

ARTIFICIAL INTELLIGENCE ENABLED WIRELESS NETWORKING FOR 5G AND BEYOND: RECENT ADVANCES AND FUTURE CHALLENGES

Cheng-Xiang Wang, Marco Di Renzo, Sławomir Stańczak, Sen Wang, and Erik G. Larsson

ABSTRACT

5G wireless communication networks are currently being deployed, and B5G networks are expected to be developed over the next decade. AI technologies and, in particular, ML have the potential to efficiently solve the unstructured and seemingly intractable problems by involving large amounts of data that need to be dealt with in B5G. This article studies how AI and ML can be leveraged for the design and operation of B5G networks. We first provide a comprehensive survey of recent advances and future challenges that result from bringing AI/ML technologies into B5G wireless networks. Our survey touches on different aspects of wireless network design and optimization, including channel measurements, modeling, and estimation, physical layer research, and network management and optimization. Then ML algorithms and applications to B5G networks are reviewed, followed by an overview of standard developments of applying AI/ML algorithms to B5G networks. We conclude this study with future challenges on applying AI/ML to B5G networks.

INTRODUCTION

Global demand for wireless communication networks continues to increase, mainly due to the ever-growing numbers of wireless users and new emerging wireless services. The fifth generation (5G) and beyond 5G (B5G) wireless networks are expected to be developed in the future and offer higher data rates, improved coverage, better cost efficiency, resource utilization, security, adaptability, and scalability [1]. Artificial intelligence (AI) technologies have the potential to efficiently solve unstructured and seemingly intractable problems involving large amounts of data that need to be dealt with in the design and optimization of 5G and B5G wireless networks.

AI is “the simulation of human intelligence processes by machines, especially computer systems” [2]. It is usually defined as the science of making computers perform tasks that require intelligence like humans. Whereas AI is a broader concept of machines being able to carry out tasks smartly, machine learning (ML) is a current (probably

the most popular) application of AI that enables machines to learn from large amounts of data and act accordingly without being explicitly programmed. As a special type of ML, deep learning studies artificial neural networks (ANNs) that contain more than one hidden layer to “simulate” the human brain. Currently, deep learning is one of the most widespread ML methods as it has successfully been applied to different fields such as computer vision, speech recognition, and bioinformatics.

AI technologies will not only reduce or even replace manual efforts for network development, configuration, and management, but also deliver better system performance, reliability, and adaptability of communication networks by making real-time robust decisions based on predictions of the networks’ and users’ behavior. ML, as a typical AI technology, is widely expected to rapidly become a key component of B5G communication networks. It will make full use of big data to overcome the challenges of designing and operating B5G networks. Potential benefits of introducing ML into communication systems include the following. First, channel and interference models are extremely complicated in reality due to the dynamic nature of wireless communication channels, especially in B5G scenarios. ML techniques may automatically extract the unknown channel information by learning from the communication data and prior knowledge. Second, as the density of wireless access points continues to increase, there is an urgent need for global optimization of communication resources and fine tuning of system settings. However, the enormous amount of resources, system parameters to be optimized, and their coupled correlations render these tasks notoriously difficult to solve using existing approaches. In contrast, sophisticated ML algorithms (e.g., deep learning and probabilistic learning methods) may be able to model the highly nonlinear correlations and estimate (sub-)optimal system parameters. Last, ML will realize learning-based adaptive configuration of networks by finding out behavioral patterns and responding quickly and flexibly to various scenarios, for example, anticipating traffic and planning ahead rather than simply reacting to unexpected events.

Current cellular networks designed and operated based on previous postulates may systematically fail to enable future communication services, since

Cheng-Xiang Wang (corresponding author) is with Southeast University and also with the Purple Mountain Laboratories; Marco Di Renzo is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes; Sławomir Stańczak is with Heinrich Hertz Institute, and also with the Technical University of Berlin; Sen Wang is with Heriot-Watt University; Erik G. Larsson is with Linköping University.

Digital Object Identifier:
10.1109/MWC.001.1900292

they cannot keep pace with the data explosion and the underlying complexity of the generated data while guaranteeing the required capacity, reliability, and adaptability. Thus, the network cannot quickly react to and anticipate events that might deteriorate communication services in real time. However, as most AI algorithms and applications are not specifically designed for wireless communication networks, it is hard to directly apply existing AI algorithms to B5G networks.

Compared to earlier survey papers such as [2], this article aims to explore the advantages of combining AI technologies with B5G wireless networks, leveraging the potential of AI technologies to tackle challenges that cannot be efficiently addressed using conventional communication technologies. This article focuses on five aspects that bring AI technologies into B5G wireless networks, as shown in Fig. 1. Accordingly, the remainder of the article is organized as follows. In the following section, we discuss channel measurements, modeling, and estimation for B5G networks using AI technologies. We then study physical-layer research for B5G networks using AI technologies. Following that, we provide a survey of network management and optimization for B5G networks using AI technologies. AI algorithms and applications to B5G networks are then given. We then provide an overview of standard developments of AI technologies and ML algorithms for B5G networks. Conclusions are drawn and future challenges are discussed in the final section.

