ is the predicted DL parameter value of the bth sample.

A standardized model, the 3GPP TR 38.901 channel model, is used as one of the comparison methods. We calculate PL using its urban macro (UMa) propagation model with the non-line-of-sight (NLoS) condition.

BPNN is also used as a comparison method. The rectified linear unit function is selected as

Table 2 Parameter configurations of the simulation environment

Parameter	Value/Description
Environment	Ottawa urban scenario
Central frequency	26 GHz
Bandwidth	500 MHz
Transmit power	40 dBm
BS height	30 m
UE height	1.5 m
Rx/Tx beamwidth	120°/30°, 120°/10°, 60°/30°, 60°/10°
Number of BS locations	1
Number of UE locations	(2518, 1703, 1414)*
Height distribution of buildings	(9–48, 9–45, 9–51) m*
Average height of buildings	(20.23, 19.23, 18.67) m*

* The three values in the brackets refer to the parameter values of sector 1, 2, and 3, respectively. BS: base station; UE: user equipment.

the activation function. A three-layer feed-forward structure is employed and the optimal number of neurons in the hidden layer is 15.

In the following analysis, we compare the prediction results of the proposed method with the 3GPP TR 38.901 channel model and the non-ensemble learning method BPNN. Both prediction accuracy and running time are evaluated.

5.2 Evaluation of the two-step feature selection algorithm

To verify that the proposed two-step feature selection method could achieve balance between the prediction accuracy and efficiency of DL channel parameter prediction, we compare the prediction accuracy, number of features, and training time of the proposed two-step feature selection method with those of a variance filter (Wang F et al., 2018; Zhao et al., 2018; Huang et al., 2019) and Lasso regularization (Gallieri and Maciejowski, 2012; Yao et al., 2015). The variance filter is used to remove the features with low variance. The results of feature selection using the variance filter are obtained by setting the threshold at 0.16 and discarding all features with variance less than the threshold. Lasso regularization performs regularization and feature selection of the given data. Lasso regularization places a limit on the sum of features and the results of feature selection consist of the features that are not reduced to zero after reduction. In the comparison simulation, the UL beamwidth is set to 120° and the DL beamwidth is set to 30°. The proposed models are trained using the feature sets obtained from different feature selection methods to predict multiple DL channel parameters. A comparison of different feature selection methods is shown in Table 3.

According to the comparison results of the predicted MAE, the number of selected features, and the model training time of different feature selection methods, it is shown that different numbers of features are selected, with the number ranging from 7 to 10. The feature selection based on the variance filter considers only the variance of the features, so the selection is the same for different prediction targets, and the number of features is 2. Although the reduction in the number of features reduces the model training time, the prediction accuracy is significantly reduced. For the feature selection based on Lasso regularization, different feature sets are se-

lected depending on the prediction target, but the number of selected features is larger than that of the proposed method and is between 10 and 13. The increased number of features results in a longer training time. However, instead of improving the prediction accuracy of the model, the increase in the number of features enlarges the MAE compared to that of the proposed method. This indicates that the inclusion of less relevant features not only increases the training time of the model but also reduces the prediction accuracy. The feature set obtained by the proposed method can achieve accurate and efficient predictions.

5.3 Performance analysis of downlink parameter prediction in the same condition

To verify the proposed method in DL channel parameter predictions in the same condition, there are 2518 samples corresponding to different UE locations in sector 1. To verify the prediction ability of the proposed prediction algorithm, we randomly divide the collected data into two sets. Eighty percent of the data are used as the training set, and the remaining, as the testing set. Table 4 shows the predicted MAEs of different methods. The proposed method with three different boosting methods has higher prediction accuracy than BPNN. Furthermore, the prediction results when XGBoost is chosen as the boosting module are the worst. When AdaBoost or LightGBM is employed, AdaBoost has higher prediction stability and gives accurate prediction results for each DL channel parameter. Light-GBM has slightly lower prediction accuracy than AdaBoost for some parameters, but faster model training and prediction.

In addition, the Tx and Rx beamwidths of the BS have an impact on the model prediction accuracy. When the beamwidths of Tx and Rx differ greatly, the intrinsic relation between UL and DL channel parameters drops drastically, which causes increase of MAE. Finally, Table 4 illustrates the MAEs of the proposed method and BPNN for different DL parameters.

