Thirty Years of Machine Learning:
The Road to Pareto-Optimal Wireless Networks

Jingjing Wang, *Member, IEEE*, Chunxiao Jiang, *Senior Member, IEEE*,
Haijun Zhang, *Senior Member, IEEE*, Yong Ren, *Senior Member, IEEE*,
Kwang-Cheng Chen, *Fellow, IEEE,* and Lajos Hanzo, *Fellow, IEEE*

Abstract—Future wireless networks have a substantial potential in terms of supporting a broad range of complex compelling applications both in military and civilian fields, where the users are able to enjoy high-rate, low-latency, low-cost and reliable information services. Achieving this ambitious goal requires new radio techniques for adaptive learning and intelligent decision making because of the complex heterogeneous nature of the network structures and wireless services. Machine learning (ML) algorithms have great success in supporting big data analytics, efficient parameter estimation and interactive decision making. Hence, in this article, we review the thirty-year history of ML by elaborating on supervised learning, unsupervised learning, reinforcement learning and deep learning. Furthermore, we investigate their employment in the compelling applications of wireless networks, including heterogeneous networks (HetNets), cognitive radios (CR), Internet of things (IoT), machine to machine networks (M2M), and so on. This article aims for assisting the readers in clarifying the motivation and methodology of the various ML algorithms, so as to invoke them for hitherto unexplored services as well as scenarios of future wireless networks.

Index Terms—Machine learning (ML), future wireless network, deep learning, regression, classification, clustering, network association, resource allocation.

NOMENCLATURE

5G	The 5th Generation Mobile Network
AI	Artificial Intelligent
AMC	Automatic Modulation Classification
ANN	Artificial Neural Network
AP	Access Point
AWGN	Additive White Gaussian Noise
BBU	BaseBand processing Unit
BS	Base Station
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Network
CogNet	Cognitive Network
CoMP	Coordinated Multiple Points
CR	Cognitive Radio
C-RAN	Cloud Radio Access Network
CRN	Cognitive Radio Network
CSI	Channel State Information
CSMA/CA	Carrier-Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier-Sense Multiple Access with Collision Detection
C-S Mode	Client-Server Mode
D2D	Device to Device
DBN	Deep Belief Network
DNN	Deep Neural Network
DQN	Deep Q-Network
EA	Energy Awareness
EE	Energy Efficiency
EH	Energy Harvesting
ELP	Exponentially-weighted algorithm with Linear Programming
EM	Expectation Maximization
eMBB	enhanced Mobile Broad Band
ERM	Empirical Risk Minimization
EXP3	EXPonential weights for EXPloration and EXPloitation
FANET	Flying Ad Hoc Network
FDA	Fisher Discriminant Analysis
FDI	False Data Injection
FSMC	Finite State Markov Channel
GMM	Gaussian Mixture Model
HetNet	Heterogeneous Network
HMM	Hidden Markov Model
ICA	Independent Component Analysis
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
ITS	Intelligent Transportation System
KNN	K-Nearest Neighbors
LED	Light Emitting Diode
LOS	Line of Sight
LS	Least Square
LSTM	Long Short Term Memory

J. Wang and Y. Ren are with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China. E-mail: chinaeephd@gmail.com, reny@tsinghua.edu.cn.

C. Jiang is with Tsinghua Space Center, Tsinghua University, Beijing, 100084, China. E-mail: jchx@tsinghua.edu.cn.

H. Zhang is with Institute of Artificial Intelligence, Beijing Advanced Innovation Center for Materials Genome Engineering, Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, University of Science and Technology Beijing, Beijing 100083, China. E-mail: haijunzhang@ieee.org.

K.-C. Chen is with the Electrical Engineering Department, University of South Florida, Tampa, FL 33620, USA. Email: kwangcheng@usf.edu.

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK. Email: lh@ecs.soton.ac.uk.

LTE	Long Term Evolution
M2M	Machine to Machine
MANET	Mobile Ad Hoc Network
MAP	Maximum a Posteriori
MDP	Markov Decision Process
MIMO	Multiple-Input and Multiple-Output
ML	Machine Learning
MLE	Maximum Likelihood Estimation
mMTC	massive Machine Type of Communication
NB-IoT	NarrowBand Internet of Things
NB-M2M	NarrowBand Machine to Machine
NFV	Network Function Virtualization
NLOS	Non-Line of Sight
NOMA	Non-Orthogonal Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OSPF	Open Shortest Path First
P2P	Peer to Peer
PCA	Principal Component Analysis
POMDP	Partially Observable Markov Decision Process
PU	Primary User
QoE	Quality of Experience
QoS	Quality of Service
RAT	Radio Access Technology
RBM	Restricted Boltzmann Machine
RBF	Radial Basis Function
RFID	Radio Frequency IDentification
RNN	Recurrent Neural Network
RRU	Remote Radio Unit
SDA	Stacked Denoising Auto-encoder
SDN	Software Defined Network
SDR	Software Defined Radio
SE	Spectrum Efficiency
SG	Stochastic Geometry
SRM	Structural Risk Minimization
STBC	Space Time Block Code
SU	Secondary User
SVM	Support Vector Machine
TAS	Transmit Antenna Selection
TCP	Transmission Control Protocol
TD	Temporal Difference
TOA	Time of Arrival
UAV	Unmanned Aerial Vehicle
UDN	Ultra Dense Network
uRLLC	ultra-Reliable Low-Latency Communication
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
V2X	Vehicle to Everything
VANET	Vehicular Ad Hoc Network
VLC	Visible Light Communication
VR	Virtual Reality
WANET	Wireless Ad Hoc Network
WBAN	Wireless Body Area Network
WLAN	Wireless Local Area Network
WiMAX	Worldwide Interoperability for Microwave Access
Wi-Fi	Wireless Fidelity
WMAN	Wireless Metropolitan Area Network
WPAN	Wireless Personal Area Network
WSN	Wireless Sensor Network
WWAN	Wireless Wide Area Network

I. INTRODUCTION

WIRELESS networks have supported a variety of military services, intelligent transportation, healthcare, etc. To elaborate briefly, next-generation mobile networks are expected to support high date rate communication [1]. As a complement, wireless sensor networks (WSN) support sustained monitoring in unmanned or hostile environments relying on widely dispersed operating sensors [2]. Furthermore, the popular Wi-Fi network provides convenient Internet access for various devices in indoor scenarios [3]. With the rapid proliferation of portable mobile devices and the demand for a high quality of service (QoS) and quality of experience (QoE), future wireless networks will continue to support a broad range of compelling applications, where the users benefit from high-rate, low-latency, low-cost and reliable information services.

A. Motivation

In contrast to the conventional wireless networks, future wireless networks have the following evolutionary tendency [4], [5]:

- *Network Scale*: The network is associated with a tremendous network size including all kinds of entities, each of which has different service capabilities as well as requirements. Furthermore, interactions among these entities result in a diverse variety of traffic, such as text, voice, audio, images, video, etc.
- *Network Structure*: On one hand, the future wireless network tends to have a self-configuring element, where each entity cooperatively completes tasks. This characteristic is termed as "being ad hoc". On the other hand, the future wireless network is heterogeneous and hierarchical, having different network slices[1]. Furthermore, the mobility of entities results in a complex time-variant network structure, which requires dynamic time-space association.
- *Network Control*: Future wireless networks facilitate convenient reconfiguration by software-based network management, hence improving network flexibility and efficiency.

Machine learning (ML) was first introduced as a popular technique of realizing artificial intelligence in the late 1950's [6]. ML algorithms can learn from training data without being explicitly programmed. It is beneficial for classification/regression, prediction, clustering and decision making [7]–[9], whilst relying on the following three basic elements [10]:

- *Model*: Mathematical or signal models are constructed from training data and expert knowledge, in order to statistically describe the characteristics of the given data set. Then again, relying on these trained models, ML can be used for classification, prediction and

---

[1]In our paper, network slices are multiple logical networks running on the top of a shared physical network infrastructure and operated by a control center.

decision making. In case the appropriate models are not available, techniques on the feature extraction or knowledge discovery can be developed to achieve the same goal.

- *Strategy*: The criteria used for training mathematical models are called strategies. How to select an appropriate strategy is closely associated with training data. Empirical risk minimization [11] and structural risk minimization [12] constitute a pair of fundamental strategies, where the latter can beneficially avoid the notorious "over-fitting" phenomenon.

- *Algorithm*: Algorithms are constructed to find solutions based on predetermined model and strategy selected, which can be viewed as an optimization process. A powerful algorithm can find a globally optimal solution with high probability at a low computational complexity and storage.

In the last thirty years, ML has been successfully applied to the field of computer vision [13], automatic control [14], bioinformatics [15], etc. Considering the aforementioned characteristics of future wireless networks, data-driven ML is also likely to become a powerful technique of network association for substantially improving the network performance. This is achieved by accurately learning the near-real-time physical operating scenario, which allows them to outperform the traditional model-driven optimization algorithms based on more-restrictive assumptions detailed in [16]. More specifically,

- The wireless data torrance may be conveniently managed by the *big data processing capability* of ML [17]. For example, the tele-traffic volume generated by on-demand information and entertainment is predicted to substantially increase over the next decade, and an average smart phone may generate as much as 4.4 GB data per month by the year 2020 [18]–[20]. The massive amount of data constitutes a large training set, which can be statistically exploited for data-mining as well as for classification and for prediction with the aid of ML algorithms.

- Future wireless networks require both individual node intelligence and swarm intelligence [21]. Moreover, as for resource allocation and management, we tend to strike a trade-off among numerous factors, such as the capacity, power consumption, latency, complexity, etc. rather than only considering a single-component objective function. Thanks to learning from trial and error experiments, ML is conducive to supporting *intelligent multi-objective decision making* in the context of multi-agent collaborative network management. Future wireless networks may hence be expected to benefit from intelligent multi-agent systems.

- *Modeling and parameter estimation* play an important role in wireless networks. For instance, in massive multiple-input and multiple-output (MIMO) systems, an accurate estimate of the channel state information (CSI) potentially allows us to approach the system's capacity. Given that traditional mathematical models often fail to accurately describe the system in typical time-varying scenarios, ML provides an alternative technique of adaptive modeling and parameter estimation relying on learning from the recorded history.

- Future wireless networks are also expected to take into account the *human behavior*, for example by taking into account the geographic position of access points (AP) in an ultra dense network (UDN), where user-centric designs have been conceived for reducing the cluster-edge effects. By mimicking human intelligence, ML may be deemed to be the most appropriate tool for adapting the network's structure to the human behavior observed [22], [23].

Next-generation wireless network optimization relying on ML has emerged as an important research topic, so much so that the standard body ITU-T has formed a dedicated focus group to study this subject from 2018 to 2020. When incorporating ML functionalities into the network architecture, there are two fundamental mechanisms of exploiting ML algorithms, namely online and offline ML. The online ML family represents ML functionalities embedded into networking algorithms or protocols. By contrast, offline ML may be executed by a co-located computing facility connected to the corresponding network entities. However, offline ML can also be supported by remote computing facilities.

In recent years, a range of surveys have been conceived on ML paradigms. Some of them focused their scope on a specific wireless scenario, such as WSNs [24], [25], cognitive radio networks (CRN) [26]–[28], Internet of Things (IoT) [29], wireless ad hoc networks (WANET) [30], self-organizing cellular networks [31], etc. Specifically, Alsheikh *et al.* [24] provided an extensive overview of ML methods applied to WSNs which improved the resource exploitation and prolonged the lifespan of the network. Kulkarni *et al.* [25] surveyed some common issues of WSNs solved by computational intelligence algorithms, such as data fusion, routing, task scheduling, localization, etc. Moreover, Bkassiny *et al.* [26] investigated decision-making and feature classification problems solved by both centralized and decentralized learning algorithms in CRN in a non-Markovian environment. Gavrilovska *et al.* [27] studied the nature of the CRN's capability of reasoning and learning. Park *et al.* [29] reviewed a range of learning aided frameworks designed for adapting to the heterogeneous resource-constrained IoT environment. Forster [30] portrayed the advantages of using ML for the data routing problem of WANETs. Furthermore, a detailed literature review of the past fifteen years of ML techniques applied to self-configuration, self-optimization and self-healing, was provided by Klaine *et al.* [31].

Some of the literature were restricted to a specific application [32]–[38], whilst others considered a single learning technique [39]–[44]. To elaborate, Al-Rawi *et al.* [32] presented an overview of the features, methods and performance enhancement of learning-assisted routing schemes in the context of distributed wireless networks. Additionally, Fadlullah *et al.* [33] provided an overview of the state-of-the-art in learning aided network

TABLE I: The Topics of Survey Papers on Different ML Paradigms in Wireless Networks

Application-oriented	ML-oriented
[24],[29] resource allocation	[27] capability of learning
[25],[30],[32] routing schemes	[45] supervised learning
[33] traffic control	[39] unsupervised learning
[26],[34],[37] traffic classification	[44] artificial neural networks
[35] intrusion detection	[40] reinforcement learning
[36] software defined networking	[41],[42],[43] deep learning
[31],[38] networking techniques	

traffic control schemes as well as in deep learning aided intelligent routing strategies, while Nguyen *et al.* [34] focused their attention on the ML techniques conceived for Internet traffic classification. ML and data mining assisted cyber intrusion detection were surveyed in [35], including the complexity comparison of each algorithm and a set of recommendations concerning the best methods applied to different cyber intrusion detection problems. Moreover, ML techniques applied to software defined networking (SDN) were investigated in [36], from the perspective of traffic classification, routing optimization, resource management, etc. Pacheco *et al.* [37] surveyed the ML techniques based on several steps to achieve traffic classification. Sun *et al.* [38] focused on the the recent advances of ML techniques in the MAC layer, network layer, and application layer.