CHANNEL MEASUREMENTS, MODELING, AND ESTIMATION FOR B5G NETWORKS USING AI TECHNOLOGIES

CHANNEL MEASUREMENT DATA PROCESSING AND CHANNEL MODELING

For B5G wireless communication systems, the diversity of frequency bands, including sub-6 GHz, millimeter-wave (mmWave), terahertz (THz), and optical bands, has made channel modeling more complex. To address the B5G channel modeling requirements, the existing channel models are extended with much higher computational complexity. When modeling for new scenarios, channel measurements must be conducted to understand new channel characteristics, which is a time-consuming task. Apart from a recent work [3], there are very few investigations that have studied the benefits of applying AI to channel modeling, and most existing works only employ very simple AI techniques on a very limited part of the channel modeling process. There is no work that comprehensively investigates the application of AI technologies to channel measurements.

Due to the high complexity of modeling the signal propagation in diverse scenarios, conventional methods make many assumptions and approximations to simplify the processing and modeling methods. Wireless channel features can be extracted from the huge amount of existing measurement data, and, at the same time, the channel modeling problem can be tackled in a data-driven manner, seamlessly integrating with model-based methods. A good balance of the accuracy-complexity trade-off of both processing and modeling techniques will be maintained.

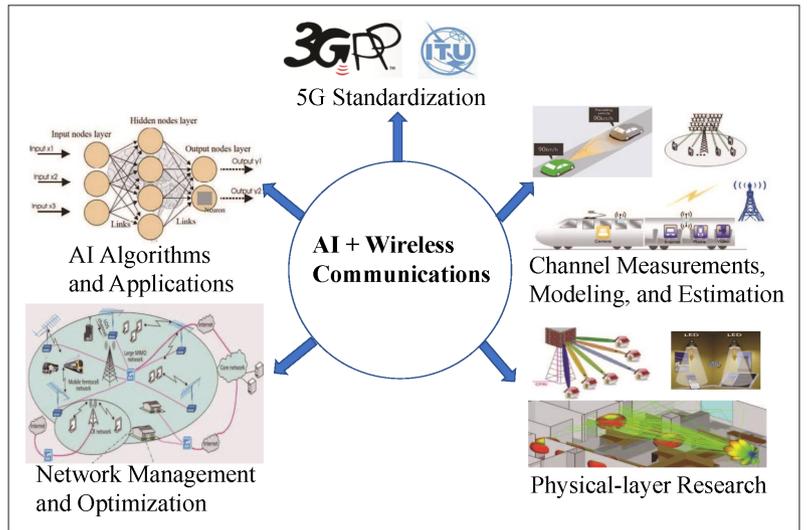


FIGURE 1. Research aspects that bring AI technologies into B5G wireless networks.

ML can be utilized to channel features prediction, channel impulse response (CIR) modeling, multipath component (MPC) clustering, channel parameter estimation, and scenario classification, based on channel measurement data and environment information. The authors of [4] proposed a big-data-enabled channel model based on both a feed-forward neural network (FNN) and a radial basis function neural network (RBF-NN). It can predict channel statistical properties including the received power, root mean square (RMS) delay spread (DS), and RMS angle spreads (ASs) with input parameters of transmitter (Tx) and receiver (Rx) coordinates, Tx-Rx distance, and carrier frequency. The performance of FNN and RBF-NN were fully compared based on both real channel measurement data and synthetic data. An example of the measured and predicted path loss and RMS DS is shown in Fig. 2. Both FNN and RBF-NN show good potential for channel modeling. In [5], the ANN was applied to remove the noise from measured CIR, and the principal component analysis (PCA) was utilized to exploit the features and structures of the channel and model the CIR. In [6], several clustering algorithms were investigated for MPC clustering and tracking, including K-means, fuzzy C-means (FCM), and density-based spatial clustering of applications with noise (DBSCAN). In [7], a convolutional neural network (CNN) was used to automatically identify different wireless channels and help decide which relevant wireless channel features should be used. The MPC parameters like amplitude, delay, and Doppler frequency were extracted and used as input parameters in the CNN, and the output of the CNN was the class of the wireless channels.

CHANNEL ESTIMATION ASSOCIATED WITH ML

In wireless communications, the channel state information (CSI) can be acquired through blind and pilot-based channel estimation techniques. However, blind channel estimation extracts statistical properties by using abundant received symbols. For the pilot-based technique, with the deployment of 5G key technologies, pilot overhead, nonlinear channel, high-mobility channel,

LARGE-SCALE SENSING VIA MASSIVE RADIO INTERFACES

The use of large antenna arrays offers not only unprecedented performance in terms of reliable and high-rate communications (as exploited in massive MIMO), but also provides enormous amounts of baseband-level data that can be used to make inferences about the environment. More novel emerging use cases include inference problems, for example, detection of the presence of moving objects, estimation of the amount of traffic on a road, counting of the number of persons in a room, and guarding against intrusion in protected spaces. Particular technical challenges that lie within reach are the sensing of open spaces, indoor venues, and even through walls. There are emerging commercial use cases and also many applications in security, surveillance, and monitoring.

