Deep Reinforcement Learning for Resource Allocation in Multi-Agent Wireless Networks

Qiushi Xu, Yiyi Li

***Abstract*—Due to the non-convex characteristics of multi-agent wireless network, it is challenging to obtain an optimal strategy for the resource allocation issue. Deep Reinforcement Learning (DRL) becomes a potential method to be employed to find the optimal solution of this problem. In this paper, based on the idea of dueling DRL, to maximize the network utility, we proposed a new resource allocation strategy using a dueling deep neural network named as Dueling Deep-Q-Full-Connected-Network (DQFCNet). The simulation result shows that, compared with the proposed DQFCNet, the traditional water-filling power allocation strategy, and the original Q-Learning method, this strategy achieves the best performance in convergence speed and network utility.**

***Index Terms*—Deep Reinforcement Learning, Resource Allocation, Wireless Network**

I. Introduction

Power and wireless resource allocation problems have been studied since Long-Term Evolution (LTE) system and 5G mobile communication system. However, with the rapid growth of network scale and the dramatic increase of the number of devices, the problem of optimal power allocation becomes particularly acute.[1] The simply optimization, such as water filling model, however, can hardly meet the demands and obtain an optimal strategy for the resource allocation issue due to the non-convex and combinatorial characteristics of this problem.

Inspired by a design of CSMA mechanism for finding optimal throughput with Markov approximation, M.Chen et al. [2] applied the Markov approximation to synthesize new distributed algorithms, and studied both the path selection problem in multipath wire-line network and the frequency channel assignment to wireless LANs. S.Bayat et al. [3] designed another algorithm for finding the optimal and stable outcomes of source allocation by using matching algorithm with relatively small number of iterations. Ahmed R. Elsherif et al. [4], from another perspective, formulated this resource allocation optimization problem as an efficient and low complexity linear programming, and modified it to the maximization of the revenue of mobile network operators.

Those above proposed optimization algorithms, however, can hardly achieve the optimal solutions without such complete information. In [5], researchers indicate that deep reinforcement learning becomes a promising solution, which can can achieve the better performance than the normal reinforcement learning method with the help of deep neural network and consider the maximum long-term rewards at the same time, compensating for the shortcomings of other mentioned algorithms. Recently, DRL-based approach has been applied to several research fields, such as mobile offloading, dynamic channel access, fog radio access networks, etc..[6]

In this paper, inspired by the Deep-Q-Full-Connected-Network (DQFCNet) and dueling double deep Q-network (D3QN) strategy, which are proposed by Y.Zhang et al.[7] and N.Zhao et al.[6] respectively, we try to design a new DRL strategy, which can be used to find the optimal resource allocation in multi-agent wireless network. In this strategy, we tried to embed the D3QN in to the DQFCNet, combining the advantages of these two techniques together.

The remainder of this paper is organized as follows. In Section II, we will review the resource allocation model. The DRL model and dueling neural network will be described in Section III and IV with details. Then we will analyze the simulation results in Section V. Section VI will concludes our works in this project and elaborate some future works.

II. Resource Allocation Model

The system model in our research describes a downlink multi-agent OFDM cellular environment with densely deployed Base Stations(BSs) whose geographical distribution is followed by the Poisson point process model and a huge set of User Equipments(UEs) which adopts Random walk model. Every BS fully reuses the K orthogonal subcarriers. When a user connects to the BS, the BS is activated and one subcarrier can only be assigned to the user with a transmit power. The system performance is analyzed in terms of the overall capacity measured in (bps/Hz). The interference between users and base stations mainly comes from inter-cell interference. The link rate is affected by the signal

strength from the attached base station to the user, and the inter-cell interference strength. At the initialization stage, we assign the downlink subcarriers of one base station to all attached users equally for the sake of fairness and the power on the subcarrier is allocated according to the water-filling algorithm. We aim to adjust the transmission power of base stations on each subcarrier to increase the overall capacity of the entire network.

The above problem is a multi-objective non-convex optimization problem. The traditional method to solve this problem are heuristic search algorithms, most of which are very inefficient. Therefore, we resort to reinforcement learning method for achieving a potential optimization.

III. Deep Reinforcement Learning

BAsed on [7], a solution is presented for Deep Reinforcement Learning, which more specifically is multi-agent Q-learning based on continuously interacting with the environment. Q-learning, as a model-free reinforcement learning algorithm based on value function, has a natural advantage in solving the dynamic wireless networks environment problem. This behavioral value function is generally represented by the Q-value which represents the expected value of the cumulative reward of action at state over an infinite time horizon.