As for exploring learning techniques, Usama *et al.* [39] provided an overview of the recent advances of unsupervised learning in the context of networking, such as traffic classification, anomaly detection, network optimization, etc. Yau *et al.* [40] investigated the employment of reinforcement learning invoked for achieving context awareness and intelligence in a variety of wireless network applications such as data routing, resource allocation and dynamic channel selection. The authors of [41] and [42] focused their attention on the benefit of deep learning in wireless multimedia network applications, including ambient sensing, cyber-security, resource optimization, etc. Mao *et al.* [43] provided a comprehensive survey of the applications of deep learning algorithms in terms of different network layers, including physical layer, data link layer and routing layer. Additionally, Chen *et al.* [44] overviewed the artificial neural networks based ML algorithms conceived for various wireless networking problems. The main contributions of the existing ML aided wireless networks survey and tutorial papers are contrasted in Fig. 1 and Table I to this survey.

B. Contributions

Hence, our focus is on the comprehensive survey of ML aided wireless networks. Inspired by above-mentioned challenges, in this article we review the development of ML aided wireless networks. We commence by investigating a series of popular learning algorithms and their compelling applications in wireless networks and then provide some specific examples based on some recent research results, followed by a range of promising open issues in the design of future networks. Our original contributions are summarized as follows:

- We critically review the thirty-year history of ML. Depending on how we use training data, we classify ML algorithms into three categories, i.e. supervised learning [45], unsupervised learning [46] and reinforcement learning [47]. In addition, we highlight the family of deep learning algorithms, given their success in the field of signal processing.
- The development of wireless networks is reviewed from their birth to the future wireless networks. Moreover, we summarize the evolution of wireless networking techniques, and characterize a variety of representative scenarios for future wireless networks.
- We appraise a range of typical supervised, unsupervised, reinforcement learning as well as deep learning algorithms. Moreover, their compelling applications in wireless networks are surveyed for assisting the readers in refining the motivation of ML in wireless networks, all the way from the physical layer to the application layer.
- Relying on recent research results, we highlight a pair of examples conceived for wireless networks, which can help the readers to gain the insight into hitherto unexplored scenarios and into their applications in wireless networks.

C. Organization

The remainder of this article is outlined as follows. In Section II, we provide a brief overview of the history of ML and of the development of wireless networks. In Section III, we introduce a range of typical supervised learning algorithms and highlight their compelling applications in wireless networks. In Section IV, we investigate the family of unsupervised learning algorithms and their related applications. Some popular reinforcement learning algorithms are elaborated on in Section V. Moreover, we present two examples of how these reinforcement learning algorithms can improve the performance of wireless networks. In Section VI, we introduce some typical deep learning algorithms and their applications in wireless networks. Some future research ideas and our conclusions are provided in Section VII. The structure of this treatise is summarized at a glance in Fig. 2.

II. A Brief Overview of Machine Learning and Wireless Networks

A. The Thirty-Year Development of Machine Learning

The term "machine learning" was first proposed by Arthur Samuel in 1959 [6], which referred to computer systems having the capability of learning from their large amounts of previous tasks and data, as well as of self-optimizing computer algorithms. Hard-programmed algorithms are difficult to adapt to dynamically fluctuating demands and constantly renewed system states. By contrast, relying on learning from previous experiences, ML aided algorithms are beneficial for scientific decision making and task prediction, which is achieved by constructing a self-adaption model from sample inputs. To elaborate a little further, as for the concept of "learning", Tom M.

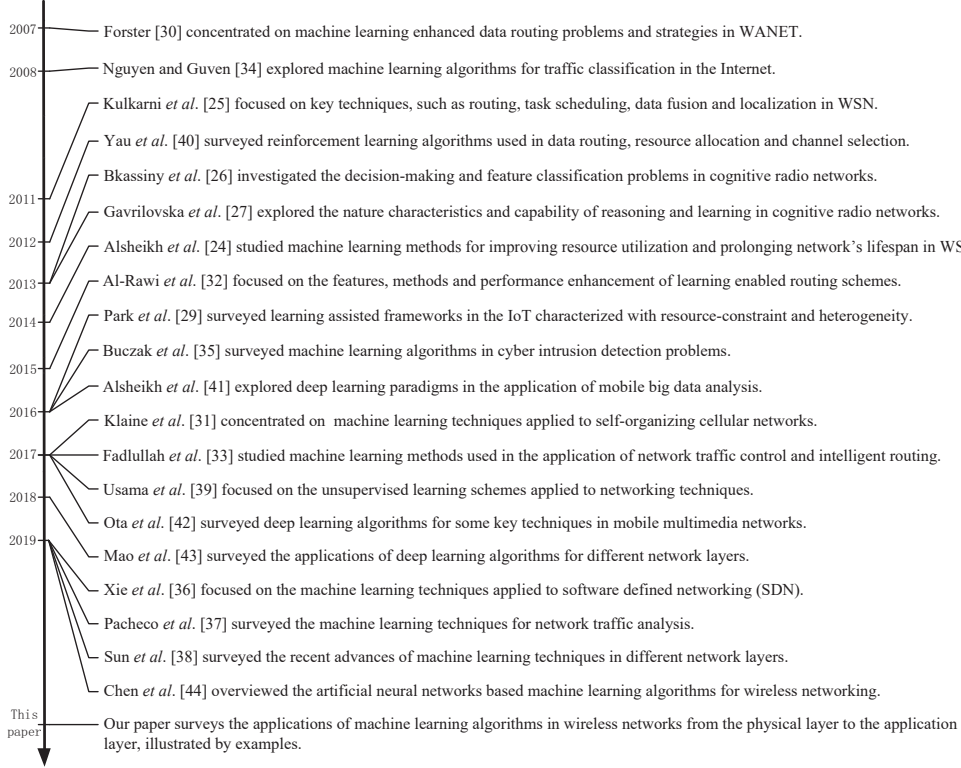|	
2007	Forster [30] concentrated on machine learning enhanced data routing problems and strategies in WANET.
2008	Nguyen and Guven [34] explored machine learning algorithms for traffic classification in the Internet.
	Kulkarni *et al*. [25] focused on key techniques, such as routing, task scheduling, data fusion and localization in WSN.
	Yau *et al*. [40] surveyed reinforcement learning algorithms used in data routing, resource allocation and channel selection.
	Bkassiny *et al*. [26] investigated the decision-making and feature classification problems in cognitive radio networks.
2011	Gavrilovska *et al*. [27] explored the nature characteristics and capability of reasoning and learning in cognitive radio networks.
2012	Alsheikh *et al*. [24] studied machine learning methods for improving resource utilization and prolonging network's lifespan in WSN.
2013	Al-Rawi *et al*. [32] focused on the features, methods and performance enhancement of learning enabled routing schemes.
2014	Park *et al*. [29] surveyed learning assisted frameworks in the IoT characterized with resource-constraint and heterogeneity.
2015	Buczak *et al*. [35] surveyed machine learning algorithms in cyber intrusion detection problems.
	Alsheikh *et al*. [41] explored deep learning paradigms in the application of mobile big data analysis.
2016	Klaine *et al*. [31] concentrated on machine learning techniques applied to self-organizing cellular networks.
2017	Fadlullah *et al*. [33] studied machine learning methods used in the application of network traffic control and intelligent routing.
2018	Usama *et al*. [39] focused on the unsupervised learning schemes applied to networking techniques.
2019	Ota *et al*. [42] surveyed deep learning algorithms for some key techniques in mobile multimedia networks.
	Mao *et al*. [43] surveyed the applications of deep learning algorithms for different network layers.
	Xie *et al*. [36] focused on the machine learning techniques applied to software defined networking (SDN).
	Pacheco *et al*. [37] surveyed the machine learning techniques for network traffic analysis.
	Sun *et al*. [38] surveyed the recent advances of machine learning techniques in different network layers.
	Chen *et al*. [44] overviewed the artificial neural networks based machine learning algorithms for wireless networking.
This paper	Our paper surveys the applications of machine learning algorithms in wireless networks from the physical layer to the application layer, illustrated by examples.

Fig. 1: The timeline of survey papers on the application of different ML paradigms in wireless networks.

Mitchell [48] provided the widely quoted description: "*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.*"

ML began to flourish in the 1990s [8]. Before this era, logic- and knowledge-based schemes, such as inductive logic programming, expert systems, etc. dominated the artificial intelligence scene relying on high-level human-readable symbolic representations of tasks and logic. Thanks to the development of statistics theory and stochastic approximation, ML schemes regained researchers' attention leading to a range of beneficial probabilistic models. Researchers embarked on creating date-driven programs for analyzing a large amount of data and tried to draw conclusions or to learn from the data. During this era, ML algorithms such as neural networks as well as kernel methods became mature. During the 2000s, researchers gradually renewed their interest in deep learning with the aid of the advances in hardware-based computational capability, which made ML indispensable for supporting a wide range of services and applications.

Given the development of progressive learning techniques [49], at present, the research focus of ML has shifted from "learning being the purpose" to "learning being the method". Specifically, ML algorithms no longer blindly pursue to imitate the learning capability of human beings, instead they focus more on the task-oriented intelligent data-driven analysis. Nowadays, thanks to the abundance of raw data and to the frequent interaction

between exploration and exploitation, ML algorithms have prospered in the fields of computer vision, data mining, intelligent control, etc. Future wireless networks aim for providing ubiquitous information services for users in a variety of scenarios. However, the rapid growth in the number of users and the resulted explosive growth of tele-traffic data pushes the limits of network-capacity. As a remedy, ML aided network management and control can be viewed as a corner stone of future wireless networks in view of their limited power, spectrum and cost.

B. Classifying Machine Learning Techniques

1) A General Taxonomy: Again, depending on how training data is used, ML algorithms can be grouped into three categories, i.e. supervised learning, unsupervised learning and reinforcement learning [50], [51]. In the following, we will provide a brief description of the three types of algorithms.