ML algorithms are particularly suitable to analyze the vast amounts of data generated by large antenna arrays, especially massive MIMO arrays, as typically parametric models are unavailable or inaccurate; hence, classical estimation/detection algorithms are inapplicable. More specifically, in terms of algorithmic approaches, deep learning networks and methodology from image processing and video analytics may offer the most promising path. It is important to note the distinction from conventional radar imaging, where the objective is to create an image or map of the environment, whereas the goal of emerging large-scale sensing is to extract specific features of the dynamics of the environment and make inferences about specific phenomena [11].

Important future research directions should include both pertinent physical modeling work and the construction of an algorithmic foundation that exploits relevant ML tools. Trained deep neural networks represent an important technology component in this regard, but also various forms of dictionary learning might be used. Simulated channel models should be used for evaluation, along with experimentally obtained real data. Through the use of these techniques, research along this direction could significantly advance the state of the art in sensing of open spaces, indoor venues, and through the wall, and accomplish inference tasks that are impossible with conventional model-based signal processing. Another application that may benefit from the technology is gesture recognition, especially when implementing sensing at higher frequencies.

SIGNAL PROCESSING

Massive MIMO technologies have been adopted in 5G communication systems. It is one of the obvious use cases where AI can be deployed. Although massive MIMO has many advantages, such as spectrum efficiency, energy efficiency, security, and robustness, it can produce a large volume of data. For example, in channel measurements, a massive MIMO system with 32×56 antennas and 100 MHz bandwidth can produce data larger than 32 GB. Both detection and channel estimation for massive MIMO systems are usually time-consuming processes and require great

Massive MIMO technologies have been adopted in 5G communication systems. It is one of the obvious use cases where AI can be deployed. Although massive MIMO has many advantages, such as spectrum efficiency, energy efficiency, security, and robustness, it can produce a large volume of data.

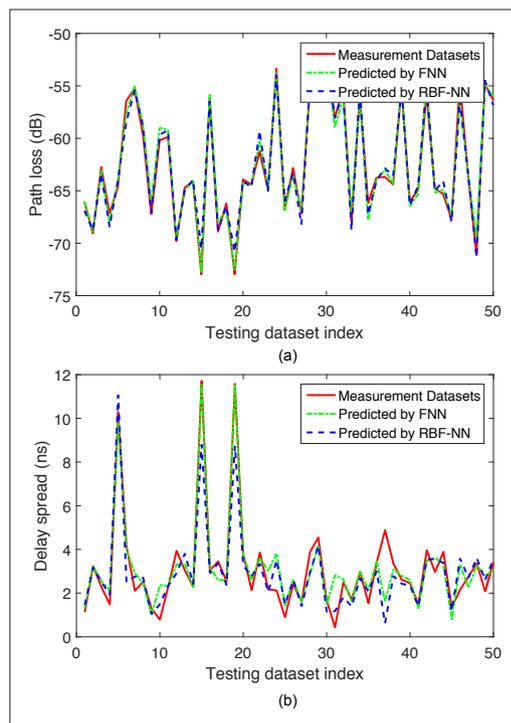


FIGURE 2. a) Measured and predicted path loss; b) measured and predicted RMS DS.

and so on are challenges to be conquered in channel estimation. For example, the pilot overhead can be intolerable for massive multiple-input multiple-output (MIMO) channels and ultra-dense networks (UDNs). Thus, the trade-off between pilot length and channel estimation accuracy should be considered. The nonlinear characteristics of visible light communication (VLC) channel and mmWave channel make it hard to get accurate CSI. In addition, with the development of high-speed railways (HSRs), accurate channel estimation is important to guarantee the quality of service (QoS) and efficient information transmission.

Researchers resort to ML techniques to solve the above mentioned problems. To address the channel estimation in a fast fading time-varying multipath channel, a 2D nonlinear complex support vector regression (SVR) based on an RBF kernel was proposed to achieve accurate channel estimation [8]. In [9], a deep-learning-based channel estimation algorithm was proposed for beam-space mmWave massive MIMO systems. It can learn channel structure and estimate channel from a large number of training data. In [10], an off-grid sparse Bayesian learning-based channel estimation algorithm was proposed for mmWave massive MIMO uplink. It can identify the angles and gains of the scatterer paths by exploiting spatial sparse structure in mmWave channels.

For this topic, one future direction is the generalized ML-based channel estimation scheme, which can be directly used in different scenarios without further training. In order to build this generalized scheme, a vast amount of pre-collected communication data have to be used by machine/deep learning algorithms to learn the channel feature of different environments.

computational power. The big data property of massive MIMO systems makes researchers think of ML methods. In [12], the large amount of data generated from a massive MIMO system is represented by large random matrices and analyzed using the single ring law. One of the challenges for massive MIMO systems is pilot contamination, which can have a significant impact on the performance of massive MIMO systems. The pilot contamination stems from the pilot interference between adjacent cells and can limit the ability of systems to obtain accurate CSI. As the number of antennas increases, channels in beamspace are approximately sparse, that is, most of the MPC power results from a few paths gathered into clusters in the space, and the channel matrix contains a small number of nonzero elements [13]. Based on the sparsity property of channels in beamspace, the authors of [14] obtain the CSI of massive MIMO systems using the sparse Bayesian learning method. Compared to conventional CSI estimators, the Bayesian learning method can achieve better performance in terms of pilot contamination. The sparse recovery problem is an important research issue for Bayesian compressive sensing. It aims to estimate a non-negative compressible vector from a set of noiseless measurements.

