Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 39, S. 71-75, Temmuz 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 39, pp. 71-75, July 2022

Research Article

Machine Learning Based Classification Algorithm for AP Selection in Cell-Free MIMO Systems

Mert Demirel[1*] Esra Aycan Beyazıt[2]

[1*] Izmir Katip Celebi University, Faculty of Engineering and Architecture, Departmant of Electrical and Electronics Engineering, Izmir, Turkey, (ORCID: 0000-0001-7905-8456), mdemirel97@gmail.com

[2] Izmir Katip Celebi University, Faculty of Engineering and Architecture, Departmant of Electrical and Electronics Engineering, Izmir, Turkey, (ORCID: 0000-0003-1035-6695), esra.aycan@ikcu.edu.tr

(3rd International Conference on Applied Engineering and Natural Sciences ICAENS 2022, July 20-23, 2022)

(**DOI:** 10.31590/ejosat.1142546)

ATIF/REFERENCE: Demirel, M., Beyazıt, E.A. (2022). Machine Learning Based Classification Algorithm for AP Selection in Cell-Free MIMO Systems. *European Journal of Science and Technology*, (39), 71-75.

Abstract

With the rapid development of technology, cellular networks in wireless networks are insufficient to meet the demands. In order to provide a correct and good service to each user, communication systems must change. Although cell-free networks have many advantages over cellular networks, since there are too many users and access points (APs) in cell-free networks, AP selection is very important. In this thesis, the AP selection model has been studied and compared five different machine learning classification methods. The campus of Izmir Katip Celebi University has been chosen as the environment where the study will be carried out, and capacity values have been obtained from the users and APs that have been placed on the campus in the simulation environment. Numerical calculation results have been obtained from the Wireless Insite (WI) software. The AP selection to be created with the capacity values has been supported by artificial intelligence algorithm techniques. With two different data sets have been compared, better results have been tried to be obtained with different feature values. As a result of the comparisons made, the best machine learning classification technique used has been proposed.

Keywords: Cell-free, MIMO System, Access point selection, Machine Learning, Classification algorithm

Hücresiz ÇGÇÇ Sistemlerinde AP Seçimi için Makine Öğrenimi Tabanlı Sınıflandırma Algoritması

Öz

Teknolojinin hızlı gelişimi ile kablosuz ağlarda hücresel ağlar talepleri karşılamakta yetersiz kalmaktadır. Her kullanıcıya doğru ve iyi hizmet verebilmek için iletişim sistemlerinin değişmesi gerekmektedir. Hücresiz ağların hücresel ağlara göre birçok avantajı olmasına rağmen, hücresiz ağlarda çok fazla kullanıcı ve erişim noktası (AP) olduğundan, AP seçimi çok önemlidir. Bu tezde, AP seçim modeli incelenmiş ve beş farklı makine öğrenmesi sınıflandırma yöntemi karşılaştırılmıştır. Çalışmanın gerçekleştirileceği ortam olarak İzmir Katip Çelebi Üniversitesi kampüsü seçilmiş ve simülasyon ortamında kampüse yerleştirilen kullanıcı ve AP'lerden kapasite değerleri elde edilmiştir. Sayısal hesaplama sonuçları Wireless Insite (WI) yazılımından alınmıştır. Kapasite değerleri ile oluşturulacak AP seçimi yapay zeka algoritma teknikleri ile desteklenmiştir. İki farklı veri seti karşılaştırılmış, farklı öznitelik değerleri ile daha iyi sonuçlar elde edilmeye çalışılmıştır. Yapılan karşılaştırmalar sonucunda kullanılan en iyi makine öğrenmesi sınıflandırma tekniği önerilmiştir.

Anahtar Kelimeler: Hücresiz ağ, Çoklu Girişli Çoklu Çıkışlı Sistem, Erişim noktası seçimi, Makine öğrenmesi, Sınıflandırma algoritması

[*] Corresponding Author: mdemirel97@gmail.com

1. Introduction

Radio frequencies have been used in wireless communication systems to transmit data over the air. Although 5G is similar to 4G, it uses higher radio frequencies. This enables larger volumes of data to be transmitted at a faster rate. However, path loss effect has a greater impact on the transmission of signals at higher frequencies. Moreover, physical influences such as trees and buildings can easily hinder transmission. Massive multiple input and multiple output (massive MIMO) systems have been widely adressed as a solution to the aforementioned issues in recent studies [1,2].

