Guest Editorial
Special Issue on Deep Reinforcement Learning for Emerging IoT Systems

NOWADAYS we are witnessing the formation of a massive Internet-of-Things (IoT) ecosystem that integrates a variety of wireless-enabled devices ranging from smartphones, wearables, and virtual reality facilities to sensors, drones, and connected vehicles. As IoT is penetrating every aspect of people's life, work, and entertainment, an increasing number of IoT devices and the emerging IoT applications are driving exponential growth in wireless traffic in the foreseeable future. As a result, current IoT system architectures are facing significant challenges to handle millions of devices; thousands of servers; the transmission and processing of large volume of data, etc.

The growing diversity of IoT services and the complexity of mobile network architectures have made monitoring and managing a multitude of IoT elements extremely difficult. Meanwhile, deep reinforcement learning (DRL) techniques have been seen as a promising solution for building such complex IoT systems and to innovate at a rapid pace. Reinforcement learning (RL) is about agents learn optimal behaviors to maximize a numerical reward signal by interacting with the environment. DRL is a particular type of RL, with deep neural networks used for state representation or function approximation of value function, policy, transition model, and reward function. Over the last few years, DRL has achieved remarkable success in different areas, including video games, robotics, natural language processing (NLP), healthcare, etc. Researchers in IoT areas also began to recognize the power and importance of DRL and have been exploring different DRL techniques to solve problems specific to the emerging IoT systems.

Researchers have explored the inherent power of fusion between DRL technologies and IoT systems in both industrial and academic field. DRL algorithms can provide effective and smart solutions for sequential decision making, optimization and control problems, and dealing with incomplete or inconsistent information related to IoT. To combine DRL technologies with IoT systems, a diversity of research issues remains open and need in-depth investigations. The goal of this special issue is to provide researchers, developers, and practitioners from both academia and industry with the recent advances and key results on DRL for emerging IoT systems.

We received a total of 79 submissions, and after a rigorous reviewing process, 23 papers were accepted to appear in this special issue. The accepted papers cover state-of-the-art research and development in different aspects of learning-based emerging IoT systems and applications, including infrastructures and applications, vehicular networks, energy storage, healthcare, edge computing, testbed and experiment experiences, cybersecurity, smart cities, resource management, and others.

The article titled "Deep-reinforcement-learning-based autonomous UAV navigation with sparse rewards" formulates the problem of autonomous navigation of UAVs in large-scale complex environments as a Markov decision process (MDP) with sparse rewards and proposes an efficient DRL with the nonexpert helpers' algorithm. The proposed algorithm guides the agent in exploring informative states especially in the early training stage and enables the agent to learn to achieve goals in different training stages. The authors construct a simulator for UAV navigation in large-scale complex environments and compare the algorithm with several baselines.

The article titled "Deep actor–critic learning-based robustness enhancement of Internet of Things" considers IoT network topologies that provide robust communication for heterogeneous networks and study the network stability of IoT devices. The authors investigate the network robustness problem both for the network architecture and its resistance to cyberattacks. For the network architecture, they optimize the robustness of IoT network topology with a scale-free network model. For the resistance to cyberattacks, a deep deterministic policy learning algorithm is proposed to improve the network stability.

The article titled "Multiagent deep reinforcement learning for joint multichannel access and task offloading of mobile-edge computing in Industry 4.0" proposes a novel multiagent reinforcement learning algorithm for task offloading and multichannel access in Industry 4.0, which enables edge devices to cooperate with each other by utilizing historical observations and actions. The proposed solution can significantly reduce computation delay and improve the channel access success rate.

In the article titled "Cooperative computation offloading and resource allocation for blockchain-enabled mobile-edge computing: A deep reinforcement learning approach," the authors develop a cooperative computation offloading and resource allocation framework for blockchain-enabled MEC systems. A multiobjective function is designed to maximize the computation rate of MEC systems and the transaction throughput of blockchain systems by jointly optimizing

Digital Object Identifier 10.1109/JIOT.2020.2998256

offloading decision, power allocation, block size, and block interval.

