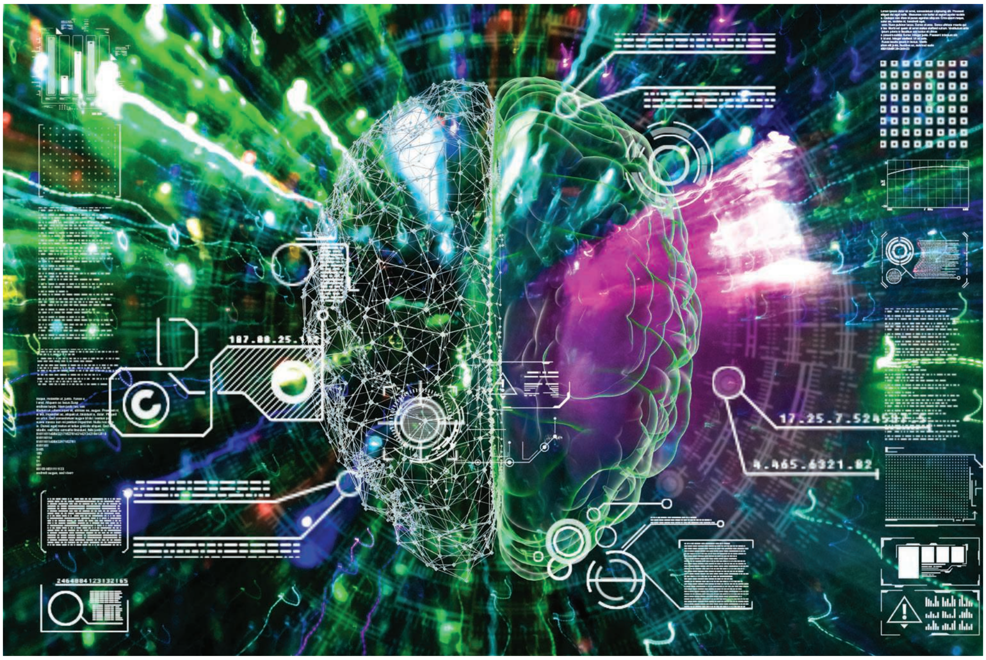
WHEN MACHINE LEARNING MEETS BIG DATA

A Wireless Communication Perspective

Yuanwei Liu, Suzhi Bi, Zhiyuan Shi, and Lajos Hanzo



We have witnessed an exponential growth in commercial data services, which has led to the so-called big data era. Machine learning, one of the most promising artificial intelligence (AI) tools for analyzing this deluge of data, has been called upon in many industry and academic research areas. In this article, we briefly review big data analysis and machine learning, along with their potential applications in next-generation (NG) wireless networks. Next, we invoke big data analysis to predict the requirements of mobile users and exploit such analysis to improve the performance of "social network-aware wireless." In particular, a unified, big data-aided machine-learning framework is proposed that consists of feature extraction, data modeling, and prediction/online refinement. The main benefits of this proposed framework are that, by relying on big data that reflects both the spectral and other challenging requirements of users, we can refine the motivation, problem formulations, and methodology of powerful machine-learning algorithms in the context of wireless networks.

To characterize the efficiency of the proposed framework, a pair of intelligent, practical applications are provided as case studies to predict 1) the positioning of drone-mounted areal base stations (BSs) according to the specific teletraffic requirements by gleaning valuable data from social networks and 2) the content-caching requirements of BSs according to users' preferences by mining data from social networks. Finally, open research opportunities are identified for motivating future investigations.

Digital Object Identifier 10.1109/MVT.2019.2953857
Date of current version: 17 December 2019

 1556-6072/20©2020IEEE

Current Challenges

The current family of NG techniques seeks to provide high-quality communication services that rely on high throughput, massive connectivity, and low delay. For cellular systems, a 1–10-Gb/s downlink transmission rate and a delay under 1 ms are expected. Meanwhile, 5G standards also allow for conventional high-speed cellular communication to coexist with machine-to-machine and Internet of Things (IoT) services, with an emphasize on wide-coverage, advanced, dense connectivity (associated with up to one million sensor connections within a square kilometer area) [1]. For instance, the IoT networks conceived for industrial control and health monitoring generate vast amounts of sensing data. In addition, autonomous connected vehicles of the near future will support millions of high-velocity devices, which will significantly increase the data traffic of the emerging NG network.

The term *artificial intelligence* was first coined by John McCarthy in 1956 [2]. As a branch of computer science, AI's promise is to enhance the intelligence of computers by imitating the actions of human beings, i.e., understanding natural language and planning/perceiving sounds and objects in problem solving and learning, as shown in Figure 1. There are numerous techniques used for formulating AI solutions. The early seminal approaches tend to explicitly program a decision system by leveraging domain-specific knowledge, leading to the concept of expert systems. Various carefully defined rules have to be structured and shaped to create such domain-specific expert programs. In contrast to expert systems, which contain millions of lines of code along with decision trees and complex rules, machine learning is of potentially lower complexity; hence, it has made tremendous progress over the last several decades [3]. The stylized relationship between machine learning and AI is illustrated in Figure 1. The core motivation behind machine learning is that of allowing autonomous learning/training based on its access to huge amounts of data rather than writing hard-coded routines of specific instructions. The resultant algorithm then offers a variety of intelligent actions, ranging from learning based on past experience, reasoning for comprehending complex ideas, and generalizing to new situations.

In this light, a natural question arises: How can big data and machine learning help to enhance the performance of future 5G and beyond wireless networks? At the time of this writing, it is not so widely understood how best to harness machine-learning techniques for solving the typical optimization problems in wireless networks scenarios that depend on big data analysis. Contributing to the solution of this problem motivates us to develop this treatise, where big data resources are utilized by analytical machine-learning tools in support of intelligent applications in wireless networks. The main contributions of this article can be summarized as follows:

1) The beneficial exploitation of big data resources in wireless networks is discussed.
2) Three classification approaches suitable for machine learning are proposed according to different criteria. The pros and cons are presented and consider compelling application scenarios.
3) A unified framework for invoking machine-learning techniques in social network-aware wireless is proposed and augmented by considering a pair of intelligent application scenarios.

