Pilot Allocation Scheme Based on Machine Learning Algorithm and Users' Angle of Arrival in Massive MIMO System

Min Yu(✉), Si Yuan Li, and Dong Feng Chen

School of Communication and Information Engineering,
Chongqing University of Posts and Telecommunications, Chong Qing, China
120588660@qq.com

Abstract. Massive MIMO system has attracted attention due to it's significant improvement in system capacity and spectrum utilization. Pilot pollution greatly limited the performance of Massive MIMO system. To optimize the pilot pollution in Massive MIMO system. In this paper, a pilot allocation scheme based on machine learning algorithm and users' angle of arrival is proposed. The scheme firstly classified all users according to whether the users' angle of arrival overlaps with each other. It randomly assigned pilot sequences to users whose angle of arrival do not overlap with each other. Secondly, it used machine learning algorithm to classify users whose angle of arrival overlap with each other into interfering group and non-interfering groups based on users' location information. We assign orthogonal pilots to users in the interfering group and randomly assign pilot sequences to users in the non-interfering group. Simulation results show that when the number of antenna reached 300, the pilot efficiency can be increased by about 11.67%. The pilot allocation scheme proposed in this paper can effectively suppress the impact of pilot pollution on the performance of Massive MIMO system, improve pilot efficiency and reduce pilot overhead.

Keywords: Pilot allocation · K-means clustering · Users' angle of arrival · Massive MIMO

1 Introduction

Massive MIMO system is one of the key technologies of Fifth Generation (5G) mobile communication systems. It is configured a large number of antennas on the base station side. The system capacity can be greatly increased. It brings high energy efficiency and high spectrum efficiency [1]. Massive MIMO technology has attracted much attention because of these advantages. However, the above-mentioned advantages of massive MIMO technology depend on being able to correctly analyze the channel state information. To accurately estimate the uplink channel, terminals should use mutually orthogonal pilot sequences. However, due to the surge in the number of antennas, the number of orthogonal pilots is limited by the channel coherence time. It leads to a limited number of orthogonal pilots. The pilot sequence is inevitably reused between neighboring cells. The interference caused by non-orthogonal pilot sequences is called

H. Gao et al. (Eds.): ChinaCom 2020, LNICST 352, pp. 753–765, 2021.
https://doi.org/10.1007/978-3-030-67720-6_52

pilot pollution. Pilot pollution greatly limits the system performance. Therefore, it is particularly important to study how to reduce the impact of pilot pollution. In order to solve the problem, many scholars have done a lot of research. N. Akbar et al. proposed to use line-of-sight interference between users to classify users, and assign non-orthogonal pilots to users with low line-of-sight interference. It can suppress pilot pollution to a certain extent [2]. Zhu, X. et al. proposed that according to certain standards, each cell is divided into a central area and an edge area, users in the edge area are assigned orthogonal pilot sequences, users in the central area reuse the pilot sequences [3]. Although pilot pollution can be reduced in this way, the gain of pilot multiplexing will become smaller. X. Nie et al. proposed a location-aware pilot assignment algorithm, which used the line-of-sight interference between users to classify users, assigned non-orthogonal pilot sequences to users with low line-of-sight interference, assigned orthogonal pilot sequences to users with high line-of-sight interference. It can suppress pilot contamination to a certain extent [4]. Kim, K. et al. proposed to distinguish users by defining users' similarity, thereby establishing the basis of distribution. They designed the angle of arrival positioning method to obtain users' location information, monitored and estimated the distance between users in different cells based on their location information to represent the potential interference of the transmission power between users. According to the similarity and distance between users, the pilot allocation strategy is implemented [5]. It can effectively reduce the problem of pilot pollution and improve pilot efficiency. X. Zhao et al. proposed to use machine learning algorithm in the pilot allocation strategy, used the exhaustive method to obtain the optimal pilot allocation scheme as the training sequence, and used the training sequence to obtain the pilot allocation model [6]. However, the complexity of this method is relatively high. Zhu Xu dong et al. proposed a low-complexity pilot allocation scheme based on matching algorithm. This scheme divided the users of the target cell into a strong user group and a weak user group and used a minimum-maximum matching method for pilot allocation. In the weak user group, the Hungarian algorithm is used for pilot allocation. The pilot allocation method is designed to ensure the fairness of strong user groups [7]. S. Ma et al. proposed a pilot allocation scheme that used the Signal to Interference Noise Ratio (SINR) of harmonic to quantify the fairness of all users in the network [8]. On the other hand, T. Van Chien et al. proposed a pilot allocation scheme, which did not model the pilot design as a combined allocation problem, but used basic pilot and treated the correlation power coefficient as a continuous optimization variable [9]. In order to reduce the pilot overhead in the massive MIMO system, X. Xiong et al. suggested that pilot is reused by using the characteristic that the energy of the channel is concentrated in a small number of specific areas of the angle domain and the delay domain. When there is pilot pollution in the system, the solution analyzed the asymptotic behavior and designed an interference cancellation (IC) precoder to ensure the user's Qos to the greatest extent [10]. C. Hu et al. combined pilot allocation and semi-blind channel estimation and proposed a sector-based pilot allocation method, which includes cross-sector pilot allocation and intra-sector pilot optimization. When there are a large number of users in each cell, the method can reduce the complexity of searching for optimal pilot allocation [11].