DATA-DRIVEN LOCALIZATION IN WIRELESS NETWORKS

Accurate positioning is valuable for context awareness and location-based network management and services. Most current wireless positioning techniques use channel information and fingerprinting to estimate locations. In reality, the channel information needs to be frequently updated to reflect the true channel characteristics since they are susceptible to a variety of dynamic and time-varying transmission impediments (e.g., path loss, interference, and blockage). This periodic and long-term maintenance is time-consuming and labor-intensive, especially for large-scale B5G systems. As the amount and diversity of sensing and communication data dramatically increase in the B5G systems, data-driven localization is a promising solution, that is, positioning devices and users through learning from raw sensing and communication data with ML algorithms. The data-driven localization algorithms will not only be self-adaptive to real-time dynamic transmission impediments, but also evolve over time by consistently learning from data. The wireless channel locations can be continuously updated and improved by automatically learning from the crowd-sourced big data from a vast number of mobile devices. Benefiting from the accurate localization results, users will enjoy better location-based services in return.

NETWORK MANAGEMENT AND OPTIMIZATION FOR B5G NETWORKS USING AI TECHNOLOGIES

FROM MODEL-BASED TO DATA-DRIVEN OPTIMIZATION OF UDNS

The current approach to the management and optimization of cellular networks is based on “models.” This approach is used across all the network functionalities, from the design of the physical layer to the deployment of network infrastructure. However, it is insufficient for the

design of future networks, which will be based on multiple and diverse radio access technologies, ultra-densely deployed, and have to serve a broad class of applications and requirements. Such a complicated network ecosystem cannot be optimally designed and orchestrated based on “models” that reproduce actual network deployments only in part and are not accurate in practice. For example, let us consider a typical on-demand network deployment that relies on the use of unmanned aerial vehicles (UAVs) for rescue operation, disaster recovery, and so on. Such a network needs to be deployed in an ad hoc manner and cannot rely on models that do not even exist for the specific case of interest. Based on these premises, there is a compelling need for radically changing the way future networks will be engineered and optimized. The complexity of future networks and the broad set of requirements that they need to fulfill necessitate that they go beyond the concept of models for network design and exploit the large availability of “data.” In this context, a paradigm shift for the efficient design of B5G networks is necessary to leverage AI and ML in order to take advantage of big data analytics to enhance the situational awareness and overall network operation of future networks. AI can parse through massive amounts of data generated from multiple sources such as wireless channel measurements, sensor readings, and drones and surveillance images to create a comprehensive operational map of the massive number of devices within the network. It can be exploited to optimize various functions, such as fault monitoring and user tracking, across the wireless network. Resource management mechanisms based on AI will be able to operate in a fully online manner by learning the states of the wireless environment and the network’s users in real time. Such mechanisms will be able to continuously improve their own performance over time, which in turn will enable more intelligent and dynamic network decision making.

In order to substantiate the potential of using ANNs for the design of communication networks, we consider the optimization of a typical threshold-based demodulator for Poisson channels, which find applications in optical and molecular communication networks that are known to be difficult to model in the presence of inter-symbol interference (ISI). We compare the typical approach employed in communications, where the system/channel model is assumed to be perfectly known, and the optimal demodulation threshold is obtained by minimizing the analytical expression of the error probability that accounts for the ISI, against a data-driven approach, where nothing is known about the channel model and ANNs are used in order to learn the best demodulator, and thus the optimal demodulation threshold without prior information on the system model. As far as the data-driven approach is concerned, the ANN is trained by using supervised learning. In particular, we use the Bayesian regularization back propagation method, which updates the weights and biases of the ANN by using the Levenberg-Marquardt optimization algorithm. Figure 3 shows the optimal demodulation threshold for two values of ISI (small and large values) in different signal-to-noise ratios (SNRs). In both cases, we note that a data-driven

Resource management mechanisms based on AI will be able to operate in a fully online manner by learning the states of the wireless environment and the network’s users in real time. Such mechanisms will be able to continuously improve their own performance over time which in turn will enable more intelligent and dynamic network decision making.

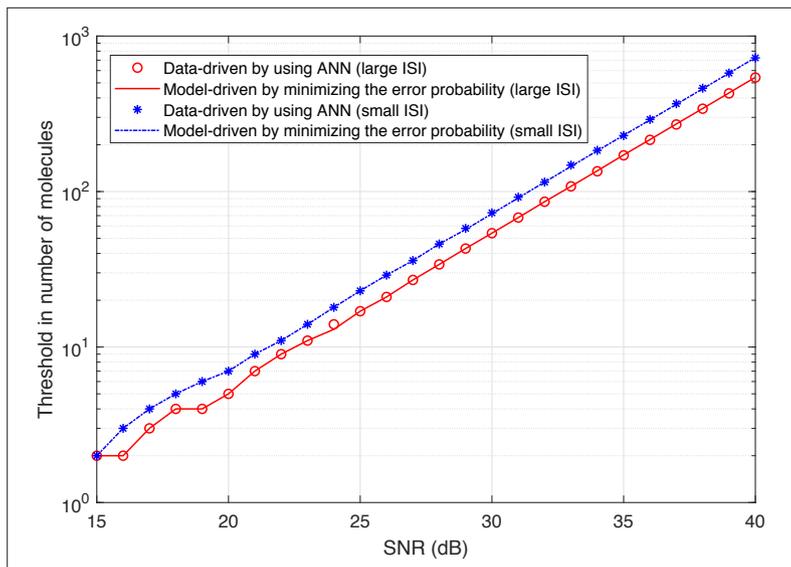


FIGURE 3. Optimal demodulation threshold for small and large values of ISI.

approach provides the same demodulation threshold and consequently the same error probability, by dispensing the system designer to perfectly know the system/channel model.