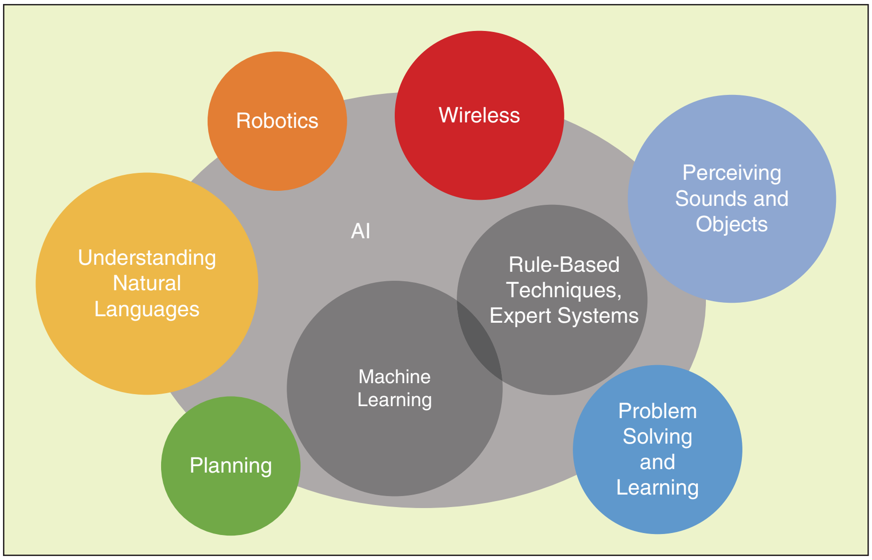


FIGURE 1 The relationships between AI and machine learning. In a rule-based system, static knowledge is explicitly represented as a set of rules that can be thought of as a collection of facts, constraints, or regulators. An expert system constitutes one of the classic examples of a rule-based system and makes use of various domain-specific expert programs. The implementation and integration of heuristic knowledge also provide a basis for transparency and flexibility. By contrast, machine learning tends to automatically learn computational models based on its access to huge amounts of data without relying on hard-coded routines of specific instructions.

Big Data in Wireless Networks

In this section, we classify the family of existing data sources into three broad categories, i.e., general wireless data, social network-aware data (social data), and cloud data, as shown in Figure 2. The corresponding application scenarios are discussed in the following sections.

Wireless Data

The big data generated by wireless users contain useful information about their activity patterns versus time, frequency, and space. For instance, from the data traffic/demand variation over time, we can infer the interference power at different frequencies, the congestion-level distribution at different locations, and so on. By making use of these spectral patterns, we can efficiently manage wireless resources to improve the system's spectral efficiency and enhance user quality of service (QoS). As displayed in Figure 2, one of these intelligent applications is load balancing, which relies on proactive wireless resource allocation. In this context, the operator can adjust the transmit power, frequency, or direction (e.g., through sectorized antennas) of different BS transmitters that rely on estimating the mobile users' distributions. Furthermore, the operator can dispatch mobile BSs in advance when a surge of regional data traffic is anticipated.

Another important application is constituted by wireless security surveillance. Given the spectral activity patterns inferred, we can detect anomalies in the radio environment (i.e., the perspective of rogue BSs) based on atypical, measured, real-time spectrum usage. The key challenge of the aforementioned applications is to derive such a "radio map" from the vast amount of noisy wireless big data so that we can accurately characterize the spectral usage patterns in different dimensions and scales [4].

Social Data

The main cause of the soaring data volume on the Internet is online social networks. The penetration of the mobile Internet into our daily lives makes convenient multimedia communications ubiquitously accessible for everyone. The amount of social data has reached an astonishing magnitude and is set to exhibit an increasing trend in the next few years. In 2017, on average, more than 500 million Twitter messages were generated per day, over 80% of them initiated from mobile terminals (data from online). On the one hand, social network data feature strong ties to public events occurring in the physical world. For instance, an important football game or political event may inspire heated online discussions that last for days; meanwhile, the frequent sharing of high-score online evaluations about a movie premiere or a newly opened restaurant may attract a large number of customers in the real world. On the other hand, mobile social network data contain rich information about the contexts/preferences of individuals or social groups. As an example, we can infer from tweets that mobile users visiting a famous tourist site are unhappy about its wireless services. To address this shortcoming, we can improve the typical tourist experience by temporarily allocating more bandwidth to the nearby BSs. As a result, a social network-aware wireless concept can be understood. Here, the major technical difficulty is to surmise the true "meaning" of the users' messages and take the appropriate actions. Given the vast amount of social network data and the diverse nature of information conveyed, e.g., text or multimedia, this may be achievable through advanced machine learning.

Cloud Data

Another major cause of the data tsunami in wireless networks is the transmission of multimedia content stored on cloud servers. It is predicted that, by 2019, more than 80% of the world's Internet traffic will be videos, i.e., YouTube short clips, Netflix long clips, and Facebook live streaming (data from online). Meanwhile, online audio streaming services also contribute a large portion of the remaining 20% data traffic. A unique feature of cloud-based big data is that the users' preferences concerning specific content are often similar and correlated. For instance, 3% of YouTube videos account for more than 90% of its total views (data from online). In other words,

