Fast Beam Switching Based on Machine Learning for MmWave Massive MIMO Systems

Kean Chen[1], Danpu Liu[2](✉), and Xingwen He[3]

[1] Beijing Laboratory of Advanced Information Networks, Beijing, China
[2] Beijing Key Laboratory of Network System Architecture and Convergence, Beijing, China
dpliu@bupt.edu.cn
[3] Beijing University of Posts and Telecommunications, Beijing 100876, People's Republic of China
hexingwen@bupt.edu.cn

Abstract. Millimeter wave (mmWave) and massive multiple-input-multiple-output (MIMO) systems are two key technologies for 5G. Beamforming based on massive MIMO can produce high directional beams with array gain, and thus effectively compensate for the high path loss of mmWave. As the number of antennas increases, the beams become increasingly narrow, resulting in large overhead and high latency in the initial access and handover of the beams. For high-speed mobile scenarios, beam switching becomes more challenging since the traversal search among a large number of beams cannot be completed in a short period of time. To address this problem, this paper proposes a method based on machine learning to predict the optimal Base Station (BS) and the optimal beam pair at the successive instant for the User Equipment (UE) in motion. More specifically, a Random Forest (RF) classification model is trained to learn the channel's features in a multi-cell scenario, and complete the nonlinear modeling of the propagation environment. Furthermore, this model is used to predict the future optimal BS and the optimal beam pair for a moving UE based on the present UE's location, BS beam index and RSRP value. The simulation results show that the prediction accuracy is greater than 90% in most situations, thus the latency and the consumption of signaling resources for beam switch is reduced significantly while the loss in spectral efficiency is little.

Keywords: 5G mobile communication · Beam switching · Machine learning · Random forest

1 Introduction

mmWave is an important frequency band that will be used in 5G communication. To compensate the high attenuation of mmWave, massive MIMO and beamforming technologies are used in the 5G system to generate a highly directional beam and provide high beamforming gain. As the antenna scale increases, the number of candidate beams at the transceiver also increases, and the beams become narrower. Since beams have

X. Wang et al. (Eds.): AICON 2021, LNICST 396, pp. 19–29, 2021.
https://doi.org/10.1007/978-3-030-90196-7_2

strong directivity, User Equipment (UE) is accompanied by frequent beam switching as they move. However, the traditional method to find the optimal beam in next time step is to perform exhaustive searching among all candidate beams, which results in high beam sweeping complexity and signaling overhead in massive MIMO scenario. Moreover, if the UE moves too quickly and the beam switches not fast enough, the beam after switching is not the optimal beam pair at the current location. This beam misalignment inevitably results in the loss of spectral efficiency. Therefore, quicker and more accurate beam switching methods in high-speed scenario are needed.

Many approaches to improve the speed and efficiency of beam switching have been proposed. These algorithms can be divided into three categories. In the first category, the beam search process combined with codebook design is optimized. More specifically, a fast beam search algorithm based on a new codebook [1] is proposed to improve the speed of the beam search, and a hierarchical beam search algorithm [2] is proposed to simplify the complexity of the beam search. Besides, a dynamic iterative beam search algorithm is proposed in [3] to reduce the time delay of the beam search. In [4], the features of the Poisson distribution are used before and after the beam is blocked, and a scheme of dynamic beamforming is proposed. The second approach is to perform beam switching based on location information. Owing to the development of positioning technology, the UE's location can be obtained when the UE moves from one location to another; and the search range of the beam can be greatly narrowed down with the help of location information. For example, GPS positioning data is used in [5], and demonstrates good performance in the absence of line-of-sight propagation. A beamforming method based on image tracking and positioning in the LOS scenario is proposed in [6]. In [7], the Extended Kalman Filter is used to predict the location which is used to optimize beamforming. In [8], a 3D beamforming technology based on positioning assistance in mobile scenario is proposed to improve the beam switching speed and spectrum efficiency. Besides, both channel and location information are used to estimate the DOA angle in [9]. This method is more suitable for UAV beam switching scenarios with three-dimension (3D) mobility and rapid velocity changes. The third approach is to use machine learning methods to optimize beam management. In [10], a deep neural network is utilized, and the optimal beam combination can be output by inputting the channel vector with transmit power. In [11], the gNB learns the UE mobility information, signal-to-noise ratio and current beam information, and predicts whether the beams are aligned or not. In addition, Q-Learning is used for beam selection in NLOS scenarios in [12].