In massive MIMO system, pilot pollution greatly limited the performance of the massive MIMO system. The methods to mitigate pilot pollution are mainly from the following three aspects: pilot allocation, precoding and channel estimation. Pilot allocation is one of the mainstream methods.

The main contributions in this paper are listed as follows:

a. Different from the previous way, the solution proposed in this article grouped users multiple times. According to whether the users' angles of arrival overlapped with each other, users were divided into n + 1 categories. The first n categories represented users whose angles of arrival did not overlap, and the n + 1-th category represented users whose angles of arrival partially overlapped or completely overlapped.
b. Because of the large number of users in massive MIMO systems, it used an unsupervised machine learning algorithm, which suitable for massive data processing—K-means clustering algorithm to group the n + 1th users into two groups according to the interference cost. The function measured the amount of interference between users. It divided users with high interference into interfering group and users with low interference into non-interfering group. Finally performed different pilot allocation schemes for users in different groups.
c. Grouping users multiple times to reduce the number of users processed in each step. Compared with the previous pilot allocation scheme, it reduced pilot overhead to a certain extent.

The remainder of this paper is organized as follows. Section 2 describes the system model. Section 3 introduces pilot allocation scheme. Section 4 analyzes experimental results. Finally, conclusions are given in Sect. 5.

2 System Model

It is configured a large number of antennas on the base station side. We consider the UL of a multi-cell Massive MIMO system with L cells. Each cell consists of a BS equipped with M antennas that serves K single-antenna users. $\mathbf{h}_{(j,k)l} \in C^{M\times 1}$ is the channel vector from the k-th user in the j-th cell to the BS in the l-th cell. The channel is modeled using shadow fading in a uniform linear massive MIMO antenna array as:

$$\mathbf{h}_{(j,k)l} = \sqrt{\frac{\beta_{(j,k)l}}{p}} \sum_{p=1}^{P} a(\theta^{p}_{(j,k)l}) \partial^{p}_{(j,k)l} \tag{1}$$

Where P is the number of multipath bars $a(\theta) = [1, e^{-2\pi jd/\lambda_c \sin\theta}, \ldots, e^{-j2\pi d(M-1)/\lambda_c \sin\theta}]^T$ is the antenna array vector in the direction θ. $\partial^{p}_{(j,k)l} \sim CN(0,1)$ is the complex Gaussian gain on the P-th path of the signal. d is the spacing between antennas λ_c is signal wavelength. $\beta_{(j,k)l}$ is a large-scale fading factor, which includes path loss and shadow fading. It can be expressed as:$\beta_{(j,k)l} = z_{(j,k)l}/r^{\partial}_{(j,k)l}.r_{(j,k)l}$ is the distance between the k-th user in the j-th cell and the base station that in the l-th cell. ∂ is path loss fading coefficient. $z_{(j,k)l}$ is shadow fading factor.