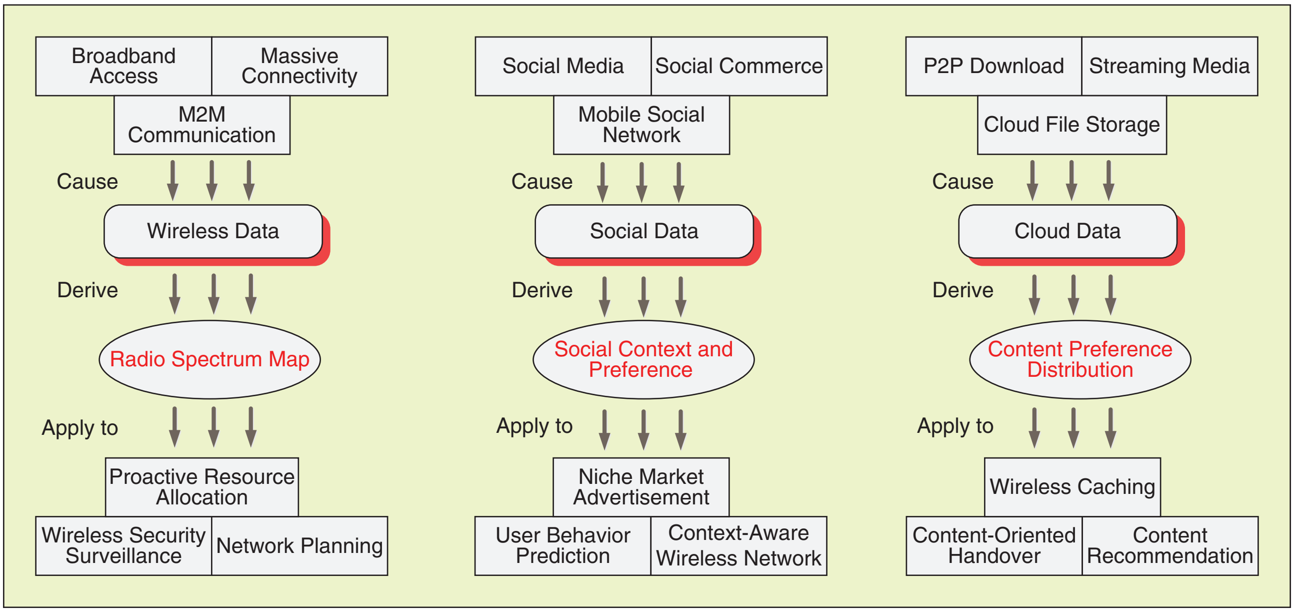


FIGURE 2 A flowchart of the classifications of wireless big data: its source, the information hidden in it, and its applications. M2M: machine-to-machine; P2P: point-to-point.

most of the content transferred over the Internet is based on its popularity. We can, therefore, reduce the teletraffic of the system by exploiting users' preference for specific cloud content. For instance, we can precache the most popular videos at edge servers so that no real-time backhaul data downloading is needed for frequent requests. Moreover, by caching the video content at multiple nearby BSs, we can form a virtual antenna array and support seamless handovers for mobile terminals. Additionally, each individual user preference for specific multimedia content, if available, can be used for predicting the user's future demand. The network operator can then perform prefeeding or recommendation actions.

Naturally, the main technical challenge is to accurately predict users' preference distribution. Another challenge is to correctly label the lengthy number of videos for future reference/searching. In this case, using manual labeling is infeasible for high-population networks; hence, AI techniques are needed for designing content-oriented, intelligent wireless transmissions.

Machine Learning in Wireless Networks

In this section, we detail how machine-learning techniques are classified and how they can be applied to wireless communications. The pros and cons of each machine-learning family are discussed and displayed in Figure 3.

Human Supervision Requirement

According to whether or not the algorithms require human supervision, machine-learning models may be grouped into the following three general categories [5], [6]:

1) *Supervised learning*: Long considered a major branch of machine learning, supervised learning has been extensively studied and developed. Large quantities of human-labeled data should always be readily available for learning a functional mapping between the training samples observed and the desired output. The advantage of supervised learning is that both the convergence speed and the action quality are high, although they typically require a large amount of data to be labeled manually, thus making the data processing

Machine Learning		Pros	Cons	Algorithms	Applications
Supervision?	Supervised	Task Specific, Effective	Requires Large Labeled Training Data	Linear Regression, Random Forest, Support Vector Machines	Wireless Resource Allocation, Encoder and Decoder Design
	Unsupervised	Utilize Unlabeled Data	Implicit, Reduced Accuracy	*K*-Means, Principal Component Analysis, Latent Variable Models	User Association, Attack Detection From Malicious User, User Grouping for NOMA
	Reinforcement	Interactive Learning, No Direct Supervision	Interpretability Requires Substantial Computing Resources	Q-Learning, SARSA, A3C, Deep-Q Network, DDPG, PPO	UAV, Autonomous Driving, MEC, Wireless Caching
Incremental?	Batch Processing	One-Off Training	Less Representative of Streaming Data	Mini-Batch Gradient Descents	Wireless Caching, Offloading
	Online	Continuously Updating, Memory Efficient	Vulnerable to Noisy Data	Stochastic Gradient Descent, Online Convex Optimization	Cognitive Radio, UAV Movement
Generalize?	Model Based	Computational Efficiency	Coarse Function Approximations	Neural Network, Decision Tree, Probabilistic Graphical Models	Trajectory Prediction for Mobile Users, Channel Modeling for mm-Wave
	Instance Based	Learning Stage Not Required	Excessive Memory Requirement	*K*-Nearest Neighbor Algorithm	Power Spectrum Density Estimation, User Demand Prediction

FIGURE 3 The classifications of machine learning, each classification's corresponding pros and cons, and some application scenarios for each. The intent of this figure is to provide a systematic review of practical applications for modern machine-learning methods in wireless communication. By categorizing the various approaches into three principal groups according to 1) how much human supervision they require, 2) whether they are able to perform incremental learning, and 3) how they generalize large-scale training/testing scenarios, we explicitly demonstrate the pros and cons of each scheme and highlight the most appropriate applications. NOMA: nonorthogonal multiple access; UAV: unmanned aerial vehicle; MEC: mobile edge computing; mm-wave: millimeter-wave; SARSA: state–action–reward–state–action; AC-3 algorithm: arc-consistency algorithm 3; DDPG: deep deterministic policy gradient; PPO: proximal policy optimization.

more complex. Attractive scenarios for applying supervised learning are wireless resource allocation and encoder and decoder design, where the objective function definition of the application is clear and collecting sufficient training data is relatively easy and less costly.

