Power Control in 5G Heterogeneous Cells Considering User Demands Using Deep Reinforcement Learning

Anastasios Giannopoulos, Sotirios Spantideas, Christos Tsinos, Panagiotis Trakadas

Power Control in 5G Heterogeneous Cells considering User Demands using Deep Reinforcement Learning

Anastasios Giannopoulos[1], Sotirios Spantideas[1], Christos Tsinos[1] and Panagiotis Trakadas[1]

[1] National and Kapodistrian University of Athens, Psachna, Evia, 34400, Greece
angianno@uoa.gr, sospanti@uoa.gr, chtsinos@gmail.com, ptrakadas@uoa.gr

Abstract. Heterogeneous cells have been emerged as the dominant design approach for the deployment of 5G wireless networks. In this context, inter-cell interferences are expected to drastically affect the 5G targets, especially in terms of throughput experienced by the mobile users. This work proposes a novel Deep Reinforcement Learning (DRL) scheme, targeting at minimizing the difference between the allocated and requested user throughput through power regulation. The developed algorithm is employed in heterogeneous cells that are controlled in a centralized manner and validated for 5G-compliant channel models. First, the proposed learning framework of the DRL method is presented, mainly including the stabilization of the learning-related hyperparameters. Then, the DRL method is evaluated for several simulation scenarios and compared to well-established optimization methods for power allocation, namely the Water-filling and Weighted Minimum Mean Squared Error (WMMSE) algorithms, as well as a fixed power control scheme. The evaluation outcomes demonstrate the ability of the DRL framework in accurately approaching the user requirements, whereas the Water-filling and WMMSE solutions present large deviations from the user demands since they aim at the total network-wide throughput maximization.

Keywords: 5G, Heterogeneous Cell, Power Control, Deep Q-Learning, Reinforcement Learning, Radio Resource Management.

1 Introduction

Next-generation broadband wireless networks are characterized by an ever-growing need for high-volume investments targeting at the delivery of rich-content and low-latency services [1–3]. In the 5G era, the configuration of the wireless networks has become an increasingly important issue, since it is strongly related to the 5G network design specifications. In order to cope with these demanding specifications, the design requirements of 5G networks inherently involve self-configuration and optimization capabilities of the network parameters [4], enabling autonomous decision-making processes. Thus, there is an ever-increasing need for the deployment of intelligent, adaptive and scalable optimization algorithms with respect to the radio resource man-

agement (RRM). In this context, increased network densification has been introduced in 5G-enabled systems to maximize the total network-wide throughput, embracing the concept of Ultra Dense Networks (UDNs) [5]. Nevertheless, the dense deployment of the network elements, along with the large number of mobile users lead to a significant increase in the complexity of the wireless environment, which, in turn, has a significant impact on the degradation of the network performance and reduction of the experienced Quality of Service (QoS) [5].

In 5G wireless networks, the RRM usually involves the solution of non-convex optimization problems. For purposes of finding optimal or sub-optimal solutions to these problems, deterministic, as well as stochastic methods have been established [5–7]. Although these algorithms perform well for small-scale cellular systems, they fail to provide adequate solutions when high-dimensional wireless environments are considered. To this end, the rapid progress in machine learning (ML), and specifically reinforcement learning (RL), has led to intelligent and automatic approaches towards the solution of complex optimization problems [8, 9]. Several CCO techniques employing ML methods have been proposed [10–12]. One of the major challenges arising in UDNs is the power allocation [11, 13]. Since multiple Radio Units (RUs) operate simultaneously in the same coverage network area, the inter- and intra-cell interference should be carefully regulated. Several joint optimization schemes have been proposed, including the channel/power allocation and power and user association approaches [11–18]. In particular, a multi-agent DRL framework was proposed in [13] for dynamic power allocation based on cross-cell channel state information (CSI). Finally, a fair power control scheme for UDNs employing RL was proposed in [16], whereas a joint consideration of power/user association was investigated in [18], targeting to optimize the long-term total network utility in heterogeneous networks, using DRL.

