A Survey on Applications of Deep Reinforcement Learning in Resource Management for 5G Heterogeneous Networks

Ying Loong Lee*,† and Donghong Qin*
* Guangxi University for Nationalities, Guangxi, China
E-mail: donghong_qin@163.com
† Universiti Tunku Abdul Rahman, Selangor, Malaysia
E-mail: leeyingl@utar.edu.my

***Abstract*—Heterogeneous networks (HetNets) have been regarded as the key technology for fifth generation (5G) communications to support the explosive growth of mobile traffics. By deploying small-cells within the macrocells, the HetNets can boost the network capacity and support more users especially in the hotspot and indoor areas. Nonetheless, resource management for such networks becomes more complex compared to conventional cellular networks due to the interference arise between small-cells and macrocells, which thus making quality of service provisioning more challenging. Recent advances in deep reinforcement learning (DRL) have inspired its applications in resource management for 5G HetNets. In this paper, a survey on the applications of DRL in resource management for 5G HetNets is conducted. In particular, we review the DRL-based resource management schemes for 5G HetNets in various domains including energy harvesting, network slicing, cognitive HetNets, coordinated multipoint transmission, and big data. An insightful comparative summary and analysis on the surveyed studies is provided to shed some light on the shortcomings and research gaps in the current advances in DRL-based resource management for 5G HetNets. Last but not least, several open issues and future directions are presented.**

I. Introduction

The explosive growth of data traffics and multimedia services has led to an intensive research and development for fifth generation (5G) communications. 5G systems are expected to achieve up to 20 Gb/s peak data rates, three-times spectral efficiency and 100-times energy efficiency compared to the current fourth generation (4G) systems, as specified in the International Mobile Telecommunications (IMT)-2020 [1]. One of the key technologies to achieve the IMT-2020 specifications is heterogeneous networks (HetNets). The main feature of this technology is that low-power small-cell base stations (SBSs) are deployed at the hotspot or indoor areas within macrocells to enhance the data capacity in the areas [2]. Thanks to this feature, HetNets have been studied intensively, along with other technologies such as energy harvesting, network slicing, cognitive radio, coordinated multipoint transmission (CoMP) and caching. However, the introduction of small-cells within macrocells makes the resource management in the HetNets more complex and challenging compared to conventional cellular networks due to the fact that the HetNets need to take into account interference between small-cells and macrocells while guaranteeing quality of service (QoS) for the mobile users [2].

A. Motivation

Recently, several successes have been witnessed in the field of machine learning such as the AlphaGo [3]. This has inspired numerous researchers to investigate the applications of machine learning on resource management for 5G systems. In particular, deep reinforcement learning (DRL) has received much attention in the field due to its powerful optimization and convergence properties [4], [5]. In DRL, an *agent* learns the optimal policy, that is, determine the optimal *action* for each possible *state* (i.e., condition) of the environment, by estimating the immediate *reward* generated from performing an action after observing the state in every time epoch and by calculating the long-term discounted reward for each state-action pair. Then, for each possible state, the optimal policy can be obtained by observing the action which corresponds to the highest long-term discounted reward. To estimate the long-term discounted reward for each state-action pair, deep neural networks (DNNs) are used and can be trained online to learn approximating the reward values. The DNNs can thus save the memory required for storing and learning the long-term discounted reward, especially for environments with a large action and state spaces. For mathematical details regarding the DRL, the reader is suggested to refer to the comprehensive tutorial on DRL given in [5].

Intrigued by the powerful features of DRL, we are motivated to investigate the applications of DRL in resource management

This work is supported in part by the National Natural Science Foundation of China under grant numbers 61462009 and 61862007, in part by the Natural Science Foundation of Guangxi under grant numbers 2018GXNSFAA281269 and 2018GXNSFAA138147, in part by the Universiti Tunku Abdul Rahman Research Fund (UTARRF) under grant number IPSR/RMC/UTARRF/2018-C2/L02, and in part by the Ministry of Education (MOE) Malaysia through the Fundamental Research Grant Scheme (FRGS) under the project entitled "Intelligent and Energy-Efficient Network Slicing for Sustainable Virtualized Multi-Tenant 5G Systems" with grant number FRGS/1/2019/ICT05/UTAR/02/1.

research for 5G HetNets in various domains, and to explore possible future research directions for further development of 5G and beyond.