PROACTIVE WIRELESS NETWORKING FOR ONLINE SOFTWARE NETWORK ORCHESTRATION

Recently, the industry has witnessed the increasing maturity of software-defined networking (SDN) and network function virtualization (NFV), which constitute fundamental enabling technologies to realize the 5G Public-Private Project (5G PPP) vision of software/programmable networking. With the evolution of SDN and NFV, 5G cellular networks have advocated a revolutionary concept called network slicing (NS). Instead of building dedicated networks for different services, NS allows operators to intelligently create customized network pipes to provide optimized solutions for different services that require diverse functionalities, performance metrics, and isolation criteria. Specifically, mobile edge computing, edge caching, and so on have potentially evolved to replace its forwarding-only functionality to an area equipped with storage, memories, and computational power capabilities. Enabling future cellular architectures with NS is a fundamental necessity for optimal network orchestration and for offering services with such diverse requirements, notably enhanced mobile broadband (eMBB), which needs bandwidth-consuming and throughput-driving to new services such as ultra-reliable and low-latency communications (uRLLC) and massive machine-type communications (mMTC). Today's networks and even 5G networks are conceived, designed, and optimized based on the reaction principle, which passively responds to incoming demands and serves them when requested. This principle is not adequate for the new service capabilities that future networks need to provide. The future networks will be heterogeneous software-defined networks. Even different services are logically independently operated; their data traffic will finally be mixed, which can result in a highly dynamic and unmanageable manner to the reaction principle. Besides, for future net-

works, the requirements of some services cannot be fulfilled by the current reaction principle. For example, uRLLC applications may not accept the delay associated with this reaction principle. On the other hand, future networks need prediction capabilities, which enable them to anticipate the future and proactively allocate network resources. In a proactive approach, rather than passively responding to incoming demands and serving them when requested, network architectures with NS can predict traffic patterns and determine future off-peak times on different spectrum bands so that incoming traffic demands can be properly allocated over a given time window. Predicting users' behaviors will result in better utilization of network resources and will allow us to optimally allocate end-to-end network slices in an online fashion. This paradigm shift from reactive to proactive network design can only be made possible with the aid of AI and ML techniques.

AI ALGORITHMS AND APPLICATIONS FOR 5G NETWORKS

DISTRIBUTED ML ALGORITHMS FOR 5G NETWORKS

In current communications applications, signal processing and ML algorithms are typically executed centrally. An archetype architecture is the cloud radio access network (C-RAN), where joint estimation and data processing for all network devices is performed at a central unit (e.g., the cloud). In the presence of a large number of devices and communication limitations on the fronthaul/backhaul links, however, various network functions should be executed locally or with minimal information exchange with the cloud. Therefore, of central importance is the provisioning of a decentralized functional architecture, which adapts dynamically to network requirements [15]. As shown in Fig. 4, a lightweight deep learning model can be applied to cloud, fog, and edge computing networks. The cloud network is the data and computing center, the fog network includes many nodes, and the edge network contains enormous numbers of end users and devices. In parallel, there is the need for decentralized learning, classification, and signal processing algorithms, which seamlessly adapt to the number and the type of the information sources, considering the available communication bandwidth. In the presence of a dynamic edge computing architecture, the advantages of decentralized and centralized algorithms should be combined, thereby trading off complexity, latency, and reliability. This requires integration and further development of methods for data fusion, compression, and distributed decision making.

In the distributed setting, there is also the need to develop solutions that are capable of learning the relationships between the network entities and their time evolution. Since dynamic network inference is a complex task in general, scalable solutions are required. Therefore, it is necessary to evaluate the potential of online learning methods, including kernel-based adaptive filters, high-dimensional set-theoretic algorithms, and other robust statistical estimation methods. Additional examples include Bayesian approaches in conjunction with approximate inference methods, such as approximate message passing and generalizations therein.

ML ALGORITHMS FOR ULTRA-FAST TRAINING AND INFERENCE

ML algorithms are mostly designed for systems and applications that do not need to achieve high-frequency performance. Unfortunately, this is not the case in the context of B5G networks, which must ensure high data processing rates for ultra-low latency. This imposes a strict requirement on the speeds of training and, particularly, inference of ML models. Therefore, one big challenge is to develop ML algorithms with ultra-fast training and inference capability for future wireless communications.

There are two potential directions to accelerate ML training speed. One is implementing ML algorithms in hardware, which should result in low power consumption and high efficiency. The other option is to reduce the complexity of ML algorithms while keeping a reasonable accuracy.