2) *Unsupervised learning*: This model relies on vast amounts of unlabeled data for inferring the underlying information structure without depending on external resources and human supervision. The advantage of unsupervised learning is that no prior knowledge is required; however, this comes at the cost of potentially reducing its accuracy. Another disadvantage is that the automatically discovered data are not always representative of real-world conditions. Given its unique features, unsupervised learning is suitable for solving the problems of user association, user grouping for hybrid multiple access, attack detection of malicious users, and so on.

3) *Reinforcement learning*: Initially designed to discover optimal action spaces through adaption and interactions in uncertain time-varying environments, this model provides another way of learning from unlabeled data as long as either positive or negative feedback can be gleaned during the learning process by trial and error. Reinforcement learning does not require direct supervision. In fact, the interactive learning paradigm is capable of learning to act so that it may prepare itself for achieving an ever-improving performance. The disadvantage of reinforcement learning is, however, that it relies on huge amounts of resources. Another drawback is that the resultant high-performance solutions often lack plausible physical interpretations. Nonetheless, successful application scenarios for reinforcement learning have been found in unmanned aerial vehicle (UAV) communications, autonomous driving, mobile edge computing (MEC), and wireless caching placement.

Learning Capability

Based on its learning capability, we can classify the family of machine-learning models into the following two subsets [7]:

1) *Batch learning-based algorithms* typically train a model after a sufficiently large amount of training data have been collected prior to the learning task. This offline learning procedure has the advantage of allowing for more involved machine learning based on all the available data, where the model tends to be updated infrequently in general. The main advantage of batch learning is that the convergence speed is high; however, batch learning may not be suitable for real-time learning of rapidly fluctuating processes subject to stringent delay requirements. As a result, its beneficial application scenarios are wireless caching or wireless offloading.

2) *Online learning* enables a model to learn from a stream of data instances arriving sequentially, which is achieved by continuously changing and adapting its structure and parameters. The “on-the-fly” learning scheme holds the promise of being both memory efficient and highly scalable in solving large-scale learning problems. A particularly remarkable advantage of online learning is that it is eminently suitable for real-time processing. Nevertheless, its convergence speed is typically slow; therefore, it can be used for cognitive radio networks or UAV movement scheduling.

Generalization Requirement

On the basis of how the methods generalize their findings from the training data to hitherto unseen examples, machine-learning models can be divided into the following two parts [8]:

1) *Instance-based learning algorithms* [9] tend to use the whole set of instances found in the training data stream for constructing inference structures and predicting unseen instances. Although they are quite capable of competent generalization, this is achieved at the cost of extended search time and high memory requirements. An especially remarkable advantage of instance-based learning is that it does not require any prior-model assumptions. Therefore, instance-based learning is suitable for complex wireless scenarios, such as power spectrum density estimation or user demand prediction. However, large data sets are required for high-quality instance-based learning.

2) *Model-based learning* aims to find the optimal parameters of the algorithms designed [10] so as to optimize the objective function and maximize the generalization capability in the face of previously unseen testing data. Such models usually suffer from limited accuracy but offer higher computational efficiency. Hence, the key advantage of model-based learning is that its implementation cost is low. Given these attributes, model-based learning can be readily used for the movement trajectory prediction of mobile users, channel modeling estimation in millimeter-wave communications, and so on.

A Unified, Big Data-Aided AI Framework

In this section, we propose the new, unified machine-learning framework presented in Figure 4. According to social, cloud, or wireless data, we implement advanced machine-learning approaches that analyze the data used to extract useful information. More particularly, the proposed framework includes the following three stages, as illustrated in Figure 4:

1) *Feature extraction*: During this stage, we extract features from data sets that include social, cloud, or wireless data, as depicted in Figure 2. Let us consider as an example social data, which include social media, social commerce, or mobile social networks. We can

derive social context and user preferences with the help of social data by invoking natural language processing or other techniques. The detailed procedure includes a pair of steps, the first of which is syntax processing. Here, its input is the tremendous amount of social data, which are subjected to natural language processing techniques, e.g., language identification, tokenization, lemmatization, stemming, etc. The output should be some useful geo- and time-tagged keywords according to the specific application, e.g., crime prevention. The second step is semantic analysis, in which the input is determined by the keywords subtracted during the last stage through the use of sophisticated techniques, e.g., lexical semantics, clustering content, sentiment analysis, and summarization. The outputs are expected to be context-aware user preferences. A suitable example is to apply Twitter data to predict topics of high popularity for content caching in wireless networks.

2) *Data modeling*: The feature extraction stage mentioned previously can be regarded as the predata processing stage. This data modeling stage constitutes the core process of establishing a machine-learning model. More particularly, the feature we extracted during the last stage, i.e., user mobility and the content popularity of users, represents the data set we seek for modeling the problem. Typically, we first formulate an objective function [e.g., maximizing the quality of experience (QoE), reducing delay, or improving energy efficiency] for enhancing the performance of the wireless networks considered. Then, we utilize the modeling strategies presented in the data modeling stage of Figure 4 to mathematically represent the formulated problems.