B. Related Work

A number of surveys on applications of machine learning techniques in communications and networking can be found in the literature [4]-[10]. In [4] and [5], the authors review the applications of DRL in communications and networking. The survey in [6] investigates various machine learning applications in wireless networks. In [7] and [8], the surveys focus on deep learning in mobile and wireless networking. The survey in [9] emphasizes on machine learning for massive machine type communications in HetNets. In [10], the survey reviews the applications of machine learning in data-driven wireless networks.

Despite the fact that a number of related surveys [4]-[10] have been done, the scope of these surveys is too wide, and discussion and analysis on each specific topic covered in the surveys, especially on resource management for 5G HetNets, is not sufficiently in-depth. In this paper, we survey the applications of DRL in resource management for 5G HetNets in various domains. Unlike the surveys in [4]-[10], our survey is more focused and specific to resource management for 5G HetNets, which provides a quicker review in this area. Moreover, we relate the discussion on the survey with 5G specifications, which are more useful to industry researchers.

C. Contributions

In this paper, a survey of the applications of DRL in resource management for 5G HetNets is presented. The contributions of this paper are threefold and can be summarized as follows:

1) A survey on applications of DRL in resource management for 5G HetNets in various domains is provided. These domains include energy harvesting, network slicing, cognitive HetNets, CoMP and big data.
2) A comparative summary and analysis on the surveyed DRL-based resource management schemes in terms of resource management functions, 5G design aspects and practicality is presented.
3) Based on the aforementioned comparative analysis, several open issues and future directions, which are critical for further exploration, are identified and highlighted.

D. Organization

The remainder of this paper is organized as follows: Section II provides an overview of 5G HetNets. In Section III, DRL-based resource management schemes for 5G HetNets in different domains are reviewed and discussed. A comparative summary and analysis on the surveyed schemes is provided in Section IV. Several open issues and future directions are presented in Section V. Finally, Section VI concludes the paper.

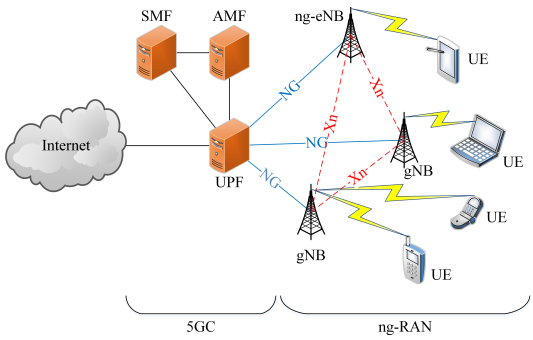

Fig. 1. 5G architecture.

II. Overview of 5G HetNets

Major architectural changes have been introduced to the core network and radio access network sides for 5G cellular systems. In this section, an overview on 5G architectures and HetNets is given, followed by an introduction to 5G resource management functions.

A. 5G Architecture

Fig. 1 shows the overall architecture of a 5G system consisting of a core network known as 5G Core Network (5GC) and a radio access network known as Next Generation Radio Access Networks (NG-RAN), interconnected through the so-called NG interfaces. In the 5GC, there are three main components: Access and Mobility Management Function (AMF), User Plane Function (UPF) and Session Management Function (SMF). The AMF is responsible for non-access stratum (NAS) security and idle state mobility, the UPF handles protocol data unit (PDU) and mobility anchoring, and the SMF allocates Internet protocol (IP) address for user equipment (UE) and controls PDU sessions. For further details about 5GC, we refer the reader to [11] and [12].

On the other hand, the NG-RAN consists of two types of base stations (BSs): next-generation NodeB (gNB) and next-generation evolved NodeB (ng-eNB). Both gNB and ng-eNB manage the downlink and uplink data transmission between the 5GC and the UE in the 5G network. The difference between gNB and ng-eNB is that the former serves the UE based on the 5G New Radio (NR) protocol whereas the latter is based on the Long Term Evolution (LTE) protocol. Nevertheless, both types of BSs are responsible for radio resource management and can be interconnected via so-called Xn interfaces in the 5G network. We refer the reader to [11] for further details about NG-RAN.