- *Supervised Learning*: The algorithms are trained on a certain amount of labeled data [45]. Both the input data and its desired label are known to the computer, resulting in a data-label pair. Their goal is to infer a function that maps the input data to the output label relying on the training of sample data-label pairs. Specifically, considering a set of N sample data-label pairs in the form of $\{(x_1, y_1), (x_2, y_2), \ldots (x_N, y_N)\}$, where x_n is the n-th sample input data and y_n represents its label. Let $\mathbb{X} = \{x_1, x_2, \ldots, x_N\}$ denote the input data set and $\mathbb{Y} = \{y_1, y_2, \ldots, y_N\}$ represent
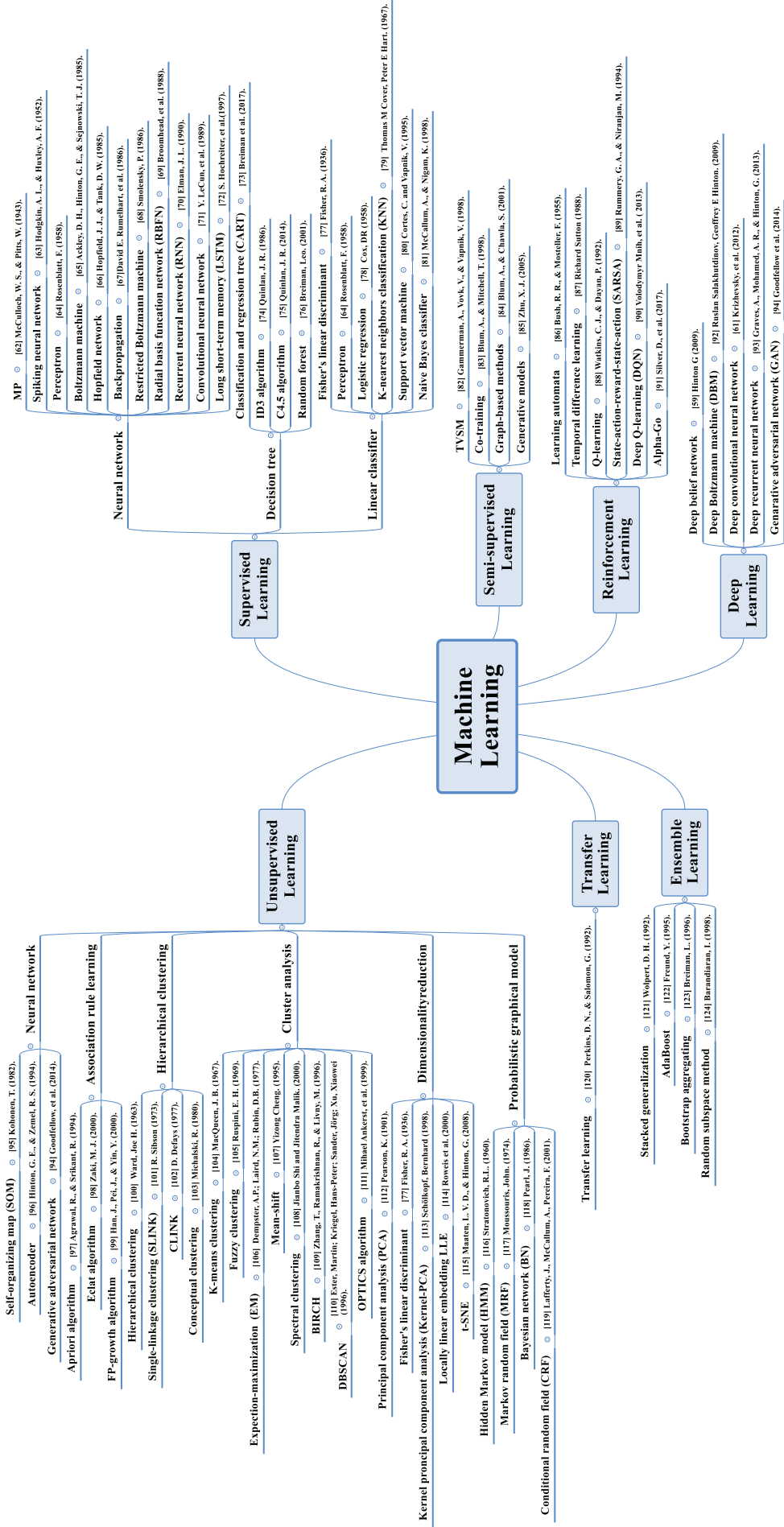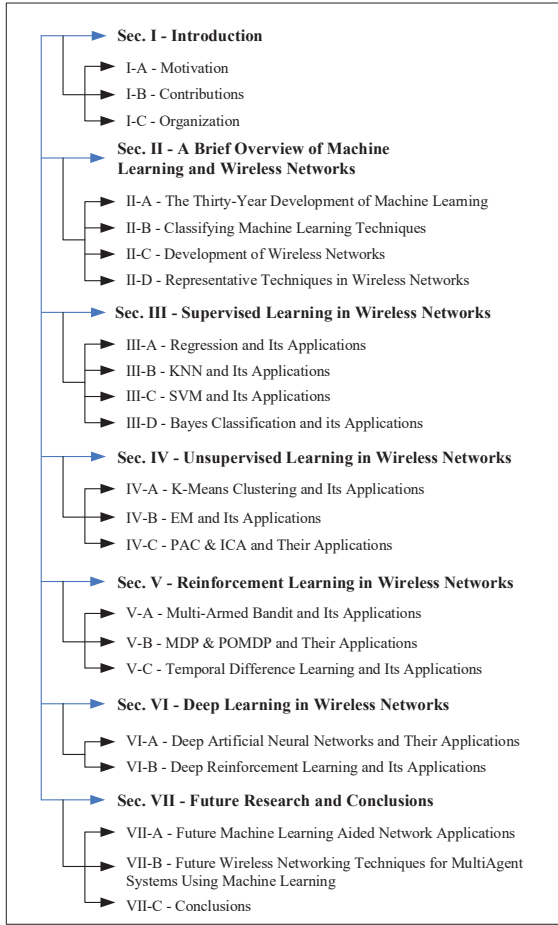
Fig. 3: A comprehensive survey of ML algorithms.

Fig. 2: The structure of this treatise.

the output label set. Usually, these sample pairs are independent and identically distributed (i.i.d.). The learning algorithms aim for seeking a function $g(x)$ that yields the highest value of the score function $f(x, y)$, hence we have $g(x) = \arg\max_{y} f(x, y)$. As a special case, if only part of the sample data-label pairs are known to the computer and some of the desired output labels of input data are missing, the corresponding learning algorithms are termed as semi-supervised learning[2]. These supervised learning algorithms can be widely used in the context of classification, regression and prediction.

- *Unsupervised Learning*: Relying on unlabeled input data, unsupervised learning algorithms try to explore the hidden features or structure of the data [46], [52]. Given the lack of sample data-label pairs, there is no standard accuracy evaluation for the output of unsupervised learning algorithms, which is the main difference compared to its supervised learning aided counterpart. By analyzing N input data $\mathbb{X} = \{x_1, x_2, \ldots, x_N\}$, a pair of popular methods has been conceived for revealing the underlying

unknown features of N input data, namely density estimation [53] as well as feature extraction [54]. To elaborate, density estimation aided methods are characterized by explicitly building statistical models of how the underlying features might create the input. By contrast, feature extraction based techniques aim for directly extracting statistical regularities or even sometimes irregularities from the input data set.

- *Reinforcement Learning*: In contrast to the aforementioned two learning techniques, reinforcement learning algorithms are conceived for decision making by learning from interaction with the environment, which are trained by the data on the basis of trial and error [47], [55]. They neither try to identify a category as supervised learning algorithms do, nor do they aim for finding hidden structures as unsupervised learning algorithms do. Specifically, at each time step, the system or environment is in some state S, and the agent selects a legitimate action A. The system responds at the next time step by moving into a new state S' with a certain probability influenced both by the specific action chosen as well as by the system's inherent transitions. Meanwhile, the agent receives a corresponding reward $r(S, A)$ from the system, as time evolves. Reinforcement learning algorithms aim for learning how to map situations S into actions A in order to attain the maximal cumulative weighted reward within the horizon in such a closed-loop fashion.

2) Deep Learning: As an important member of the ML family, *deep learning* has been booming since 2010, because it was found to be capable of handling the soaring growth of training data volume facilitated by the rapid development of computing hardware [56], [57]. Deep learning algorithms rely on a multiple-layer "network" consisting of inter-connected nodes for feature extraction and transformation, which is inspired by the biological nervous system, namely the neural network. Each layer utilizes the output of the previous layer as its input. The term "deep" refers to having multiple layers in the network. Generally, relying on the way the training data is exploited, deep learning algorithms can also be classified into deep supervised learning, deep unsupervised learning as well as deep reinforcement learning [56]. Moreover, some deep learning architectures, such as deep neural networks (DNN) [58], deep belief networks (DBN) [59], recurrent neural networks (RNN) with LSTM units [60] and convolutional neural networks (CNN) [61], have had success in a range of fields including computer vision, natural language processing (NLP), speech recognition, etc. They have also been invoked in compelling applications of wireless networks.

Fig. 3 comprehensively summarizes ML algorithms even including above-mentioned supervised learning, unsupervised learning, reinforcement learning and deep learning algorithms as well as some non-mainstream branches [62]–[124]. Furthermore, Fig. 4 shows the involvement of ML in wireless networks based on the aforementioned four categories. Below we list a variety

---

[2]In this paper, semi-supervised learning algorithms are viewed as a specific category of supervised learning algorithms. However, in some of the literature, semi-supervised learning is listed as a separate member of the ML family

Fig. 4: The categories, characteristics and applications of ML algorithms.

of popular learning algorithms and highlight their applications in future wireless networks.

C. Development of Wireless Networks

1) The Birth and Development of Wireless Networks: Just as the terminology implies, wireless networks connect various network nodes via electromagnetic waves. Relying on their coverage, wireless networks can be roughly classified into four categories, such as wireless personal area networks (WPAN) [125], wireless local area networks (WLAN) [126], wireless metropolitan area networks (WMAN) [127] and wireless wide area networks (WWAN) [128]. Correspondingly, a family of networking standards and their variants that cover most of the physical layer specifications have been established by the IEEE 802 Working groups, including the IEEE 802.15 for W-PAN, IEEE 802.11 for WLAN, IEEE 802.16 for WMAN and IEEE 802.20 for WWAN standards. Furthermore, when considering the network's functions, some popular representatives of wireless networks include cellular networks [129], WSNs [130], WANETs [131], wireless body area networks (WBAN) [132], etc.

The first wireless network, namely ALOHANET, was developed at the University of Hawaii in 1969 and came into operation in 1971, which for the first time transmitted wireless data packets over a network [133]. The first commercial wireless network was the WaveLAN product family designed by the NCR Corporation in 1986. In 1997, the first IEEE 802.11 protocol was released for WLAN [126]. Afterwards, the emergence and progress of reliable and low-cost Wi-Fi marked the maturity of wireless networking technologies at the end of the 20th century, which facilitated Internet access for a range of Wi-Fi compatible devices including personal computers, smart phones, etc. Future wireless networks aim for providing high-rate, low-latency, full-coverage and low-cost yet reliable information services. Compared to traditional wireless networks connecting humans and their devices, future
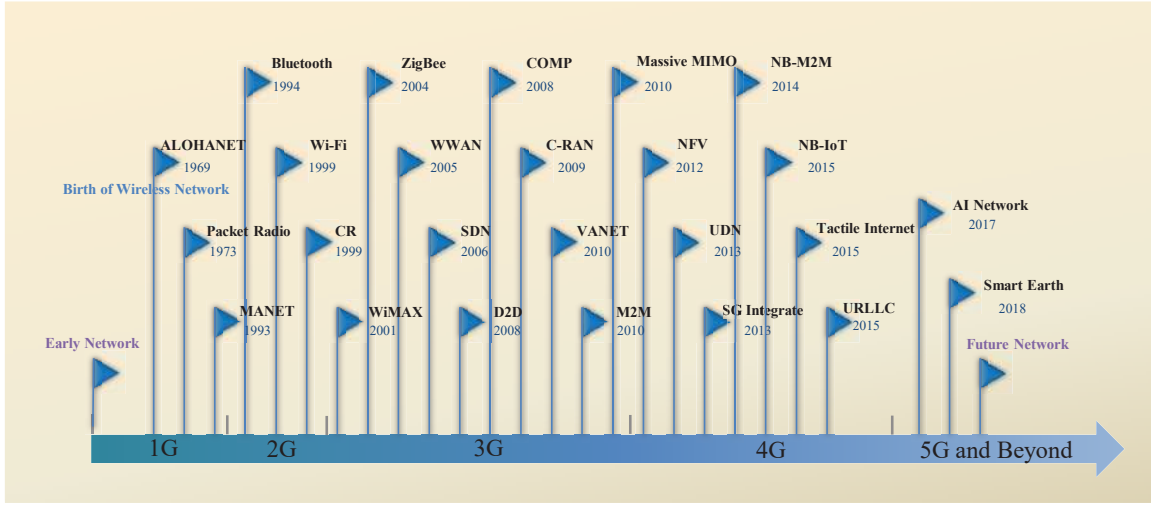
Fig. 5: The development of wireless networks interspersed with ever-accelerated milestone techniques.

wireless networks are expected to interconnect everything under the umbrella of the 'Internet of Everything'. Fig. 5 demonstrates the development of wireless networks in terms of their milestone techniques.

Wireless networks have evolved from the simple client-server (C-S) mode to the distributed dense multi-layer C-S mode, and finally to the ad hoc peer-to-peer (P2P) mode. The decentralization of network architectures grant more freedom both for the network nodes and their protocols, which requires more sophisticated techniques for supporting efficient and reliable implementations. Furthermore, the soaring growth of both the type and the amount of data provides a promising field of applications for ML algorithms, which are beneficial for self-organized and self-adaptive network architectures.

2) Pareto-Optimal Future Wireless Networks: The challenging real-world optimization problems encountered in future wireless networks have to meet multiple objectives in order to arrive at an attractive solution [134]. In contrast to conventional single-objective optimization where we find the global optimum relying on a single metric, multi-objective optimization aims for finding the globally optimal solution relying on the notion of Pareto optimality [135]. The aim of multi-objective optimization in wireless networks is that of generating a diverse set of Pareto-optimal solutions, where by definition it is only possible to improve any of the metrics considered at the cost of degrading at least one of the others. The collection of Pareto-optimal points is referred to as the Pareto front. Fig. 6 shows a simple Pareto optimization problem having a pair of objective parameters, where the light blue circles represent legitimate operating points, while the dark blue circles denote the approximated Pareto optimal points, which are not dominated by any other solutions. The arrow intimates that the complexity of optimization is gradually increased as we gradually approach the full-search complexity, which defines the 'true Pareto front' of Fig. 6.

More specifically, let us briefly reminisce by recalling the past few decades of wireless history. Explicitly, the

wireless community has invested decades of research efforts into making near-capacity single-user operation a reality [136], which is however only possible at the cost of an ever-increasing delay, complexity and power consumption. When designing a powerful wireless network, which includes the physical layer, MAC layer, network layer, transmission layer as well as application layer, we face substantial challenges, since we often have to meet conflicting design objectives. Moreover, these conflicting optimization metrics are often coupled with each other. It may also be a substantial challenge to carry out a fair comparison among locally optimal solutions, hence we often strike a tradeoff [137]. For example, in future wireless networks we would like to be more ambitious than 'only' optimize the network's capacity - for delay-sensitive services we would like to reduce the latency and/or reduce the total energy consumption, as well as to improve the system's reliability and the user's QoS. By contrast, in wireless sensor networks we may concentrate on optimizing both the connectivity and the network's life time, just to name a few of these conflicting practical design objectives.