2.1 Uplink Pilot Transmission Process

During the uplink pilot transmission process, the signal $\mathbf{y}_l \in \mathbb{C}^{M\times\tau}$ received by the base station in the l-th cell can be expressed as:

$$\mathbf{y}_l = \sqrt{\rho_r \tau} \sum_{l=1}^{L} \sum_{k=1}^{K} \varphi_k \mathbf{h}_{(i,k)l}^T + \mathbf{n}_l \tag{2}$$

Where ρ_r is the average power sent by the user. $\mathbf{h}_{(i,k)l}$ is the channel vector from the kth user in the i-th cell to the base station in the l-th cell. $l = 1, 2, \ldots, L$. $n_l \in \mathbb{C}^{M\times\tau} \sim CN(0, \delta^2)$ is additive white gaussian noise. $\varphi_k \in \mathbb{C}^{\tau\times 1}$ is pilot sequence, which satisfy $\varphi_k^H \varphi_{k'} = \boldsymbol{\sigma}_{kk'}$ and $\boldsymbol{\sigma}_{kk'} = \begin{cases} 1, k=k' \\ 0, k\neq k' \end{cases}$ When $\sqrt{\rho_r \tau} = 1$, the base station received signal in the l-th cell can be expressed as:

$$\mathbf{y}_l = \mathbf{P}_l \mathbf{h}_{i,l} + \mathbf{n}_l \tag{3}$$

Where $\mathbf{P}_l = [\varphi_1, \varphi_2, \ldots, \varphi_K]^T \in C^{\tau\times K}$ is the pilot sequence sent by the users in the l-th cell $\mathbf{h}_{i,l} \in C^{K\times M}$ represents the channel state information matrix. In the massive MIMO system, the signal received by the base station in the l-th cell can be expressed as:

$$\mathbf{y}_l = \mathbf{P}_l \mathbf{h}_{l,l} + \sum_{i=1, i\neq l}^{L} \mathbf{P}_l \mathbf{h}_{i,l} + \mathbf{n}_l \tag{4}$$

Using the Least squares (LS) channel estimation algorithm to obtain the channel estimation matrix of the base station which in the l-th cell for the users of this cell:

$$\hat{\mathbf{h}}_{l,l} = \mathbf{P}_l^{-1} \mathbf{y}_l = \sum_{k=1}^{K} \mathbf{h}_{(l,k)l} + \sum_{k=1}^{K} \sum_{i=1, i\neq l}^{L} \mathbf{h}_{(i,k)l} + \mathbf{n}'_l \tag{5}$$

It can be seen from formula (5) that the base station in l-th cell obtained the CSI of the users in the l-th cell and the CSI of the users in other cell who sent the same pilot. Pilot pollution is caused by users in other cells which sent the same pilot.

3 Pilot Allocation Scheme Based on Machine Learning and User Arrival Angle

We proposed a massive MIMO pilot allocation scheme based on machine learning algorithm and users' angle of arrival. Firstly, it according to whether the users' angle of arrival overlaps to classify users. The users whose angles of arrival are non-overlapping were classified into the first n sets. The remaining users were classified into the n + *1*-th set. For the n + *1*-th set, the unsupervised machine learning algorithm– K-means clustering algorithm was used to cluster users according to their location information.

It further divided users into interfering group and non-interfering group. Finally, different pilot allocation schemes were performed for users in different groups. For the first n sets, the pilot sequences were randomly allocated. Pilot sequences were multiplexed in each user set. Orthogonal pilot sequences were allocated for users in the interfering group and non-orthogonal pilot sequences were randomly allocated for users in the non-interfering group.

3.1 Classify Users Based on Their Angles of Arrival

Firstly we classified users according to whether their angles of arrival overlapped with each other. The model of users' angle of arrival was shown in the Fig. 1:

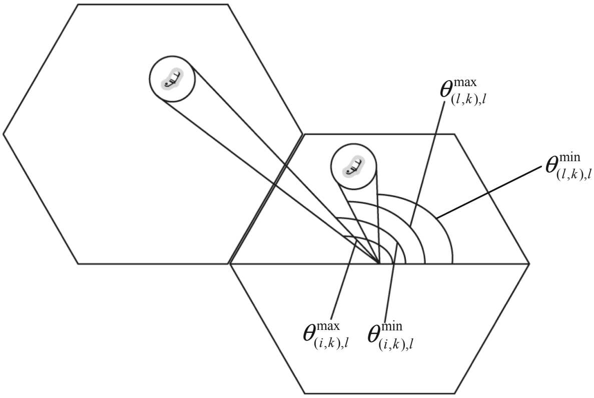

Fig. 1. User angle of arrival

$$\theta^{\mu}_{(i,k),l} = \arctan(\frac{[x_{(i,k)}]_2 - [x_l]_2}{[x_{(i,k)}]_1 - [x_l]_1}) \tag{6}$$

$$\theta^{\delta}_{(i,k),l} = \arcsin(\frac{r}{\left\| x_{(i,k)} - x_l \right\|}) \tag{7}$$