5G systems adopt a flexible radio resource structure based on scalable orthogonal frequency division multiplexing (OFDM) numerologies. The channel bandwidth is divided into smaller subchannels known as the resource blocks (RBs)[1], each consists of 12 consecutive frequency subcarriers where

[1]An RB is a subchannel which is the smallest unit of frequency band allocated by BSs to their associated UEs for data transmission

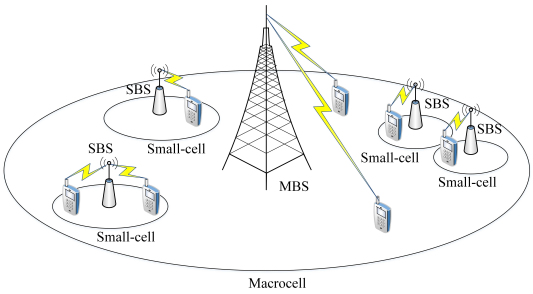

Fig. 2. HetNet architecture.

the subcarrier spacing can be scalable. Unlike LTE, 5G systems support subcarrier spacing of 15 kHz (same as LTE), 30 kHz, 60 kHz, 120 kHz and 240 kHz for each RB. The transmission timeslot size for the RB also varies as 1 ms, 0.5 ms, 0.25 ms, 0.125 ms and 0.0625 ms, respectively in the same order [13]. Such a resource structure allows 5G systems to flexibly support services with diverse requirements. For example, RBs with a large subcarrier spacing and short timeslot size are suitable for high-bandwidth, low-latency applications, whereas RBs with a short subcarrier spacing and long time timeslot size are suitable for short-packet, delay-tolerant applications.

B. Heterogeneous Networks

Although 5G specifications for small-cells have not been provided yet, it has been widely agreed by both academia and industry that the small-cell technology will be essential in 5G networks [14]. Fig. 2 shows a 5G HetNet consisting of several small-cells laying within the macrocell. It is worth noting that in the context of 5G, the macrocell BS (MBS) refer to the gNB or ng-eNB. The small-cell base stations (SBSs) basically behave similarly as the MBS, except that their transmission power and computational capacity are lower compared to those of the MBS. We can further classify small-cells based on their transmission power levels such as picocells (∼30 dBm) and femtocells (∼20 dBm). Picocell base stations (PBSs) are usually deployed by the network operator at hotspot areas where the UE density is high. On the other hand, femtocell base stations (FBSs) are user-deployed at indoor areas via the broadband connections such as the digital subscriber line (DSL) to the core network.

C. Basic Resource Management Functions for 5G HetNets

Here, we review several essential resource management functions for 5G HetNets, as follows.

1) Link adaptation and power control: These two functions actually reside in the physical (PHY) layer of each BS and are responsible to optimize the data transmission between the BSs and UEs. Link adaptation is responsible for selecting the appropriate modulation and coding scheme (MCS) for data transmission given the channel condition of the RBs selected for the transmission, in order to ensure a target bit error rate. The MCSs supported in 5G systems include quadrature PSK (QPSK), 8-quadrature amplitude modulation (QAM), 16QAM, 64QAM, and 256QAM [11]. In 5G systems, the same MCS must be applied to all RBs assigned to a UE within one transmission duration [11]. On the other hand, power control allows adjusting the transmission power level of the BS on the selected RBs to improve the channel quality.

2) Scheduling: This function is located in the medium access control (MAC) sublayer in Layer 2 of each BS, and is responsible for allocating RBs to the UEs associated with the BS. The RB allocation decision can be made based on several factors including the RB availability, the channel condition (measured and reported by the UEs) and quality of service (QoS) required by the UEs. In the 5G specifications stated in [11] and [15], 5G systems allows RB scheduling in units of slots, and support type-0 and type-1 scheduling in the frequency domain. In the type-0 scheduling, any number of groups of consecutive RBs, where each RB group may consist of two or more consecutive RBs depending on the *bandwidth part*[2] size configured for the UE, can be allocated to the UE; whereas the type-1 scheduling allocates any number of consecutive RBs within the bandwidth part configured for the UE.