In this paper, a DRL algorithm for power allocation in heterogeneous cells is proposed on 5G-compliant network configurations. The need for interference mitigation involves not only the inter-microcell interference control, but also the power regulation of the respective macrocell, which is responsible to cover the white spaces of the macro-area and complement the cell coverage. This two-fold power control framework is addressed in the present work, along with simultaneous consideration of the user demands, aiming to the optimal satisfaction of the user requests through power adjustment of the RUs. The main contributions of this paper are identified as: **(i)** The proposed scheme follows a demand-driven approach, by incorporating the difference between the requested and the allocated throughput in the rewarding system **(ii)** As opposed to deep learning models [9], this approach does not require training data or any prior knowledge of the telecommunication environment, providing robustness against the dynamic nature of the wireless systems **(iii)** The proposed algorithm is tested in a realistic 5G-compliant wireless environment in a generic manner that can be easily modified to incorporate different network settings (number of RUs, number of available resource blocks, etc.) and **(iv)** An alternative state space modeling of the telecommunication environment is determined, including a three-fold information for each user inside the network area: associated RU, the associated resource block (RB) and whether the user is satisfied or not. The values of these three parameters are acknowledged to the DRL agent, offering the flexibility to train for different user association realizations.

2 System Model and Algorithm Description

2.1 System Model

A heterogeneous network area that includes a macrocell (MaC) and M microcells (MiC, $m = 1,2,\dots,M$) is considered. The available number of RBs at each RU is F ($f = 1,2,\dots,F$), depending on the both the operational 5G frequency band and the employed 5G numerology scheme. It is assumed that the total bandwidth is equally distributed amongst the F RBs, resulting in a single-RB bandwidth of $B_1 = B_2 = \cdots = B$. Each RU n ($n = 1,2,\dots,M+1$; $n = 1$ is the MaC ID, while $n = 2,3,\dots,M+1$ is the MiC IDs) transmits over RB f with a specific power $P_{n,f}$. A minimum power level P_{min} is defined for each RB specific transmission to indicate signaling processes. A sum power constraint is also established for each RU, separately for MaC (P_{max}^{MaC}) and MiC (P_{max}^{MiC}), for power budget purposes, i.e. $\sum_{f=1}^{F} P_{1,f} \leq P_{max}^{MaC}$, for the MaC and $\sum_{f=1}^{F} P_{n,f} \leq P_{max}^{MiC}, \forall n > 1$ for the MiCs.

The mobile user $u \in \{1,2,\dots,U\}$ (i.e. U is the total number of users) located inside the network area can occupy a single RB f of a particular RU n, whereas multiple users can be associated with a specific RU. An allocation matrix A (with elements $a_{n,f,u} \leftarrow 1$) is additionally defined to denote whether user u is connected to the RB f of RU n (or 0 otherwise). Moreover, each user u requests a service s from a set of available service classes S, corresponding to realistic throughput requirements in order to ensure adequate QoS. Thus, a demand vector D_u ($D_1, D_2, \dots, D_U$) is introduced to designate the requested service class of user u, expressed in terms of throughput. Importantly, a capacity overflow in a particular MiC occurs when all of its available RBs have been occupied by F mobile users.

As already mentioned, the dense wireless environment is related with the superposition of multiple interference signals both from the operating RU of the MaC and the operating RUs of the MiCs that are located in close proximity. The signal-to-interference-plus-noise ratio (SINR) received by a user u that is linked to RB f of RU n is given by:

$$SINR_u^{n,f} = \frac{P_{n,f} \cdot G_{n,f,u}}{(\sum_{n' \neq n}^{M+1} P_{n,f} \cdot G_{n',f,u}) + N_0}, \quad (1)$$

where $P_{n,f}$ stands for the operating power of the RU n over RB f, $G_{n,f,u}$ stands for the channel gain from RU n to user u and $N_0{}'$ denotes the noise power density at the receiver. The channel gain reflects the propagation losses of the wireless environment (e.g. urban, rural, etc.) and also depends on the distance between the RU and the user [21]. The channels models are according to the 5G specifications detailed in [21] and different models for path loss are employed for the MaC (UMa channel model) and MiC (UMi channel model) depending on the user association, but also on the respective distance between the user u and RU n. Finally, the downlink user data rate can be expressed by ($\beta = 1$, a Bit Error Rate (BER) threshold of $BER = 10^{-6}$ is assumed):