In most communication scenarios, the location of scatters such as buildings and vegetation are relatively fixed, and the line-of-sight path also exists, which will result in the correlated channel between adjacent time steps. However, the correlation between channels is nonlinear and difficult to be analyzed by deterministic mathematical models. Machine learning is well suited to deal with nonlinear problems, it can accomplish the prediction by extracting nonlinear information from a large amount of data. More specifically, the nonlinear features of the channel are expressed by some information obtained by the receiver easily, such as the received signal strength, the beam direction, and UE's location. With these available information, the channel features of the current moment can be extracted through machine learning, and further used to accurately predict the channel features of the next moment. Therefore, it is possible to achieve a much faster beam switching owing to the removal of time-consuming beam search.

Based on above consideration, we propose a method to optimize beam switching speed in high-speed mobile scenario by using a Random Forest (RF) multi-classification model in this paper. The model is firstly trained to learn the nonlinear features of the channel, and then it can quickly predict which beam pairs are optimal for the UE to switch to at next moment. Based on the results obtained by simulation, the model proposed in this paper successfully predicts the desired beam to switch to for each UE in a short time with a high prediction accuracy above 90% in most situations. Moreover, although the beam predicted by the model is not always the optimal, nearly the same spectral efficiency can be achieved. In addition, in order to make a compromise between the training cost and the prediction accuracy, several combinations of features are tested and the most suitable one for the scenario is found out. To sum up, the method proposed in this paper reduces complex signaling interactions, and improves the beam switching efficiency greatly in high-speed mobile scenario.

2 System Model

2.1 Scenario Description

Figure 1 shows a mobile mmWave communication scenario. A number of cars, i.e., the UEs are moving within an area consisting of multiple cells, and each cell is served by one Base Station (BS). Both BSs and the UEs are equipped with multiple antennas, each BS has Nt antennas, each UE has Nr antennas, and there is only one RF link between each BS and each UE. That means analog beamforming is applied at both BS and UE, and each UE is only served by one BS at each time slot. The analog precoder and combiner are composed of a set of phase shifters, which can offset the phase of the input signal [13].

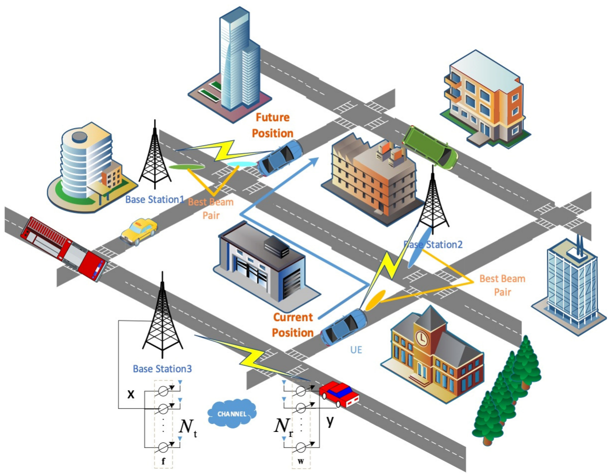

Fig. 1. Mobile mmWave communication scenario with analog beamformer.

In this paper, we focus on the downlink, so the signal is sent at the k-th BS and received at the served u-th UE. The received signal at the u-th UE can be expressed as [14]:

$$y_u = \boldsymbol{w}_u^H \boldsymbol{H}_{u,k} \boldsymbol{f}_{u,k} x_u + \boldsymbol{w}_u^H \boldsymbol{n}_u \tag{1}$$

where x_u denotes the transmitted signal from the BS to the u-th UE, and its average power is $P_t = E\left\{|x_u|^2\right\}$. $\boldsymbol{f}_{u,k}$ denotes the analog beamforming vector with $N_t \times 1$ dimension, $\boldsymbol{H}_{u,k}$ denotes the channel matrix with $N_r \times N_t$ dimension between the u-th UE and the k-th BS, $\boldsymbol{w}_u$ denotes the analog combination vector with $N_r \times 1$ dimension at the u-th UE, $\boldsymbol{n}_u$ denotes Gaussian white noise with $N_r \times 1$ dimension.