The agents, states, actions are defined as follows:

- Agents: BSs
- States: The state information contains set of pairs of each UE with its connected BS on specific subcarrier and the transmit power emitted on this channel. Every UE can measure the received power from the associated BS and then reports its own current state to its current associated BS. By the message passing among the BSs through the backhaul communication link, the global state information of all UEs is obtained. In order to reduce the complexity of the state space of the network, the power of the BS is discretized to selected thresholds.
- Actions: Action made by agents denotes the adjustment value of the power on the k-th subcarrier of the n-th BS. During the learning procedure, the tradeoff of exploration-exploitation is considered in the action selection mechanism. In order to balance the exploitation of the current best Q-value function with the exploration of the better option, we adopt ϵ-greedy policy where an action is selected at random with probability ϵ.

However, due to the dynamic nature of the environment, the state space of the environment is relatively large despite the discretization of transmission power. Besides, the wireless networks environment has noise, incomplete information, and soft boundaries, which are difficult to map this information to Q-Learning. Therefore, we introduce Deep Neural Network(DNN) into the framework of Q-learning to speed up the convergence of the Q-value, which is a deep full-connection neural network(DQFCNet). The structure of Deep Neural Network applied in our research project is shown in Fig.1

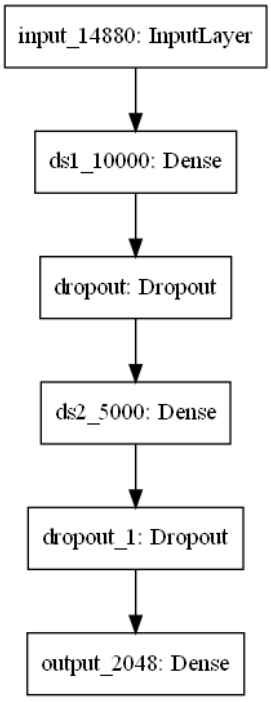

Fig. 1: Original DQFC Q-learning Network

IV. Dueling Deep Neural Network

IN standard Q-learning and DQN, the same values are used both to select and to evaluate an action by the max operator, which makes it more likely to select overestimated values, resulting in overoptimistic value estimates. [8] Dueling Deep Neural Network(DDQN) decouples these two processes by using the Q-network for action selection and the target Q-network for action evaluation. This reduces the likelihood of overestimating Q-values, leading to more accurate value estimates and better policy learning. Furthermore the avoidance of overestimation, DDQN often converges to Nash equilibrium faster than the standard DQN algorithm with fewer training iterations and reducing the computational resources and time for training.

Therefore, in our research project, to get a better evaluation without overestimation, we firstly split the

original DQN into two subnetworks to estimate the value and the advantage functions, $V(s)$ and $A(s, a_i) = Q_i(s, a_i) - V(s)$, separately. The advantage function describes the advantage of the action a_i compared with the other possible actions. By combining the value and advantage together, the action value function $Q_i(s, a_i)$, then can be estimated. The procedure of DRL with DDQN strategy for this research project is shown in Fig.2

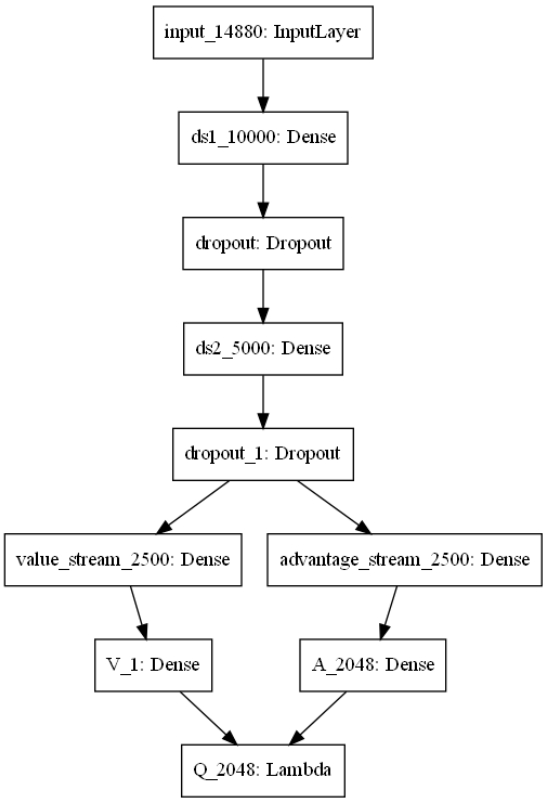

Fig. 2: Dueling Deep Q-Network (DDQN) model used in this paper

In next section, after constructing this DDQN, and formulating the system by python, we will analyze the convergence of our proposed algorithm, which embedded the dueling deep Q-network into the DQFC network, and compare the performance of this new DRL strategy on resource allocation with the traditional resource allocation strategy (i.e. Water-filling) and the DRL strategy using original Q-learning with or without DQFC network.