Where $\theta^{\min}_{(i,k),l} = \theta^{\mu}_{(i,k),l} - \theta^{\delta}_{(i,k),l}$, $\theta^{\max}_{(i,k),l} = \theta^{\mu}_{(i,k),l} + \theta^{\delta}_{(i,k),l}$. And r is the user's scattering radius. From theorem 1 of [6], we can see that when the angle of arrival of the interfering user $[\theta^{\min}_{(i,k),l}, \theta^{\max}_{(i,k),l}]$ does not overlap with the angle of arrival of the expected user $[\theta^{\min}_{(l,k),l}, \theta^{\max}_{(l,k),l}]$ at all. Even if the used pilot sequence is non-orthogonal, there will be no interference. Users are classified according to whether the angles of arrival overlap with each other. There are n + 1 categories, in which $C_1, C_2, \ldots, C_n$ is the users whose angles of arrival do not overlap each other and C_{n+1} is the users whose angles of arrival overlap partially or completely overlap.

3.2 User Classification Based on Machine Learning Algorithm

For users in C_{n+1}, the K-means clustering algorithm was further used to classify users based on users' location information. The users in C_{n+1} users were further divided into interference groups and non-interference groups.

- Set the cost function

We assumed that the angle of arrival of the target user is θ_a, the angle of arrival of the interfering user is θ_b, the signal wavelength is λ, the number of antennas at the base station is M and the antenna spacing is D. The target user's guidance vector of the Pth path is:

$$\partial(\theta_a^p) = \sum_{m=1}^{M} \exp(2\pi i(m-1)\frac{D}{\lambda}\cos(\theta_a^p)) \tag{8}$$

The interfering user's guidance vector of the Pth path is:

$$\partial(\theta_b^p) = \sum_{m=1}^{M} \exp(2\pi i(m-1)\frac{D}{\lambda}\cos(\theta_b^p)) \tag{9}$$

The distance between the target user and the interfering user is:

$$d_{ab} = \|x_a - x_b\| \tag{10}$$

We defined the interfering cost function as:

$$\begin{aligned} J_{ab} &= \frac{1}{d_{ab}M}\partial(\theta_a^p)\partial(\theta_b^p) \\ &= \frac{1}{d_{ab}M}\sum_{m=1}^{M} \exp(2\pi i(m-1)\frac{D}{\lambda}(\cos(\theta_a^p) - \cos(\theta_b^p))) \end{aligned} \tag{11}$$

When the target user's angle of arrival and the interfering user's angle of arrival partially overlapped or completely overlapped. There was $\cos(\theta_a^p) = \cos(\theta_b^p)$. When the number of antenna port on the base station side was large enough: there was $\lim_{M\to\infty} J_{ij} = \frac{1}{d_{ij}}$. Therefore, when the interfering user's angle of arrival and the target user's angle of arrival partially overlapped or completely overlapped, the amount of interference from the interfering user to the target user was inversely proportional to its relative distance. We assumed that: $\cos(\theta_i^p) = \cos(\theta_j^p) = 1$, $d \in [10, 2000]$, $M = 64$. The simulation results were as shown in the Fig. 2:

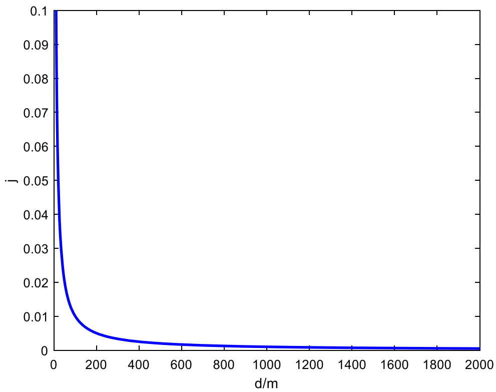

Fig. 2. Interference cost function

It can be seen from the figure that when $d \approx 1000$ and $J \approx 0.002$, the amount of interference between users is almost negligible. Based on the above results, for set C_{n+1} users, it further used the unsupervised machine learning algorithm–K-means clustering algorithm to classify users according to users' location information. The users were further divided into K clusters. These clusters with a distance between cluster centers greater than 1000 m were divided into non-interfering group, and these clusters with a distance between cluster centers less than 1000 m were divided into interfering groups.