The article titled “Deep-learning-based joint optimization of renewable energy storage and routing in vehicular energy network” studies a vehicular energy network (VEN) with time-varying point-to-point traffic flow and adjustable energy storage capacity at stations. The authors design a dynamic allocation strategy to guarantee maximum benefit from the fixed size of total storage capacity and this strategy is considered in conjunction with the routing of energy-enabled electric vehicles. The joint storage allocation and routing method can be derived through linear programming defined over a time-expanded network topology graph.

The article titled “Deep-reinforcement-learning-based QoS-aware secure routing for SDN-IoT” proposes a DRL-based Quality-of-Service (QoS)-aware secure routing protocol (DQSP) to defend against the internal attacks and DDoS attacks during SDN-IoT routing. The proposed method can extract knowledge from history traffic demands by interacting with the underlying network environment and dynamically optimize the routing policy to guarantee QoS. The DQSP is capable of finding secure routing policies with competitive performance.

In the article titled “Deep-learning-based joint resource scheduling algorithms for hybrid MEC networks,” the authors introduce a hybrid deep-learning-based online offloading (H2O) framework to minimize the sum energy consumption of user equipment in a hybrid mobile-edge computing platform. The framework combines three AI algorithms. The first is a large-scale path-loss fuzzy c-means (LSFCM) algorithm. The second is a fuzzy membership matrix U-based particle swarm optimization (U-PSO) algorithm, and the third is utilized to make the task admission and resource allocation decisions.

The article titled “Deep reinforcement learning for partially observable data poisoning attack in crowdsensing systems” studies the data poisoning problem against the TruthFinder in crowdsensing systems. They modeled the data poisoning under a partially observable environment and designed a DRL-based attack method to optimize the attack strategy. The method leverages experiences from continuous attack attempts to find effective attack strategies, which helps attackers hide themselves and influence the TruthFinder. They conduct experiments on real-life data sets to verify the effectiveness of the proposed method.

A DRL-based unsupervised wireless-localization method is proposed in the article titled “Deep reinforcement learning (DRL): Another perspective for unsupervised wireless localization.” The authors model a continuous wireless localization process as an MDP and process it within a DRL framework. To alleviate the challenge of obtaining rewards when using unlabeled data (e.g., daily-life crowdsourced data), they present a reward-setting mechanism, which extracts robust landmark data from unlabeled wireless received signal strengths (RSSs). To ease requirements for model retraining when using DRL for localization, RSS measurements together with agent location is used to construct DRL inputs.

The article titled “Deep reinforcement learning for economic dispatch of virtual power plant in Internet of Energy,” proposes a DRL algorithm for the optimal economic dispatch in virtual power plants explicitly incorporating the stochastic characteristics of distributed renewable power generation. The authors further utilize an edge computing-based framework such that the optimal dispatch solution can be achieved with a reasonably low computation complexity. The experimental results show the effectiveness of the proposed algorithm.

The article titled “Deep-learning-based SDN model for Internet of Things: An incremental tensor train approach” designs a novel approach for intelligent software-defined networking. It shows that an order-based generalization mechanism can be used to minimize the flow table lookup time and reduce the storage occupancy in SDN. The proposed approach works in four phases: 1) tensor representation; 2) deep Boltzmann machine-based classification; 3) subtensor-based flow matching process; and 4) incremental tensor train network for flow table synchronization. The proposed model has been extensively tested to show its improvements in terms of delay, throughput, storage space, and accuracy.

In the article titled “Energy-efficient mobile crowdsensing by unmanned vehicles: A sequential deep reinforcement learning approach,” the authors consider navigating multiple unmanned vehicles to perform sensing tasks with the presence of multiple charging stations. They propose a novel sequential DRL model called “PPO+LSTM,” containing a sequential module LSTM for modeling history and trained with proximal policy optimization (PPO) for vehicle action decision making. Results show that the proposed solution outperforms others in terms of energy efficiency, data collection ratio, and geographic fairness.