In conventional cellular networks, a macro base stations (BS) serves many users under its coverage area. The farther users get from the macro BS, the lower the signal quality has been served. However, with the increasing demands for higher signal quality, user numbers and users mobility have been increasing. Due to the increasing number of user demands, conventional cellular networks are insufficient. Cell-free MIMO systems have been investigated in order to solve the problems in cellular networks. In principle, Cell-Free MIMO systems include multiple dispersed, low-cost and low-power multi-antenna APs which are connected to a network controller. The number of antennas should be greater than the number of customers. On the contrary of cellular networks, there are no cell boundaries and each user is jointly served by all APs [3,4]. The obtained results show that the cell-free networks have many advantages over cellular networks at many points [5-7]. The first benefit of the Cell-free architecture is that it achieves a higher and more uniform signal-to-noise ratio (SNR) within the coverage area than conventional cellular networks. In addition, mobile users who are switching between cells are exposed to disconnections while trying to establish a new connection with the new base station which is called handover. In order to overcome the handover issue, the concept of Cell-Free topology can overcome the excessive handover issue in small-cell systems by removing the cell boundaries. Another benefit is the ability to manage interference across multiple APs versus cellular networks with an equally dense AP distribution.

Scalability requires a system to accommodate growing demands gracefully. Scalability is critical motivation for user-centric, cell-free, spatially-distributed, MIMO networks [8]. Although cell-free massive MIMO has showed great promise for next generation wireless networks, such as 5G and beyond, understanding how to design algorithms for a low-cost and scalable system is critical.

Proper resource allocation improves the performances of both the associated system and the network, and also helps in avoiding the different kinds of transient bottlenecks involved in the network. Therefore, various algorithms have been developed for resource allocation in studies carried out so far [9-11]. Although algorithms are constantly being developed, they are insufficient to meet the demand. For this reason, especially when 6G is transitioned, resource allocation will now be made by using artificial intelligence methods. With the artificial intelligence methods that will be developed ,the user will be able to communicate with the base stations in the most accurate way.

In [12], CAPS (Cluster Based AP Selection) a new AP selection algorithm for cell-free Massive MIMO that aims to reduce computation workload and pilot contamination has been presented by introducing a machine learning algorithm for clustering, which in this case is the K-means++ clustering algorithm. In [13], a Cell-Free MIMO system, a deep learning (DL) based power control technique is suggested to overcome the max-min user fairness problem. In a cell-free massive MIMO uplink configuration, the max-min rate optimization issue is posed, where user power allocations are adjusted to maximize the minimal user rate.

In this study, Three-dimensional modeling of Izmir Katip Çelebi University (IKCU) has been made in the WI simulation program, and AP selection has been recommended in a Cell-Free MIMO system at 1.9 GHz. In the proposed AP selection scenario, five different supervised machine learning techniques have been applied and compared.

2. Material and Method

2.1. System Model

In this section, the Cell-Free MIMO system has been created with M APs and K users equipment in TDD ($M >> K$). The scenario parameters have been used in WI have been given in Table 1. In Fig.1, the IKCU campus has been drawn in three dimensions. APs are evenly distributed over the entire area. Each AP has four directional antennas. Each user has two isotropic antennas. Users have been placed on a grid, with a distance of five meters between them. In order to create the data set, the capacity values have been collected by constantly changing the positions of the users.

Fig. 1. AP and user equipment distribution in IKCU

Table 1. SIMULATION PARAMETERS OF THE SYSTEM

Parameters	***Values***
Area	$1\ km^2$
Carrier Frequency	1.9 GHz
User Height	1.64 m
AP Height	15m
Power density	-174 dBm/Hz
Noise figure	7dB
Bandwidth	20 MHz
No. of APs (M)	25
No. of Users (K)	10
No. of Antennas (L)	100
Transmitter input power	23 dBm

The signal at the user is computed as:

$$y = Hx + n$$

Where x is the $N_t \times 1$ vector containing the AP signal, y is the $N_r \times 1$ vector containing user signal, n is a vector of noise, and H is the $N_t \times N_r$ matrix of complex channel gains. N_r is the number of user antennas, and N_t is the number of AP antennas.