- K-means clustering algorithm

The K-means clustering algorithm included: a user space position acquisition unit, a cluster center calculation unit and a convergence decision unit. Firstly, the central control unit obtained the user's location information. Secondly, the cluster center calculation unit calculated the cluster center based on the position information. Finally, the convergence decision unit judges whether to output the clustering result. The K-means clustering algorithm is shown in Table 1:

The steps of K-means clustering algorithm are as follows:

1. Initial input: K is the number of clusters. ε is a minimum value used to determine when the clustering ends. C_{n+1} is the user's location information who need to cluster.
2. First step, randomly selected the location information of K users as the initial clustering center. And the preferred value of K was set to 6, because the number of neighboring cells of each cell is 6. (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3), (x_4, y_4, z_4), (x_5, y_5, z_5), (x_6, y_6, z_6), as the initial clustering center.
3. According to the distance between the set C_{n+1} users and the cluster center, $d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$, the users were clustered into the cluster with the closest cluster center. And update the cluster center: $C_j = \overline{D_j}$.
4. Third step calculated the objective function to determine whether the two clustering centers have converged. If they have converged, then end to the clustering;

Table 1. K-means clustering algorithm

	K-means clustering algorithm
1:	**input**: K, ε, C_{n+1}
2:	**Parameter initialization**: Clustering center $C_1=(X_1,Y_1,Z_1),\ldots,C_K=(X_K,Y_K,Z_K)$
3:	**Calculate the distance of each user relative to the cluster center**: i is the cell index, j is the user index $d[i][j]=\sqrt{(X_i-X_j)^2+(Y_i-Y_j)^2+(Z_i-Z_j)^2}$
4:	**Attributing the user with the smallest distance to a cluster**: $D_j=\min(d[i])\bigcup(X_j,Y_j,Z_j)$
5:	**Update cluster center**: $C_j=\overline{D_j}$
6:	**Calculate decision parameters** E_i: $E_i=\sum_{i=1}^{6}\sum_{x\in C_i}\lVert x-u_i\rVert^2 \quad u_i=\frac{1}{\lvert C_i\rvert}\sum_{x\in C_i}x$
7:	**Stop iteration condition**: $\lvert E_{i+1}-E_i\rvert<\varepsilon$
	output: C_j

otherwise, update the clustering center and return to step 2 to re-cluster according to the current clustering center. The objective function is: $E=\sum_{i=1}^{6}\sum_{x\in C_i}\lVert x-u_i\rVert^2$, $u_i=\frac{1}{\lvert C_i\rvert}\sum_{x\in C_i}x$. Where u_i is the mean vector of cluster C_i and the distance from each sample point to the mean point. Determine whether the last two adjacent objective functions $\lvert E_2-E_1\rvert<\varepsilon$ hold. Where ε is a minimum value. And the difference between the criterion functions of two adjacent iterations is less than a minimum value, which indicated that the sum of squared errors within the cluster has converged and the clustering end.

- Pilot allocation

Pilot sequences were randomly allocated to the first n types users and each group of pilot sequences was multiplexed in each user set. The users in the interfering group were assigned strictly orthogonal pilot sequences and the users in the non-interfering group were randomly assigned pilot sequences.