In this context, the family of ML techniques may be viewed as an attractive set of optimization tools for finding Pareto-optimal solutions of multi-objective optimization problems in future wireless networks, which tend to have a large search-space. To expound a little further, it is plausible that every time we incorporate an additional parameter into the objective function, the search-space is expanded and the surface of optimal solutions may exhibit numerous locally optimal solutions. Hence traditional gradient-based techniques routinely fail to find the global optimum. In this context Fig. 7 portrays some popular metrics commonly used in constructing multi-objective optimization problems in future wireless networks.

D. Representative Techniques in Wireless Networks

As shown in Fig. 8, we first of all portray the representative application scenarios and techniques of future wireless networks. In the following, we will briefly introduce a
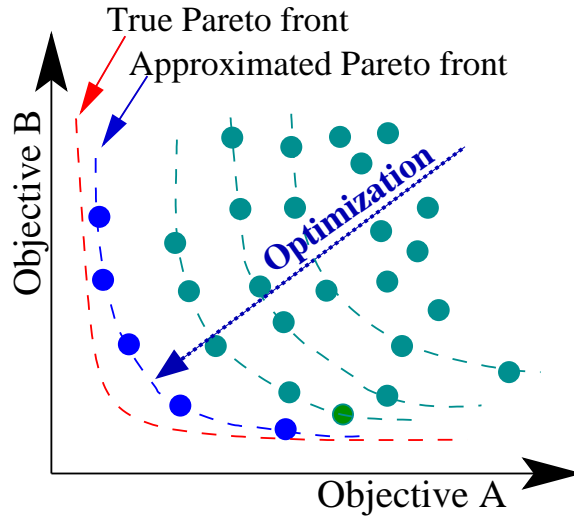
Fig. 6: A simple example of a two-parameter Pareto-optimization problem.
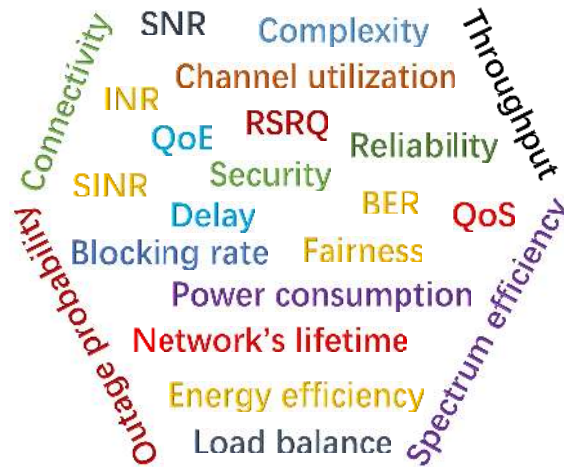

Fig. 7: Useful metrics commonly used in constructing multi-objective optimization problems.

range of compelling techniques and their development trends in future wireless networks, which is summarized in Fig. 9.

1) From MIMO to Massive MIMO: The MIMO technology relying on multiple antennas in both the transmitter and receiver can be viewed as a breakthrough in terms of multiplying the capacity of a radio link compared to the single-transmit single-receive antenna aided wireless system having a variety of cost, technology and regulatory constraints [138]. Both single-user MIMO (SU-MIMO) and Multi-user MIMO (MU-MIMO) schemes have been proposed. To elaborate, multiple data streams of the same source are sent to a single user in SU-MIMO, while a transmitter simultaneously serves multiple users on the same channel resource in MU-MIMO [139], [140].

Generally, invoking more antennas is beneficial in terms of improving the data rate and/or link reliability [4], [141]–[143]. However, the performance degradation caused by inaccurate CSI and the high computational complexity of channel estimation constitute substantial challenges [144].

2) From D2D, M2M to IoT: In the spirit of direct communication between nearby mobile devices without traversing base stations (BS) or core networks, device-to-device (D2D) communication networks have been widely investigated in recent years, which can be deemed to be important milestones on the road towards self-organization and P2P collaboration. In D2D networks, the same resource slots can be reused both by the D2D links as well as the cellular links, which is capable of substantially improving the network capacity. Moreover, it is potentially beneficial in terms of enhancing the energy efficiency (EE), also reducing the transmission delay and improving the network's fairness to users [145], [146], which is also closely related to machine-to-machine (M2M) communications. The corresponding massive machine type of communication (mMTC) [147] mode of the 5G network[3] is capable of supporting sensing, transmitting, fusion and processing sensory data. Furthermore, M2M is also capable of supporting the smart home [148], smart grid [149], etc.

[3]In 2015, International Telecommunication Union (ITU) officially defined three application scenarios of 5G network, i.e. enhanced mobile broad band (eMBB), massive machine type communication (mMTC) and ultra reliable low latency communication (uRLLC).
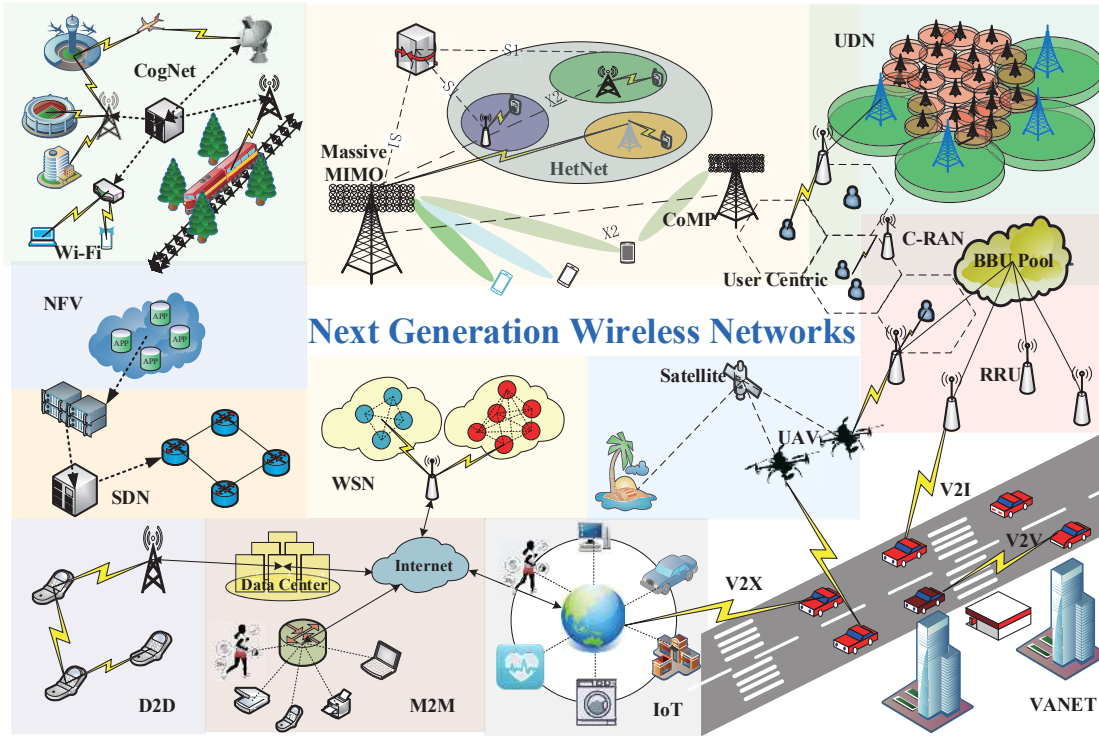
Fig. 8: The representative application scenarios and related techniques of future wireless networks.

Aiming for "connecting everything", IoT was first defined for enabling objects to connect and exchange data in 1999 [150]. Furthermore, the IoT allows objects to be sensed and controlled remotely, creating opportunities for direct interaction between the physical world and computer-based virtual systems, which is beneficial in terms of improving operational efficiency and of reducing human intervention. Both WSNs and M2M communications can be viewed as a part of the IoT. Although the IoT faces a range of reliability, robustness and security challenges, there is no doubt that it will make our world ever smarter [151], [152].

3) From UDN to HetNet: In order to meet the demand of supporting massive data traffic, the so-called UDN architecture has been defined where the density of BSs or APs potentially reaches or even exceeds the density of users [153], [154]. The UDN architecture is conducive to increasing the network capacity as well as simultaneously improving the user experience. However, the interference encountered in UDNs tends to be more severe and of higher volatility than that in traditional cellular networks because of the dense deployment of BSs and APs. Hence, the joint consideration of resource allocation, interference management and traffic routing are essential for UDNs [129], [155].

Considering a wide area network scenario, heterogeneous networks (HetNet) are characterized by the employment of multiple types of radio access technologies (RAT) [156]. Upon combining macrocells, microcells, picocells [157] and femtocells [158], [159], HetNets are

capable of providing a seamless wireless coverage ranging from outdoor environments to office buildings and even to underground areas by selecting another RAT when a RAT fails, and HetNets can also provide load-balancing in the face of non-uniform spatial distribution of users [160].

4) From DBS to C-RAN: Compared to the traditional BS, which integrates baseband processing units (BBU) and remote radio units (RRU)[4] in a single cabinet, distributed base station (DBS) aided systems separate the BBU as well as the RRU and connects them with optical fiber. The DBS system allows more flexibility in network planning and deployment, where RRUs can be placed a few hundred meters or a few kilometres away for enhancing network's edge-coverage.

Cloud-radio access networks (C-RAN) can be viewed as an evolution of the aforementioned DBS system, which is a centralized processing and cloud computing aided radio access network architecture [161]. The principle of C-RAN relies on gathering the BBUs from several BSs into a centralized BBU pool, whilst allowing hundreds of RRUs to connect to the centralized BBU pool [162]. Hence, resources can be allocated to each user based on joint dynamic scheduling. By exploiting coordination and virtualization, the spectral efficiency (SE), the system's flexibility and the load balancing capability are substantially improved. Moreover, the centralized management of resources reduces the cost of the system's operation and maintenance.

---

[4]In some works, RRU is also called remote radio head (RRH)

5) From SDN to NFV: Software-defined networking (S-DN) is employed as a programmable network architecture in order to achieve cost-effective dynamic network configuration and monitoring [163], [164]. The SDN philosophy suggests to centralize network intelligence in a single network component by decoupling the control plane and the data plane, which disassociates network control and its forwarding functions. The two planes can communicate with the aid of the OpenFlow protocol[5], and the network resources can be managed logically and efficiently. A SDN connects decentralized users to cloud computing through a "network pipeline" [165] [166].

Relying on IT virtualization techniques, network function virtualization (NFV) transforms the entire set of network node functions into different building blocks, which separates the networking functions from specific hardware blocks [167]. Hence, NFV is eminently suitable for service diversification and promotes the standardization of networking equipment [168]. Explicitly, NFV can be viewed as a beneficial hardware-agnostic design in the application layer of SDN architectures.

6) From EH to EA: Energy harvesting (EH) is an environmentally friendly process, which captures and stores ambient energy, such as solar power, thermal energy, wind energy, etc. for low-power wireless devices [169], especially in WSNs and WBANs, for example.

In future wireless networks, energy optimization is a significant concern motivated by mitigating climate change. However, energy consumption is related to both the network's throughput and to its entire lifetime with a trade-off between them. As a remedy, instead of only focusing on EH, energy awareness (EA) at every stage of the network's design and management is the most promising approach to striking a trade-off amongst the conflicting objectives of reducing energy consumption, improving the system's throughput as well as prolonging its lifetime, especially in energy-constrained networks [170], [171].

7) From CR to CogNet: Cognitive radio (CR) constitutes a technique that allows us to dynamically and efficiently exploit the wireless spectral resources [172]–[175]. By relying on spectrum sensing, CR is capable of achieving dynamic spectrum access and spectrum sharing. Specifically, in the process of spectrum sensing, the secondary user (SU) detects an empty slicer of spectrum, for example, based on energy detection schemes. Then, in the process of spectrum access, power control is invoked by the SU for maximizing its capacity, whilst observing the interference power constraint in order to protect the primary user (PU). As a benefit, CR dynamically and flexibly exploits the scarce wireless spectral resources, hence substantially improving the spectrum efficiency [176].

In contrast to CR techniques, which only deal with the issues of physical-layer spectrum sensing and data link-layer access, cognitive networks (CogNet) are characterized by a cognitive cross-layer process according to their end-to-end goals, where the overall network conditions are monitored, and then decisions are made based on the perceived conditions as well as on the feedback and experience gleaned from previous actions [177]. The network's cognitive capability relies on a range of advanced techniques, such as knowledge representation and ML, which exploit a wealth of information generated within the network improving both the network management, the resource efficiency [178] and the energy efficiency [179].

8) Interference Management: Interference constitutes the fundamental limiting factor of the overall wireless system performance, hence it is a key challenge faced by designers. Therefore substantial efforts have been dedicated to exploiting the communication channel's state information (CSI) either at the transmitter (CSIT) or at the receiver (CSIR) for mitigating the effects of interference. Hence diverse time/frequency/space division multiple access based resource allocation schemes have been conceived for avoiding interference by creating orthogonal resource units [180]–[182]. Creative efforts have also been dedicated to the conception of non-orthogonal access systems, as exemplified by a large variety of cognitive radio [183] and non-orthogonal multiple access (NOMA) schemes [184] relying on sophisticated transceiver designs. Additionally, multi-antenna based techniques, such as joint/partial pre/postcoding and antenna selection, have also been proposed for ameliorating the effects of interference by exploiting the benefits of spatial diversity [185].