The article titled “Reinforcement-learning-empowered MLaaS scheduling for serving intelligent Internet of Things” proposes a region-based reinforcement learning (RRL) scheduling framework for ML serving in IoT applications that can efficiently identify optimal configurations under dynamic workloads. The authors theoretically show that the RRL approach can achieve fast convergence speed at the cost of performance loss. Therefore, they propose an adaptive RRL algorithm based on Bayesian optimization to balance the convergence speed and the performance. Extensive experimental results show that the proposed approach can outperform state-of-the-art approaches.

A DRL-based storm surge flood simulation approach is proposed in the article titled “Effective IoT-facilitated storm surge flood modeling based on deep reinforcement learning,” to simulate the flood situation under specific levels of the storm surge. By analyzing real-time urban flood data and weather data collected by the IoT system, the model of the urban flood can be constructed and refined. The proposed approach can effectively estimate the degradation of the underground pipelines through exploring the power of the IoT system. The results demonstrate that the proposed method can effectively simulate the storm surge flood with high accuracy.

The article titled “FDC: A secure federated deep learning mechanism for data collaborations in the Internet of Things” presents a secure data collaboration framework based on federated deep learning technology. The proposed framework can realize the secure collaboration of multiparty data computation on the premise that the data does not need to be transmitted out of its private data center. This framework is empowered by public data center, private data center, and the blockchain technology. A real-world IoT scenario is used to verify the effectiveness of the proposed framework.

The article titled "A learning-based incentive mechanism for federated learning" proposes a new incentive mechanism for federated learning to motivate edge nodes to contribute model training. Specifically, a DRL-based (DRL) incentive mechanism has been designed to determine the optimal pricing strategy for the server and the optimal training strategies for edge nodes. Numerical experiments have been implemented to evaluate the efficiency of the proposed DRL-based incentive mechanism for federated learning.

The article titled "Deep reinforcement learning for throughput improvement of the uplink grant-free NOMA system" introduces DRL in the decision making for grant-free nonorthogonal multiple access (NOMA) systems, to mitigate collisions and improve the system throughput in an unknown network environment. Subchannel and device clustering and discrete uplink power control are adopted to reduce collisions and computational complexity of DRL. Then, the long-term cluster throughput maximization problem is formulated as a partially observable MDP and solved by the proposed DRL-based grant-free NOMA algorithm.

In the article titled "Deep-reinforcement-learning-based mode selection and resource allocation for cellular V2X communications," the authors propose a DRL-based transmission mode selection and resource allocation approach for cellular vehicle-to-everything (V2X) communications. It aims to maximize the sum capacity of vehicle-to-infrastructure (V2I) users while guaranteeing the latency and reliability requirements of vehicle-to-vehicle (V2V) pairs. Moreover, a federated DRL algorithm is developed to help obtain robust models. The simulation results show that the proposed algorithms outperform other baselines.

The article titled "Deep reinforcement learning for resource protection and real-time detection in IoT environment" designs a DRL-based detection algorithm for virtual IP watermarks to ensure the security of low-level hardware in intelligent manufacturing of IoT environments. The deep Q-learning (DQN) algorithm is used to generate the watermarked position adaptively, making the watermarked positions secure yet close to the original design. An artificial neural network (ANN) model is used for training the position distance characteristic vectors of the IP circuit, where the characteristic function of the virtual position for IP watermark is generated after training.

The article titled "A deep-reinforcement-learning-based recommender system for occupant-driven energy optimization in commercial buildings" presents recEnergy, a DRL-based recommender system to reduce energy consumption in commercial buildings with human-in-the-loop. The authors define four different types of energy saving recommendations: 1) move; 2) shift schedule; 3) coerce; and 4) reduce personal resources. The recEnergy learns actions with high energy saving potential, actively distribute recommendations to occupants in a commercial building, and utilizes feedback from the occupants to learn better energy saving recommendations.