$$R_u^{n,f} = B \cdot log\left(1 + \beta \cdot SINR_u^{n,f}\right), \quad (2)$$

2.2 Problem Formulation

The algorithm aims at minimizing the difference between the allocated and the requested throughput (Eq. 4) for each individual user by regulating appropriately the power levels of each RB of the $M+1$ RUs, thus ensuring that the user requirements are fulfilled. Formally, the non-convex *optimization problem (P)* may be defined as:

$$(P) \qquad min \sum_{u=1}^{U} \left(D_u - R_u^{n,f}\right) \tag{3}$$

s.t.:

$$(C1) \qquad \sum_{n=1}^{M+1} \sum_{f=1}^{F} a_{n,f,u} \leq 1\,, \forall u = 1, \dots, U \tag{4}$$

$$(C2) \qquad \sum_{f=1}^{F} P_{1,f} \leq P_{max}^{MaC}, \sum_{f=1}^{F} P_{n,f} \leq P_{max}^{MiC}\,, \forall n = 2, \dots, M+1 \tag{5}$$

$$(C3) \qquad P_{n,f} \geq P_{min}, \forall n = 1, \dots, M+1; f = 1, \dots, F \tag{6}$$

$$(C4) \qquad \sum_{f=1}^{F} a_{n,f,u} \leq F\,, \forall n = 1, \dots, M+1 \tag{7}$$

$$(C5) \qquad R_u^{n,f} \leftarrow min\{D_u, R_u^{n,f}\}, \forall u = 1, \dots, U \tag{8}$$

The optimization problem constraints can be described as follows: $(C1)$ guarantees that each user can only be associated to one RB of a single RU (either MaC or MiC), $(C2)$ ensures that the power limitations are satisfied for the MaC and MiC RUs respectively, $(C3)$ secures a minimum power level for each RB intended for transmission of signaling processes (sleep mode), $(C4)$ designates that a user cannot establish any connection link due to channel capacity restrictions (in case that all RBs of an RU are occupied by other mobile users) and, finally, $(C5)$ ensures that the allocated throughput is upper bounded by the user demands.

The conventional approaches to solve problems similar to P involve either their non-convexity relaxation [19], which results in traditional solutions for convex optimization problems, or finding sub-optimal solutions in an iterative manner due to their NP-hard nature [20]. However, these methods suffer from significant challenges in large-scale optimization, computational complexity and time required for convergence. On the other hand, DRL exhibits the following advantages [12, 14, 15]: (i) the optimal policies can be captured without requiring analytical environment description, (ii) the algorithm can perform well even when large state/action spaces are considered, (iii) the training phase does not require previously gathered samples; instead, the knowledge is extracted in a trial-and-error basis through interaction with the wireless environment and (iv) pre-trained RL agents can be inferred in near-real time without requiring extensive re-programming.

2.3 Deep Q-Learning Framework and Algorithm Outline

In the presented DRL framework (see Fig. 1), a central network entity observes the telecommunication environment and regulates the power of the RBs of all RUs in the network area in order to minimize the difference between the allocated and the requested throughput (optimization problem *P*). Initially, the controller has no prior knowledge of the environment and regulates the power levels randomly (performs random actions during the exploration phase). Then, the controller evaluates the impact of the performed actions through positive (e.g. increase in sum-rate of users) or negative feedback (e.g. interference increase) from the environment. As the proposed scheme unfolds, the agent/controller begins to utilize its past experience and gradually exploits the actions that have beneficial outcomes. The Q-values are estimated by using a neural network attached in the agent's side. In principle, a DRL scheme involves the following:

State space: The wireless environment is effectively acknowledged to the agent via 3 parameters for each mobile user: (i) the number of the associated RU, (ii) the number of the RB that the user currently occupies and (iii) a binary value v specifying whether the user requirements are satisfied or not. The state transition sequence is expressed by $S = \{S_1, \dots, S_t, \dots, S_T\}$, where the system state at a given time t is given by $S_t = [(RU_1, RB_1, v_1), \dots, (RU_U, RB_U, v_U)]$. The user u is associated with the RU/RB that provides the best-quality signal (maximum throughput association).