A closely related issue in future wireless networks is interference management, which is a particularly critical task in ultra-dense networks in the face of their stringent throughput, delay and reliability specifications. Hence sophisticated resource allocation and interference management schemes are required. Therefore a range of ML algorithms have also been invoked for interference management relying on their environmental awareness and learning capability [186]–[188].

III. SUPERVISED LEARNING IN WIRELESS NETWORKS

Having covered the networking basics, in this section, we will introduce some rudimentary supervised learning algorithms, such as regression, K-nearest neighbors (KNN), support vector machines (SVM) and Bayes classification including their applications in wireless networks. Table II summarizes some of the typical applications of the above-mentioned four supervised learning algorithms in wireless networks.

A. Regression and Its Applications

1) Methods: Regression analysis is capable of estimating the relationships among variables. Relying on modeling the functional relationship between a dependent variable (objective) and one or more independent variables (predictors), regression constitutes a powerful statistical tool of predicting and forecasting a continuous-valued objective given a set of predictors.

In regression analysis, there are three variables, namely the

- *Independent variables* (predictors): X
- *Dependent variable* (objective): Y

---

[5]The OpenFlow protocol is a communication protocol that gives access to the forwarding plane of a switcher or router over the network.
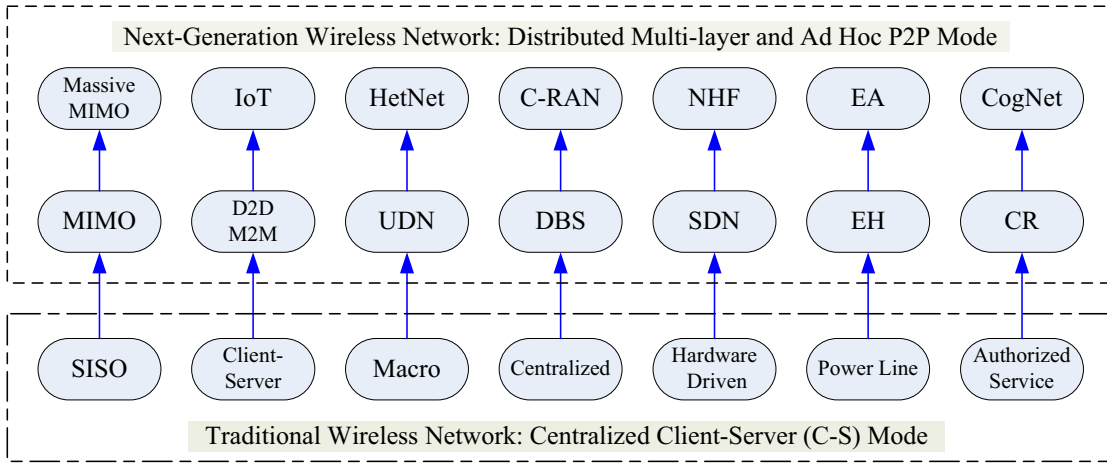
Fig. 9: The development trends of wireless networks.

- *Other unknown parameters* that affect the estimated value of the dependent variable: ε

The regression function f models the functional Y vs X relationship perturbed by ε, which can be formulated as: $Y = f(X, \varepsilon)$. Usually, we characterize this relationship in terms of a specific regression function with the aid of its probability distribution. Moreover, the approximation is often modeled as $E = [Y \mid X] = f(X, \varepsilon)$. When conducting regression analysis, first of all we have to determine the specific form of the regression function f, which relies on both the common knowledge about the dependent vs independent variables as well as on its convenient evaluation. Based on the specific form of regression function, regression analysis methods can be classified as ordinary linear regression [189], logistic regression [190], polynomial regression [191], etc.

In linear regression, the dependent variable is a linear combination of the independent variables or unknown parameters. Let us assume having N random training samples and M independent variables, formulated as $\{y_n, x_{n1}, x_{n2}, \ldots, x_{nM}\}, n = 1, 2, \ldots, N$. Then the linear regression function can be formulated as:

$$y_n = \varepsilon_0 + \varepsilon_1 x_{n1} + \varepsilon_2 x_{n2} + \cdots + \varepsilon_M x_{nM} + e_n, \quad (1)$$

where ε_0 is termed as the regression intercept, while e_n is the error term and $n = 1, 2, \ldots, N$. Hence, Eq. (1) can be rewritten in the form of a matrix as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\varepsilon} + \boldsymbol{e}$, where $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^T$ is an observation vector of the dependent variable and $\boldsymbol{e} = [e_1, e_2, \ldots, e_N]^T$, while $\boldsymbol{\varepsilon} = [\varepsilon_0, \varepsilon_2, \ldots, \varepsilon_M]^T$ and $\boldsymbol{X}$ represents the observation matrix of independent variables, given by:

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1M} \\ 1 & x_{21} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{NM} \end{bmatrix}.$$

Linear regression analysis [189] aims for estimating the unknown parameter $\widehat{\boldsymbol{\varepsilon}}$ relying on the least squares (LS) criterion. The corresponding solution can be expressed as:

$$\widehat{\boldsymbol{\varepsilon}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}. \quad (2)$$

By contrast, in logistic regression [190], the dependent variable is binary. In order to facilitate our analysis, in the following we consider the case of a binary dependent variable, for example. The goal of the binary logistic regression is to model the probability of the dependent variable having the value of 0 or 1, given the training samples. To elaborate a little further, let the binary dependent variable y depend on M independent variables $\boldsymbol{x} = [x_1, x_2, \ldots, x_M]$. The conditional distribution of y under the condition of $\boldsymbol{x}$ obeys a Bernoulli distribution. Hence, the probability of $\Pr(y = 1 \mid \boldsymbol{x})$ can be expressed in the form of a standard logistic function[6], also termed as a sigmoid function:

$$P \triangleq \Pr(y = 1 \mid \boldsymbol{x}) = \frac{1}{1 + e^{-g(\boldsymbol{x})}}, \quad (3)$$

where $g(\boldsymbol{x}) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_M x_M$ and $\boldsymbol{w} = [w_0, w_1, \ldots, w_M]$ represents the regression coefficient vector. Similarly, we have:

$$\Pr(y = 0 \mid \boldsymbol{x}) = 1 - P = \frac{1}{1 + e^{g(\boldsymbol{x})}}. \quad (4)$$

Relying on the aforementioned definitions, we have $g(\boldsymbol{x}) = \ln(\frac{P}{1-P})$. Hence, for a given dependent variable, the probability of its value being y_n can be expressed by $P(y_n) = P^{y_n}(1 - P)^{1-y_n}$. Given a set of training samples $\{y_n, x_{n1}, x_{n2}, \ldots, x_{nM}\}, n = 1, 2, \ldots, N$, we are capable of estimating the regression coefficient vector $\boldsymbol{w} = [w_0, w_1, \ldots, w_M]$ with the aid of the maximum likelihood estimation (MLE) method. Explicitly, logistic regression can be deemed to form a special case of the generalized linear regression family using kernel model.

Furthermore, there exist numerous other useful regression models [191]–[194]. When the dependent variable is a polynomial function of the independent variables, we refer to it as polynomial regression [191], where the best-fit line is a curve. Moreover, ridge regression [192], least absolute shrinkage and selection operator (LASSO) regression [193] and ElasticNet regression [194] are widely

---

[6]The logistic function is a common "S" shape function, which is the cumulative distribution function (CDF) of the logistic distribution.

applied, when independent variables are of multi-collinear nature and highly correlated. Fig. 10 demonstrates the basic flow of a regression model.

2) Applications: The regression models can be used for estimating, detecting and predicting physical layer radio parameters related to wireless network scenarios. Specifically, Chang *et al.* [195] proposed a novel regression-aided interference model, which characterized the relationship between the SINR and the packet reception ratio, and evaluated its accuracy relying on the statistics. Based on this model, they constructed an analytic framework for striking a trade-off between the overhead imposed and the accuracy of interference measurement attained. In [196], Umebayashi *et al.* used regression analysis for formulating a deterministic-stochastic hybrid model for detecting the spectrum usage by PUs, which had a reduced number of parameters and yet maintained a high detection accuracy. In [197], Al Kalaa *et al.* used logistic regression for estimating the likelihood of Wi-Fi and ZigBee wireless co-existence in the context of medical devices. Furthermore, Xiao *et al.* [198] constructed a logistic regression-aided physical layer authentication model for detecting spoofing attacks in wireless networks without relying on a known channel model, which exhibited a high detection accuracy, despite its low computational complexity.

The regression models can also be employed for solving both estimation and detection problems in the upper layers of the seven-layer OSI model. For example, Chang *et al.* derived a regression-based analytical model for the sake of estimating the contention success probability considering heterogeneous sensor-traffic demands, which beneficially improved the channel's exploitation in IoT [199]. Moreover, in [200], Chen *et al.* employed a regression model for reconstructing the radio map with the aid of signal strength models for the path planning and UAV-location design in UAV-assisted wireless networks. As a further advance, Lei *et al.* [201] employed a logistic regression classifier for device-free localization relying on fingerprint signals, which yielded a low localization error.

B. KNN and Its Applications

1) Methods: KNN constitutes a non-parametric instance-based learning method, which can be used both for classification and regression. Proposed by Cover and Hart in 1968, the KNN algorithm is one of the simplest of all ML algorithms. By relying on the distance between the object and training samples in a feature space, the KNN algorithm determines which class of the object belongs to. Specifically, in a classification scenario, an object is categorized into a specific class by a majority vote of its K nearest neighbors. If $K = 1$, the category of the object is the same as that of its nearest neighbor. In this case, it is termed as the one nearest neighbour classifier. By contrast, in a regression scenario, the output value of the object is calculated by the average of the value of its K nearest neighbors. Fig. 11 shows the illustration of the unweighted KNN mechanism associated with $K = 4$.

Let us assume that there are N training sample pairs of $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)\}$, where y_n is the property value or class label of the sample $\boldsymbol{x}_n$, $n = 1, 2, \ldots, N$. Typically, we use the Euclidean distance or the Manhattan distance [202] for calculating the similarity between the object $\overline{\boldsymbol{x}}$ and the training samples. Let $\boldsymbol{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nM}]$ contain M different features. Hence, the Euclidean distance between $\overline{\boldsymbol{x}}$ and $\boldsymbol{x}_n$ can be expressed by:

$$d_e = \sqrt{\sum_{m=1}^{M} (\overline{x}_m - x_{nm})^2}, \qquad (5)$$

while their Manhattan distance is calculated as [202]:

$$d_m = \sum_{m=1}^{M} |\overline{x}_m - x_{nm}|. \qquad (6)$$

Relying on the associated similarity, the class label or property value of $\overline{\boldsymbol{x}}$ can be voted on or first weighted and then voted on by its K nearest neighbors, which is formulated:

$$\overline{y} \leftarrow \text{VOTE} \{K \text{ nearest } (\boldsymbol{x}_k, y_k)\}. \qquad (7)$$

The performance of the KNN algorithm critically depends on the value of K, whilst the best choice of K hinges upon the training samples. In general, a large K is conducive to resisting the harmful influence of noise, but it fuzzifies the class boundary between different categories. Fortunately, an appropriate value of K can be determined by a variety of heuristic techniques based on the true characteristics of the training data set.

2) Applications: In KNN, an object can be classified into a specific category by a majority vote of the object's neighbours, with the object being assigned to the class that is the most common one among its K nearest neighbors. Hence, as a kind of simple and efficient classification algorithms, KNN is beneficial in terms of, for example, traffic prediction [203], anomaly detection [204], [205], missing data estimation [206], modulation classification [207], interference elimination [208], etc.