The article titled "Reinforced spatiotemporal attentive graph neural networks for traffic forecasting" proposes a novel graph-based model (RSTAG) for road traffic prediction, which consists of a graph neural network and two independent attention mechanisms. The proposed method captures dynamic spatial correlations through diffusion network graphs, while temporal dependencies are represented through the sequence-to-sequence model with an attention mechanism. In addition, the authors utilize the policy gradient to update model parameters while largely alleviating the exposure bias issue that exists in previous traffic prediction models.

A deep learning enhanced human activity recognition (HAR) method in the Internet of Healthcare Things (IoHT) environments is proposed in the article titled "Deep-learning-enhanced human activity recognition for Internet of Healthcare Things." To solve the problem of inadequately labeled sample data, an intelligent auto-labeling scheme based on DQN is developed. A multisensor-based data fusion mechanism is then developed to seamlessly integrate the on-body sensor data, context sensor data, and personal profile data. Moreover, an LSTM-based classification method is proposed to identify fine-grained patterns according to the high-level features contextually extracted from sequential motion data.

In the article titled "A deep-reinforcement-learning-based approach to dynamic eMBB/URLLC multiplexing in 5G NR," the authors investigate dynamic multiplexing of enhanced mobile broadband (eMBB) and ultrareliable and low-latency communications (URLLCs) on the same channel in 5G NR via a preemptive puncturing mechanism. The authors present a model-free DRL-based multiplexer (DEMUX) for eMBB and URLLC services, which offers a solution to the URLLC preemption problem without a prior explicit problem formulation. The experimental results show that DEMUX significantly outperforms state-of-the-art algorithms proposed in the 3GPP and the literature.

We express our gratitude to all the authors for their contributions to this special issue, and all the reviewers for their hard work in giving timely and rigorous reviews. We are especially grateful to Prof. X. Shen and Prof. H. Wang, the former and current Editors-in-Chief of the IEEE INTERNET OF THINGS JOURNAL for approving this special issue and for their help along the process of its preparation. We hope that this special issue will serve as a good reference for researchers, engineers, and practitioners in the field of intelligent IoT systems and applications.

JIA HU, *Guest Editor*
Department of Computer Science
University of Exeter
Exeter EX4 4RN, U.K.

PENG LIU, *Guest Editor*
School of Computer Science
Hangzhou Dianzi University
Hangzhou 310018, China

HONG LIU, *Guest Editor*
School of Computer Science
and Software Engineering
East China Normal University
Shanghai 200026, China

OBINNA ANYA, *Guest Editor*
Cloud Artificial Intelligence
and Industry Solutions
Google
Mountain View, CA 94043 USA

YAN ZHANG, *Guest Editor*
Department of Informatics
University of Oslo
0373 Oslo, Norway

Jia Hu received the B.Eng. and M.Eng. degrees in electronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2006 and 2004, respectively, and the Ph.D. degree in computer science from the University of Bradford, Bradford, U.K., in 2010.

He is a Lecturer of computer science with the University of Exeter, Exeter, U.K. His research has been funded by U.K. EPSRC, EU, China NSFC, and industry. He has published over 60 research papers within these areas in prestigious international journals (e.g., the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, the IEEE INTERNET OF THINGS JOURNAL, and the *IEEE Communications Magazine*) and at reputable international conferences (e.g., Globecom and ICC). His research interests include performance evaluation, edge-cloud computing, resource optimization, applied machine learning, and network security.

Dr. Hu has been a recipient of the Best Paper Awards at IEEE SOSE'16 and IUCC'14. He serves on the editorial board of *Computers & Electrical Engineering* (Elsevier) and has guest edited 12 special issues in major international journals, such as the IEEE INTERNET OF THINGS JOURNAL, *Ad Hoc Networks* (Elsevier), and *Computing* (Springer). He has served as the General Co-Chair/Program Co-Chair of many international conferences, such as IEEE ScalCom'19, SmartCity'18, and CIT'15.