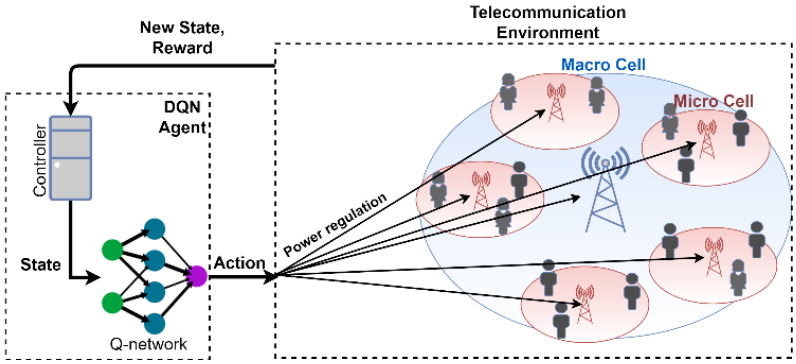

Fig. 1. The interaction cycle between the DQN agent and telecommunication environment.

Action space: The DQL agent performs a sequence of actions $\{A_1, \dots, A_t, \dots, A_T\}$. At a specific time instance t, the agent selects a single RB of each RU and then regulates its power, i.e. $A_t = [(f_1, a_1), \dots, (f_{M+1}, a_{M+1})]$ and the adjustment value of the power on the f-th RB of the n-th RU is expressed by $a_n \in \{P_S, 0, -P_S\}$, where the power step P_S is a constant value. Thus, the agent selects an RB from each RU and then applies either a power step increase or a power incremental decrease or keeps the power fixed for this RB.

Reward system: Once the agent performs an action, a new network system state is triggered, leading to different user RB/RU association configuration and throughput allocation. The response of the wireless environment may be expressed as:

$$r_t = \begin{cases} I = \sum_{u=1}^{U} min\{D_u, R_u^{n,f}\}_t - \sum_{u=1}^{U} min\{D_u, R_u^{n,f}\}_{t-1}, \;\; if\; I > 0 \\ \sum_{u=1}^{U} \{D_u\}_t, \;\; if\; \{R_u^{n,f}\}_t \geq \{D_u\}_t, \;\; \forall u \\ 0, \;\; otherwise \end{cases} \tag{9}$$

Intuitively, this rewarding system definition (*Case 1*) reflects the main objective of the proposed algorithm; it is beneficial to allocate a throughput vector uniformly to all users that is as close as possible to their demand vectors, rather than aiming at the total network throughput maximization. Furthermore, the high-valued positive reward (*Case 2*) implies that the agent will prefer actions that lead to complete fulfillment of all users (when possible). The progression of the DRL algorithm can be described in the following steps:

Step 1: The cognitive controller associates each user with a specific RU/RB pair based on the maximum throughput criterion. Each RU/RB is initialized to transmit with random power level before the agent starts to explore the environment.

Step 2: An RB is selected for each RU and its power is regulated depending on the operational mode of the algorithm: in exploration phase, an RB is randomly selected and its power is either increased, decreased or kept fixed, whereas in exploitation phase the action is estimated by the *Q*-network.

Step 3: The environment provides feedback to the DRL agent regarding the performed action (immediate reward) and the next state. This procedure involves the update of RU/RB association and calculation of the allocated throughput for each user with respect to the new power levels.

Step 4: The system stores the experience tuple $(s_t, a_t, r_{t+1}, s_{t+1})$ into the replay memory. In case that the memory is full, the least recently used tuple is replaced. Noteworthy, the memory is initially filled with 1000 experience tuples corresponding to random actions.

Step 5: A batch of experience tuples N_B is randomly selected from the memory. The current state s_t batch elements are forward-passed through the *Q*-network to predict the Q-values of all actions. The weights of the *Q*-network neurons are adjusted through the back-propagation method (Stochastic Gradient Descent).

Step 6: Every N_u steps, the weights of the *Q*-network model are cloned to the *target Q*-network model.

Step 7: The agent reduces the value of ε (linear decay) to get closer to the exploitation mode and repeats steps 1-6 until convergence.