In practice, $\boldsymbol{w}_u$ and $\boldsymbol{f}_{u,k}$ are codewords selected from a DFT beam codebook, which is a pre-defined matrix. To facilitate the traversal search among a combination of candidate beams, the beam direction is usually quantized into a vector in the beam codebook, and each column in the codebook corresponds to a beam direction. The values in each vector in the codebook represent the phase shift applied to each antenna. The expression for the DFT codebook [15] is defined as:

$$\boldsymbol{W}(m, n) = exp(i\frac{2\pi mn}{N}), n = 0, 1, ..., N-1; m = 0, 1, ..., M-1, \tag{2}$$

where M is the number of antennas and N is the number of beams. Generally, $M = N$ in the DFT codebook because the beam vectors generated in this way are orthogonal to each other.

Given the received signal at the UE in (1), the Spectral Efficiency (SE) is defined as:

$$SE = log_2\left(1 + \frac{P_t}{\delta_{nu}^2}\frac{\boldsymbol{w}_u^H \boldsymbol{H}_{u,k} \boldsymbol{f}_{u,k} \boldsymbol{f}_{u,k}^H \boldsymbol{H}_{u,k}^H \boldsymbol{w}_u}{\boldsymbol{w}_u^H \boldsymbol{w}_u}\right) \tag{3}$$

where δ_{nu}^2 is the power of Gaussian white noise.

2.2 Beam Switching

As shown in (3), the SE is the highest when the UE connects to the BS with the best channel quality and the beam direction at the transmitter is aligned with the receiver. However, when the UE keeps moving, the optimal serving BS and the optimal beam pair may continuously change. For example, when a blue car moves along the blue line in Fig. 1, initially it is served by BS2, but the optimal beam pair between them will vary with the car's location. Then, after the car passes through a specific location, the optimal BS will change from BS2 to BS1, and the best beam pair between the UE and BS1 will continue to vary. Therefore, beam switching including BS handover will frequently take place in order to maintain connection and high SE in high-speed mobile scenario, and how to identify the optimal serving BS and the beam pair for the UE in time and efficiently becomes challenging.

3 Beam Switching Based on Machine Learning

3.1 Feasibility of Machine Learning Prediction

As we all know, exhaustive beam sweeping method leads to high beam searching overhead which hardly meet the low latency requirements in high-speed scenario, hence it is necessary to design faster and accurate beam switching algorithms to ensure the optimal communication quality. In a given scenario, the locations of many scatters such as buildings, vegetation, and ground are fixed as well as line-of-sight path exists, which makes the channel correlated between adjacent time steps. The channel features cannot be directly measured, but indirectly reflected by measurable information, such as the received signal strength, beam index, BS and UE's locations, etc. This provides the possibility to predict the channel features of the next moment by the channel features of the previous moment. However, the correlation between channels is nonlinear, and it is difficult to extract the nonlinear features of channels by deterministic mathematical modeling methods. Given the outstanding advantage of machine learning for dealing with nonlinear problems, we attempt to use this approach to extract the nonlinear features of the channel and complete the prediction of the BS index, BS beam index and UE beam index at the next moment. With respect to the model for machine learning, the Random Forest (RF) Multi-Classification model is chosen owing to its good performance at multi-classification tasks with simple logic and extremely fast training speed.