To compare the prediction performance between the proposed method, 3GPP TR 38.901 channel model, and BPNN, we select PL prediction results of route 4 and plot them in Fig. 5. The horizontal coordinate indicates the indices of the test samples, which are given in the order of their positions. The

Table 3　Comparison of different feature selection methods

Feature selection method	Prediction target	MAE	Number of features	Training time (s)
Proposed method	PL	1.03 dB	7	4.69
	DS	13.44 ns	10	5.46
	PN	0.20	8	5.27
	AASA	0.34°	10	5.30
	EASA	2.77°	7	4.88
	AASD	0.09°	9	5.17
	EASD	0.37°	10	6.13
Variance filter	PL	2.73 dB	2	2.66
	DS	68.73 ns	2	2.74
	PN	0.98	2	1.83
	AASA	1.85°	2	2.05
	EASA	32.61°	2	2.21
	AASD	2.64°	2	1.96
	EASD	2.78°	2	2.25
Lasso regularization	PL	1.08 dB	10	6.45
	DS	13.99 ns	12	6.68
	PN	0.22	13	7.37
	AASA	0.37°	13	7.09
	EASA	2.83°	13	6.45
	AASD	0.13°	12	5.62
	EASD	0.40°	13	6.34

MAE: mean absolute error

vertical coordinate indicates the PL values. From the prediction results of ray tracing in the figure, the PL values on the street fluctuate greatly due to the presence of building obstruction in the urban scenario. The proposed method can accurately predict the PL fluctuation. The MAE of the proposed method is 1.01 dB. In contrast, 3GPP TR 38.901 can predict only the trend of PL based on the propagation distance. The MAE of the 3GPP TR 38.901 channel model is 26.91 dB. BPNN can also predict PL fluctuations, and the MAE of BPNN is 4.21 dB, which is larger than that of our proposed method.

In addition, we compare the running time of each method and list the results in Table 5. Ray tracing relies on the electromagnetic map, and the calculation time is almost hundreds of times that of our method with AdaBoost and almost thousands of times that of our method with LightGBM. In addition, when using the proposed method for DL parameter prediction of the channel in the same condition, offline training is available and the model needs to be trained only once. After the model is trained, the actual time required is only the prediction time. BPNN takes less time than ray tracing, but its training time and prediction time are much longer than those of our proposed method. Among our proposed method with different boosting methods, LightGBM
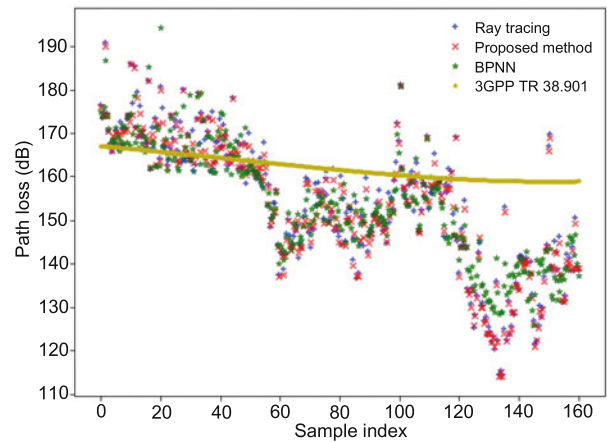
Fig. 5　Path loss prediction values of the proposed method with AdaBoost and other methods

takes the least time, whereas AdaBoost takes the longest time for training and prediction.

5.4 Performance analysis of downlink parameter prediction in new conditions

In this subsection, we consider two transfer cases of new conditions. The first case is to predict the channel parameters with a new Tx beamwidth. To verify our proposed method, the source domain is selected as the channel with 120° Rx beamwidth and 30° Tx beamwidth in sector 1. The target domain