3 Simulation Results

In this section, simulation results regarding the stabilization of the proposed DRL method's hyperparameters are provided. The algorithm was implemented in Python 3.8 (Keras and Tensorflow 2.0 for the DQN). To this end, the proposed DRL scheme is tested by considering a 5G-compliant network that includes a single MaC containing 4 MiCs, operating at 3.5 GHz. Moreover, two separate maximum power limitations were established for MaC (80 W) and MiCs (25 W) respectively, while the min-

imum power level per RB is set to 0.1 W. Moreover, 5G numerology 4 was considered in all simulation scenarios, resulting in 6 available RBs per MaC/MiC cell (the total bandwidth of each RU is 20 MHz). For this reason, 7 users are randomly placed within the radius of each MiC (100 m), thus ensuring that at least one residual user will be associated with the MaC. The interference mitigation challenge in this configuration setup lies in the fact that the MaC inevitably causes serious interferences in all MiCs, as an attempt to cover the residual (out of MiC capacity) user demands.

The optimization of the DQN hyperparameters included simultaneous consideration of the learning rate (a), the discount factor (γ) and the power step (P_S). Towards this direction, several values of these parameters were tested in terms of the accumulated average reward. The learning curves of the DQN agent are sequentially illustrated in Figs. 2 – 4. Evidently, the values of the hyperparameter triplet (a, γ, P_S) were stabilized at (10^{-3}, 0.7, 1) as the optimal values of throughput increment convergence.

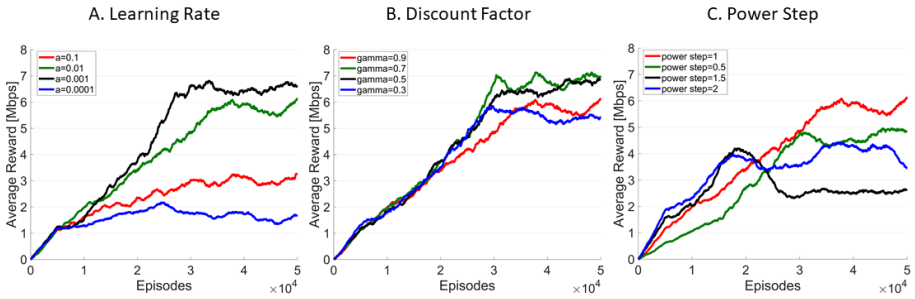

Fig. 2. Accumulated reward for different values of the learning rate α (panel A), discount factor γ (panel B) and power step P_S (panel C).

Furthermore, the performance of the DRL algorithm was tested in three validation scenarios and compared to three baseline power allocation methodologies, namely the Water-filling algorithm [6], the Weighted Minimum Mean Squared Error (WMMSE) method [7] and the fixed average power allocation. According to the latter method, all RUs transmit with the median power level, equally distributed amongst their RBs, as a trade-off between maximization of the coverage radius and minimization of the harmful interference.

The validation scenarios include: (i) all users request service class 2 (1 Mbps), (ii) all users request a random service class uniformly selected from service set S and (iii) all users request the most demanding service class 3 (2.5 Mbps). All evaluation scenarios were conducted by employing the DRL learning hyperparameters and network simulation parameters, summarized in Table I.

The DRL algorithm was tested by inferring N_i = 1000 exploitative solutions (Monte-Carlo simulations), where in each realization, the users are randomly placed inside the network area. For each compared methodology, the Average (across N_i simulations) Variation from Demands (AVD) is defined as:

$$AVD = \frac{\sum_{i=1}^{N_i} \sum_{u=1}^{U} \left| R_u^i - D_u^i \right|}{N_i}, \tag{10}$$

where R_u^i and D_u^i are the allocated and requested throughput of user u, respectively, in the i-th simulation setup. AVD reflects the average residual throughput relative to the total user demands.

Table 1. Simulation setup parameters.