3) *Prediction/online refinement*: After establishing the machine-learning model, we proceed to either the prediction or the online refinement stage. For the prediction stage, it is implied that the established model can be used only for predicting user behavior/network performance without updating the model invoked. During the online refinement stage, the established model can be refined based on changes to the real-time data input. The intuitive difference between prediction and online refinement is that the system provides a flexible mechanism for inference testing, relying on either an offline or periodically refined model. The selection of these two variants ultimately depends on the particular context of a specific application. For the use case when the data distribution does not show large variations over the pertinent time scales, predictions that depend on an offline trained model are always the best choice for ease of computation. Conversely, online refinement must be undertaken to improve local errors when the data are intrinsically time varying; however, the latter is more computationally demanding. Note that a significant advantage of this framework is that the machine-learning model depends not only on historical data but also on real-time data. For example, we could periodically use the past two weeks of data to predict the next day's trends. By exploiting these time-sequence characteristics, the prediction capability of this model becomes quite accurate and timely. The techniques displayed in the prediction/online refinement stage of Figure 4 can all be beneficially applied.

Case Study: Social Network-Aware Wireless

Having introduced the three key stages of the proposed framework, the next important step is to identify suitable

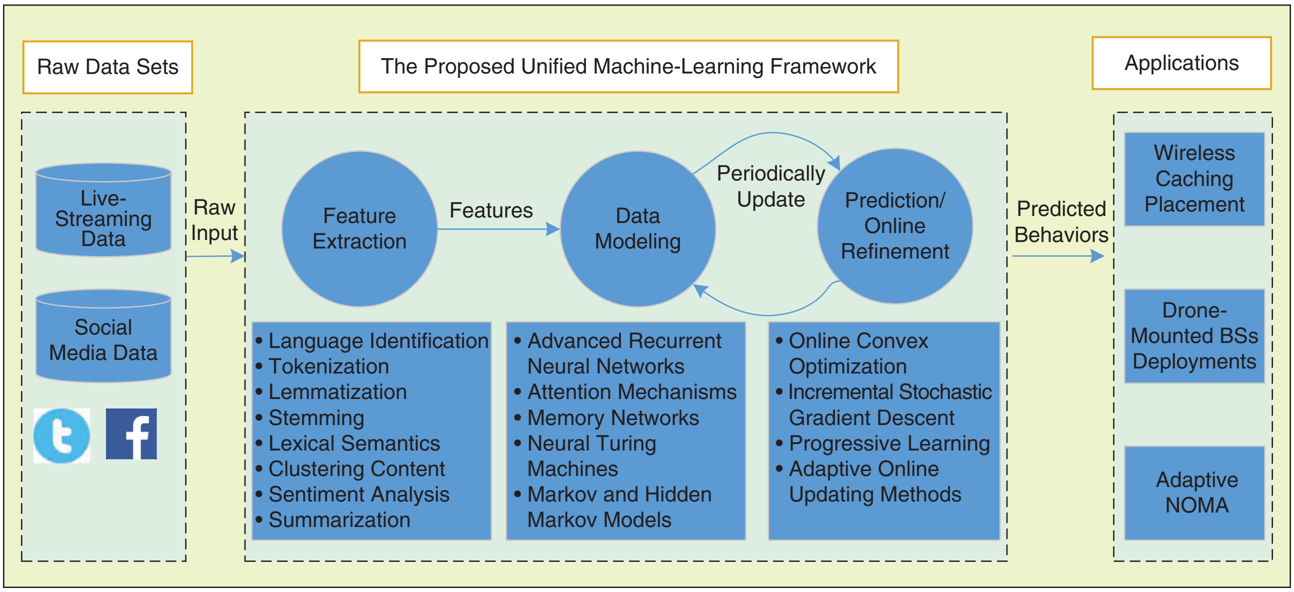


FIGURE 4 A unified big data-aided machine-learning framework.

application scenarios. In this section, we consider a pair of intelligent applications as a way of demonstrating how best to apply our big data-aided machine-learning framework for fostering intelligent wireless networks. In both case studies, we develop the position information collected using Twitter's application programming interface to infer the periodic behaviors of users in social networks. We propose efficient algorithms based on neural networks for predicting the user's mobility or content popularity, both of which belong to the feature extraction stage. Then, reinforcement learning is used to solve the problems formulated in the data modeling stage, either for cooperative caching or UAV deployment, for example.

Dynamic UAV Deployment and Movement Design

Thanks to the rapid capability upgrade of drones, BSs on the fly have emerged as an efficient solution for improving radio coverage conditions in the face of rapidly fluctuating traffic demands by appropriately adjusting the position of drones. Because of the availability of drone-mounted BSs, how best to position them for satisfying dynamically evolving data requests, and so maximize their benefits to operators, becomes one of the most challenging and critical problems. In such a scenario, the strategies for deploying drone-mounted BSs can be categorized into the two following types:

1) *Predeploy UAVs before the requests arrive*: Based on the proposed framework, historical social data can be used for predicting forthcoming data requests in a specific area. If the outputs of data mining indicate that data requests may be expected to exceed the capability of fixed BSs covering that area, drone-mounted BSs can be deployed to the area in advance. The goal of this strategy is to provide an improved user experience by enhancing the network's capability before congestions happen.
2) *Move UAVs based on real-time requests*: Naturally, historical social data cannot exactly predict user requests, as they fluctuate dynamically. Therefore, a real-time UAV action is desired for improving the network's capability to enhance user experience.

Real-time social data related to the area of interest, i.e., tweets complaining about poor network quality, can be further analyzed, and more drone-mounted BSs can be dispatched to the hotspot area. In particular, by analyzing the critical tweets, we first decide the satisfaction level of users about the current transmit rate based on the QoE model, which identifies potential teletraffic congestion events. Next, the required number of UAVs as well as their movements can be determined for maximizing the users' sum mean opinion score (MOS) by adopting machine-learning approaches.