Table 4 MAEs of different methods for each Downlink parameter

| Prediction target	Rx/Tx beamwidth	MAE			
		Our method (AdaBoost)	Our method (LightGBM)	Our method (XGBoost)	BPNN
PL (dB)	120°/30°	1.03	1.28	2.44	3.42
	120°/10°	2.10	2.29	5.07	7.71
	60°/30°	0.96	1.02	2.01	3.20
	60°/10°	2.17	2.21	4.96	7.62
DS (ns)	120°/30°	13.44	18.66	58.24	53.39
	120°/10°	25.35	24.98	79.3	60.45
	60°/30°	9.78	18.69	50.27	51.61
	60°/10°	22.55	24.75	76.29	53.54
PN	120°/30°	0.20	0.48	0.61	0.66
	120°/10°	0.33	0.50	0.71	0.90
	60°/30°	0.21	0.50	0.53	0.60
	60°/10°	0.32	0.54	0.73	0.87
AASA (°)	120°/30°	0.34	0.29	0.86	1.36
	120°/10°	0.72	0.48	1.36	1.85
	60°/30°	0.29	0.24	0.70	1.26
	60°/10°	0.74	0.45	1.34	1.70
EASA (°)	120°/30°	2.77	3.97	13.11	17.98
	120°/10°	6.08	6.29	20.78	19.27
	60°/30°	2.01	3.73	10.29	18.05
	60°/10°	5.21	5.72	20.36	27.57
AASD (°)	120°/30°	0.09	0.07	0.16	0.18
	120°/10°	0.07	0.05	0.14	0.21
	60°/30°	0.08	0.07	0.13	0.30
	60°/10°	0.07	0.05	0.13	0.19
EASD (°)	120°/30°	0.37	0.44	1.59	0.72
	120°/10°	0.24	0.25	0.77	0.93
	60°/30°	0.32	0.39	0.13	1.85
	60°/10°	0.23	0.24	0.76	0.89

is the channel with 120° Rx beamwidth and 10° Tx beamwidth in the same sector. The prediction target is the DL PL with the new Tx beamwidth. The total number of samples in the target domain is 2518. We choose only 10–250 samples for training. The training set is composed of a total of 2518 samples in the source domain and the chosen samples in the target domain. We compare the performance of using the proposed algorithm with that of using the method without the transfer learning algorithm.

Table 5 Running time of different methods

Prediction method	Training time (s)	Prediction time (s)
Our method with		
AdaBoost	4.56	0.13
LightGBM	0.45	0.02
XGBoost	1.34	0.05
BPNN	105	32
Ray tracing	–	1703

Fig. 6 shows a comparison of the results, showing that the participation of instance transfer improves the accuracy of DL channel parameter prediction. As the number of samples in the target domain increases, the prediction accuracy of the proposed model gradually increases. When the number of samples in the target domain reaches 150, the predicted MAE is only 3.54 dB. Meanwhile, the MAE of the method without transfer learning is 14.25 dB. Note that the proposed method can also be extended to the prediction of other channel parameters.

The second case is to predict the channel parameters in different sectors. We select the source domain as the channel with 120° Rx beamwidth and 30° Tx beamwidth in sector 1. The target domain is the channel with the same beamwidth pair in sector 2 or 3, and the prediction target is the DL PL. The total number of samples in sector 2 is 1703 and that in sector 3 is 1414. We choose only 50–200
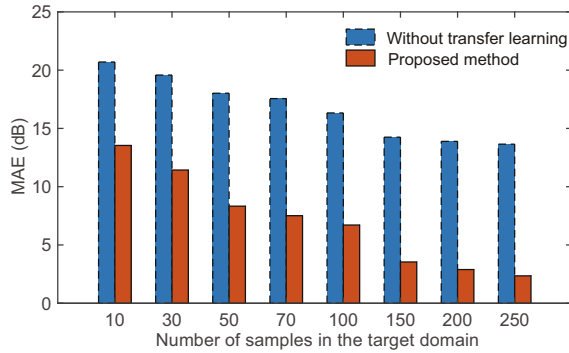
Fig. 6 Comparison of the path loss prediction accuracy of the proposed method and the method without transfer learning when the downlink beamwidth changes from 30° to 10° (MAE: mean absolute error)

samples of the target domain for training. The training set is composed of a total of 2518 samples in the source domain and the chosen samples in the target domain.

Fig. 7 shows the comparison between the proposed method and the method without transfer learning. The results of the proposed method are always better than those of the method without transfer learning. In sector 2, when the number of samples in the target domain reaches 150, the MAE of the proposed method has fallen to 6.31 dB and is 19.01 dB less than that of the method without transfer learning. Moreover, the prediction accuracy is related to the similarity between the original sector and the new sector. As shown in Table 2, the distribution and the average height of the buildings in sector 2 are more similar to those in sector 1, and the prediction error of sector 2 is a bit lower than that of sector 3.
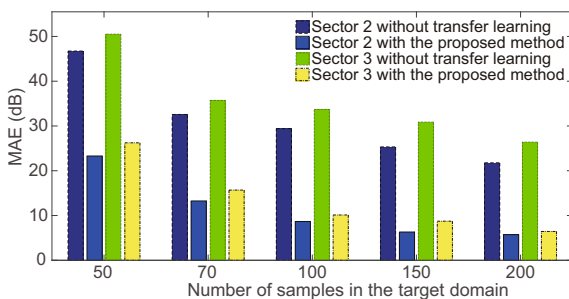
Fig. 7 Comparison of the path loss prediction accuracy of the proposed method and the method without transfer learning when the sector changes (MAE: mean absolute error)