V. Experimental Statistics

A. Simulation Settings

To find the optimal solution of resource allocation in Multi-Agent wireless network, we need to construct a network simulation first. In this research, we constructed a $150 * 150m^2$ network with 20 BSs and 300 UEs in it. The total Bandwidth (BW) of each base station is 10MHz, while the initial power is 10dbm. The other parameter of this simulated wireless network can be found in Table I. The topology of BSs and UEs in this network is also shown in Figure 3, while this topology will update at each step due to the node mobility.

TABLE I: Wireless Network Simulation

Parameter	Value
Bandwidth	10 MHz
Number of BSs	32
Number of UEs	200
Radius	30 m
Initial Power	10 dbm
Subcarrier	64
Network Size	150*150 m^2
White Noise	-70 dbm

After the wireless network construction and the establishment of Q-Learning environment, we created a deep full-connection neural network (DQFCNet) introduced in Section III. The number of neurons per layer is configured as [14880,10000,5000,2048]. More specifically, for each layer we choose $ReLU$ as activation function and dropout rate is 0.5. The Optimizer is Adam Optimization Algorithm with learning rate being 0.0001. To compare the performance of Dueling DQN and the original DQN, we also formulated a Dueling Deep Neural Network. As we discussed in Section IV, we split the last layer into value stream and advantage stream to estimate the value function and advantage function separately. Note that the layer sizes of both value stream and advantage stream are 2500 neurons. The output size of value function for $V(s)$ is 1, while advantage function for each state $A(s, a_i)$ should keep the same size with the original output of DQFC, which is 2048. Eventually, the output layer merge two parts by calculating the Q-Value using the equation $Q(s, a) = V(s) + A(s, a)$. The rest of the layer organization is identical with DQFC Network.

B. Performance Comparison

Figure 4 shows the spectral efficiencies, which represent the performance of each model in different resource allocation strategies, including the traditional algorithm(i.e. Water-filling model), original Q-learning, deep Q-learning, Deep Learning as well as those strategies with Duel Network.

Among all the strategies, shown in the graph, the Deep Learning with both Neural Networks achieve the highest frequency efficiency as the model is trained to predict the optimal situation according to the existing action and Q-Value data set. In aspect of the Q-learning methods, we can see the performance of Deep Q-Learning stays

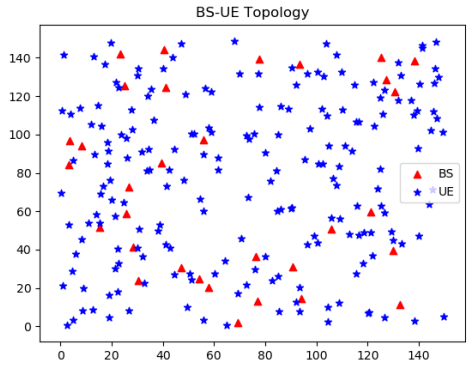

Fig. 3: Topology of UEs and BSs in Simulated Network

close to Q-learning, which indicates that the role of Deep Learning for this scenario seems to become saturated. However, for Dueling Q-Learning, despite its overall outstanding performance, one outlier is that the fluctuation shown by the curve is out of our expectation as stability should be the feature of the Dueling Network.

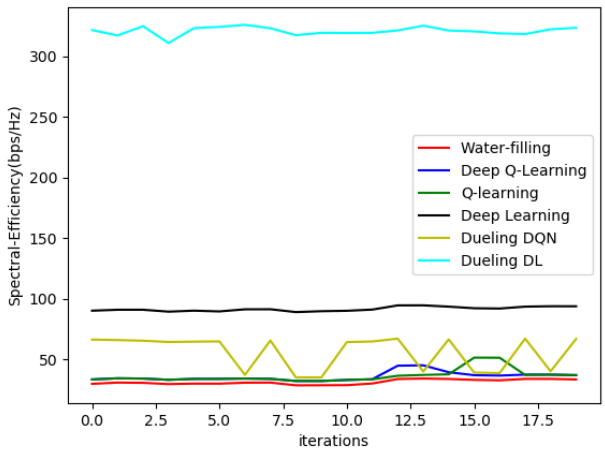

Fig. 4: Comparison of spectral efficiency with different strategies

C. Converges

The comparison of convergence speeds of networks is illustrated by Figure 5 in vertical and horizontal views, i.e. comparing the convergence speed of dueling DQFC-Network with Q-learning and original DQFCNetwork, respectively.