3) User association: This function is actually performed via the cell selection, admission control and handover mechanisms located in the radio resource control (RRC) layer of each BS, which allows UEs to establish connections with targeted BSs for data transmission. User association can be implemented based on various objectives such as load balancing and QoS satisfaction, which are similar to those for scheduling. It is noteworthy that small-cells can operate in different modes where UEs can have different priorities for access and association [2], [16].

4) Cell Activation/Deactivation: This function is responsible to turn BSs into active or sleep modes. BSs can be turned to sleep mode when they do not serve any users, thus reducing the power consumption.

III. DRL-Based Resource Management for 5G HetNets

In this section, we survey the DRL-based resource management studies for 5G HetNets. In particular, we categorize the surveyed schemes into the following areas: Conventional HetNets, HetNets with energy harvesting, HetNets with network slicing, cognitive HetNets, CoMP-enabled HetNets, and HetNets with big data.

A. Conventional HetNets

In [17], the authors study the user association and resource allocation problem for a three-tier HetNet consisting of MBSs, PBSs and FBSs. The study formulates an optimization problem that aims to simultaneously maximize the network throughput and minimize the network power consumption by means of user association and RB allocation. To approach this problem, the authors model it as a Markov decision process

[2]In the 5G context, a bandwidth part refers to a subset of contiguous RBs within a carrier bandwidth.

(MDP) and solve it using a multi-agent DRL technique known as double deep Q-network (DDQN) [18]. In the MDP, the UEs behave as the agents with their states, actions, and rewards modeled as their QoS satisfaction level, user association and RB allocation decisions, and the objective function value in the aforementioned optimization problem subtracted by an amount of cost after the agent selects an action, respectively. In the DDQN, Q-learning (QL) [19] is used to solve the MDP by finding the optimal policy through iteratively updating the Q-value[3] for each state-actionh pair. Further, two DNNs are used, with one (known as the *online network*) being used to determine the best action that leads to the highest Q-value and another (known as the *target network*) to approximate the Q-value. Subsequently, the transition, i.e., the state change, the action performed and the reward are stored in a *replay memory*. The transitions stored in the memory will then be used to train the DNNs by randomly sampling the minibatches of transitions in the memory. The proposed DDQN-based scheme is shown to outperform the conventional QL and deep QL (DQL)[4] techniques in terms of achievable network throughput, however the network power consumption is not evaluated. It is noteworthy that the RB allocation mechanism in [17] restricts each UE to be assigned at most one RB, unlike 5G that can allocate several consecutive RBs in groups, as mentioned in Section II-C.

B. HetNets with Energy Harvesting

Energy harvesting wireless communications have been gaining interest in both academia and industry. The key idea of energy harvesting is to install renewable (off-grid) energy sources such as photovoltaic solar panels to help power up wireless networks. As such, greenhouse emissions and the on-grid energy costs can be reduced. In [20], the application of DRL on resource management for HetNets equipped with both on- and off-grid energy sources is investigated. The study considers a HetNets with only SBSs (i.e., without MBSs) and a cloud processor, which is responsible to control the activation of all SBSs in the network. A single-objective optimization problem is formulated to maximize the network energy efficiency and minimize the total traffic delay at a particular time instance, by means of small-cell activation. To solve the problem, DQL is implemented with the cloud processor behaving as the learning agent. Here, the state of the network is modeled as the SBSs' harvested energy levels, total energy levels, traffic loads, throughput values and traffic delays. The actions of the agent are the SBS activation decisions. After the agent performs an action, the reward is calculated as a long-term discounted version of the objective function in the optimization problem. The solution to the problem is then found by performing QL through iteratively updating the Q-value for each state-action pair. In each update, the Q-value is approximated using a DNN and the new transition generated will be stored in the replay memory for training the DNN. Simulation results show that the proposed DQL-based SBS activation scheme attains substantial energy efficiency gain and traffic delay reduction compared to the conventional QL-based scheme. Despite such promising results, the study is lacking of considering the MBS in the network, since an MBS consumes significantly more power compared to an SBS.