6 Conclusions

In this paper, an ensemble-transfer-learning-based channel parameter prediction method was proposed for asymmetric massive MIMO systems with different Tx/Rx beam patterns. The UL channel parameters and environment features participated in the model training. A two-step feature selection algorithm was designed to improve the prediction accuracy and efficiency of the model. In addition, the proposed method could predict DL parameters in the new propagation condition with the help of the instance transfer algorithm, even when the new training samples were insufficient. The proposed method could be used to predict PL, PN, DS, and angular spread. Its performance was compared with those of the 3GPP TR 38.901 channel model and BPNN. Simulation results showed that the prediction accuracy of the proposed method was better than those of the compared methods. When the Tx beamwidth or the sector changed, the instance-transfer-based DL parameter prediction method provided higher prediction accuracy than the method without the transfer learning algorithm with a small number of new samples. The proposed prediction method could be useful in analyzing the effect of beamwidth on channel parameters and for developing practical applications of asymmetric systems, thereby reducing the burden of data processing and hardware costs in 6G mobile networks.

Contributors

Zunwen HE, Yan ZHANG, and Wancheng ZHANG proposed the ideas and designed the simulations. Yue LI processed the data and completed the simulations. Zunwen HE, Kaien ZHANG, Liu GUO, and Haiming WANG drafted, revised, and finalized the paper.

Compliance with ethics guidelines

Zunwen HE, Yue LI, Yan ZHANG, Wancheng ZHANG, Kaien ZHANG, Liu GUO, and Haiming WANG declare that they have no conflict of interest.

Data availability

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not available.

References

3GPP, 2020. Study on Channel Model for Frequencies from

0.5 to 100 GHz. TR 38.901 V16.1.0. 3rd Generation Partnership Project (3GPP), France.

Albreem MA, Juntti M, Shahabuddin S, 2019. Massive MIMO detection techniques: a survey. *IEEE Commun Surv Tut*, 21(4):3109-3132.
https://doi.org/10.1109/COMST.2019.2935810

Albreem MA, Habbash AH, Abu-Hudrouss AM, et al., 2021. Overview of precoding techniques for massive MIMO. *IEEE Access*, 9:60764-60801.
https://doi.org/10.1109/ACCESS.2021.3073325

Bazzi A, Slock DTM, Meilhac L, 2016. Detection of the number of superimposed signals using modified MDL criterion: a random matrix approach. Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, p.4593-4597. https://doi.org/10.1109/ICASSP.2016.7472547

Chen TQ, Guestrin C, 2016. XGBoost: a scalable tree boosting system. Proc 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining, p.785-794.
https://doi.org/10.1145/2939672.2939785

Chen YS, Lai FP, You JW, 2019. Analysis of antenna radiation characteristics using a hybrid ray tracing algorithm for indoor WiFi energy-harvesting rectennas. *IEEE Access*, 7:38833-38846.
https://doi.org/10.1109/ACCESS.2019.2906646

Erden F, Ozdemir O, Guvenc I, 2020. 28 GHz mmWave channel measurements and modeling in a library environment. Proc IEEE Radio and Wireless Symp, p.52-55.
https://doi.org/10.1109/RWS45077.2020.9050106

Gallieri M, Maciejowski JM, 2012. ℓasso MPC: smart regulation of over-actuated systems. Proc American Control Conf, p.1217-1222.
https://doi.org/10.1109/ACC.2012.6315171

Gao DY, Hu BQ, Qin FJ, et al., 2021. A real-time gravity compensation method for INS based on BPNN. *IEEE Sens J*, 21(12):13584-13593.
https://doi.org/10.1109/JSEN.2021.3069960

Han Y, Li MY, Jin S, et al., 2020. Deep learning-based FDD non-stationary massive MIMO downlink channel reconstruction. *IEEE J Sel Areas Commun*, 38(9):1980-1993. https://doi.org/10.1109/JSAC.2020.3000836