In this case study, multiple UAVs are used as flying BSs serving ground users by considering both of the aforementioned types of UAV deployments. Each UAV is capable of roaming in a free 3D space. Because ground users are intended to move flexibly, UAVs can fly dynamically according to the users' real-time positional information. Figure 5(a) shows the four steps of UAV deployment and movement design that rely on the proposed framework. Step 1 is considered the feature extraction stage. Here, the movement of users is the key feature we would like to extract. Multiagent Q-learning is the core algorithm used during the data modeling stage, which is employed in steps 2 and 4 for designing the specific deployment and movement of UAVs. Figure 5(b) illustrates how a UAV moves from an initial position to its final destination through the use of reinforcement learning for formulating its trajectory. In our scenarios, multiple UAVs move simultaneously according to the predicted movement of ground users; this can be termed the *prediction* stage. By doing so, the predeployment of UAVs can be achieved for enhancing the energy efficiency of networks.

Figure 6 compares the throughput between the moving UAV solution (both a predeployed and real-time movement solution) and the static one. It can be observed that the instantaneous transmit rate decreases as time elapses. This is because users are roaming during each time slot and, when user density is reduced, the instantaneous sum of the transmit rate is affected. It can also be observed that the redeploying of UAVs based on the movement of users is an efficient method of mitigating the downward trend compared to the static scenario.

Wireless Caching Placement and Resource Allocation

Due to the rapid development of mass-storage techniques, storage becomes an increasingly low cost resource, while the opposite trend prevails for spectral resources. The caching process has two parts: caching placement and cached content delivery. In contrast to most existing research contributions, which have assumed that the caching placement process follows a specified distribution, in this case study, the content caching at each BS is dynamically varied according to user requirements. More particularly, with the aid of social data within a given area and the corresponding location information of BSs, the caching placement can be improved considerably. Typically, the main objective of caching placement is to identify the most popular K files to be stored in the BS during a particular period to reduce the potential latency when users download these files. To efficiently allocate wireless content, e.g., video clips to BSs, some prior information on user mobility and content popularity is required. However, the complicated relationship between historical and future information makes conventional approaches less applicable. AI algorithms are capable of predicting network demands in the near future, which helps with efficiently caching the most popular content when the

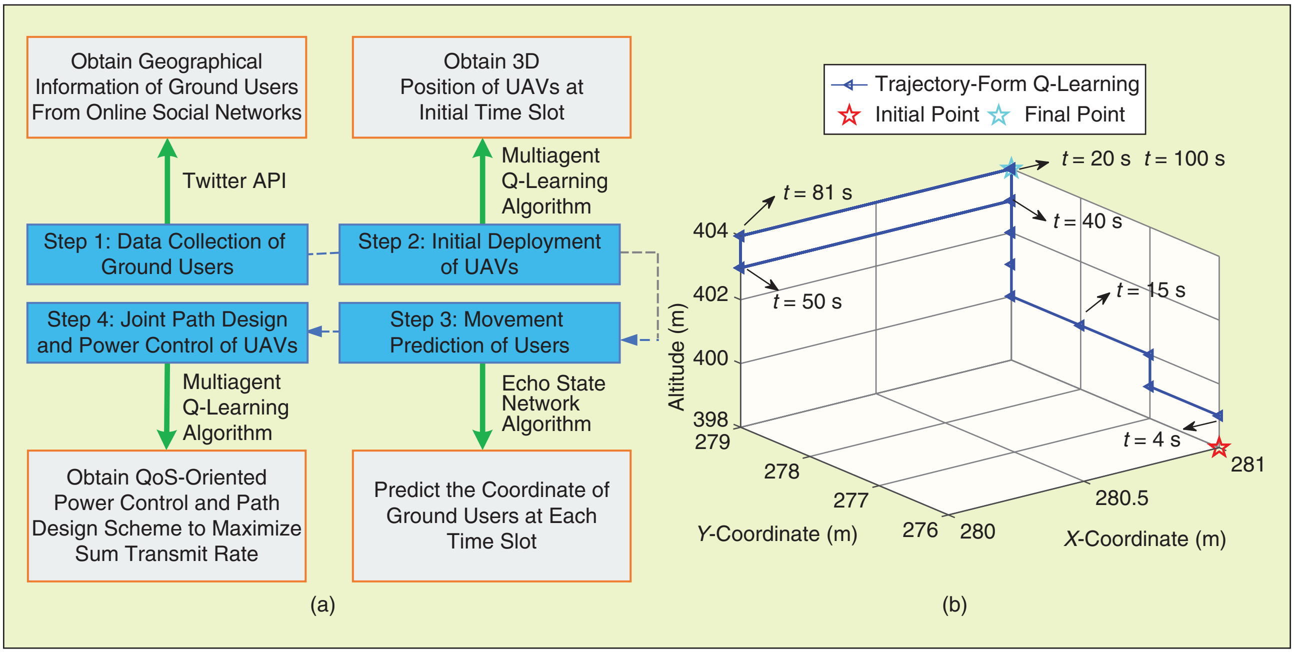


FIGURE 5 A case study of the deployment and movement design for multiple UAV networks [14]. (a) The procedure for UAV deployment and movement and (b) the trajectory for UAV deployment and movement. API: application programming interface.

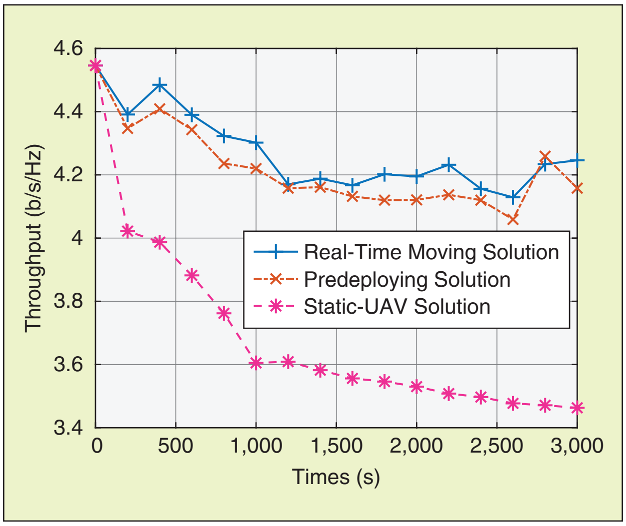


FIGURE 6 A comparison of throughput between a moving-UAV and static-UAV solution.

communication resources (e.g., bandwidth, storage capacity, and computing speed, etc.) are finite.