C. HetNets with Network Slicing

Network slicing has originally emerged as a network sharing technology to allow multiple mobile operators sharing the same physical network infrastructure. The underlying principle of network slicing is that the physical network infrastructure and resources are "sliced" into a number of *network slices* with each consisting of a set of virtual network functions and resources, and each of the operators operate on one of these network slices [21]. Currently, the network slicing technology not only can slice the physical network based on the number of mobile operators sharing the network, but can also slice according to the number of different services provided to the UEs in the network [4]. In [22], the authors focuses on radio access network slicing in a 5G HetNet, specifically the slicing of RBs. Considering the heterogeneous backhaul capacity of the HetNet, a RAN slicing framework is developed to slice the RBs according to the service types currently available in the network. In this framework, DQL is applied to autonomously refine the slicing process. Assuming the presence of a controller which behaves as the agent, the state for the DQL is modeled as the QoS utility, fraction of RBs allocated to each slice and RB utilization of each slice. The available actions of the agent are a set of increment/decrement percentages for increasing/decreasing the fraction of resources allocated to each slice. The reward for each state-action pair is calculated as the weighted sum of the QoS utility and slice resource utilization. The Q-values are approximated using a DNN trained the with experience replay strategy. The authors implemented the proposed DQL in two modes, i.e., allocating and reserving the resource fractions of each slice. Performance results show that substantial QoS satisfaction and resource utilization improvements are attained over the conventional network slicing schemes. Nevertheless, the channel model considered in [22] appears to have omitted the interference between the MBS and the SBSs. Hence, the performance results in an interference-accounted scenario, especially the QoS satisfaction, may vary significantly. Moreover, the RB slicing and allocation mechanisms are not compliant with the 5G specifications for RB scheduling. Therefore, the implementation of the proposed scheme in an actual 5G system may be impractical.

D. Cognitive HetNets

Cognitive radio [23] has been applied to SBSs in HetNets to address the spectrum scarcity issues. The CR-enabled SBSs, also known as the *secondary transmitter*, can detect and exploit under-utilized spectrum in the *primary band* belonging to the MBS for data transmission. One interesting study in [24] has developed a link adaptation scheme for such

[3]The Q-value of a state-action pair represents the long-term discounted reward of the pair.

[4]QL with one DNN for approximating the Q-value.

networks based on DRL. The main objective of this study is to allow the MBS to learn and maximize its transmission rate by choosing the appropriate MCS based on the historical interference information under a scenario when secondary transmissions are ongoing. Here, the MBS behaves as the agent and the state of the DRL framework consists of the action taken, reward yielded, signal-to-interference-plus-noise ratio (SINR) received and SINR per bit received by the MBS. The actions of the MBS are the different choices of MCSs and the reward for each state-action pair is calculated by the amount of bits transmitted using the selected MCS minus the MCS switching cost. Similar to [17], DDQN is applied to iteratively update the Q-value for each state-action pair to find the optimal MCS selection. The proposed MCS selection scheme is shown to achieve a reasonably low system overheads and high transmission rate. However, the proposed scheme appears to only consider a single primary data transmission, which is a less realistic case.

E. CoMP-Enabled HetNets

The main purpose of CoMP is to boost the transmission rate of UEs as it allows each UE to have concurrent data transmissions with multiple MBSs and SBSs within a transmission duration. Since HetNets consist of densely deployed multiple MBSs and SBSs, CoMP can be fully leveraged to enhance the network capacity. Exploiting this fact, the study in [25] design a joint user association and power allocation scheme based on DQL for CoMP-enabled HetNets. In particular, the authors aims to address the non-line-of-sight (NLOS)[5] problem for the networks. An optimization problem is formulated to maximize the network throughput by means of user association and power control, subject to the transmission power and SINR constraints. The DQL is implemented as follows: The UEs behave as the agents having states defined as the discretized channel gain, actions as the user association and power control decisions, and the reward function as the total network throughput. The proposed DQL scheme is shown to outperform an existing scheme in terms of throughput for the network.