LIGHTWEIGHT ML ALGORITHMS FOR UNIVERSAL EMBEDDED SYSTEMS

Existing ML algorithms mainly focus on computer vision, natural language processing, and robotics with powerful graphics processing unit (GPU) or central processing unit (CPU) enabled computing to operate in real time. However, communication systems are full of resource-constrained devices, for example, embedded and Internet of Things (IoT) systems. Therefore, the ML algorithms for communications should not only learn complex statistical models that underlie networks, consumers, and devices, but also effectively work with embedded devices having limited storage capabilities, computational power, and energy resources. It is challenging but highly rewarding to develop lightweight ML algorithms, especially deep learning models, for embedded systems.

In this aspect, one potential direction will be the combination of ML and distributed computing frameworks such as fog computing and edge computing. Another important research direction is the investigation of a high-level ML development library and toolbox.

AI/ML FOR B5G NETWORKS IN STANDARDS AND STUDY GROUPS

Although the convergence of AI/ML and communication networks is rapidly progressing, it is still in an early stage. As the various sensors, devices, applications, and systems connected in B5G networks will produce a variety of formats and sizes of data to be transmitted, it is extremely complex to standardize ML algorithms for B5G networks. No standard or baseline ML algorithm has been established, and it is unclear for the whole communication community which types of ML algorithms suit the B5G systems best.

Recently, there have been some preliminary works on applying AI/ML to B5G networks in standards organizations including the International Telecommunication Union (ITU) and Third Generation Partnership Project (3GPP), as well as other study groups such as FuTURE, telecom infra project (TIP), and 5G PPP, as shown in Table 1. ITU started a focus group on "Machine learning for future networks including 5G (ML5G)" (<https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx>, accessed 26 June 2018) by ITU-T

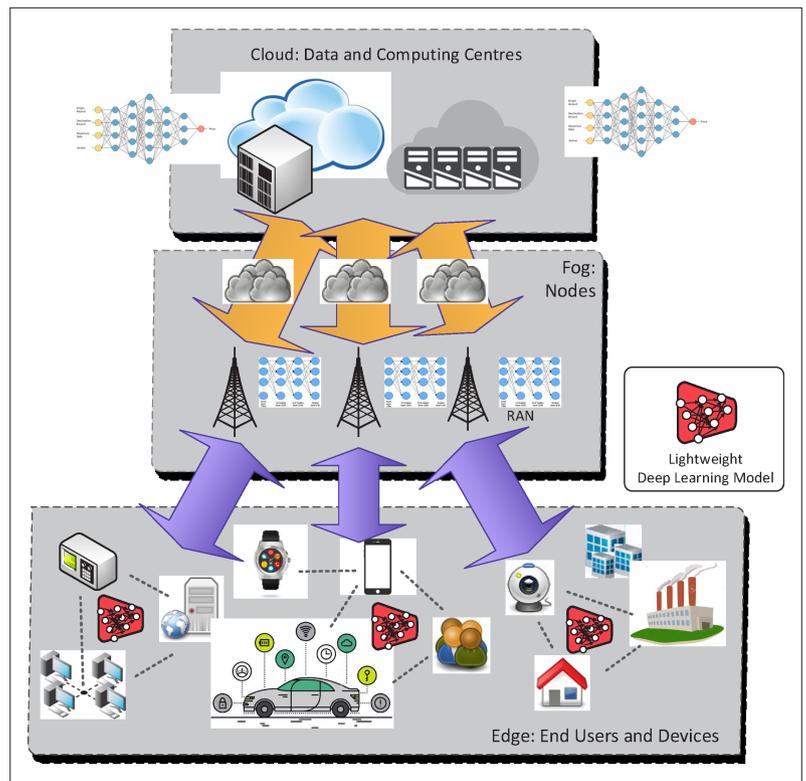


FIGURE 4. Application of deep learning in cloud, fog, and edge computing networks.

Study Group 13 at its meeting in Geneva, 6–17 November 2017. The focus group will draft technical reports and specifications for ML for future networks, including interfaces, network architectures, protocols, algorithms, and data formats. The three working groups are "Use cases, services and requirements," "Data formats & ML technologies," and "ML-aware network architecture." A 3GPP standards group developed an ML function that could allow 5G operators to monitor the status of a network slice or third-party application performance on "Zero Touch & Carrier Automation Congress" (<http://www.tech-invite.com/3m29/tinv-3gpp-29-520.html>, accessed 26 June 2018) in Madrid, 22 March 2018. The network data analytics function (NWDAF) forms a part of 3GPP's 5G standardization efforts and could become a central point for analytics in the 5G core network. Note that the NWDAF is still in the early stages of standardization, but could become an interesting place for innovation. A white paper named "Wireless Big Data for Smart 5G" (<http://www.future-forum.org/en/>, accessed 26 June 2018) was published on the FuTURE Forum in November 2017. This white paper is a collection of pioneering research works on big data for 5G in China, in both academia and industry. It proposed the concept of "smart 5G" and believed that the 5G network needs to embrace new and cutting-edge technologies such as wireless big data and AI to efficiently boost both spectrum efficiency and energy efficiency, improve the user experience, and reduce the cost. TIP launched a project group, "AI and applied machine learning" (<http://telecominfraproject.com/introducing-the-tip-artificial-intelligence-and-applied-machine-learning-project-group/>, accessed