3.2 Beam Prediction Based on RF

The RF algorithm is a bagging ensemble learning algorithm whose base evaluator is a decision tree [16]. RF increases the differences between each classification model (decision tree) by constructing different training sets, thus improving the prediction performance of the combined training model (forest). It can be used to complete multi-classification tasks [17]. As shown in Fig. 2. the procedure of RF algorithm is as follows: firstly, n sample sets $[T_1, T_2, ..., T_n]$ are selected from the original training set T using bagging sampling; secondly, n decision tree models $\left[h_{T_1}, h_{T_2}, ..., h_{T_n}\right]$ are built for each of the n sample sets; thirdly, the same validation data x are used to test each of the n decision tree models to obtain n classification results $[R_1(x), R_2(x), ..., R_n(x)]$. Finally, the final classification is voted on according to the n classification results, following the principle of majority rule. The flow of RF model can be represented by the following equation:

$$H(x) = arg\,\max_{Y} \sum\nolimits_{i=1}^{n} I\left(h_{T_n}(x) = Y\right) \tag{4}$$

where $H(x)$ represents the final classification results from the vote, $I(^\circ)$ represents a schematic function, h_{T_n} represents a single decision tree classification model, x represents the validation set, and Y represents output classification.

When the RF model is used to predict the optimal BS and beam pair for the next moment, the prediction process can be expressed specifically as:

$$f[\theta, x_{test}] = \left[(I_{BS})_{opt}, \left(I_{beam}^{BS}\right)_{opt}, \left(I_{beam}^{UE}\right)_{opt}\right] \tag{5}$$

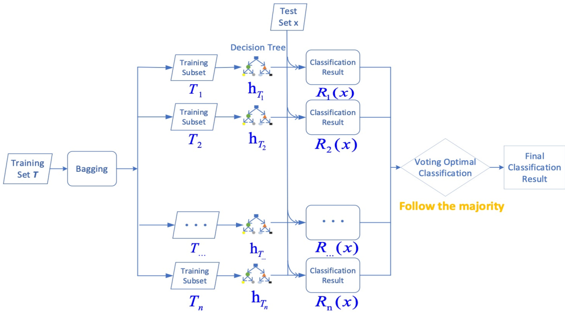

Fig. 2. RF processing flow

where θ represents some parameters of the model, such as the number of decision trees, the number of samples per node, etc. x_{test} represents the feature of validation data on the model input. Considering the difficulty of feature acquisition and the influence of each feature on the prediction effect, we investigate six candidate features, which are UE's location (L_{UE}); the beam index of the BS $\left(I_{beam}^{BS}\right)$; the RSRP value of the received signal $(RSRP_{rx})$; the distance between the UE and the BS $\left(d_{UE}^{BS}\right)$; the BS index (I_{BS}); the beam index of the UE $\left(I_{beam}^{UE}\right)$ [18]. Some of these features may contain similar nonlinear channel features, so the contribution of certain features to the model prediction accuracy may be the same, thus, the combination of some features in the feature set can be selected to reduce the training cost. The effect of different feature combinations on the model prediction accuracy is tested through simulation, and the feature combinations that correspond to higher prediction accuracy and smaller number of features are found out. Specifically, the combination of L_{UE}, I_{beam}^{BS}, $RSRP_{rx}$ performs better. The output of the model includes three prediction targets, i.e., the optimal BS index $\left((I_{BS})_{opt}\right)$, the optimal BS beam index $\left(\left(I_{beam}^{BS}\right)_{opt}\right)$, and the optimal UE beam index $\left(\left(I_{beam}^{UE}\right)_{opt}\right)$. These three targets will be transmitted to the BS and the UE, and used in the following beam switching process.

3.3 Generation of Data Sets

In order to complete the training and testing of the RF model, a large number of channel-related data sets are needed. Given that real channel measurement data during communication is difficult to obtain, some theoretical channel models are applied to generate channel samples close to real scenarios. Currently, there are two main approaches for mmWave channel modeling: Extended Saleh-Valenzuela (eSV) channel modeling and ray tracing channel modeling. The eSV channel model adopts the statistical modeling method and takes into account the clustering characteristics of mmWave and multi-antenna array structure, while the ray-tracing method calculates the amplitude, phase, delay, and polarization of each ray based on electromagnetic wave propagation theory. Given that the ray tracing modeling method is closer to the actual communication environment, Wireless Insite simulation software based on ray-tracing is used in this paper to generate the data sets of channel samples. Specifically, the data sets consist of numerical values, each data including six candidate features and three labels. All the BSs and all beam directions between BS and UE are traversed at each location of the UE, and the beam pair corresponding to the maximum RSRP at the UE is selected as the optimal beam pair which is used as the label of the dataset. To generate the data sets as close as possible to the real scenario, a random positioning error within 10 m is added to each L_{UE} sample.