Hong W, Zhou JY, Chen JX, et al., 2020. Asymmetric full-digital beamforming mmWave massive MIMO systems for B5G/6G wireless communications. Proc IEEE Asia-Pacific Microwave Conf, p.31-32.
https://doi.org/10.1109/APMC47863.2020.9331559

Hong W, Jiang ZH, Yu C, et al., 2021. The role of millimeter-wave technologies in 5G/6G wireless communications. *IEEE J Microw*, 1(1):101-13584.
https://doi.org/10.1109/JMW.2020.3035541

Hu WM, Hu W, Maybank S, 2008. AdaBoost-based algorithm for network intrusion detection. *IEEE Trans Syst Man Cybern B Cybern*, 38(2):577-583.
https://doi.org/10.1109/TSMCB.2007.914695

Huang L, Long T, Mao EK, 2009. Robust estimation of the number of sources using an MMSE-based MDL method. Proc IET Int Radar Conf, p.1-4.
https://doi.org/10.1049/cp.2009.0262

Huang L, Fu QB, Li GF, et al., 2019. Improvement of maximum variance weight partitioning particle filter in urban computing and intelligence. *IEEE Access*,

7:106527-106535.
https://doi.org/10.1109/ACCESS.2019.2932144

Jiang T, Zhang JH, Tang P, et al., 2021. A study of uplink and downlink channel spatial characteristics in an urban micro scenario at 28 GHz. *Front Inform Technol Electron Eng*, 22(4):488-502.
https://doi.org/10.1631/FITEE.2000443

Joo J, Park MC, Han DS, et al., 2019. Deep learning-based channel prediction in realistic vehicular communications. *IEEE Access*, 7:27846-27858.
https://doi.org/10.1109/ACCESS.2019.2901710

Kim MD, Liang JY, Lee J, et al., 2016. Directional multipath propagation characteristics based on 28GHz outdoor channel measurements. Proc 10th European Conf on Antennas and Propagation, p.1-5.
https://doi.org/10.1109/EuCAP.2016.7481755

Li SD, Liu YJ, Chen ZP, et al., 2017. Measurements and modelling of millimeter-wave channel at 28 GHz in the indoor complex environment for 5G radio systems. Proc 9th Int Conf on Wireless Communications and Signal Processing, p.1-6.
https://doi.org/10.1109/WCSP.2017.8170925

Lin B, Gao FF, Zhang S, et al., 2021. Deep learning-based antenna selection and CSI extrapolation in massive MIMO systems. *IEEE Trans Wirel Commun*, 20(11):7669-7681.
https://doi.org/10.1109/TWC.2021.3087318

Lin CH, Chi CY, Chen LL, et al., 2018. Detection of sources in non-negative blind source separation by minimum description length criterion. *IEEE Trans Neur Netw Learn Syst*, 29(9):4022-4037.
https://doi.org/10.1109/TNNLS.2017.2749279

Liu XM, Tang JS, 2014. Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method. *IEEE Syst J*, 8(3):910-920.
https://doi.org/10.1109/JSYST.2013.2286539

Lundberg SM, Lee SI, 2017. A unified approach to interpreting model predictions. Proc 31st Int Conf on Neural Information Processing Systems, p.4768-4777.
https://doi.org/10.5555/3295222.3295230

Luo XR, Zhang Y, He ZW, et al., 2019. A two-step environment-learning-based method for optimal UAV deployment. *IEEE Access*, 7:149328-149340.
https://doi.org/10.1109/ACCESS.2019.2947546

Međedović P, Veletić M, Blagojević Ž, 2012. Wireless in-site software verification via analysis and comparison of simulation and measurement results. Proc 35th Int Convention MIPRO, p.776-781.

Pan X, Yan CX, Zhang JK, 2019. Nonlinearity-based single-channel monopulse tracking method for OFDM-aided UAV A2G communications. *IEEE Access*, 7:148485-148494.
https://doi.org/10.1109/ACCESS.2019.2946960

Pardoe D, Stone P, 2010. Boosting for regression transfer. Proc 27th Int Conf on Machine Learning, p.863-870.
https://doi.org/10.5555/3104322.3104432

Qiu JH, Xu K, Xia XC, et al., 2022. Secure transmission scheme based on fingerprint positioning in cell-free massive MIMO systems. *IEEE Trans Signal Inform Process Netw*, 8:92-105.
https://doi.org/10.1109/TSIPN.2022.3149112

