IEEE TCCN Special Section Editorial: Deep Reinforcement Learning for Future Wireless Communication Networks

Shimin Gong, Hoang Thai Dinh, Dusit Niyato, Ahmed El Shafie, Antonio De Domenico, Emilio Calvanese Strinati, and Jakob Hoydis

We are delighted to introduce the readers to this special section of the IEEE Transactions on Cognitive Communications and Networking (TCCN), which aims to explore recent advances and address practical challenges in the applications of DRL in modern wireless networks. We have received a total number of 23 submissions, and after a rigorous review process, 13 articles have been selected for publication, which are briefly discussed as follows.

The first article is entitled "Deep Reinforcement Learning for Adaptive Caching in Hierarchical Content Delivery Networks", authored by Sadeghi *et al.* In this article, a decentralized caching scheme for optimally distributing the limited storage capacity across network entities is invesitgated in a hierarchical network system. The two-way interactions between caching decisions at the parent and leaf nodes are modeled in a reinforcement learning framework. Due to the large continuous state space, a scalable deep reinforcement learning approach based on hyper deep Q-networks (DQNs) is proposed to learn the Q-function, and thus the optimal caching policy, in an online fashion. Numerical results verify that the proposed caching policy for the parent node can adapt itself to local policies of leaf nodes and the dynamics of file requests, leading to remarkable caching performance improvment.

In the second article, entitled "DeepChunk: Deep Q-Learning for Chunk-based Caching in Wireless Data Processing Networks", a novel DQN base framework, namely DeepChunk, is proposed for chunk-based caching in wireless data processing networks. By jointly learning limited network, data streaming, and processing statistics at runtime, DeepChunk makes caching decisions to optimize multiple objectives including chunk hit ratio, processing stall time, and download time while being self-adaptive with time-varying workload and network conditions. A prototype of DeepChunk is implemented and tested with real-world WiFi and 4G traces. The experiments demonstrate significant improvement, i.e., 52% increase in total reward and 68% decrease in processing stall time, over a number of baseline policies.

The third article, entitled "Contextual Multi-Armed Bandit for Cache-Aware Decoupled Multiple Association in UDNs: A Deep Learning Approach", by Dai *et al.*, investigated the challenging user association problem in heterogeneous ultra dense networks, due to densely deployed base stations (BSs) with highly variable transmit powers and topologies. A novel cache-aware association mechanism is proposed for full-duplex ultra dense networks, which allows a user to associate with multiple BSs in uplink and downlink in a decoupled manner. To obtain the optimal strategy, a DQN based algorithm is used to solve a contextual multi-armed bandit problem with unknown and large-scale network states. The convergence of the algorithm is proven. Simulation results validate that the system performance can be improved by about 39% to 42% compared with the existing association mechanisms.

The fourth article, entitled "Deep Reinforcement Learning for Intelligent Internet of Vehicles: An Energy-Efficient Computational Offloading Scheme", by Ning *et al.*, focuses on the fulfillment of demanding service requirements for vehicular entertainment. A three-layer offloading framework is proposed for the intelligent Internet of Vehicles (IoV) to minimize its overall energy consumption. Due to the high computational complexity, the energy minimization problem is decomposed into two parts: flow redirection and offloading decision, solved by a deep reinforcement learning approach. Performance evaluations based on real-world traces of taxis demonstrate that the average energy consumption can be decreased by around 60% compared with the baseline algorithm.

In the fifth article, entitled "Reinforcement Learning Based Vehicle-cell Association Algorithm for Highly Mobile Millimeter Wave Communication", by Khan *et al.*, the problem of vehicle-cell association in millimeter wave communication networks is investigated in the paradigm of vehicle-to-everything (V2X) communications. The aim is to maximize the time average rate per vehicular user while ensuring a target minimum rate for all users with low signaling overhead. The user association problem is firstly formulated as a discrete non-convex optimization. Then, the optimization is solved by a low-complexity approximate algorithm devised based on the distributed deep reinforcement learning (DDRL) and the asynchronous actor critic algorithm (A3C). This allows each road side unit (RSU) to make local decision independently and forward it to a central entity, which computes the global reward and then feeds it back to individual RSUs. This algorithm is shown to reduce control overhead and computational complexity compared to an online algorithm for the non-convex optimization problem. Simulation results show that the proposed solution achieves up to 15% gain in terms of sum rate and 20% reduction in outage performance compared to several baseline designs.