4 Simulation Results

4.1 Simulation Scenario

Figure 3 shows a 2D view of the simulation scenario, which contains many buildings, 5 BSs and 4 UE's movement paths. The 5 BSs are located in the four corners and the center of the scenario, and 4 movement paths of the UE are randomly generated. Considering the UE's moving direction influences the choice of beam pairs, the UE's moving trajectory covers the main streets in both directions. The red triangles in the Fig. 3 indicate the sampling points during the UE's movement. The sampling interval is 1 s, i.e., the channel information is measured every second. Considering a downlink scenario, we set different BS transmitted power (P_t) in the interval of −20 dBm to 20 dBm in steps of 5 dBm. A total of 25000 data samples were generated under one P_t. Given the 9 P_ts of the BS, data set includes a total of 225000 data samples, of which the validation set accounted for 30% and the training set accounted for 70%. The UE's movement paths of the validation set and the training set are the same. However, due to the randomness of the noise and the positioning error, the $RSRP_{rx}$ and the L_{UE} are different in each piece of training or validation data. The other parameters are detailed in Table 1.

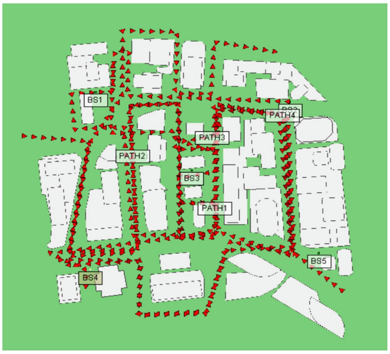

Fig. 3. A 2D view of the simulation scenario

Table 1. Simulation parameters

Parameter name	Value
Carrier type	Sine wave
Carrier frequency	28 GHz
Antenna number of base stations	128
Antenna number of users	32
Number of base stations	5
Number of user movement paths	4
User movement speed	15 m/s(54 km/h)
Sampling interval	15 m
Channel bandwidth	100 MHz
Antenna array	ULA
Noise power	−90 dBm

4.2 Results and Analysis

The performance metrics of the proposed method mainly include the model's prediction accuracy and the spectral efficiency (SE). Prediction accuracy is defined as:

$$Prediction\ accuracy = \frac{\sum_{k=1}^{N(Validationset)} I_k\left\{(index)_{pre} = (index)_{opt}\right\}}{N(Validation\ set)} \tag{6}$$

where $N(Validation\ set)$ represents the size of validation set. $I_k\{\}$ is an indicative function of the k-th sample of validation set, where 1 represents the conditions in the brackets are met, and 0 otherwise. $(index)_{pre}$ represents RF model's predicted index,$(index)_{opt}$ represents optimal index acquired by traversal, both of which include three prediction targets, as expressed in Eq. (5).

At first, we investigate the effect of different channel feature combinations on the model's prediction accuracy, where $\left(I_{beam}^{BS}\right)_{opt}$ is selected as the prediction target of the model. From Fig. 4, it can be seen that no matter which combination of features is chosen, model's prediction accuracy is improved with the increase of the P_t. Moreover, the more features selected for training, the better the prediction performance of the model. However, the difficulty of learning and training time will greatly go up with the increase of the number of the used features. Therefore, it is necessary to make a compromise between training cost and prediction accuracy. As shown in the figure, the model trained with the feature combination $(L_{UE}, I_{beam}^{BS}, RSRP_{rx})$ or the feature combination containing the triplet demonstrates the highest accuracy, while the prediction accuracy is lower for the feature combination without the beam index. That indicates the beam index contains the majority information of channel nonlinear characteristics. However, prediction accuracy is also very low if there are only $(I_{BS}, I_{beam}^{BS}, I_{beam}^{UE})$ in the combination. That indicates the diversity of the features also needs to be considered. Therefore, the triple feature combination $(L_{UE}, I_{beam}^{BS}, RSRP_{rx})$ is used in the following simulations.