To elaborate, for the sake of capturing the dynamic characteristics of wireless resource demands, Feng *et al.* constructed a weighted KNN model by learning from a large-scale historical data set generated by cellular operators' networks, which was used for exploring both the temporal and spatial characteristics of radio resources [203]. In [204], Xie *et al.* proposed a novel KNN aided online anomaly detection scheme based on hypergrid intuition in the context of WSN applications for overcoming the 'lazy-learning' problem [209] especially when the computational resource and the communication cost quantified in terms of bandwidth and energy were constrained. Moreover, in [205], Onireti *et al.* proposed a KNN based anomaly detection algorithms for improving the outage detection accuracy in dense heterogeneous networks. As for missing data estimation, a KNN assisted missing data estimation algorithm was conceived on the basis of the temporal and spatial correlation feature of sensor data, which jointly utilized the sensor data from multiple neighbor nodes [206]. Furthermore, Aslam *et al.* [207] combined genetic programming and the KNN

(a) Raw data

(b) Build prediction functions based on the cost funtcion

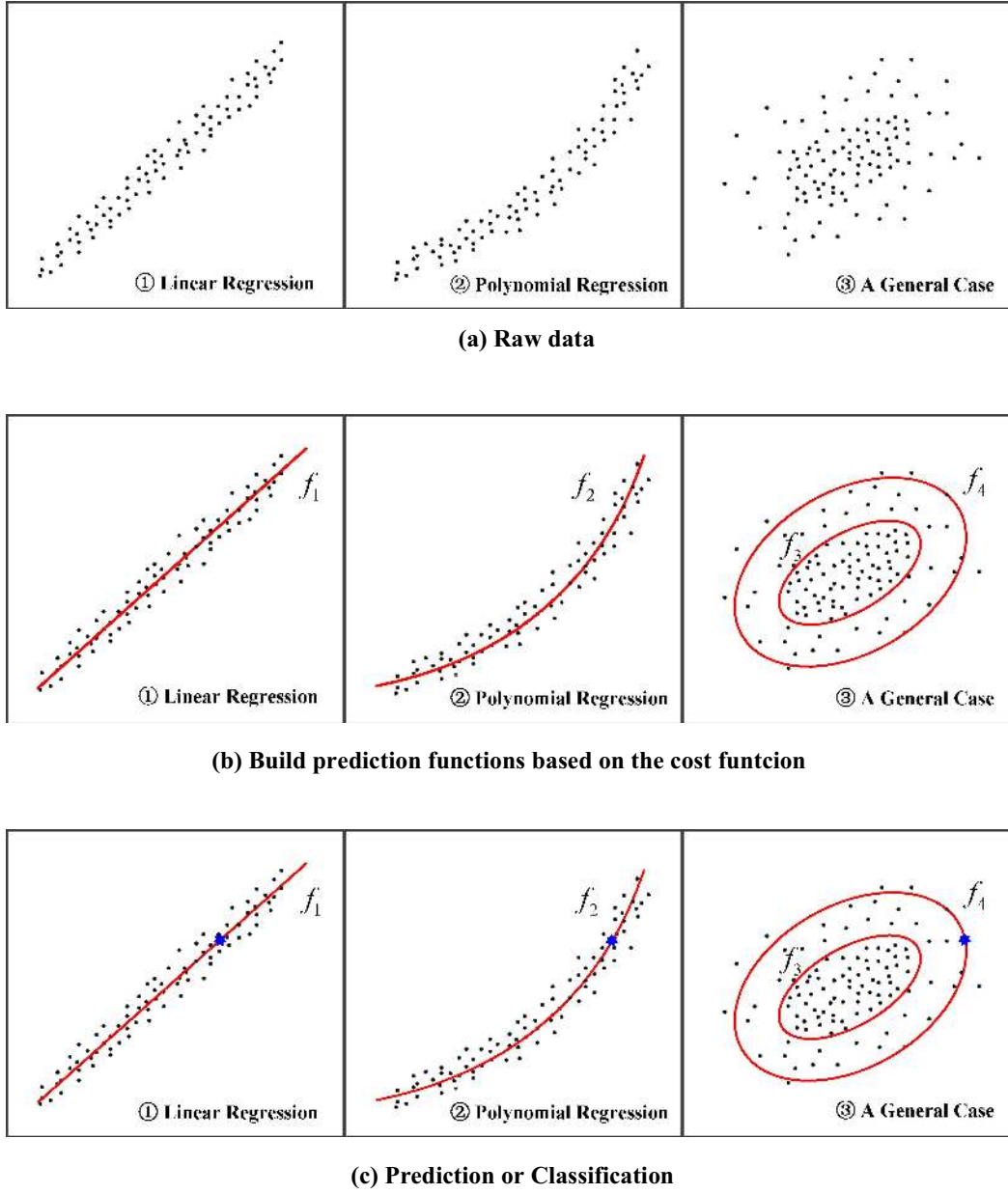**(c) Prediction or Classification**

Fig. 10: The basic flow of a regression model.

in order to improve the modulation classification accuracy, which can be viewed as a reliable modulation classification scheme for the SU in cognitive radio networks. In [208], the KNN algorithm was used both for extracting the environmental interference imposed by 5G Wi-Fi signals and for reducing the computational complexity and yet improving the performance of indoor localization.

C. SVM and Its Applications

1) Methods: Being constructed purely by mathematical theory, SVM is another supervised learning model conceived for classification and regression relying on constructing a hyperplane or a set of hyperplanes in a high-dimensional space. The best hyperplane is the one that results in the largest margin amongst the classes. However,

the training data set may often be linearly non-separable in a finite dimensional space. To address this issue, SVM is capable of mapping the original space into a higher dimensional space, where the training data set can be more easily discriminated.

Considering a linear binary SVM, for example, there are N training samples in the form of $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)\}$, where $y_n = \pm 1$ indicates the class label of the point $\boldsymbol{x}_n$. SVM aims for searching for a hyperplane having the maximum possible separation from the training samples, which best discriminates the two classes of $\boldsymbol{x}_n$ associated with $y_n = 1$ and $y_n = -1$. Here, the maximum separation implies having the maximum possible distance between the nearest point and the hyperplane. The hyperplane is
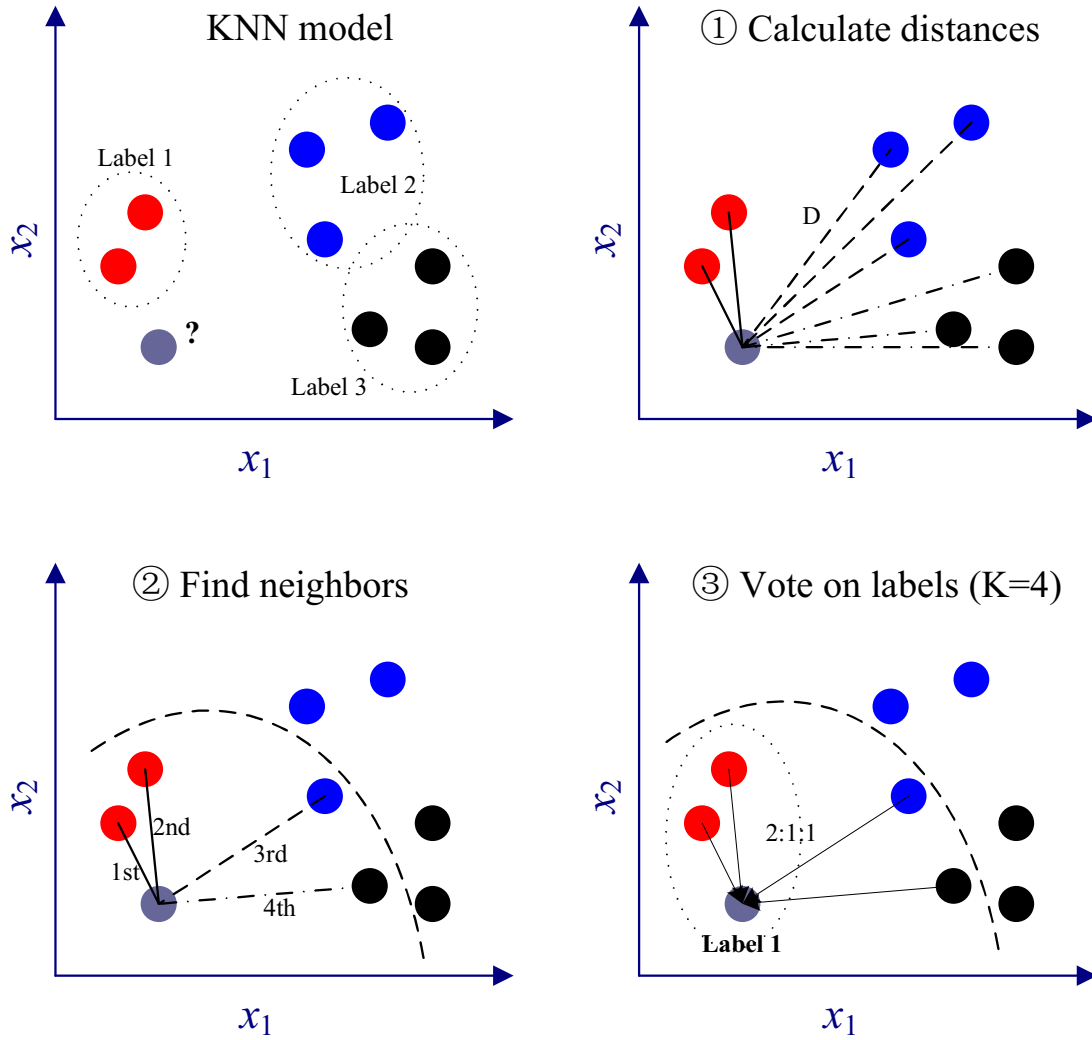
Fig. 11: The illustration of the unweighted KNN mechanism with $K = 4$, for example.

represented by:

$$\boldsymbol{\omega}^T \boldsymbol{x} + b = 0. \tag{8}$$

Hence, we can quantify the separation of the training sample $(\boldsymbol{x}_n, y_n)$ as:

$$\gamma_n = y_n(\boldsymbol{\omega}^T \boldsymbol{x}_n + b). \tag{9}$$

Moreover, we assume having the correct classification if $\boldsymbol{\omega}^T \boldsymbol{x}_n + b \geq 0$ when $y_n = 1$, while $\boldsymbol{\omega}^T \boldsymbol{x}_n + b \leq 0$ when $y_n = -1$. Because we have $y_n(\boldsymbol{\omega}^T \boldsymbol{x}_n + b) \geq 0$, a higher separation implies a more reliable classification. Again, the SVM tries to find the optimal hyperplane that maximizes the minimum separation between the training samples and the hyperplane considered. Given a set of linearly separable training samples, after the operation of normalization, the SVM based classification can be formulated as the following optimization problem:

$$\max_{\boldsymbol{\omega}, b} \min_{n=1,\dots,N} y_n \left(\left(\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|} \right)^T \boldsymbol{x}_n + \frac{b}{\|\boldsymbol{\omega}\|} \right)$$
$$\text{s.t. } y_n(\boldsymbol{\omega}^T \boldsymbol{x}_n + b) \geq \gamma, n = 1, 2, \dots, N, \tag{10}$$
$$\|\boldsymbol{\omega}\| = 1,$$

where we have $\gamma = \min_{n=1,\dots,N} y_n \left(\left(\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|} \right)^T \boldsymbol{x}_n + \frac{b}{\|\boldsymbol{\omega}\|} \right)$. After some further mathematical manipulations, the problem in (10) can be reduced to an optimization problem having a convex quadratic objective function and linear constraints, which can be expressed by:

$$\min_{\boldsymbol{\omega}, b} \frac{1}{2}(\|\boldsymbol{\omega}\|)^2$$
$$\text{s.t. } y_n(\boldsymbol{\omega}^T \boldsymbol{x}_n + b) \geq 1, n = 1, 2, \dots, N. \tag{11}$$

Problem (11) is a typical convex optimization problem. Taking advantage of Lagrange duality [210], we can obtain the optimal $\boldsymbol{\omega}$ and b.

Again, if the training samples are linearly non-separable, SVM is capable of mapping data to a high dimensional feature space with a high probability of being linearly separable. This may result in a non-linear classification or regression in the original space. Fortunately, kernel functions play a critical role in avoiding the "curse of dimensionality" in the above-mentioned dimensionality ascending procedure [211], [212]. To elaborate a little further, given the original input samples $\boldsymbol{x}$, we may be

interested in learning some features $\phi(\boldsymbol{x})$. Let us assume $\boldsymbol{x}_m, \boldsymbol{x}_n \in \mathbb{R}^n$, hence the corresponding kernel function $K(\boldsymbol{x}_m, \boldsymbol{x}_n)$ is defined as:

$$K(\boldsymbol{x}_m, \boldsymbol{x}_n) = \phi(\boldsymbol{x}_m)^T \phi(\boldsymbol{x}_n). \tag{12}$$

Fortunately, even though the high dimensional feature mapping $\phi(\boldsymbol{x}_n)$ may be expensive to calculate, the kernel function calculated relying on their inner product can be easy obtained after some further mathematics manipulations.

There are a variety of alternative kernel functions, such as linear kernel function, polynomial kernel function, radial basis kernel function, neural network kernel function, etc. Furthermore, some regularization methods haven been conceived in order to make SVM be less sensitive to outlier points.

The specific choice of the kernel function plays a key role in ML [213], hence we have to beneficially design the kernel function. The construction of kernels can be generally developed by the inner product operations of feature mappings $\phi(\boldsymbol{x}_n)$ between the input samples over the Hilbert space, whose infinite number of dimensions allow the appropriate representation of big data to exploit their geometric properties. Such a Hilbert space associated with a kernel invoked for producing functions by calculating the inner product of the feature mappings is known as the reproducing kernel Hilbert space (RKHS) [214], and has been applied in diverse learning contents [215], [216]. The RKHS therefore serves a critical foundation in statistical learning theory. Fig. 12 provides a graphical illustration of the kernel-based method.

On the other hand, we may rely on statistical learning theory for appropriately constructing the signal space in order to identify sufficient statistics for reliable signal detection and estimation in statistical communication theory [217]. Inspired by Parzen [214], Kailath observed that RKHS may also be beneficially invoked both for detection and estimation [218] by exploiting the one-to-one relationship between RKHS and finite-variance linear functionals of a random process. Corresponding to the simplest setup of signal detection in additive white Gaussian noise (AWGN) using the Karhunen-Loeve expansion [219], the RKHS representation associated with the noise covariance function is capable of providing an equivalent theoretical framework of statistical communication theory. After a series of efforts inverted into different areas of signal detection and estimation, Kailath and Poor [220] conceived the RKHS approach for the detection of stochastic signals.