Considering some recent, vile terrorist attacks as an example, typically, there are immediate reactions on social media, i.e., Facebook and Twitter. In the ensuing period, the discussions, related posts, and messages that use social data increase. By invoking natural language processing to analyze the posts, keywords can be extracted. The keywords may then be classified or mapped to different hot topics. The hot topics determine what content should be cached in particular BSs. Based on time-geotagged hot topics, machine learning may be deployed for training this learning model. The final goal is that, according to the historical keywords in a certain past period of, e.g., 24 h, the content to be cached in the next period of 2 h, e.g., can be dynamically decided.

Figure 7(a) depicts the detailed procedure of wireless caching. More specifically, we can still map the wireless caching placement to our proposed framework. In the feature extraction stage, both user mobility and content popularity must be predicted to support proactive content placement. During the data modeling stage, neutral networks and reinforcement learning may be invoked to mathematically model mobility/content popularity and cooperatively cache placement, respectively. During the prediction stage, the outputs from neural networks [13] can be used for predicting the cooperative caching allocation. Figure 7(b) shows the total MOS of users in the network versus time for different algorithms. During the initial time period, the optimal content is placed according to the users' positions. Then, because we consider a user movement scenario, the positions of users change with the time period; therefore, the content placement is not optimal. As a result, users' total MOS decreases. We can observe that the Q-learning algorithm, which is a close relative of the reinforcement learning algorithm, is capable of achieving a near-optimal performance, while outperforming global *K*-means-based caching.

Future Challenges and Concluding Remarks

In this article, the design challenges of invoking machine-learning techniques for enhancing wireless networks with the aid of big data solutions were investigated. First, we identified the key features of big data, which spawn new research opportunities for wireless networks. Then, we discussed the classification of machine-learning techniques and their corresponding application scenarios in wireless networks. Furthermore we proposed a unified, big data-aided machine-learning framework. We followed this by introducing a pair of case studies in wireless networks, i.e., UAV deployment/trajectory design and wireless caching placement.

Still, numerous open research opportunities must be pursued in the context of machine-learning-aided wireless communications:

- *Adaptive, nonorthogonal, massive multiple access*: Nonorthogonal multiple access (NOMA) is a multifaceted technique used for enhancing wireless networks. Note that each form has advantages and disadvantages. Accordingly, different NOMA techniques are suitable for various scenarios, [15, Table 9]. Motivated by this, a software-defined NOMA architecture is proposed for supporting diverse user scenarios; however, teletraffic demands can vary seasonally or due to public events. Therefore, machine learning can be used to predict data traffic. The multiple access settings can be categorized into two types: pre- and real-time settings. Batch learning can be used for presettings, while real-time settings are

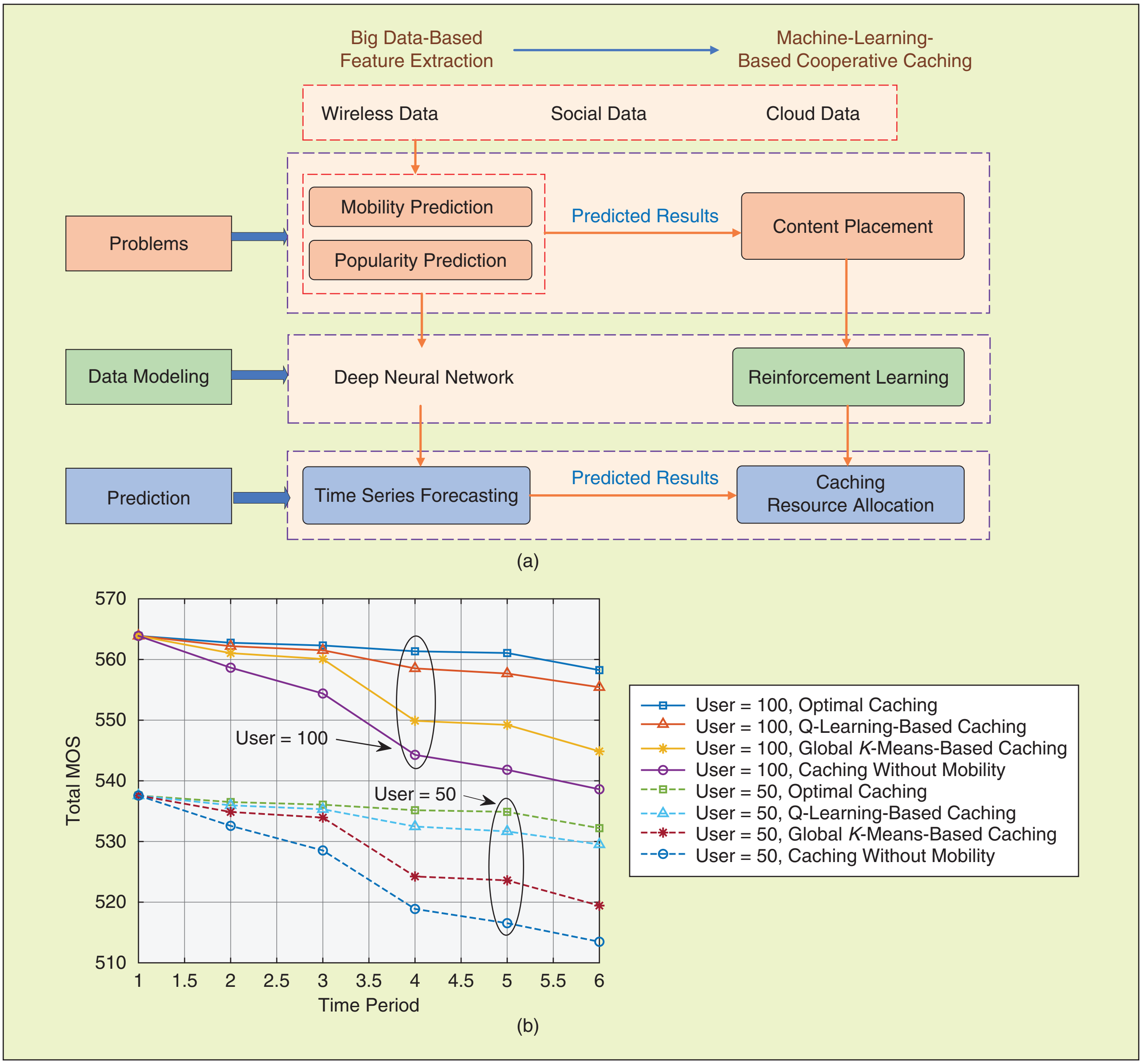
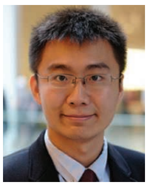


