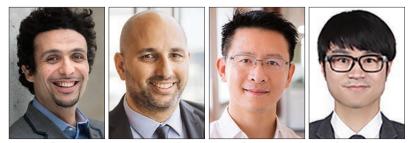# COMMUNICATION TECHNOLOGIES FOR EFFICIENT EDGE LEARNING

Mehdi Bennis   Mérouane Debbah   Kaibin Huang   Zhaohui Yang

Traditional machine learning is centralized in the cloud (data centers). Recently, the security concern and the availability of abundant data and computation resources in wireless networks are pushing the deployment of learning algorithms toward the network edge. This has led to the emergence of a fast growing area, called edge learning, which integrates two originally decoupled areas: wireless communication and machine learning. It is widely expected that the advancements in edge learning would provide a platform for implementing edge artificial intelligence (AI) in 5G-and-Beyond systems and solving large-scale problems in our society ranging from autonomous driving to personalized healthcare. A typical edge learning framework (e.g., federated learning) features distributed learning over many wireless devices as coordinated by edge servers to cooperatively train a large-scale AI model using local data and CPUs/GPUs. The iterative learning process involves repeated downloading and uploading of high-dimensional (millions to billions) model parameters or their updates by tens to hundreds of devices. This will generate enormous data traffic, placing a heavy burden on the already congested radio access networks. The training problem cannot be efficiently solved using traditional wireless techniques targeting rate maximization and decoupled from learning. Achieving the goal of edge learning with high communication efficiencies calls for the design of new wireless techniques based on a communication-and-learning integration approach.

This Feature Topic of *IEEE Communications Magazine* introduces to the communication community the latest advancements in edge learning and points readers to many promising research opportunities.

In particular, the first article, "Communicate to Learn at the Edge", co-authored by D. Gunduz *et al.*, introduces the new paradigm of joint communication, learning and inference at the edge. Various designs based on the interplay of ideas from coding and communication are discussed. Moreover, this article presents the challenges faced in achieving fully distributed edge intelligence across heterogeneous agents communicating over imperfect wireless channels.

To deploy edge AI in wireless networks, devices are required to transmit their multimedia data or local training results over unreliable wireless links. This exposes the performance of learning and inference to degradation caused by limited radio resources (e.g., power, time and bandwidth). This makes it important to jointly manage communication and computation resources for efficient and robust edge AI. Addressing this issue, the second article, "Communication-Computation Tradeoff in Resource-Constrained Edge Inference" by J. Shao and J. Zhang, introduces the existence of a fundamental tradeoff between the computation and communication cost of edge devices in a device-edge cooperative inference system. The investigation of such a tradeoff in different aspects of cooperative inference gives rise to a spectrum of research opportunities as overviewed in the article. The third article, "Edge Learning with Timeliness Constraints: Challenges and Solutions" by Y. Sun *et al.*, establishes the concept of timely edge learning, aiming at minimizing the communication and computation delay under a guarantee of training/inference accuracy. The authors discuss key challenges and propose different solution approaches from the perspectives of data, model and resource management. The fourth article, "Toward Self-learning Edge Intelligence in 6G" by Y. Xiao *et al.*, concerns automatic data learning and synthesis and proposes a promising self-learning architecture based on self-supervised Generative Adversarial Net (GAN) to support such operations in 6G networks.

Fog networking is emerging as an end-to-end architecture that aims to distribute computing, storage, control, and networking functions along the cloud-to-things continuum of nodes spreading from data centers to end users. The fifth article, "From Federated to Fog Learning: Distributed Machine Learning over Heterogeneous Wireless Networks" by S. Hosseinalipour *et al.*, presents a new learning paradigm called fog learning that features a multi-layer architecture allowing intelligent distribution of AI-model training in fog networks.

The last article, "Wireless Communications for Collaborative Federated Learning" by M. Chen *et al.*, proposes a novel federated learning framework called cooperative federated learning (CFL). While traditional federated learning requires an edge server to aggregate local models trained at devices, the CFL framework enables devices to perform federated learning even in the absence of coordination by the server. Furthermore, the authors demonstrate how to use wireless techniques to improve the performance of CFL.

We would like to thank our dedicated reviewers for their efforts on reviewing the submitted manuscripts. We also thank the Editor-in-Chief, Dr. Tarek El-Bawab for his support and guidance in the editorial process.

BIOGRAPHIES

MEHDI BENNIS is currently an associate professor with the Centre for Wireless Communications, University of Oulu, Oulu, Finland, where he is also an Academy of Finland Research Fellow and the Head of the Intelligent Connectivity and Networks/Systems Group (ICON). His current research interests include radio resource management, heterogeneous networks, game theory, and machine learning in 5G networks and beyond. He was a recipient of several prestigious awards, including the 2015 Fred W. Ellersick Prize from the IEEE Communications Society, the 2016 Best Tutorial Prize from the IEEE Communications Society, the 2017 EURASIP Best Paper Award for the *Journal on Wireless Communications and Networking*, the All-University of Oulu Award for research, and the 2019 IEEE ComSoc Radio Communications Committee Early Achievement Award. He is an editor for *IEEE Transactions on Communications* and an IEEE Fellow

MÉROUANE DEBBAH is Vice-President of the Huawei France Research Center and jointly the director of the Mathematical and Algorithmic Sciences Lab as well as the director of the Lagrange Mathematical and Computing Research Center. He is an IEEE Fellow, a WWRF Fellow, and a Membre émérite SEE. He was a recipient of the ERC Grant MORE (Advanced Mathematical Tools for Complex Network Engineering) from 2012 to 2017. He was a recipient of the Mario Boella Award in 2005, the IEEE Glavieux Prize Award in 2011, the Qualcomm Innovation Prize Award in 2012 and the 2019 IEEE Radio Communications Committee Technical

Recognition Award. He has received 20 best paper awards, including the 2007 IEEE GLOBECOM Best Paper Award and the 2019 IEEE Communications Society Young Author Best Paper Award.

KAIBIN HUANG (lead guest editor) is an associate professor with the Department of EEE, The University of Hong Kong. His research interests include mobile edge computing, distributed learning, and 5G systems. His research interests include edge computing and learning and wireless communication techniques. Currently, he is an associate editor for *IEEE Transactions on Wireless Communications* and *IEEE Transactions on Green Communications and Networking*. He received the IEEE Communication Society's 2019 Best Tutorial Paper Award and two Asia Pacific Outstanding Paper Awards in 2015 and 2019, Outstanding Teaching Award from Yonsei, Motorola Partnerships in Research Grant, the University Continuing Fellowship from UT Austin, and Best Paper Awards from IEEE GLOBECOM 2006 and IEEE/CIC ICCC in 2018. He was named a Highly Cited Researcher by Clarivate Analytics in 2019. He is an IEEE Fellow.

ZHAOHUI YANG received the B.S. degree in information science and engineering from Chien-Shiung Wu Honors College, Southeast University, Nanjing, China, in June 2014, and the Ph.D. degree in communication and information systems with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in May 2018. He is currently a postdoctoral research associate with the Center for Telecommunications Research, Department of Informatics, King's College London, U.K., where he has been involved in the EPSRC SENSE Project. His research interests include federated learning, reconfigurable intelligent surface, UAV, MEC, energy harvesting, and NOMA. He was a TPC member of IEEE ICC during 2015–2020 and Globecom during 2017-2020. He was an exemplary reviewer for *IEEE Transactions on Communications* in 2019.