$G_k[m]$ is the ratio of the power received by user antenna element k divided by the power radiated by AP antenna element m. $\theta_k[m]$ is the phase in radians of the voltage across a matched load at k under the same conditions. Note that $G_k[m]$ and $\theta_k[m]$ include all of the propagation paths in a complex multi-path environment from AP antenna element m to user antenna element k summed coherently.

The propagation factor, $g_k[m]$, is defined as:

$$g_k[m] = \sqrt{G_k[m]}\, e^{i\,\theta_k[m]}$$

Closely associated with g_k is the channel vector h_k, an N-dimensional complex column vector ($N_k \times 1$) given by $h_k = g_k^*$ where * denotes the conjugate transpose.

Maximal Ratio Combining (MRC) has been used as the combining technique. With this technique, the user optimally combines the user voltages from all antenna elements using a weighting vector that adjusts both the phase and the magnitude to maximize the total SNR.

The optimal weighting vector is linearly proportional to h:

$$w = h/norm$$

where norm is a normalization factor that scales the weighting vector such that the sum of the squares of the magnitudes is equal to N_r.

Interference power is defined as:

$$P_{I,avg} = \sum_{N_{t,m}}^{M-1} \frac{P_{t,m}}{N_{t,m}} \left[\sum_{k=0}^{N_r-1} \sum_{i=0}^{N_t-1} \left[H_{m,k,i}\right]^2 \right] / N_r$$

Noise is defined as:

$$P_N = w^T w \sigma^2$$

Total interference power is defined as:

$$P_{I,total} = w^T w\, P_{I,avg}$$

Weighting vector is then applied to the h-vector to compute the total received power:

$$P_r = \frac{P_t}{N_t} \left[\sum_{k=0}^{N_r-1} \frac{|h_k|^2}{norm} \right]^2$$

The signal-to-interferer-plus-noise ratio (SINR) is the ratio of the received power from the transmitter to the sum of power from all interference sources and all noise sources. The ratio is given by:

$$SINR(dB) = 10\log_{10}\left(P_R(i)\right) - 10\log_{10}\left(P_{I_{total}}\right) - 10\log_{10}\left(P_{N_{total}}\right)$$

where P_R is the user power from the AP. $P_{I_{total}}$ is the total interference. $P_{N_{total}}$ is the total noise.

The channel capacity represents the maximum possible data transmission rate for a communication channel and is calculated using the Shannon-Hartley theorem:

$$Capacity = B \log_2(1 + SINR)$$

where *B* is the bandwidth of the channel.

2.2. Dataset Creation

When creating a data set, it is necessary to decide which outputs will be used as features. First of all, the x and y axis values of each user have been taken. Since the height value given as the z-axis is considered the same for each user, it has not been added to the dataset. The capacity values of each user during the communication with the APs have been included as 25 different features. An feature with the best capacity value has been added to monitor which AP the users connect best with.

The modulation technique information and the capacitance value in the output are not included in the data set. The reason for this is to provide the optimum value in the feature values and to get more efficient results in machine learning training. After determining which feature values are used, it has been determined under which conditions the users equipment should establish the best connection with which APs in the dataset. By comparing the capacity value of the connection have been made by the users with each APs value and the best capacity value, values between 1-25 have been given to the output column. For the user equipment who has the same capacity value with two different APs, the choice has been made according to the distance difference.

The reason why distance is not used as feature values is that users who are at a close distance due to buildings and environmental effects may have lower capacity values. Since the decrease in the capacity value has the opposite effect with respect to the distance, it has put the machine learning training in a wrong state. Various arrangements have been made in the data and cleanings have been made and the AP selection model has been used for the training phase. In addition, instead of taking the capacity value between users and each AP, only the capacity value with the best value has been taken. Whether the reduction in the number of features leads to a better result for machine learning is indicated by comparison.