On the one hand, the DQFCNetwork strategy converges in about 10 times of iterations, which costs fewer iterations than the original DQFCNetwork, indicating that the convergence speed of this new strategy is faster than the original one's. It is also obviously that, after the system converges, the dueling DQFCNetwork strategy brings better solution of resource allocation than the other two strategies.

On the other hand, unsurprisingly, Figure 5 also reveals the vulnerability of those strategies to the node mobility. As Q-Learning needs to re-calculate and update the current policies when the network topology changes in the dynamic scenario, it's reasonable to see the fluctuation of the curve during convergence stage. But it is cheerful to observe that, after the convergence happens, the spectral efficiencies getting from Dueling DQN fluctuates less, thus we could say that our improved strategy shows greater stability than the others. Interestingly, this observation is conforming to the character of dueling DRLs, but is opposite to what we observed in section V-B. More experiments are needed to determine whether dueling DRLs bring stability to the resource allocation strategy.

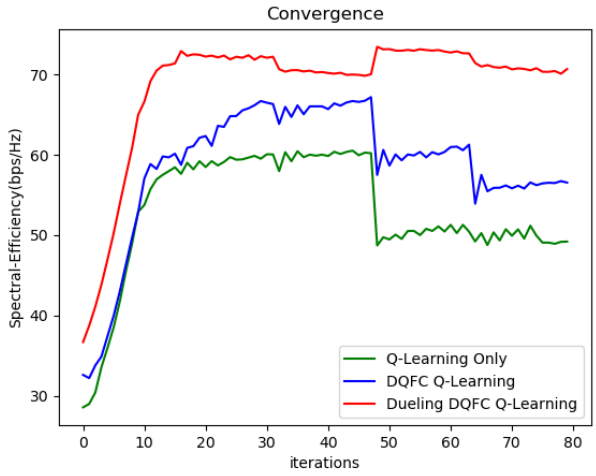

Fig. 5: Comparison of convergence speed.

VI. Conclusion

Resource allocation problem in Multi-Agent Networks has long been a research difficulty in the field of Wireless communication. On the basis of previous research, Our research project successfully reproduced the experimental scenario of multi-cell BS-UE simulation. We conducted the Deep Q-Learning mapping, and constructed the DQFC Neural Network. In addition, we applied the Dueling Network to the framework in order to expect a better policy evaluation. As a result, in comparison of traditional strategy, deep learning and deep q-learning algorithms, we do see a huge improvement in theoretical upper bound for Deep Learning, as well as slightly increased performance in terms of bandwidth efficiency and improved stability for

DQN. In the future research, we think there is still much exploration space for deep reinforcement learning based resource allocation problem, such as Double DQN(online and target network), and so on.

REFERENCES

[1] Y. Huang, J. Tan, and Y.-C. Liang, "Wireless big data: Transforming heterogeneous networks to smart networks," *Journal of Communications and Information Networks*, vol. 2, no. 1, pp. 19–32, 2017.

[2] M. Chen, S. C. Liew, Z. Shao, and C. Kai, "Markov approximation for combinatorial network optimization," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6301–6327, 2013. DOI: 10.1109/TIT.2013.2268923.

[3] S. Bayat, R. H. Y. Louie, Z. Han, B. Vucetic, and Y. Li, "Distributed user association and femtocell allocation in heterogeneous wireless networks," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 3027–3043, 2014. DOI: 10.1109/TCOMM.2014.2339313.

[4] A. R. Elsherif, W.-P. Chen, A. Ito, and Z. Ding, "Resource allocation and inter-cell interference management for dual-access small cells," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1082–1096, 2015. DOI: 10.1109/JSAC.2015.2416990.

[5] C. Chaieb, F. Abdelkefi, and W. Ajib, "Deep reinforcement learning for resource allocation in multiband and hybrid oma-noma wireless networks," *IEEE Transactions on Communications*, vol. 71, no. 1, pp. 187–198, 2023. DOI: 10.1109/TCOMM.2022.3225163.

[6] N. Zhao, Y.-C. Liang, D. Niyato, Y. Pei, M. Wu, and Y. Jiang, "Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5141–5152, 2019. DOI: 10.1109/TWC.2019.2933417.

[7] Y. Zhang, C. Kang, T. Ma, Y. Teng, and D. Guo, "Power allocation in multi-cell networks using deep reinforcement learning," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–6. DOI: 10.1109/VTCFall.2018.8690757.

[8] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, 2016.