In the sixth article, entitled "Blockchain-Based Distributed Software-defined Vehicular Networks: A Dueling Deep Q-Learning Approach", by Zhang *et al.*, network security issues are invesigated in vehicular ad hoc networks (VANETs). A novel blockchain-based distributed software-defined VANET (block-SDV) framework is introduced to build a secure architecture to fight against malicious users and avoid performance degradation. The trust features of blockchain nodes, the number of consensus nodes, trust features of each vehicle, and the computational capability of the blockchain are considered in a joint optimization problem, which is modeled in a Markov decision process and further solved by a novel dueling DQN with prioritized experience replay. Simulation results demonstrate the effectiveness of the proposed block-SDV framework.

In the seventh article, entitled "Optimal UAV Base Station Trajectories Using Flow-Level Models for Reinforcement Learning", by Saxena *et al.*, a flow-level model is proposed for UAV base station networks to realistically characterize its performance in terms of a broad range of flow- and system-level metrics. Based on the flow-level model, a deep reinforcement learning approach is used to learn in offline the optimal UAV trajectories that maximize a cumulative performance metric, given the user traffic density and starting UAV locations. The online performace is evaluated in a discrete event simulator. Simulation results show that the proposed approach achieves approximately a three-fold increase in the average user throughput compared to the initial UAV placement, while balancing traffic loads across different base stations.

In the eighth article, entitled "Online Antenna Tuning in Heterogeneous Cellular Networks with Deep Reinforcement Learning", by Balevi *et al.*, the complex optimization of antenna tilt angle, and vertical and horizontal half-power beamwidths of the macrocells in a heterogeneous cellular network (HetNet) is proposed and practically solved by a two-step compromise algorithm based on a multi-agent deep reinforcement learning approach. Simulation results verify that the two-step algorithm performs better than the classic single- and multi-agent reinforcement learning. It can also guarantee certain performance gain with high environmental dynamics.

The ninth article, entitled "A Deep Actor-Critic Reinforcement Learning Framework for Dynamic Multichannel Access", by Zhong *et al.*, proposed a actor-critic based deep reinforcement learning approach for dynamic multi-channel access to improve the efficiency of spectrum usage. A decentralized multi-agent framework is further extended for performance evaluation in multi-user networks. The probability of users' access to channels with favorable conditions and the probability of collision are analyzed. The efficiency of the actor-critic framework can be verified by computing the percentage of runtime that can be saved compared to the DQN framework.

The tenth article, entitled "Distributed Power Control for Large Energy Harvesting Networks: A Multi-Agent Deep Reinforcement Learning Approach", by Sharma *et al.*, develop a multi-agent reinforcement learning framework to obtain online power control policies for a large energy harvesting multiple access channel. The online power control problem is firstly modeled as a discrete-time mean-field game (MFG) with a unique stationary solution. Then, a deep reinforcement learning approach is proposed to learn the stationary solution of the game in a completely distributed fashion. A deep neural network (DNN) based centralized as well as distributed online power control schemes are also proposed as different benchmarks. Simulation results show that the DNN based centralized power control policy provides a very good performance for large energy harvesting networks. Besides, both of the two distributed policies can achieve a close performance to that of the centralized policy.

The eleventh article, entitled "A Novel Transmission Scheduling Based on Deep Reinforcement Learning in Software-Defined Maritime Communication Networks", by Yang *et al.*, proposed a software-defined framework to improve the quality of service (QoS) of data transmission in maritime communication. An enhanced DQN algorithm is employed to learn the optimal scheduling strategy as fast as possible and accurately after a plethora of data self-learning. Simulation results verify an enhnaced quality of service provisioning, as well as improved reliability and the timeliness of maritime communication.

The twelveth article, entitled "Deep Reinforcement Learning Based Mobility-Aware Robust Proactive Resource Allocation in Heterogeneous Networks", by Li *et al.*, proposed a mobilityCaware robust proactive resource allocation approach in heterogeneous networks to minimize service delay under constraints of different levels of quality-of-service (QoS) requirement and mobility intensity. It preCallocates resources in both time and frequency domains among mobile users based on the prediction of users' trajectories in a hidden Markov model. Based on the pricinple of distributed acting and centralized criticizing, a multiCactor deep deterministic policy gradient algorithm is designed to flexibly coordinate resource allocation among multiple mobile users over time horizon. Simulations demonstrate that the proposed approach achieves better robustness against prediction uncertainty and has good adaptiveness to users' rate requirements and mobility intensities.