4 Performance Analysis

It can be seen from formula (5) that the base station used zero-forcing detection algorithm to receive the signal can be expressed as:

$$\begin{aligned} x_{l,k} &= (h_{(l,k)l})^H (\sum_{i=1}^{L} \sum_{k=1}^{K} h_{(i,k)l} x_{i,k} + N_l) \\ &= (h_{(l,k)l})^H h_{(l,k)l} x_{l,k} + (h_{(l,k)l})^H \sum_{i=1, i \neq l}^{L} \sum_{k=1}^{K} h_{(i,k)l} x_{i,k} \\ &+ (h_{(l,k)l})^H N_l \end{aligned} \tag{12}$$

Where $(h_{(l,k)l})^H h_{(l,k)l} x_{l,k}$ is the desired signal $(h_{(l,k)l})^H \sum_{i=1, i \neq l}^{L} \sum_{k=1}^{K} h_{(i,k)l} x_{i,k}$ is interfering signal. $(h_{(l,k)l})^H N_l$ is noise. The uplink SINR can be expressed as:

$$SINR_{l,k} = \frac{|E\{(h_{(l,k)l})^H h_{(l,k)l}\}|^2}{\sum_{i=1, i \neq l}^{L} \sum_{k=1}^{K} |E\{(h_{(l,k)l})^H h_{(i,k)l}\}|^2 + \delta^2 |E\{(h_{(l,k)l})^H\}|^2} \tag{13}$$

Users' uplink reachable rate is:

$$R_{<l,k>} = E\{\log_2(1 + SINR_{(l,k)})\} \tag{14}$$

The spectrum efficiency of the lth cell can be expressed as:

$$SE_l = (1 - \mu) \sum_{k=1}^{K} R_{<l,k>} \tag{15}$$

Where $\mu = \frac{\tau}{T}$ represents the loss of frequency efficiency due to uplink pilot transmission. τ is the pilot length and T is the channel coherence time. The pilot efficiency is defined as: $\gamma = \frac{R}{O}$ $R = \sum_{l=1}^{L} \sum_{k=1}^{K} R_{<l,k>}$ is the uplink reach sum rate. O represents the number of orthogonal pilot sequences used by the algorithm in the multi-cell massive MIMO system.

5 Simulation Analysis

5.1 Simulation Parameter Setting

In order to better analyze the performance of the pilot allocation scheme proposed in this paper, we conducted simulation analysis based on the MATLAB simulation platform. The simulation used a massive MIMO system composed of L cells in TDD mode. Each cell center had a base station configured with M antennas and containing K randomly distributed users. The simulation parameters were set as follows (Table 2):

Table 2. Simulation parameters

Simulation parameters	Set value
Number of cells L	7
Number of antennas M	$100 \sim 500$
Number of users K	32
Cell radius $/m$	500
SNR $/dB$	20
Shadow fading coefficient	8
Pilot length τ	5
Path loss index	3

5.2 Analysis of Simulation Results

We assumed that the positions of users in a cell follow a random distribution. Random distribution of user positions can better simulate the actual situation and verify the universality of the algorithm. The location of users in a cell in the system is shown in Fig. 3:

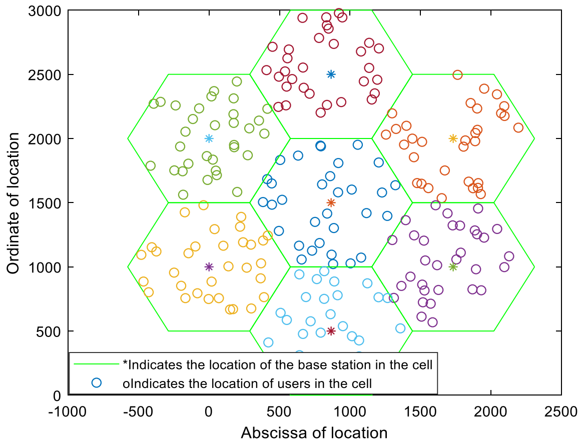

Fig. 3. User Distribution

Figure 4 shows the pilot efficiency of different pilot allocation algorithms with different numbers of antennas. It can be seen that the pilot allocation algorithm proposed in this paper further performed K-means clustering algorithm to classify C_{n+1} users from the pilot allocation scheme that based on users' arrival of angle. And different pilot allocation schemes were performed for different groups of users. Compared with the pilot allocation scheme based on users' angle of arrival, the scheme proposed in this paper has higher pilot efficiency and less pilot overhead. Compared with the pilot allocation scheme based on large-scale fading factors, the scheme proposed in this paper is more refined in user grouping. Not only does the user's location information be used for grouping, but also the user's arrival angle is used for grouping, which greatly improved the pilot efficiency.