Leading organization	Group name	Starting time	Purpose	Applications
3GPP	NWDAF	March 2018	Allow 5G operators to monitor the status of a network slice or third-party application performance	5G core network data analytics
5G PPP	CogNet	July 2015	Build an intelligent system of insights and action for 5G network management	Autonomic network management based on machine learning
FuTURE	Wireless big data for smart 5G	November 2017	A white paper that collects pioneering research works on big data for 5G in China	Boost spectrum efficiency and energy efficiency, improve the user experience, and reduce the cost
ITU	ML5G	November 2017	Identify relevant gaps and issues in standardization activities related to ML for future networks	Interfaces, network architectures, protocols, algorithms, and data formats
TIP	AI and applied machine learning	November 2017	Define and share reusable, proven practices, models, and technical requirements for applying AI and ML	ML-based network operations, optimization, and planning; Customer behavior-driven service optimization; Multi-vendor ML-AI data exchange formats

TABLE 1. A summary of AI/ML for B5G networks in standards.

26 June 2018), in November 2017. It will apply AI and ML to network planning, operations, and customer behavior identification to optimize service experience and increase automation. The objective is to define and share reusable, proven practices, models, and technical requirements for applying AI and ML to reduce the cost of planning and operating telecommunications networks, understand and leverage customer behavior, and optimize service quality for an improved experience. 5G PPP has also launched its efforts on combining AI with wireless communications, such as CogNet (<http://www.cognet.5g-ppp.eu/cognet-in-5gpp/>, accessed 16 September 2019). It aims to build an intelligent system of insights and action for 5G network management. These developments in standards and study groups aim to use AI for physical layer and network management, which will greatly boost the performance of wireless networks.

CONCLUSIONS AND FUTURE CHALLENGES

In this article, we have investigated how AI and ML can be used to efficiently solve the unstructured and seemingly intractable problems in future B5G wireless communication networks. A comprehensive survey of recent advances on the combination of AI/ML and wireless networks has been provided, including channel measurements, modeling, and estimation, physical layer research, and network management and optimization. Challenges and potential future research directions have been discussed. AI algorithms and their applications to B5G networks have been introduced. An overview of developments for applying AI/ML to B5G systems carried out by standard organizations and study groups has also been provided.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China under Grant 2018YFB1801101, the National Natural Science Foundation of China (NSFC) under Grant 61960206006, the High Level Innovation and Entrepreneurial Talent Introduction Program in Jiangsu, the Research Fund of National Mobile Communications Research Laboratory, Southeast University, under Grant 2020B01, the Fundamental Research Funds for the Central Universities

under Grant 2242019R30001, and the EU H2020 RISE TESTBED2 project under Grant 872172.

REFERENCES

- [1] C.-X. Wang *et al.*, "Cellular Architecture and Key Technologies for 5G Wireless Communication Networks," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 122–30.
- [2] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless Communications in the Era of Big Data," *IEEE Commun. Mag.*, vol. 53, no. 10, Oct. 2015, pp. 190–99.
- [3] S. M. Aldossari and K.C. Chen, "Machine Learning for Wireless Communication Channel Modeling: An Overview," *Wireless Personal Commun.*, vol. 106, no. 1, 2019, pp. 41–70.
- [4] J. Huang *et al.*, "A Big Data Enabled Channel Model for 5G Wireless Communication Systems," *IEEE Trans. Big Data*, vol. 6, no. 1, Mar. 2020.
- [5] X. Ma *et al.*, "Data Scheme-Based Wireless Channel Modeling Method: Motivation, Principle and Performance," *J. Commun. Info. Net.*, vol. 2, no. 3, Sept. 2017, pp. 41–51.
- [6] R. He *et al.*, "Clustering Enabled Wireless Channel Modeling Using Big Data Algorithms," *IEEE Commun. Mag.*, vol. 56, no. 5, May 2018, pp. 177–83.
- [7] H. Li *et al.*, "Wireless Channel Feature Extraction via GMM and CNN in the Tomographic Channel Model," *J. Commun. Info Net.*, vol. 2, no. 1, Mar. 2017, pp. 41–51.
- [8] A. Charrada and A. Samet, "Joint Interpolation for LTE Downlink Channel Estimation in Very High-Mobility Environments With Support Vector Machine Regression," *IET Commun.*, vol. 10, no. 17, Nov. 2016, pp. 2435–44.
- [9] H. He *et al.*, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, Oct. 2018, pp. 852–55.
- [10] H. Tang, J. Wang, and L. He, "Off-Grid Sparse Bayesian Learning Based Channel Estimation for mmWave Massive MIMO Uplink," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, Jan. 2019, pp. 45–48.
- [11] Y. Li *et al.*, "Large-Scale Remote Sensing Image Retrieval by Deep Hashing Neural Networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, Feb. 2018, pp. 950–65.
- [12] P. Sure, C. N. Babu, and C. M. Bhuma, "Applicability of Big Data Analytics to Massive MIMO Systems," *Proc. IEEE INDI-CON '16*, Bangalore, India, Dec. 2016, pp. 1–6.
- [13] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-Complexity Technologies to Enable Millimeter-Wave MIMO with Large Antenna Array for 5G Wireless Communications," *IEEE Commun. Mag.*, vol. 56, no. 4, Apr. 2018, pp. 211–17.
- [14] C. K. Wen *et al.*, "Channel Estimation for Massive MIMO Using Gaussian-Mixture Bayesian Learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, Mar. 2015, pp. 1356–68.
- [15] Y. Fu *et al.*, "Artificial Intelligence to Manage Network Traffic of 5G Wireless Networks," *IEEE Net.*, vol. 32, no. 6, Dec. 2018, pp. 58–64.