Finally, in the last article, "DeepVR: Deep Reinforcement Learning for Predictive Panoramic Video Streaming", by Xiao *et al.*, investigated tile-based adaptive streaming for predictive panoramic video delivery to minimize the requirement on bandwidth. A long short-term memory (LSTM) model is proposed to predict user's field of view (FoV) and a deep reinforcement learning approach is used to adapt the bitrate according to the environmental dynamics. The proposed quality adaptation policy is validated in a prototype system and shown to achieve a superior performance in terms of the quality of experience (QoE) score, outperforming existing panoramic video streaming frameworks.

Our Guest Editor team is pleased with the technical depth and span of this Special Section in IEEE TCCN. We also recognizes that it cannot cover all emerging DRL issues in wireless communications and networking. We sincerely thank

all the authors and reviewers for their efforts, and the Editor-in-Chief and Staff Members for their gracious support. We hope that the readers will enjoy this special section.

Shimin Gong (M'15) received the B.E. and M.E. degrees in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2012, respectively, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in 2014. He is currently an Associate Professor with the School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China. Before that, He was an associated researcher with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. He was a visiting scholar at the Chinese University of Hong Kong, Shatin, Hong Kong in 2011, and the University of Waterloo, Waterloo, ON, Canada, in 2012. He is a recipient of the Best Paper Award on MAC and Cross-layer Design in IEEE WCNC 2019. His research interests include wireless powered IoT, deep reinforcement learning, backscatter communications and networking.

Dinh Thai Hoang (M'16) is currently a faculty member at the School of Electrical and Data Engineering, University of Technology Sydney, Australia. He received his Ph.D. in Computer Science and Engineering from the Nanyang Technological University, Singapore, in 2016. His research interests include emerging topics in wireless communications and networking such as ambient backscatter communications, vehicular communications, cybersecurity, IoT, and 5G networks. Currently, he is an editor of IEEE Wireless Communications Letters and IEEE Transactions on Cognitive Communications and Networking.

Dusit Niyato (M'09-SM'15-F'17) is currently a professor in the School of Computer Science and Engineering and, by courtesy, School of Physical & Mathematical Sciences, at the Nanyang Technological University, Singapore. He received B.E. from King Mongkuk's Institute of Technology Ladkrabang (KMITL), Thailand in 1999 and Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada in 2008. He has published more than 380 technical papers in the area of wireless and mobile networking, and is an inventor of four US and German patents. He has authored four books including "Game Theory in Wireless and Communication Networks: Theory, Models, and Applications" with Cambridge University Press. He won the Best Young Researcher Award of IEEE Communications Society (ComSoc) Asia Pacific (AP) and The 2011 IEEE Communications Society Fred W. Ellersick Prize Paper Award. Currently, he is serving as a senior editor of IEEE Wireless Communications Letter, an area editor of IEEE Transactions on Wireless Communications (Radio Management and Multiple Access), an area editor of IEEE Communications Surveys and Tutorials (Network and Service Management and Green Communication), an editor of IEEE Transactions on Communications, an associate editor of IEEE Transactions on Mobile Computing, and IEEE Transactions on Cognitive Communications and Networking. He was a guest editor of IEEE Journal on Selected Areas on Communications. He was a Distinguished Lecturer of the IEEE Communications Society for 2016-2017. He was named the 2017, 2018 highly cited researcher in computer science. He is a Fellow of IEEE.

Ahmed El Shafie is an IEEE senior member. He received the B.Sc. degree (Hons.) in electrical engineering from Alexandria University, Alexandria, Egypt, in 2009, the M.Sc. degree in communication and information technology from Nile University, Cairo, Egypt, in 2014, and the Ph.D. degree from The University of Texas at Dallas, Richardson, TX, USA, in 2018. Since 2018, he is a Senior Systems Engineer at Qualcomm Technologies, San Diego, CA, USA. He was a recipient of the David Daniel Best Doctoral Thesis Award, in 2018, and the Jonsson School Industrial Advisory Council Fellowship Award, in 2017. He received the IEEE Transactions on Communications Exemplary Reviewer in 2015, 2016, and 2017. He also received the IEEE Communications Letters Exemplary Reviewer, in 2016. He is nominated for the 2018 CGS/ProQuest Distinguished Dissertation Award. He currently serves as an Editor for the IEEE Communications Letters, Physical Communications, and Transactions on Emerging Technologies in Telecommunications.