The correlation matrix has been created with 3 features is given in Table 2. Correlation matrix is a table that shows the correlation coefficients for various variables. The correlation between all potential pairings of values in a table is shown in the matrix. When two variables have a positive correlation, their values rise or fall together. One of the two variables that are negatively associated increases in value while the other decline. Negative correlation has been observed between the y-axis and the output. It has been observed that the feature correlations in the correlation matrix with 28 features are low. With low correlation values, machine learning has been given better results.

Table 2. *Simulation parameters of the system*

Model (3 features)	X	Y	Capacity Best	OUTPUT
X	1	0.02	-0.07	0.06
Y	0.02	1	-0.08	-0.72
Capacity Best	-0.07	-0.08	1	0.27
OUTPUT	0.06	-0.72	0.27	1

2.3. AP Selection Model

In this section, five different machine learning classification techniques have been used. Classification techniques are K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Gaussian Naive Bayes (GNB) and Decision Tree (DT). Two different scenarios have been created for each classification technique. One of the scenarios has been included the capacity values from all APs as a feature, and the other hasn't been included in the feature. The paired distribution of APs has been given in Fig. 2.

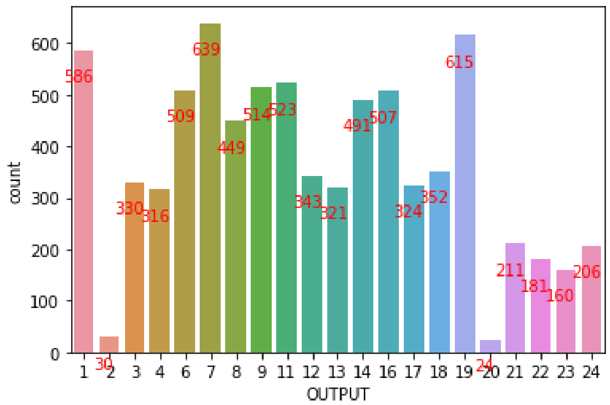

Fig. 2. Dataset output distribution

Considering the K-NN machine learning classification technique first, correlation matrix has been taken into account to determine the number of neighbors to be used. Since no connection could be established with four APs, the number of neighbors (n-neighbors) value has been chosen 25. standardization technique has been used as feature scaling in all models. 25% of the data set has been reserved as test data and 75% as training data.

A kernel is a function used in SVM for helping to solve problems. Shortcuts have been provided to avoid complex calculations. An infinite number of dimensions have been created using kernels. In SVM, Radial Basis Function (RBF) has been selected as the kernel. Gini impurity has been used in DT to divide the data into several branches. Classification have been accomplished using DT. Impurity has been used in DT to choose the optimal characteristic at each. Singular Value Decomposition (SVD) has been applied as a solver in LDA. SVD does not calculate the Covariance matrix, so this solver is recommended for data with a large number of features.

Considering the accuracy distributions, it has been observed that better results have been obtained in terms of estimation, although the number of features increases when 28 features have been used. The detailed comparison between the models created has been given in the numerical analysis section.

3. Results and Discussion

TABLE 3. *Classification Techniques Models Outputs*

Model (28 features)	*Precision*	*Recall*	*f1 score*	*Accuracy*
K-NN	0.94	0.94	0.94	0.938
SVM	0.97	0.96	0.97	0.968
LDA	0.93	0.91	0.91	0.911
GNB	0.71	0.66	0.63	0.656
DT	0.97	0.98	0.98	0.975
Model (3 features)	***Precision***	***Recall***	***f1 score***	***Accuracy***
K-NN	0.75	0.70	0.72	0.769
SVM	0.58	0.58	0.57	0.695
LDA	0.55	0.49	0.49	0.567
GNB	0.60	0.59	0.58	0.602
DT	0.64	0.59	0.60	0.676

The outputs have been used in the comparisons are precision, recall, f1 score and accuracy. Precision is an indicator of the performance of a machine learning model. The quality of a positive prediction has been made by the model. Precision refers to the number of true positives divided by the total number of positive predictions. Recall is the measure of how accurately the model identifies true positives. The f1 score is the harmonic mean of precision and recall. The f1 score is used when both precision and recall are equally important. Likewise, accuracy is the measure of the rate of predictions that the model makes correctly.