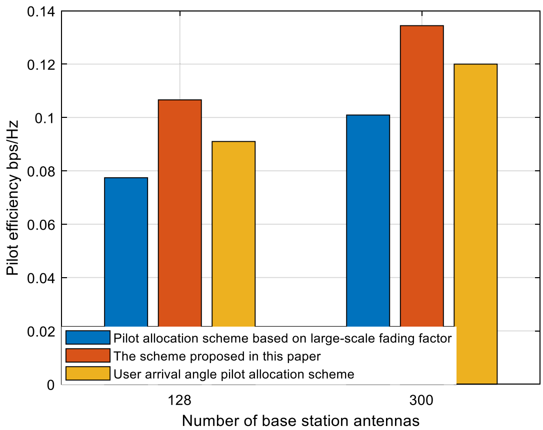

Fig. 4. Pilot efficiency of different pilot allocation algorithms with different numbers of antennas

Figure 5 shows that the users' minimum uplink reachable rate varies with the numbers of antennas on the base station side. The simulation experiments in this paper compared the simulation results of the pilot allocation scheme with the random pilot allocation scheme and the pilot allocation scheme based on user arrival angle. When the number of antennas is about 100 to 200, the users' minimum uplink reachable rate increases almost linearly with the number of antennas. When the number of antennas gradually increases, the growth rate of the users' minimum uplink reachable rate slows down.

The performance of the massive MIMO system will not be infinitely improved with the increase of the number of antennas at the base station side and the performance of the massive MIMO system will approach a fixed value. Under the condition of the same number of antennas, the minimum uplink achievable rate of the randomly allocated pilot scheme is the smallest, followed by the pilot allocation scheme proposed in this paper, and the minimum uplink achievable rate of the pilot allocation scheme based on user arrival angle is the largest. Because of the scheme proposed in this paper further clustered C_{n+1} users based on the users' grouping result of the pilot assignment algorithm based on the user arrival angle. And it further divided the C_{n+1} into interfering group and non-interfering group. The users in the interfering group multiplexed the pilot sequence. On the one hand, compared with the pilot allocation scheme based on user arrival angle, the scheme proposed in this paper further clustered users and further performed pilot multiplexing to reduce pilot overhead. On the other hand, compared with the random pilot allocation scheme, the pilot allocation scheme proposed in this paper greatly increased the minimum uplink reachable rate of users, reduced the impact of pilot pollution on system performance and improved system performance. The pilot allocation scheme proposed in this paper makes a compromise between improving system performance and reducing pilot overhead. Under the condition of acceptable system performance loss, the pilot efficiency is effectively improved and the pilot overhead is reduced.