Antonio De Domenico received the M.Sc. degree in telecommunication engineering from the University of Rome La Sapienza in 2008 and the Ph.D. degree in telecommunication engineering from the University of Grenoble in 2012. Since 2009, he has been with the CEA LETICMINATEC, Grenoble, France, as a Research Engineer. His research topics are cloud-enabled heterogeneous wireless networks, millimeter-wave-based communications, machine learning, and green communications. Antonio has more than fifty publications in international conferences, journals, and books chapters, and he is the main inventor or co-inventor of thirteen patents. He was a recipient of the CEA Enhanced Eurotalents Programme in 2017 and thanks to this scholarship, he is a Visiting Researcher with the Communications Group, Department of Electrical and Computer Engineering, in the University of Toronto. Antonio served as TPC co-chair of the IEEE CSCN 2017, publication chair of the IEEE OnlineGreenComm 2015, and he is currently an academic editor of the Hindawi Wireless Communications and Mobile Computing.

Emilio Calvanese Strinati obtained his Engineering Master degree in 2001 from the University of Rome 'La Sapienza' and his Ph.D in Engineering Science in 2005 on Radio link control for improving the QoS of wireless packet transmission. He then started working at Motorola Labs in Paris in 2002. Then in 2006 he joint CEA/LETI as a research engineer. From 2007, he becomes a PhD supervisor. From 2010 to 2012, Dr. Calvanese Strinati has been the co-chair of the wireless working group in GreenTouch Initiative which deals with design of future energy efficient communication networks. From 2011 to 2016 he was the Smart Devices & Telecommunications European collaborative strategic programs Director. Since December 2016 he is the Smart Devices & Telecommunications Scientific and Innovation Director. In December 2013 he has been elected as one of the five representative of academia and research center in the Net!Works 5G PPP ETP. From 2017 to 2018 he was one of the three moderators of the 5G future network expert group. Between 2016 and 2018 he was the coordinator of the H2020 joint Europe and South Korea 5GCHAMPION project. Since July 2018, he is the coordinator of the H2020 joint Europe and South Korea 5GAllStar project. Since 2018 he holds the French Research Director Habilitation (HDR). He has published around 100 papers in international conferences, journals and books chapters, given more than 70 international invited talks and keynotes and is the main inventor or co-inventor of more than 60 patents. He has organized more than 70 international conferences, workshops, panels and special sessions on green communications, heterogeneous networks and cloud computing hosted in international conferences as IEEE GLOBCOM, IEEE PIMRC, IEEE WCNC, IEEE VTC, EuCnC, IFIP, EUCnC and European Wireless.

Jakob Hoydis (S'08-M'12) received the diploma degree (Dipl.-Ing.) in electrical engineering and information technology from RWTH Aachen University, Germany, and the Ph.D. degree from Supéléc, Gif-sur-Yvette, France, in 2008 and 2012, respectively. He is a member of technical staff at Nokia Bell Labs, France, where he is investigating applications of deep learning for the physical layer. Before this position he was co-founder and CTO of the social network SPRAED and worked for Alcatel-Lucent Bell Labs in Stuttgart, Germany. His research interests are in the areas of machine learning, cloud computing, SDR, large random matrix theory, information theory, signal processing, and their applications to wireless communications. He is a co-author of the textbook "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency" (2017). He is recipient of the 2018 Marconi Prize Paper Award, the 2015 Leonard G. Abraham Prize, the IEEE WCNC 2014 best paper award, the 2013 VDE ITG Förderpreis, and the 2012 Publication Prize of the Suplc Foundation. He has received the 2018 Nokia AI Innovation Award and has been nominated as an Exemplary Reviewer 2012 for the IEEE Communication Letters. He is currently chair of the IEEE COMSOC Emerging Technology Initiative on Machine Learning for Communications.