Peng Liu received the B.S. and M.S. degrees in computer science and technology from Hangzhou Dianzi University, Hangzhou, China, in 2001 and 2004, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, in 2007.

He was a Visiting Scholar/Research Fellow with the University of Bradford, Bradford, U.K., in 2008; Temple University, Philadelphia, PA, USA, in 2014; the University of Waterloo, Waterloo, ON, Canada, in 2017; and Singapore University of Technology and Design, Singapore, in 2018. He is currently an Associate Professor with Hangzhou Dianzi University. He has published over 30 research papers and severed as a committee member of some leading conferences. His research interests include wireless sensor networks, delay-tolerant networks, and mobile computing.

Hong Liu received the Ph.D. degree from Beihang University, Beijing, China.

She is an Associate Professor with East China Normal University, Shanghai, China. She also works as a CTO with Shanghai Trusted Industrial Control Platform Company Ltd., Shanghai. Her research interests include the security and privacy issues in vehicular-edge computing, and industrial Internet of Things.

Dr. Liu has published 37 SCI papers, and her Google Scholar citations are 2000 times, including one IEEE Best Paper Award, one IEEE Popular Article, and two ESI "Highly Cited Papers," and serves as the program committee (PC) member and workshop chair in several international conferences. She is an Associate Editor of the IEEE INTERNET OF THINGS JOURNAL and *Security and Privacy* (Wiley). She serves as the Workshop Chair of the 12th IEEE International Conference on Ubiquitous Intelligence and Computing; and a PC Member of the IEEE International Conference on Communication, the IEEE Conference on Communications and Network Security, and the International Conference on Ambient Systems, Networks and Technologies.

Obinna Anya received the B.Sc. degree in computer science from the University of Nigeria, Nsukka, Nigeria, and the M.Sc. degree in distributed systems and the Ph.D. degree in computer science from the University of Liverpool, Liverpool, U.K.

He is a UX Research Scientist with Google, Mountain View, CA, USA. He was a Research Scientist with Liverpool Hope University, Liverpool, and a Research Staff Member with IBM Research Center, San Jose, CA, USA. His work has been published in some high-quality peer-reviewed journals and academic conferences. His research interests include human–computer interaction, big data analytics, enterprise collaborative systems, and usable machine learning.

Dr. Anya serves on the editorial boards of the *Journal of Future Generation Computer Systems* and the *Journal of Reliable Intelligent Environments*, and on the program committees of leading international conferences.

Yan Zhang (Fellow, IEEE) received the Ph.D. degree from the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore.

He is currently a Full Professor with the Department of Informatics, University of Oslo, Oslo, Norway. His research interests include next-generation wireless networks leading to 5G beyond/6G, and green and secure cyber–physical systems (e.g., smart grid and transport).

Prof. Zhang was a recipient of the global Highly Cited Researcher Award (Web of Science top 1% most cited worldwide) from 2018 to 2019. He is an Editor for several IEEE publications, including the *IEEE Communications Magazine*, the *IEEE Network Magazine*, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE COMMUNICATIONS SURVEY & TUTORIALS, the IEEE INTERNET OF THINGS JOURNAL, the IEEE SYSTEMS JOURNAL, the *IEEE Vehicular Technology Magazine*, and IEEE Blockchain Technical Briefs. He is a Symposium/Track Chair in a number of conferences, including IEEE ICC 2021, IEEE Globecom 2017, IEEE PIMRC 2016, and IEEE SmartGridComm 2015. He is an IEEE Vehicular Technology Society Distinguished Lecturer from 2016 to 2020, and he is named as a CCF Distinguished Speaker in 2019. He is the Chair of the IEEE Communications Society Technical Committee on Green Communications and Computing. He is an Elected Member of CCF Technical Committee of Blockchain.