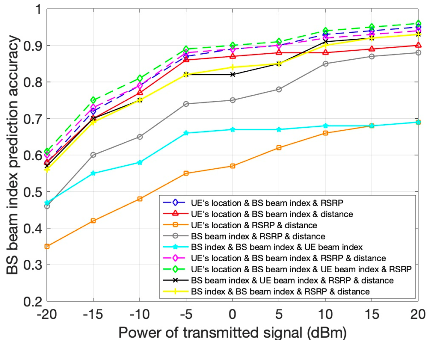

Fig. 4. Combinations of features

Figure 5 shows that prediction accuracy of more than 90% for 3 prediction targets when the P_t of BS is greater than 5_dBm. When the P_t is low, i.e., the signal-to-noise ratio (SNR) is low, a small increase of the P_t is a great boost to the prediction accuracy. However, when the P_t becomes higher, this boost effect gradually wears off. Among the three prediction targets, the model demonstrates the highest prediction accuracy for the $(I_{BS})_{opt}$ because the total number of categories for the I_{BS} is the least. On the contrary, the $\left(I_{beam}^{BS}\right)_{opt}$ prediction is the least accurate because it has the highest number of categories. The prediction accuracy of the $\left(I_{beam}^{UE}\right)_{opt}$ is in the middle. In addition, the size of the training set affects the performance. The prediction accuracy increases with the increasing size of training data and stabilizes at over 90% with more than 40,000 samples.

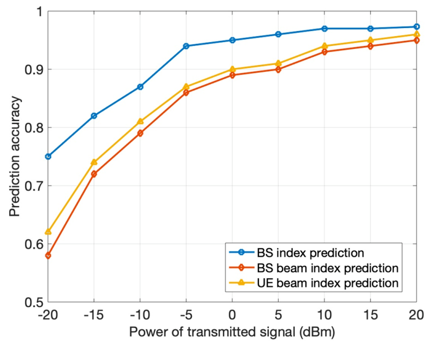

Fig. 5. The prediction accuracy for $(I_{BS})_{opt}$, $\left(I_{beam}^{BS}\right)_{opt}$, and $\left(I_{beam}^{UE}\right)_{opt}$

Figure 6 shows that the average SE corresponding to the beam pair via traversal search and RF-based prediction. Given a specific noise power, with the increase in P_t, the SE of the proposed method gradually approximates the optimal beam. Furthermore, the SE gap between the proposed method and the exhaustive beam sweeping is small, although the accuracy of prediction at low P_t is not very high as shown in Fig. 5. These results are reasonable since the received SNR difference between the suboptimal and optimal beam pairs is tiny in many cases owing to the narrow beam-width and clustering characteristics of mmWave channel. The prediction based on RF may be not optimal, but still suboptimal and supports a relatively high received SNR.

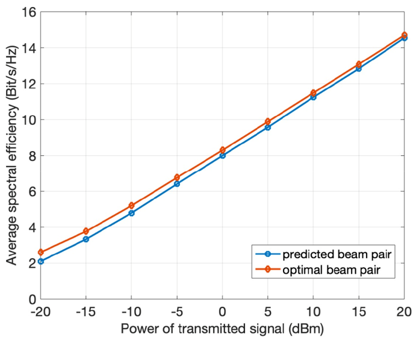

Fig. 6. Average SE comparison

To sum up, the accuracy of the model in predicting the beam depends mainly on the signal strength and the combination of features used to train the model. Moreover, although the predicted beam pair are not optimal at low transmit power, the achieved SE is still quite close to the upper bound.

5 Conclusion

In this paper, we propose a method to optimize the efficiency of beam switching in high-speed mobile scenarios by using a RF model. The core idea is to learn the nonlinear features of the channel by machine learning based on the correlation between the adjacent time steps of the channel in the fixed scenario. Therefore, the beam pair at the next moment can be predicted based on the information of the previous moment.