F. HetNets with Big Data

One particular study in [26] has investigated the application of networking, caching and computing on HetNets for energy-efficient communications. The authors consider a HetNet, where each BS can cache a certain amount of content (e.g., text or video files) and a computing server is installed at the MBS, and that the channel bandwidth is equally distributed among all UEs. A DQL-based user association scheme is developed to connect the UEs to the SBSs or the MBS for obtaining the content requested by the UEs, depending on the availability of the content in each of the BSs. For the proposed DQL-based scheme, the state of the network is modeled as the channel condition of each SBS, the computational capability of the computing server in the network, the computational capability of the UEs, the cache state of each UE and the file matching parameter. The actions of the network are the user association decisions of each UE. The reward function is modeled as the energy consumption required for data transmission and computing. Significant energy reduction is observed across different cache sizes in the SBSs for the proposed DQL-based user association scheme compared to existing schemes. However, it is unclear whether the MBS or the computing server will behave as the agent executing the proposed scheme.

[5]A NLOS link refer to a partially-obstructed transmission link between a transmitter and a receiver.

IV. Comparative Summary and Analysis

In this section, the DRL-based resource management schemes surveyed in Section III are compared and summarized in terms of resource management functions, 5G design aspects and practicality, as shown in Table I.

For the resource management functions, we observe from Table I that link adaptation, power control and cell activation have not much been jointly considered in DRL-based resource management for 5G HetNets. In particular, these resource management functions, which can have great impact to energy efficiency, have not been incorporated in [17], [20] and [26] that focus on energy efficiency.

For the 5G design aspects, we discover that only one study, i.e., [24] has considered analyzing the system overheads incurred by the implementation of the DRL. The analysis of system overheads is essential as a large amount of overheads can degrade the throughput and energy efficiency performance. On the other hand, we notice that the study in [22] has omitted the interference aspect, which is a vital issue that can profoundly affect the throughput performance of the HetNet.

For the practicality, we identify the implementation framework and 5G-compatibility of the surveyed schemes. We find that the DRL-based resource management schemes proposed in [17], [24] and [25] are implemented in a decentralized framework, which results in a low-complexity operation since the optimization tasks have been distributed and executed concurrently. Nonetheless, these studies assume that the agents have a sufficient computational capability for executing the distributed tasks and the resource management operation is perfectly synchronized. Meanwhile, the DRL-based resource management schemes in [20], [22] and [26] are based on a centralized implementation. The optimization tasks are all taken care of by the central processor. However, the centralized schemes may fail to perform if the central processor malfunctions or breaks down due to unforeseen circumstances. As such, a backup processor may be required. We also note that the studies that involve RB allocation, i.e., [17] and [22] are not 5G-compatible, as a result of the fact that their RB allocation mechanisms are not compliant with the 5G specifications.

V. Open Issues and Future Directions

Based on the comparative analysis from Section IV, several open issues and possible future directions for resource management in 5G HetNets are identified.

TABLE I
COMPARISON BETWEEN VARIOUS SURVEYED DRL-BASED RESOURCE MANAGEMENT SCHEMES FOR 5G HETNETS

Domain		Conventional	Energy Harvesting	Network Slicing	Cognitive	CoMP	Big Data
Study		[17]	[20]	[22]	[24]	[25]	[26]
Resource Management Functions	Link Adaptation				✓		
	Power Control					✓	
	Scheduling (RB Allocation)	✓		✓			
	User Association	✓		✓		✓	✓
	Cell Activation		✓				
Design Aspects	Energy Efficiency	✓	✓				✓
	QoS	✓	✓	✓		✓	
	Resource Utilization			✓	✓		
	Interference	✓	✓		✓	✓	✓
	System Overheads				✓		
Practicality	Implementation Framework	Decentralized to UEs	Centralized to a cloud processor	Centralized to a controller	Decentralized to the MBS	Decentralized to UEs	Centralized
	5G-Compatibility	No	Yes	No	Yes	Yes	Yes

A. DRL-Based Joint Resource Management Functions for 5G HetNets

From our survey, only a few studies [17], [22], [25] have applied DRL to jointly perform multiple resource management functions for 5G HetNets. The performance of the networks can be further optimized by incorporating other resource management functions as part of the actions of the DRL agents in the proposed schemes. For instance, power control and cell activation can jointly performed together with user association and scheduling in the DRL-based resource management scheme proposed in [17] to further enhance the network energy efficiency.