Parameter	Value	Parameter	Value
Central Frequency f_c	3.5 GHz	Memory size	5000
Bandwidth B	20 MHz	Batch size N_B	64
Number of RBs F	6	Loss function	Huber loss
5G numerology	4	Service class S	{0.1, 1, 2.5} Mbps
Number of RUs M	4	Noise Power Density	-174 dBm/Hz
Power step P_S	1 W	Learning rate α	10^{-3}
Maximum power P_{max}^{MaC}	80 W	Discount factor γ	0.7
Maximum power P_{max}^{MiC}	25 W	Number of hidden layers	3
Minimum power P_{min}	0.1 W	Activation function of input and hidden layers	ReLu
MiC radius	100 m	Activation function of output layer	Linear
Update target frequency N_u	100	Monte-Carlo simulations N_i	1000

The resulting evaluation metrics are depicted in Fig. 5 for all the validation scenarios. As expected, Water-filling and WMMSE algorithms outperform both DRL and Fixed Average methods in terms of total network-wide utility. This is attributed to the inherent optimization objective of these algorithms, which involves the direct maximization of the sum-rate at the cost of unbalanced throughput allocation amongst the users, presumably resulting in over- and under- satisfaction of several users' requirements. For instance, both Water-filling and WMMSE algorithms attempt to benefit from good channel conditions, producing power configurations with enhanced power levels in the favorable (in terms of SINR level) channels. On the contrary, AVD results (lower plane of Fig. 5) illustrate the demands-driven approach, followed by the proposed DRL framework.

In this context, the proposed DRL scheme aims at satisfying the user requirements regardless of their channel condition, thus ensuring minimization between the requested and the allocated throughput. The proposed method achieves a total network throughput solution comparable to the Fixed Average scheme, also allocating approximately 80% total network throughput with respect to the WMMSE algorithm (about 270 Mpbs). However, concerning the user satisfaction (AVD performance metric), the proposed DRL method outperforms all the other baselines, achieving significantly lower values (AVD = 0.6, 1.6 and 4.3 for validation scenarios i, ii and iii, respectively), as depicted in the lower panel of Fig. 5. Evidently, as the user requirements increase in terms of the requested throughput, the difference between their demands and allocated throughput becomes more noticeable.

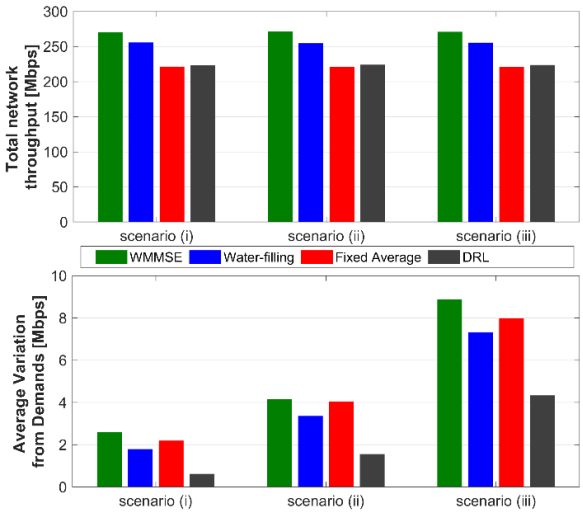

Fig. 3. Comparison amongst methods: DRL performance against *Waterfilling*, *WMMSE* and *Fixed Average* methods for the different validations scenarios in terms of total network throughput (up) and AVD (down) across 1000 network realizations.

4 Conclusion

In this work, a DRL framework for power regulation in heterogeneous cells is proposed and applied in 5G-compliant simulation scenarios. The objective of the proposed algorithm is the minimization of the difference between allocated and requested throughput of the mobile users through appropriate adjustment of both MaC and MiC transmit power. The assessment results clearly indicate that the presented DRL method effectively minimizes the variation between the user demands and the allocated throughput as compared to the Water-filling and WMMSE algorithms, as well as a fixed average power allocation scheme.

5 Acknowledgment

This work has been partially supported by the Affordable5G project, funded by the European Commission under Grant Agreement H2020-ICT-2020-1, number 957317 through the Horizon 2020 and 5G-PPP programs (www.affordable5g.eu/).