A model has been created for all classification techniques and the outputs have been given in Table 3. According to the numerical calculation results, when 28 features have been selected, SVM and DT methods have been given approximately equal and best results. The classification method with the worst mean for this study is GNB.

When 3 features have been selected, K-NN has been given the best results. Although 25 features have been removed, a good result has been obtained with an accuracy average of 76%. Since AP selection must have been made quickly in communication systems, a faster system recommendation has been made with less features value, although there is a margin of error.

4. Conclusions and Recommendations

In this paper, five different machine learning techniques have been compared in order to make the best choice between user equipment and APs installed on the IKCU campus. While making AP selection, classification has been made by creating a data set with the location and capacity values of each user. As a result of the comparisons, the most efficient machine learning technique is the DT classifier method. In addition, other machine learning techniques besides GNB have yielded good results.

References

[1] Zhang, J., Chen, S., Lin, Y., Zheng, J., Ai, B., & Hanzo, L. (2019). Cell-free massive MIMO: A new next-generation paradigm. IEEE Access, 7, 99878-99888.

[2] Interdonato, G. (2020). Cell-Free Massive MIMO: Scalability, Signal Processing and Power Control (Vol. 2090). Linköping University Electronic Press.

[3] Ngo, H. Q., Ashikhmin, A., Yang, H., Larsson, E. G., & Marzetta, T. L. (2017). Cell-free massive MIMO versus small

cells. IEEE Transactions on Wireless Communications, 16(3), 1834-1850.

[4] E. Nayebi, A. Ashikhmin, T. L. Marzetta and H. Yang, "Cell-Free Massive MIMO systems," 2015 49th Asilomar Conference on Signals, Systems and Computers, 2015, pp. 695-699, doi: 10.1109/ACSSC.2015.7421222.

[5] Björnson, E., & Sanguinetti, L. (2019, July). Cell-free versus cellular massive MIMO: What processing is needed for cell-free to win?. In 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (pp. 1-5). IEEE.

[6] Buzzi, S., D’Andrea, C., Zappone, A., & D’Elia, C. (2019). User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach. IEEE Transactions on Wireless Communications, 19(2), 1250-1264.

[7] Interdonato, G., Björnson, E., Quoc Ngo, H., Frenger, P., & Larsson, E. G. (2019). Ubiquitous cell-free massive MIMO communications. EURASIP Journal on Wireless Communications and Networking, 2019(1), 1-13.

[8] Chen, S., Zhang, J., Björnson, E., Zhang, J., & Ai, B. (2020). Structured massive access for scalable cell-free massive MIMO systems. IEEE Journal on Selected Areas in Communications, 39(4), 1086-1100.

[9] Shakya, I. L., & Ali, F. H. (2020). Joint Access Point Selection and Interference Cancellation for Cell-Free Massive MIMO. IEEE Communications Letters, 25(4), 1313-1317.

[10] Buzzi, S., & D’Andrea, C. (2017). Cell-free massive MIMO: User-centric approach. IEEE Wireless Communications Letters, 6(6), 706-709.

[11] E. Bjornson and L. Sanguinetti, “Making cell-free massive MIMO competitive with MMSE processing and centralized implementation,” IEEE Trans. Wireless Commun., vol. 19, no. 1, pp. 77–90, Jan. 2020.

[12] Biswas, S., & Vijayakumar, P. (2021, March). AP selection in Cell-Free Massive MIMO system using Machine Learning Algorithm. In 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 158-161). IEEE.

[13] Rajapaksha, N., Manosha, K. S., Rajatheva, N., & Latva-Aho, M. (2021, June). Deep Learning-based Power Control for Cell-Free Massive MIMO Networks. In ICC 2021-IEEE International Conference on Communications (pp. 1-7). IEEE.