B. Multi-Objective DRL-Based Resource Management

Since 5G systems are expected to excel in various aspects (e.g., high spectral and energy efficiencies, ultra-low delays and high peak rates), multi-objective resource management is crucial. However, multi-objective resource management is very challenging for 5G HetNets as many of the objectives or design aspects contradict among each other [27]. For example, high resource utilization may lead to high interference. Several of the surveyed studies [17], [20], [22] have considered such resource management for 5G HetNets. However, the consideration of the multiple objectives in these schemes is not sufficiently comprehensive to meet the 5G expectations.

C. Flexible Resource Management Design for 5G and Beyond

In 5G, the subcarrier spacing and timeslot size of RBs can be flexibly scaled to support the applications, depending on the requirements and characteristics of the applications. However, such flexibility is largely omitted in the RB scheduling design in the literature (e.g., [17] and [22]) for 5G HetNets. As next-generation wireless networks will support applications with highly diverse requirements, intelligent resource management is necessary to determine and flexibly scale the resource structure to accommodate the applications.

D. DRL-Based Load Balancing for 5G HetNets

Despite that a number of 5G design aspects have been considered for HetNets, DRL-based load balancing for HetNets has not been investigated under the 5G context. The main objective of load balancing is to shift some of the UEs connected to the congested BSs to other underloaded BSs, thereby enhancing BS utilization and leading to better QoS provisioning for UEs. This area has been regaining the interest for 5G system development and some related studies such as [28] have been carried out recently. Nevertheless, the existing related studies are still lacking in terms of autonomous adaptation and optimization capabilities, which are vital to realize a self-organizing 5G network, hence the need for machine learning techniques such as DRL.

VI. CONCLUSIONS

In this paper, we have conducted a survey of the application of DRL in resource management for 5G HetNets. Firstly, an overview of 5G HetNets has been provided to introduce the 5G architecture, HetNets and several essential resource management functions for HetNets. Then, we have reviewed several DRL-based resource management schemes for 5G HetNets in different domains such as energy harvesting, network slicing, cognitive HetNets, CoMP-enabled HetNets, and big data. Specifically, we have revealed the DRL framework design of each of these schemes in terms of states, actions and rewards in meeting the designated resource management requirements and objectives of the studies. Next, a comparative summary and analysis on the surveyed studies in terms of resource management functions, 5G design aspects and practicality is provided. From the comparative analysis, we have drawn and highlighted several open issues and future directions, which are critical for future development of DRL-based resource management for 5G HetNets.

REFERENCES

[1] "IMT vision—Framework and overall objectives of the future development of IMT for 2020 and beyond," International Telecommun. Union, Geneva, Switzerland, ITU-Recommendation M.2083-0, Sep. 2015.

[2] Y. L. Lee, T. C. Chuah, J. Loo, and A. Vinel, "Recent Advances in Radio Resource Management for Heterogeneous LTE/LTE-A Networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2142–2180, 4th Quarter, 2014.

[3] (2016, Jan.) Google achieves AI "breakthrough" by beating Go champion. BBC. [Online]. Available: https://www.bbc.com/news/technology-35420579

[4] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang and L.-C. Wang, "Deep Reinforcement Learning for Mobile 5G and Beyond: Fundamentals, Applications, and Challenges," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 44-52, June 2019.

[5] N. C. Luong et al., "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," *IEEE Commun. Surveys Tuts*, pp. 1-43, May 2019 (Early Access).

[6] Y. Sun, M. Peng, Y. Zhou, Y. Huang and S. Mao, "Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues," *IEEE Commun. Surveys Tuts*, pp. 1-37, June 2019 (Early Access).

[7] C. Zhang, P. Patras and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Commun. Surveys Tuts*, pp. 1-67, March 2019 (Early Access).