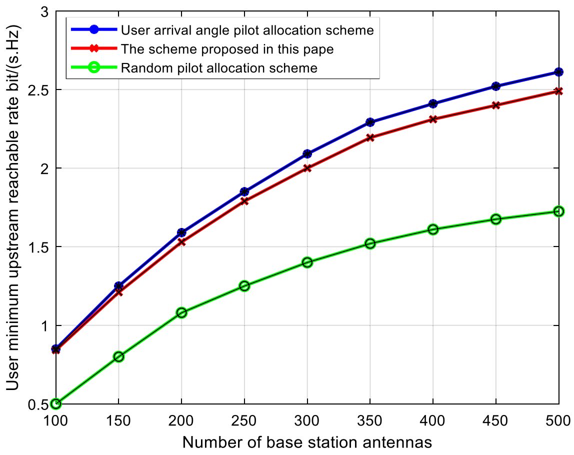

Fig. 5. Users' minimum uplink reachable rate varies with the numbers of antennas

6 Conclusion

This paper proposed a pilot allocation scheme based on machine learning algorithm and users' angle of arrival in a massive MIMO system. The scheme includes: user arrival angle grouping, K-means clustering grouping and pilot allocation. The scheme firstly classifies users according to whether their angles of arrival overlap with each other. It classifies non-overlapping users into the first n sets and classifies the remaining users into n + 1th sets. For users in the n + 1 set, the unsupervised machine learning algorithm is used to cluster users according to their location information. The users in the n + 1 set are further divided into interfering group and non-interfering group. Finally, different pilot allocation schemes are performed for users in different groups. For the first n sets, pilot sequences are randomly allocated and pilot sequences are multiplexed in each set. Orthogonal pilots are allocated to users in the interfering group. Non-orthogonal pilot sequences are randomly allocated to users in the non-interfering group. Simulation results show that, on the one hand, the pilot allocation scheme based on machine learning algorithm and users' angle of arrival proposed in this paper can effectively suppress the impact of pilot pollution on the performance of massive MIMO systems and effectively improve system performance compared with random pilot allocation schemes. On the other hand, compared with the pilot allocation scheme based on user arrival angle, the pilot allocation scheme based on machine learning algorithm and users' angle of arrival proposed in this paper can effectively improve the pilot efficiency and reduce the pilot overhead. It is suitable for massive MIMO systems with extremely tight pilot resources.

References

1. Rajmane, R.S., Sudha, V.: Spectral efficiency improvement in massive MIMO systems. In: 2019 TEQIP III Sponsored International Conference on Microwave Integrated Circuits, Photonics and Wireless Networks (IMICPW), Tiruchirappalli, India, pp. 357–360 (2019)
2. Akbar, N., Yan, S., Yang, N., Yuan, J.: Location-aware pilot allocation in multicell multiuser massive MIMO networks. IEEE Trans. Veh. Technol. **67**, 7774–7778 (2018)
3. Zhu, X., Wang, Z., Qian, C., et al.: Soft pilot Reuse and multi-cell block diagonalization precoding for massive MIMO systems. IEEE Trans. Veh. Technol. **65**(5), 3285–3298 (2015)
4. Nie, X., Zhao, F.: Joint pilot allocation and pilot sequence optimization in massive MIMO systems. IEEE Access **8**, 60637–60644 (2020)
5. Kim, K., Lee, J., Choi, J.: Deep learning based pilot allocation scheme (DL-PAS) for 5G massive MIMO systems. IEEE Commun. Lett. **22**(4), 828–831 (2018)
6. Zhao, X., Yao, Y., Xu, L., et al.: A Low-complexity and high-efficient pilot allocation scheme based on user grouping in massive MIMO system. In: 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, pp. 658–663 (2019)
7. Xudong, Z., Lingdong, D., Zhaocheng, W., et al.: Weighted-graph-coloring-based pilot decontamination formulticell massive MIMO systems. IEEE Trans. Veh. Technol. **66**(3), 2829–2834 (2017)
8. Ma, S., Xu, E.L., Salimi, A., Cui, S.: A novel pilot assignment scheme in massive MIMO networks. IEEE Wirel. Commun. Lett. **7**(2), 262–265 (2018)
9. Van Chien, T., Björnson, E., Larsson, E.G.: Joint pilot design and uplink power allocation in multi-cell massive MIMO systems. IEEE Trans. Wirel. Commun. **17**(3), 2000–2015 (2018)
10. Xiong, X., Jiang, B., Gao, X., You, X.: QoS-guaranteed user scheduling and pilot assignment for large-scale MIMO-OFDM systems. IEEE Trans. Veh. Technol. **65**(8), 6275–6289 (2016)
11. Zhu, X., Dai, L., Wang, Z.: Graph coloring based pilot allocation to mitigate pilot contamination for multi-cell massive MIMO systems. IEEE Commun. Lett. **19**(10), 1842–1845 (2015)
12. Raharya, N., Hardjawana, W., Al-Khatib, O., et al.: Pursuit learning-based joint pilot allocation and multi-base station association in a distributed massive MIMO network. IEEE Access **8**, 58898–58911 (2020)
13. He, X., Song, R., Zhu, W.: Pilot allocation for distributed-compressed-sensing-based sparse channel estimation in MIMO-OFDM systems. IEEE Trans. Veh. Technol. **65**(5), 2990–3004 (2016)
14. Ruoyu, Z., Honglin, Z., Jiayan, Z., et al.: Hybrid orthogonal and non-orthogonal pilot distribution based channel estimation in massive MIMO system. J. Syst. Eng. Electron. **29**(5), 881–898 (2018)
15. Larsson, E.G., Edforsn, O., Tufvesson, F., et al.: Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)