2) Applications: As mentioned before, SVM hinges on a mapping that can transform the original training data into a higher dimension, where the events to be classified do become linearly separable. Then it searches for the optimal separating hyperplane for delineating one class from another in this higher dimension considered. As highlighted in Fig. 13, in the spirit of this, SVM aided learning models can be used for detecting and estimating network parameters, for learning and classifying environmental signals and the user's behavior, as well as for guiding decision making concerning channel selection and anomaly detection, for example [221]–[229].

As for detecting and estimating the network parameters, Feng and Chang [221] constructed a hierarchical SVM (H-SVM) structure for multi-class data estimation. The H-SVM was constructed by a number of levels and each level was composed by a finite number of SVM classifiers. Feng and Chang used their H-SVM model both for estimating the physical locations of nodes in an indoor wireless network and the Gaussian channel's noise level in a MIMO-aided wireless network. Thanks to its hierarchical structure, the H-SVM was capable of providing an efficient distributed estimation procedure. Furthermore, Tran *et al.* proposed an SVM model for estimating the geographic location of sensor nodes in WSNs whilst only relying on their connectivity information, more precisely the hop counts [222]. It yielded fast convergence in a distributed manner. The final estimation error can be upper bounded by any small threshold upon relying on a sufficiently large training dataset. Moreover, Sun and Guo [223] conceived a least square-SVM (LS-SVM) algorithm for estimating the user's position by correlating the time-of-arrival (TOA) of radio frequency signals at the BSs without any detailed knowledge about the base station's location as well as about the propagation characteristics.

SVM can also be used for learning a user's behavior and for classifying environmental signals considering the complex spatio-temporal context and the diverse selection of devices. In [224], Donohoo *et al.* studied the context-aware energy-efficiency improvement options for smart devices. These solutions may become beneficial in terms of configuring their location-specific interface for heterogeneous networks (HetNets) constituted by diverse cells. In [225], by combining the SVM and Fisher discriminant analysis (FDA) Joseph *et al.* learned the malicious sinking behavior in wireless ad hoc networks for finding the security vulnerabilities and for designing novel intrusion detection scheme. Moreover, features such as delay between data and acknowledgement, number of re-transmits, etc. gleaned from the MAC layer were jointly considered with those from other layers, which constituted a correlated feature set. Furthermore, Pianegiani *et al.* [226] proposed an SVM-based binary classification solution for classifying acoustic signals emitted by vehicles relying on spectral analysis aided feature extraction, which was beneficial in terms of improving the classification accuracy, despite reducing the implementation complexity.

As for the SVM's benefit in assisting decision making, in [227], a common control channel selection mechanism was conceived for SUs during a given frame relying on an SVM-based learning technique proposed for a cognitive radio network, which was capable of implicitly and cooperatively learning the surrounding environment cooperatively in an online way. Moreover, Yang *et al.* [228] investigated the spoofing attack detection problem based on the spatial correlation of received signal strength gleaned from network nodes, where a cluster-based SVM mechanism was developed for determining the number of attackers. Relying on carefully designed certain training data, the
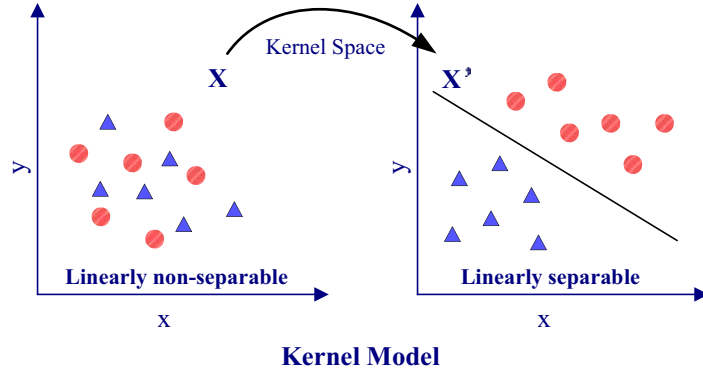
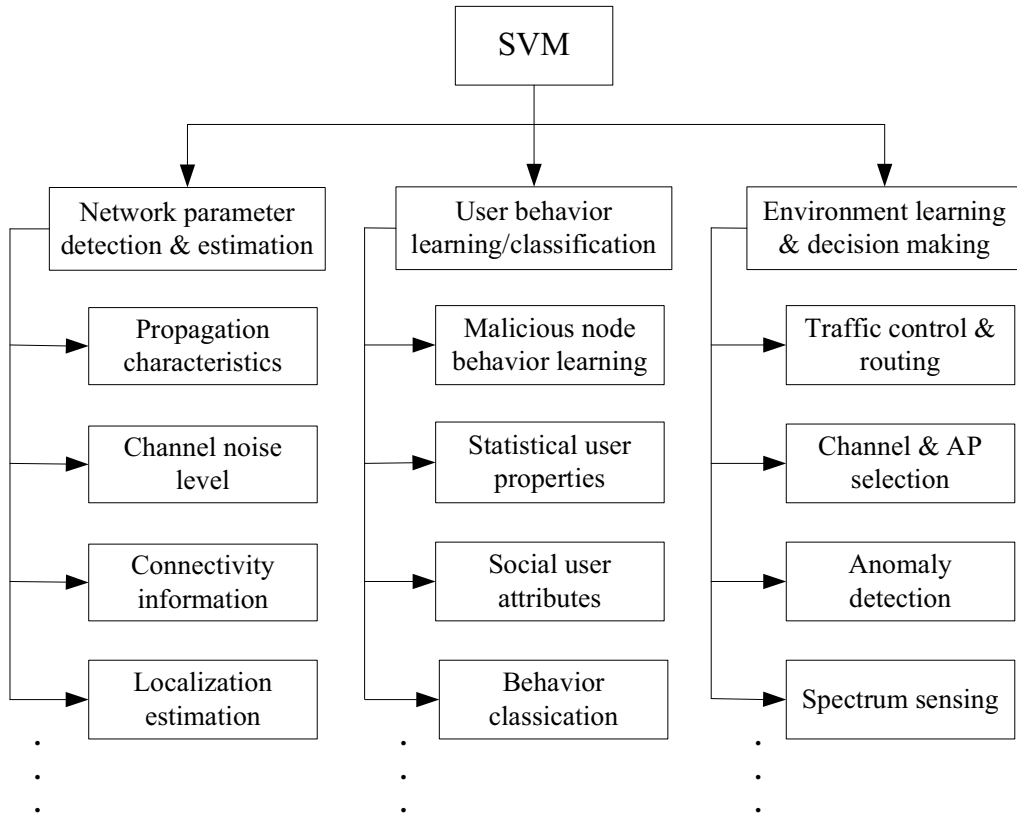Fig. 12: An illustration of the kernel-based methods.

Fig. 13: The applications of SVM aided learning models for wireless networks.

SVM algorithm employed further improved the accuracy of determining the number of attackers. Rajasegarar *et al.* [229] also investigated the malicious activity detection issues of WSNs invoking a variety of SVM based algorithms.

D. Bayes Classification and Its Applications

1) Methods: The Bayes classifier, a popular member of the probabilistic classifier family relying on Bayes' theorem, operates by computing the *posteriori* probability

distribution of the objective function values given a set of training samples. As a widely-used classification method, the naive Bayes classifier can be trained for example conditioned on a simple but strong independence assumption in features. Furthermore, the complexity of training a naive Bayes model is linearly proportional to the training set size.

To elaborate a little further, let the vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_M]$ represent M independent features for a total of K classes $\{y_1, y_2, \ldots, y_K\}$. For each of the K possible class labels y_k, we have the conditional probability

of $p = (y_k|x_1, \ldots, x_M)$. Relying on Bayes' theorem, we decompose the conditional probability to yield the form of:

$$p = (y_k|x_1, \ldots, x_M) = \frac{p(y_k)p(x_1, \ldots, x_M|y_k)}{p(x_1, \ldots, x_M)}, \quad (13)$$

where $p = (y_k|x_1, \ldots, x_M)$ is the *posteriori* probability, whilst $p(y_k)$ is the *priori* probability of y_k. Given that x_i is conditionally independent of x_j for $i \neq j$, we have:

$$p = (y_k|x_1, \ldots, x_M) = \frac{p(y_k)}{p(x_1, \ldots, x_M)} \prod_{m=1}^{M} p(x_m|y_k), \quad (14)$$

where $p(x_1, \ldots, x_M)$ only depends on M independent features, which can be viewed as a constant.

The maximum a *posteriori* probability (MAP) is used as the decision making rule for the naive Bayes classifier. Given a feature vector $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M)$, its label $\bar{y}$ can be determined according to:

$$\bar{y} = \operatorname*{arg\,max}_{y_k \in \{y_1, \ldots, y_K\}} p(y_k) \prod_{m=1}^{M} p(\bar{x}_m|y_k). \quad (15)$$

Despite idealized simplifying assumptions, naive Bayes classifiers have enjoyed popularity in numerous complex real-world situations, such as outlier detection [230], spam filtering [231], etc.

2) Applications: Based on the Bayes' theorem, Bayes classifier techniques are particularly applicable to the context where the dimensionality of the input is high. Despite their simplicity, they can often outperform other sophisticated classification methods. As for their applications in wireless networks, in the following, we will elaborate on some typical examples in different wireless scenarios, such as antenna selection, network association, anomaly detection, indoor location and QoE prediction.

Specifically, in [232], He *et al.* modeled the transmit antenna selection (TAS) problem of MIMO wiretap channels as a multi-class classification problem. Then, they used the naive Bayes-based classification scheme to select the optimal antenna for enhancing the physical layer security of the system considered. In contrast to conventional TAS schemes, simulation results showed that the proposed scheme resulted in a reduced feedback overhead at a given secrecy performance. In [233], Abouzar *et al.* proposed an action-based network association technique for wireless body area networks (WBANs). Relying on the level of received signal strength indicator of the on-body link, the naive Bayes algorithm was employed to recognize the ongoing action, which was beneficial in terms of scheduling the time slot assignment in the context of fixed power allocation on various links by the sink node under a specific data rate constraint. Moreover, Klassen *et al.* [234] used the naive Bayes classifier for detecting anomaly in *ad hoc* wireless network involving the black hole attack, the denial of service (DoS) attack and the selective forwarding attack.

Bayes classifier can also be applied to the indoor location estimation. For example, in [235], a probabilistic model was conceived for characterizing the relationship between the received signal strength and location with the aid of the naive Bayes generative learning method, which was used for learning the parameters of an initial probabilistic model, given a limited number of labeled samples. The proposed indoor location estimation method was capable of both reducing the off-line calibration efforts required, whilst maintaining a high location estimation accuracy. Furthermore, as for QoE prediction, in order to evaluate the impact of different networking and channel conditions on the QoE attained in the context of different network services, Charonyktakis *et al.* [236] proposed a modular algorithm for user-centric QoE prediction. They integrated multiple ML algorithms, including the Gaussian naive Bayes classifier and conceived a nested cross validation protocol for selecting the optimal classifier and its corresponding optimal hyper-parameter value for the sake of accurate QoE prediction.

IV. Unsupervised Learning in Wireless Networks

In this section, we will highlight some typical unsupervised learning algorithms, such as K-means clustering [237], expectation-maximization (EM) [238], principal component analysis (PCA) [239] and independent component analysis (ICA) [240] in terms of their methodology and their applications in wireless networks. Table III summarizes some typical applications of the above-mentioned unsupervised learning algorithms in wireless networks.

A. K-Means Clustering and Its Applications

1) Methods: K-means clustering is a distance based clustering method that aims for partitioning N unlabeled training samples into K different cohesive clusters, where each sample belongs to one cluster. To elaborate a little further, K-means clustering measures the similarity between two samples in terms of their distance and it has two main steps, namely assigning each training sample to one of K clusters in terms of the closest distance between the sample and the cluster centroids, and then updating each cluster centroid according to the mean of the samples assigned to it. The whole algorithm is hence implemented by repeatedly carrying out the above-mentioned pair of steps until convergence is achieved.

To elaborate a little further, given a set of samples $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$, where $\boldsymbol{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nM}]$ is a M-dimensional vector, let $\mathbb{S} = \{s_1, s_2, \ldots, s_K\}$ represent the above-mentioned cluster set, and $\boldsymbol{\mu}_k$ the mean of the samples in s_k. K-means clustering intends to find an optimal cluster-based segmentation, which solves the following optimization problem:

$$\mathbb{S}^* = \operatorname*{arg\,min}_{\{s_1, s_2, \ldots, s_K\}} \sum_{k=1}^{K} \sum_{\boldsymbol{x} \in s_k} \|\boldsymbol{x} - \boldsymbol{\mu}_k\|^2. \quad (16)$$

However, problem (16) is a non-deterministic polynomial-time hardness (NP-hard) problem [241]. Fortunately, there are a range of efficient heuristic algorithms, which converge quickly to a local optimum.