FIGURE 7 A case study of content placement in wireless cooperative caching [12]. (a) The flowchart for wireless caching and (b) the simulation results for wireless caching.

typically based on users' social media feedback regarding, e.g., the ability to dynamically change the multiple access settings on demand, which may invoke online learning, as mentioned in the "Machine Learning in Wireless Networks" section.

- *Autonomous driving in vehicle-to-everything networks*: Autonomous driving has the potential to benefit society in numerous ways, such as reducing traffic congestion and mitigating the environmental footprint of modern-day traffic. The combination of autonomous driving and vehicle-to-infrastructure (V2I) communications enables automated vehicles to receive up-to-date information about neighboring vehicles' dynamics and other traffic information, thereby enhancing both safety and traffic efficiency. The deep-reinforcement-learning models mentioned in the "Machine Learning in Wireless Networks" section, which are trained by interacting with its environment, may be utilized to optimize the behaviors of vehicles by exploring the environment in an iterative manner and learning from mistakes. With the aid of cloud or wireless data, V2I network requirements can be updated in real time. In this case, the autonomous vehicle becomes capable of safely reaching its destination, while avoiding traffic jams with the aid of up-to-date traffic information.
- *Intelligent computation offloading*: MEC is a promising technique used to meet the ever-increasing computational demands of mobile applications by providing computing capabilities at the edge of wireless networks, while migrating the computationally intensive tasks to the MEC server. The core concept of MEC is to provide abundant computing capabilities at the edges of networks and so mitigate both the backhaul and fronthaul load and reduce mobile users' energy consumption. Task offloading decisions and computational resource allocation constitute a pair of challenges in MEC. Obtaining an optimal offloading policy in such a dynamic MEC system in real time is challenging; however, machine learning is capable of intelligent inferences from historic information. By using the proposed framework, a real-time dynamic MEC system that operates with the assistance of social data can be designed utilizing machine-learning algorithms to attain significant performance-versus-complexity benefits.

Author Information



Yuanwei Liu (yuanwei.liu@qmul.ac.uk) is a lecturer (assistant professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom. His research interests include 5G/6G networks, the Internet of Things, machine learning, and stochastic geometry. He is a Senior Member of the IEEE.

Suzhi Bi (bsz@szu.edu.cn) is an associate professor with the College of Electronic and Information Engineering, Shenzhen University, China. His research interests involve optimizations in wireless information and power transfer, mobile computing, and smart power grid communications. He is a Senior Member of the IEEE.

Zhiyuan Shi (zhiyuan.shi@onfido.com) is a senior research scientist at Onfido Research London, United Kingdom. His research interests include computer vision (action recognition, object localization, and attribute learning) and machine learning (deep learning, probabilistic graphical models, and weakly supervised learning). He is a Member of the IEEE.

Lajos Hanzo (lh@ecs.soton.ac.uk) is with the School of Electronics and Computer Science, University of Southampton, United Kingdom, where he holds the chair in telecommunications. Currently, he directs an academic research team that works on a range of research projects in the field of wireless multimedia communications. He is a Fellow of the IEEE.

References

[1] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190–199, Oct. 2015.
[2] J. McCarthy and P. Hayes, "Some philosophical problems from the standpoint of artificial intelligence," in Machine Intelligence, B. Meltzer and D. Michie, Eds. Edinburgh, UK: Edinburgh Univ. Press, 1969, pp. 463–502.
[3] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks," *IEEE Access*, vol. 6, pp. 32,328–32,338, May 2018.
[4] S. Bi, J. Lyu, Z. Ding, and R. Zhang, "Engineering radio map for wireless resource management," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 133–141, Apr. 2019.
[5] N. Nasrabadi, "Pattern recognition and machine learning," *J. Electron. Imag.*, vol. 16, no. 4, pp. 049,901, 2007.
[6] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.
[7] T. Anderson, *The Theory and Practice of Online Learning*. Edmonton, AB, Canada: AU Press, 2009.
[8] C. Robert, *Machine Learning: A Probabilistic Perspective*. New York: Taylor & Francis, 2014.
[9] W. Daelemans and A. Van den Bosch, *Memory-Based Language Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[10] J. Quinlan, "Combining instance-based and model-based learning," in *Proc. 10th Int. Conf. Machine Learning*, 1993, pp. 236–243.
[11] Z. Qin, H. Ye, G. Ye Li, and B.-H. Fred Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Apr. 2019.
[12] Z. Yang, Y. Liu, and Y. Chen, "Q-learning for content placement in wireless cooperative caching," in *Proc. IEEE Global Communications Conf. (GLOBECOM'18)*, Abu Dhabi, Dec. 2018, pp. 1–6. doi: 10.1109/GLOCOM.2018.8647993.
[13] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," *IEEE Commun. Surv. Tuts.*, vol. 21, no. 4, pp. 3039–3071, Fourthquarter 2019.
[14] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7957–7969, Aug. 2019.
[15] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

VT