References

1. Andrews, J.G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A.C.K., Zhang, J.C.: What will 5G be? IEEE J. Sel. Areas Commun. 32, 1065–1082 (2014).
2. Trakadas, P., Nomikos, N., Michailidis, E.T., Zahariadis, T., Facca, F.M., Breitgand, D., Rizou, S., Masip, X., Gkonis, P.: Hybrid Clouds for Data-Intensive, 5G-Enabled IoT Applications: An Overview, Key Issues and Relevant Architecture. Sensors. 19, 3591

(2019).
3. Trakadas, P., Karkazis, P., Leligou, H.C., Zahariadis, T., Vicens, F., Zurita, A., Alemany, P., Soenen, T., Parada, C., Bonnet, J., Fotopoulou, E., Zafeiropoulos, A., Kapassa, E., Touloupou, M., Kyriazis, D.: Comparison of Management and Orchestration Solutions for the 5G Era. J. Sens. Actuator Networks. 9, 4 (2020).
4. Calabrese, F.D., Wang, L., Ghadimi, E., Peters, G., Hanzo, L., Soldati, P.: Learning Radio Resource Management in RANs: Framework, Opportunities, and Challenges. IEEE Commun. Mag. 56, 138–145 (2018).
5. Morocho-Cayamcela, M.E., Lee, H., Lim, W.: Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. IEEE Access. 7, 137184–137206 (2019).
6. Qi, Q., Minturn, A., Yang, Y.: An efficient water-filling algorithm for power allocation in OFDM-based cognitive radio systems. In: 2012 International Conference on Systems and Informatics, ICSAI 2012. pp. 2069–2073 (2012).
7. Shi, Q., Razaviyayn, M., Luo, Z.Q., He, C.: An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. IEEE Trans. Signal Process. 59, 4331–4340 (2011).
8. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018).
9. Zhang, C., Patras, P., Haddadi, H.: Deep Learning in Mobile and Wireless Networking: A Survey. IEEE Commun. Surv. Tutorials. 21, 2224–2287 (2019).
10. Balevi, E., Andrews, J.G.: A novel deep reinforcement learning algorithm for online antenna tuning. In: 2019 IEEE Global Communications Conference, GLOBECOM 2019 - Proceedings. (2019).
11. Zhang, Y., Kang, C., Ma, T., Teng, Y., Guo, D.: Power Allocation in Multi-Cell Networks Using Deep Reinforcement Learning. In: IEEE Vehicular Technology Conference. (2018).
12. Zhao, N., Liang, Y.C., Niyato, D., Pei, Y., Wu, M., Jiang, Y.: Deep Reinforcement Learning for User Association and Resource Allocation in Heterogeneous Cellular Networks. In: IEEE Transactions on Wireless Communications. pp. 5141–5152. (2019).
13. Nasir, Y.S., Guo, D.: Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks. IEEE J. Sel. Areas Commun. 37, 2239–2250 (2019).
14. Lei, L., Yuan, D., Ho, C.K., Sun, S.: Joint Optimization of Power and Channel Allocation with Non-Orthogonal Multiple Access for 5G Cellular Systems. (2016).
15. Xu, Z., Wang, Y., Tang, J., Wang, J., Gursoy, M.C.: A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs. In: IEEE International Conference on Communications. (2017).
16. Amiri, R., Mehrpouyan, H., Fridman, L., Mallik, R.K., Nallanathan, A., Matolak, D.: A Machine Learning Approach for Power Allocation in HetNets Considering QoS. In: IEEE International Conference on Communications. (2018).
17. Zhang, M., Chen, M.: Power Allocation in Multi-cell System Using Distributed Deep Neural Network Algorithm. In: International Conference on Wireless and Mobile Computing, Networking and Communications. IEEE Computer Society (2019).
18. Zhao, G., Li, Y., Xu, C., Han, Z., Xing, Y., Yu, S.: Joint Power Control and Channel Allocation for Interference Mitigation Based on Reinforcement Learning. IEEE Access. 7, 177254–177265 (2019).
19. Palomar, D.P., Chiang, M.: Alternative distributed algorithms for network utility maximization: Framework and applications. IEEE Trans. Automat. Contr. 52, 2254–2269 (2007).
20. Tang, M., Long, C., Guan, X.: Nonconvex optimization for power control in wireless CDMA networks. Wirel. Pers. Commun. 58, 851–865 (2011).
21. 3GPP: Study on channel model for frequencies from 0.5 to 100 GHz. Technical report (TR) 38.901, 3rd Generation Partnership Project (2017).