Radhakrishnan V, Taghizadeh O, Mathar R, 2021. Impairments-aware resource allocation for FD massive MIMO relay networks: sum rate and delivery-time optimization perspectives. *IEEE Trans Signal Inform Process Netw*, 7:177-191.
https://doi.org/10.1109/TSIPN.2021.3054991

Sterba J, Kocur D, 2009. Pilot symbol aided channel estimation for OFDM system in frequency selective Rayleigh fading channel. Proc 19th Int Conf on Radioelektronika, p.77-80.
https://doi.org/10.1109/RADIOELEK.2009.5158729

Tuan DM, Cheon Y, Aoki Y, et al., 2016. Performance comparison of millimeter-wave communications system with different antenna beam-width. Proc 10th European Conf on Antennas and Propagation, p.1-5.
https://doi.org/10.1109/EuCAP.2016.7482003

Wang D, Nie FP, Huang H, 2015. Feature selection via global redundancy minimization. *IEEE Trans Knowl Data Eng*, 27(10):2743-2755.
https://doi.org/10.1109/TKDE.2015.2426703

Wang DH, Yang Z, Yi Z, 2017. LightGBM: an effective miRNA classification method in breast cancer patients. Proc Int Conf on Computational Biology and Bioinformatics, p.7-11.
https://doi.org/10.1145/3155077.3155079

Wang F, Liang JL, Wang ZD, et al., 2018. A variance-constrained approach to recursive filtering for nonlinear 2-D systems with measurement degradations. *IEEE Trans Cybern*, 48(6):1877-1887.
https://doi.org/10.1109/TCYB.2017.2716400

Wang ZG, Xiao X, Rajasekaran S, 2020. Novel and efficient randomized algorithms for feature selection. *Big Data Min Anal*, 3(3):208-224.
https://doi.org/10.26599/BDMA.2020.9020005

Wax M, Ziskind I, 1989. Detection of the number of coherent signals by the MDL principle. *IEEE Trans Acoust Speech Signal Process*, 37(8):1190-1196.
https://doi.org/10.1109/29.31267

Wimalajeewa T, Varshney PK, Su W, 2017. Detection of single versus multiple antenna transmission systems using pilot data. *IEEE Trans Signal Inform Process Netw*, 3(1):159-171.
https://doi.org/10.1109/TSIPN.2016.2601023

Wu XY, Zhang Y, Wang CX, et al., 2015. 28 GHz indoor channel measurements and modelling in laboratory environment using directional antennas. Proc 9th European Conf on Antennas and Propagation, p.1-5.

Yang GS, Zhang Y, He ZW, et al., 2019. Machine-learning-based prediction methods for path loss and delay spread in air-to-ground millimetre-wave channels. *IET Microw Antenn Propag*, 13(8):1113-1121.
https://doi.org/10.1049/iet-map.2018.6187

Yang YW, Gao FF, Ma XL, et al., 2019. Deep learning-based channel estimation for doubly selective fading channels. *IEEE Access*, 7:36579-36589.
https://doi.org/10.1109/ACCESS.2019.2901066

Yao JG, Liu X, Zhu XY, et al., 2015. Control of large-scale systems through dimension reduction. *IEEE Trans Serv Comput*, 8(4):563-575.
https://doi.org/10.1109/TSC.2014.2312946

Ye H, Li GY, Juang BH, 2018. Power of deep learning for channel estimation and signal detection in OFDM systems. *IEEE Wirel Commun Lett*, 7(1):114-117.
https://doi.org/10.1109/LWC.2017.2757490

Zhang S, Liu YS, Gao FF, et al., 2021. Deep learning based channel extrapolation for large-scale antenna systems: opportunities, challenges and solutions. *IEEE Wirel Commun*, 28(6):160-167.
https://doi.org/10.1109/MWC.001.2000534

Zhang SB, Zhang S, Gao FF, et al., 2021. Machine-learning-based prediction methods for path loss and delay spread in air-to-ground millimeter wave channels. *IEEE Trans Commun*, 69(10):6691-6705.
https://doi.org/10.1109/TCOMM.2021.3097726

Zhao SY, Huang B, Liu F, 2018. Localization of indoor mobile robot using minimum variance unbiased FIR filter. *IEEE Trans Autom Sci Eng*, 15(2):410-419.
https://doi.org/10.1109/TASE.2016.2599864