Simulation results show that the spectral efficiency of the proposed method is close to the upper bound, and the prediction accuracy can reach more than 90% within a typical BS transmitted power range. What's more, the proposed method reduces frequent signaling interaction to determine the optimal beam pair, and improves the efficiency of beam switching with extremely short time delay.

Acknowledgement. This work is supported by Beijing Natural Science Foundation under Grant No. L202003, the National Natural Science Foundation of China under Grant No. 61971069, 61801051, and the Open Project of A Laboratory under Grant No. 2017XXAQ08.

References

1. Weixia, Z., Chao, G., Guanglong, D., Zhenyu, W., Ying, G.: A new codebook design scheme for fast beam searching in millimeter-wave communications. China Communications **11**(6), 12–22 (June 2014)

2. Yang, L., Ma, S., Yang, H., Tan, H.: A hierarchical beam search algorithm with betterperformance for millimeter-wave communication. In: 2019 2nd World Symposium on Communication Engineering (WSCE*)*, pp. 16–20, Nagoya (2019)
3. Lin, J., An, W.: A New Initial Beam Search Scheme in 5G New Radio. In: 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), pp. 182–186, Xiamen (2019)
4. Lin, F., Yang, J., Wang, Y., Li, J.: Dynamic Beam Search Scheme for Mobile Scenarios in Millimeter Wave Communication. In: ICC 2020 - 2020 IEEE International Conference on Communications (ICC), pp. 1–6, Dublin (2020)
5. Maiberger, R., Ezri, D., Erlihson, M.: Location based beamforming. In: 2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel, pp. 000184–000187, Eliat (2010)
6. Chen, X., Wei, Z., Zhang, X., Sang, L.: A beamforming method based on image tracking and positioning in the LOS scenario. In: 2017 IEEE 17th International Conference on Communication Technology (ICCT), pp. 1628–1633, Chengdu (2017)
7. Talvitie, J., et al.:Positioning and location-based beamforming for high speed trains in 5G NR networks. In: 2018 IEEE Globecom Workshops (GC Wkshps), pp. 1–7, Abu Dhabi (2018)
8. Lu, Y., Koivisto, M., Talvitie, J., Valkama, M., Lohan, E.S.: Positioning-aided 3D beamforming for enhanced communications in mmWave mobile networks. IEEE Access **8**, 55513–55525 (2020)
9. Miao, W., Luo, C., Min, G., Wu, L., Zhao, T., Mi, Y.: Position-Based Beamforming Design for UAV Communications in LTE Networks. ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1–6, Shanghai (2019)
10. Kwon, H.J., Lee, J.H., Choi, W.: Machine learning-based beamforming in K-user MISO interference channels. Access IEEE **9**, 28066–28075 (2021)
11. Na, W., Bae, B., Cho, S., Kim, N.: Deep-learning Based Adaptive Beam Management Technique for Mobile High-speed 5G mmWave Networks. In: 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin), pp. 149–151, Berlin (2019)
12. Wang, R., et al.: Reinforcement learning method for beam management in millimeter-wave networks. In: 2019 UK/ China Emerging Technologies (UCET), Glasgow (2019)
13. Hao, Y.: Research on beam management technology based on millimeter wave distributed antenna array. In: Master's thesis of Beijing University of Posts and Telecommunications, June 2019
14. Zhaoqiang, L.: Research on beam search algorithm in hybrid beamforming system. In: Master's thesis of Beijing University of Posts and Telecommunications, March 2018
15. Yang, D., Yang, L.L., Hanzo, L.: DFT-based beamforming weight vector codebook design for spatially correlated channels in the unitary precoding aided multiUE downlink. In: Proc. IEEE International Conference Communications, vol. 1, pp. 1–5, Cape Town (2010)
16. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
17. Cutler, A., Cutler, D.R., Stevens, J.R.: Random forests[M]//Ensemble machine learning, pp. 157–175. Springer, US (2012)
18. Ekman, B.: Machine Learning for Beam Based Mobility Optimization in NR. Thesis of Linkoping University, Swedn (2017)