[8] S. K. Sharma and X. Wang, "Towards Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions," *IEEE Commun. Surveys Tuts*, pp. 1-46, May 2019 (Early Access).

[9] Q. Mao, F. Hu and Q. Hao, "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey," *IEEE Commun. Surveys Tuts*, vol. 20, no. 4, pp. 2595-2621, Fourthquarter 2018.

[10] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu and F. Kojima, "Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks," *IEEE Access*, vol. 6, pp. 32328-32338, 2018.

[11] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2(Release 15)" Sophia-Antipolis, France, TS 38.300, June 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.300/38300-f60.zip

[12] "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2 (Release 16)" Sophia-Antipolis, France, TS 23.501, June 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/23_series/23.501/23501-g10.zip

[13] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Physical channels and modulation (Release 15)" Sophia-Antipolis, France, TS 38.211, June 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.211/38211-f60.zip

[14] P. Demestichas et al., "5G on the Horizon: Key Challenges for the Radio-Access Network," *IEEE Veh. Technol. Mag.*, vol. 8, no. 3, pp. 47-53, Sept. 2013.

[15] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Physical layer procedures for data (Release 15)" Sophia-Antipolis, France, TS 38.214, June 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.214/38214-f60.zip

[16] Y. L. Lee, J. Loo and T. C. Chuah, "Dynamic Resource Management for LTE-Based Hybrid Access Femtocell Systems," *IEEE Syst. J.*, vol. 12, no. 1, pp. 959-970, March 2018.

[17] N. Zhao, Y. Liang, D. Niyato, Y. Pei and Y. Jiang, "Deep Reinforcement Learning for User Association and Resource Allocation in Heterogeneous Networks," in *Proc. 2018 IEEE GLOBECOM*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1-6.

[18] H. van Hasselt, A. Guez and D. Silver, "Deep Reinforcement Learning with Double Q-Learning," in *Proc. 13th AAAI Conf. Artificial Intelligence*, Feb. 2016, pp. 2094 - 2100.

[19] C. J. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no.3-4, pp. 279-292, 1992.

[20] H. Li, H. Gao, T. Lv and Y. Lu, "Deep Q-Learning Based Dynamic Resource Allocation for Self-Powered Ultra-Dense Networks," in *Proc. 2018 IEEE ICC Workshops*, Kansas City, MO, 2018, pp. 1-6.

[21] Y. L. Lee, J. Loo, T. C. Chuah and L.-C. Wang, "Dynamic Network Slicing for Multitenant Heterogeneous Cloud Radio Access Networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2146-2161, April 2018.

[22] G. Sun, K. Xiong, G. O. Boateng, D. Ayepah-Mensah, G. Liu and W. Jiang,"Autonomous Resource Provisioning and Resource Customization for Mixed Traffics in Virtualized Radio Access Network," *IEEE Syst. J.*, pp. 1-12, June 2019 (Early Access).

[23] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201-220, Feb. 2005.

[24] L. Zhang, J. Tan, Y. Liang, G. Feng and D. Niyato, "Deep Reinforcement Learning-Based Modulation and Coding Scheme Selection in Cognitive Heterogeneous Networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3281-3294, June 2019.

[25] C. Luo, J. Ji, Q. Wang, L. Yu and P. Li, "Online Power Control for 5G Wireless Communications: A Deep Q-Network Approach," in *Proc. 2018 IEEE ICC*, Kansas City, MO, May 2018, pp. 1-6.

[26] Y. He, Z. Zhang and Y. Zhang, "A Big Data Deep Reinforcement Learning Approach to Next Generation Green Wireless Networks," in *Proc. 2017 GLOBECOM*, Singapore, 2017, pp. 1-6.

[27] Y. L. Lee, J. Loo, T. C. Chuah and A. A. El-Saleh, "Fair Resource Allocation With Interference Mitigation and Resource Reuse for LTE/LTE-A Femtocell Networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8203-8217, Oct. 2016.

[28] Y. L. Lee, T. C. Chuah, A. A. El-Saleh and J. Loo, "User Association for Backhaul Load Balancing With Quality of Service Provisioning for Heterogeneous Networks," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2338-2341, Nov. 2018.