BIOGRAPHIES

CHENG-XIANG WANG [S'01, M'05, SM'08, F'17] (chxwang@seu.edu.cn) received his Ph.D. degree from Aalborg University, Denmark, in 2004. He has been with Heriot-Watt University, Edinburgh, United Kingdom, since 2005, where he was promoted to professor in 2011. In 2018, he joined Southeast University, China, as a professor. He is also a part-time professor with the Purple Mountain Laboratories, Nanjing, China. He has authored three books, one book chapter, and more than 360 papers

in refereed journals and conference proceedings, including 23 Highly Cited Papers. His current research interests include wireless channel measurements and modeling, B5G wireless communication networks, and applying artificial intelligence to wireless communication networks. He is a Fellow of the IET, an IEEE Communications Society Distinguished Lecturer in 2019 and 2020, and a Highly-Cited Researcher recognized by Clarivate Analytics, 2017–2019. He is currently an Executive Editorial Committee member of *IEEE Transactions on Wireless Communications*. He has received 10 Best Paper Awards from IEEE GLOBECOM 2010, IEEE ICCT 2011, ITST 2012, IEEE VTC 2013-Spring, IWCMC 2015, IWCMC 2016, IEEE/CIC ICC 2016, WPMC 2016, and WOCC 2019.

MARCO DI RENZO [F'20] (marco.direnzo@centralesupelec.fr) received his Ph.D. degree in electrical engineering from the University of L'Aquila, Italy, in 2007, and his D.Sc. degree (HDR) from University Paris-Sud, France, in 2013. He is a CNRS research director (CNRS professor) and a faculty member of Centrale Supélec and Paris-Saclay University, France. Also, he is a Nokia Foundation Visiting Professor at Aalto University, Helsinki, Finland. He is a Highly Cited Researcher and a Distinguished Lecturer of the IEEE Communications Society (ComSoc) and IEEE Vehicular Technology Society. Currently, he serves as the Editor-in-Chief of *IEEE Communications Letters*. He has received several awards, including the SEE-IEEE Alain Glavieux Award, the IEEE Jack Neubauer Memorial Best Systems Paper Award, the IEEE ComSoc Young Professional in Academia Award, the IEEE ComSoc Best Young Researcher Award, and the Best Paper Award at IEEE ICC 2019.

SLAWOMIR STAŃCZAK [M'04, SM'11] (slawomir.stanczak@hhi.fraunhofer.de) studied electrical engineering with specialization in control theory at Wrocław University of Technology and at the Technical University of Berlin (TU Berlin). He received his Dipl.-Ing. degree in 1998 and his Dr.-Ing. degree (summa cum

laude) in electrical engineering in 2003, both from TU Berlin; his Habilitation degree (venialegendi) followed in 2006. Since 2015, he has been a full professor for network information theory with TU Berlin and the head of the Wireless Communications and Networks Department at Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI). He is a co-author of two books and more than 200 peer-reviewed journal articles and conference papers in the area of information theory, wireless communications, signal processing, and machine learning. He was an Associate Editor of *IEEE Transactions on Signal Processing* between 2012 and 2015. Since February 2018 he has been the Chairman of the ITU-T focus group on machine learning for future networks including 5G.

SEN WANG (s.wang@hw.ac.uk) is an assistant professor in robotics and autonomous systems at Heriot-Watt University and a faculty member of the Edinburgh Centre for Robotics. He is also the director of the Perception and Robotics Group in the Ocean Systems Laboratory. Previously, he was a postdoctoral researcher at the University of Oxford. His research interests include robotics, computer vision, and machine/deep learning.

ERIK G. LARSSON [S'99, M'03, SM'10, F'16] (erik.g.larsson@liu.se) is a professor at Linköping University, Sweden. He co-authored *Fundamentals of Massive MIMO* (Cambridge, 2016) and *Space-Time Block Coding for Wireless Communications* (Cambridge, 2003). Recent service includes membership in the IEEE Signal Processing Society Awards Board (2017–2019), the *IEEE Signal Processing Magazine* Editorial Board (2018–2020), and the *IEEE Transactions on Wireless Communications* Steering Committee (2019–2022). He received the *IEEE Signal Processing Magazine* Best Column Award twice, in 2012 and 2014, the IEEE ComSoc Stephen O. Rice Prize in Communications Theory 2015, the IEEE ComSoc Leonard G. Abraham Prize 2017, the IEEE ComSoc Best Tutorial Paper Award 2018, and the IEEE ComSoc Fred W. Ellersick Prize in 2019.