TABLE II: Compelling Applications of Supervised Learning in Wireless Networks

Paper	Application	Method	Description
[195]	interference estimate	regression	strike a trade-off between the overhead and accuracy of interference measurement
[196]	spectrum sensing	regression	reduce the number of parameters and maintain a high detection accuracy
[197]	wireless coexistence	regression	estimate the likelihood of the wireless coexistence of Wi-Fi and ZigBee
[198]	PHY authentication	regression	do not need the assumption on the accurate known channel model
[199]	traffic estimation	regression	estimate the contention success probability considering sensors' heterogeneous traffic demands
[200]	map reconstruction	regression	reconstruct the wireless radio map for UAV path planning and location design
[201]	wireless localization	regression	logistic regression classifier for counteracting the negative influence relying on fingerprint signals
[203]	traffic prediction	KNN	explore both the temporal and spatial characteristics of radio resources
[204]	anomaly detection	KNN	rely on the hypergrid intuition in the context of WSN applications
[206]	missing data estimation	KNN	rely on the temporal and spatial correlation feature of sensor data
[207]	modulation classification	KNN	combine the genetic programming and KNN for improving the modulation classification accuracy
[208]	interference elimination	KNN	extract environmental interference from Wi-Fi signal and reduce computational complexity
[221]	data estimation	SVM	provide an efficient estimation procedure in a distributed manner
[222]	localization estimation	SVM	yield fast convergence performance and efficiently use the communication resources
[223]	user location	SVM	without knowledge about base station location and environmental propagation characteristics
[224]	data prediction	SVM	provide location-specific interface configuration for HetNets
[225]	behavior learning	SVM	combine both the superior accuracy of SVM and fast convergence speed of FDA
[226]	signal classification	SVM	classify acoustic signals emitted by vehicles rely on feature extraction
[227]	channel selection	SVM	propose a control channel selection mechanism for a cognitive radio network
[228]	attacker counting	SVM	develop a cluster-based SVM mechanism for determining the number of attackers
[232]	antenna selection	Bayes	enhance the physical layer security relying on Bayes-based optimal antenna selection
[233]	network association	Bayes	schedule time slot assignment and fixed power allocation under data rate constraint
[234]	anomaly detection	Bayes	detect anomaly involving black hole attack, DoS attack and selective forwarding attack
[235]	indoor location	Bayes	characterize the relationship between the received signal strength and location
[236]	QoE prediction	Bayes	accurate QoE prediction by selecting optimal classifier and optimal hyper-parameter values

One of the popular low-complexity iterative refinement algorithms suitable for K-means clustering is Lloyd's algorithm [242], which often yields satisfactory performance after a low number of iterations. Specifically, given K initial cluster centroid $\boldsymbol{\mu}_k, k = 1, \ldots, K$, Lloyd's algorithm arrives at the final cluster segmentation result by alternating between the following two steps,

- Step 1: In the iterative round r, assign each sample to a cluster. For $n = 1, 2 \ldots, N$ and $i, k = 1, 2 \ldots, K$, if we have:

$$s_i^{(r)} = \{\boldsymbol{x}_n : \ \left\| \boldsymbol{x}_n - \boldsymbol{\mu}_i^{(r)} \right\|^2 \leq \left\| \boldsymbol{x}_n - \boldsymbol{\mu}_k^{(r)} \right\|^2, \forall k\}, \tag{17}$$

then we assign the sample $\boldsymbol{x}_n$ to the cluster s_i, even if it could potentially be assigned to more than one cluster.

- Step 2: Update the new centroids of the new clusters formulated in the iterative round r relying on:

$$\boldsymbol{\mu}_i^{(r+1)} = \frac{1}{|s_i^{(r)}|} \sum_{\boldsymbol{x}_j \in s_i^{(r)}} \boldsymbol{x}_j, \tag{18}$$

where $|s_i^{(r)}|$ denotes the number of samples in cluster s_i in iterative round r.

Convergence is deemed to be obtained when the assignment in Step 1 is stable. Explicitly, reaching convergence means that the clusters formulated in the current round are the same as those formed in the last round. Since this is a heuristic algorithm, there is no guarantee that it can converge to the global optimum. Hence, the result of clustering largely relies on specific choice of the initial clusters and on their centroids.

2) Applications: K-means clustering aims for partitioning N samples into K clusters. Each sample belongs to the closest cluster. The clustering algorithm proceeds in an iterative manner, where the in-cluster differences are minimized by iteratively updating the cluster centroid, until convergence is achieved.

Clustering functioning under uncertainty or incomplete information is a common problem in wireless networks, especially in the scenarios associated with numerous small traffic cells, heterogeneous large and small cell structures relying on diverse carrier frequencies, diverse time-varying tele-traffic, etc. First of all, the small cells have to be carefully clustered for avoiding excessive interference using coordinated multi-point transmission. Moreover, the devices and users should be beneficially clustered for the sake of achieving a high energy efficiency, maintaining an optimal access point association, obeying an efficient off-loading policy, and of guaranteeing a high network security. In [243], a mixed integer programming problem was formulated for jointly optimizing both the gateway deployment and the virtual-channel allocation for optical/wireless hybrid networks, where Xia *et al.* designed an efficient K-means clustering based solution for iteratively solving this problem, which beneficially reduced the delay, as well as improved the network throughput. Moreover, in [244], Hajjar *et al.* proposed a K-means based relay selection algorithm for creating small cells under the umbrella of an oversailing LTE macro cell within a multi-cell scenario under the constraint of low power clusters. Relying on the proposed relay selection algorithm, the total capacity was increased by reusing the frequency in each low power cluster, which had the benefit of supporting high data rate services. Additionally, Cabria and Gondra [245] proposed a so-called potential-K-means scheme for partitioning data collection sensors into clusters and then for assigning each cluster to a storage center. The proposed K-means solution had the advantage of both balancing the storage center loads and minimizing the total network cost (optimizing the total number of sensors). Parwez *et al.* [246] invoked both K-means clustering and hierarchical clustering algorithms for their user-activity analysis and user-anomaly detection in a mobile wireless network, which verified genuine identity of users in the face of their dynamic

spatio-temporal activities. Furthermore, El-Khatib [247] designed a K-means classifier for selecting the optimal set of features of the MAC layer bearing in mind the specific relevance of each feature, which beneficially improved the accuracy of intrusion detection, despite reducing the learning complexity.

Clustering can also be used in signal detection for the sake of both reducing the detection complexity and for improving the energy efficiency attained. In [248], the K-means clustering algorithm was invoked in a blind transceiver, where the training process was completely dispensed within the transmitter for reducing its energy dissipation, since no pilot power was required. Furthermore, Zhao *et al.* [249] conceived an efficient K-means clustering algorithm for optical signal detection in the context of burst-mode data transmission.

B. EM and Its Applications

1) Methods: The EM algorithm is an iterative method conceived for searching for the maximum likelihood estimate of parameters in a statistical model. Typically, in addition to unknown parameters whose existence has been ascertained, the statistical model also has some latent variables. In this scenario it is an open challenge to derive a closed-form solution, because we are unable to find the derivatives of the likelihood function with respect to all the unknown parameters and latent variables. The iterative EM algorithm consists of two steps, as shown in Fig. 14. During the expectation step (E-step), it calculates the expected value of the log likelihood function conditioned on the given parameters and latent variables, while in the maximization step (M-step), it updates the parameters by maximizing the specific log-likelihood expectation function considered.

More explicitly, upon considering a statistical model with observable variables $\boldsymbol{X}$ and latent variables $\boldsymbol{Z}$, the unknown parameters are represented by $\boldsymbol{\theta}$. The log-likelihood function of the unknown parameters is given by:

$$l(\boldsymbol{\theta};\boldsymbol{X},\boldsymbol{Z}) = \log p(\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\theta}). \qquad (19)$$

Hence, the EM algorithm can be described as follows [238]:

- E-step: Calculate the expected value of the log likelihood function under the current estimate of $\bar{\boldsymbol{\theta}}$, i.e.

$$Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}}) = E_{\boldsymbol{Z}|\boldsymbol{X},\bar{\boldsymbol{\theta}}}[\log p(\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\theta})]. \qquad (20)$$

- M-step: Maximize Eq. (20) with respect to $\boldsymbol{\theta}$ for generating an updated estimate of $\bar{\boldsymbol{\theta}}$, which can be formulated as:

$$\bar{\boldsymbol{\theta}}' = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}}). \qquad (21)$$

The EM algorithm plays a critical role in parameter estimation based on many of the popular statistical models, such as the Gaussian mixture model (GMM), hidden Markov model (HMM), etc. which are beneficial both for clustering and prediction.

2) Applications: The EM model can be readily invoked for a variety of parameter learning and estimation problems routinely encountered in wireless networks. Specifically, Wen *et al.* [250] estimated both the channel parameters of the desired links in a target cell and those of the interfering links in the adjacent cells relying on constructing a GMM, which was estimated with the aid of the EM algorithm. Choi *et al.* [251] modeled the cognitive radio system as a HMM, where the secondary users (SUs) estimated the channel parameters such as the primary user's (PU) sojourn time, signal strength, etc. based on the standard EM algorithm. Moreover, Assra *et al.* [252] also adopted the EM algorithm to jointly estimate the channel unknown frequency domain responses as well as the noise variance and detected the PU's signal in a cooperative wide-band cognitive system, which was shown to converge to the upper bound solution based on maximum likelihood estimation under the idealized assumption of having perfect channel parameter estimation. Additionally, Zhang *et al.* [253] proposed an EM aided joint symbol detection and channel estimation algorithm for MIMO-OFDM systems in the presence of frequency selective fading, which provided a distribution-estimate for both the hidden symbol and unknown channel parameters in an iterative manner. Li and Nehorai [254] built an asynchronous state-space model for connecting asynchronous observations with the most likely target state transition in the context of multi-sensor WSNs. Then, they adopted the EM algorithm for jointly estimating the sequential target state as well as the network's synchronization state under the assumption of knowing the temporal order of sensor clocks. Furthermore, Zhang *et al.* [255] used a variational EM iterative algorithm to recover the transmitted signals and to identify the active users in a low-activity code division multiple access based M2M communications without the knowledge of the user activity factor. The EM algorithm can also be invoked for target or source localization, which can be viewed as a joint sparse signal recovery and parameter estimation problem [256] [257].

C. PCA & ICA and Their Applications

1) Methods: Feature reduction and extraction constitutes the preparatory phase of ML, because the initially acquired raw data may contain some irrelevant or redundant features. The feature reduction and extraction processing phase results in a reduced set of features with the aid of feature transforms, conceived for reducing the feature space dimension, by eliminating correlated features. However, given the diverse definitions of "features", it is only realistic to discuss feature-extraction in specific application-oriented contexts.

PCA and ICA constitute sophisticated dimensionality reduction methods in ML, which are capable of reducing both the computational complexity and the storage requirements.

PCA utilizes an orthogonal transformation for converting a set of potentially correlated features of the training samples into a set of uncorrelated features, which are termed as the "principal components". The number of
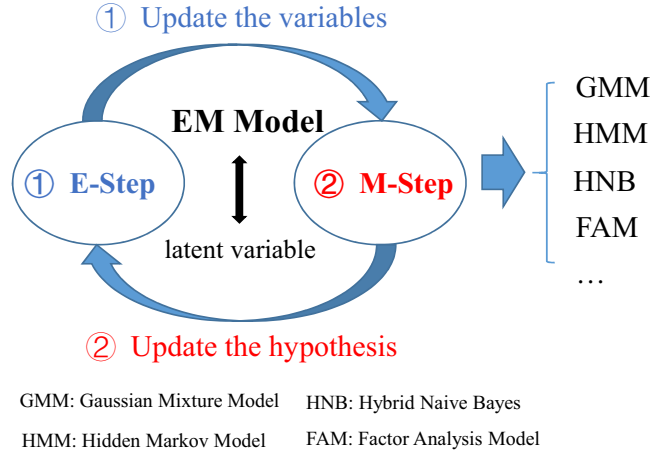
Fig. 14: The operation process of the EM model.

principal components is expected to be lower than the number of the original features of the training samples, which hence provide a more compact representation of the original samples. More explicitly, less principle components can be used for representing the original samples in the transformed domain. In PCA, the first principal component tends to have the largest variance, which indicates that it encapsulates the most information of the original features provided that these features were correlated. Similarly, each succeeding component tends to have the next highest variance. These principal components can be generated by invoking the eigenvectors of the normalized covariance matrix.

Specifically, let us consider N training samples of $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$, where $\boldsymbol{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nM}]^T$ is composed of M different features. Let us first pre-process the samples by normalizing their mean and variance. Given a unit vector $\boldsymbol{u}$, $\boldsymbol{x}_n^T\boldsymbol{u}$ can be interpreted as the length of the projection of $\boldsymbol{x}_n$ onto the direction $\boldsymbol{u}$. The PCA attempts to maximize the variance of the projections, which is formulated as:

$$\max_{\boldsymbol{u}} \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n^T\boldsymbol{u})^2 = \max_{\boldsymbol{u}} \boldsymbol{u}^T \left(\frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n\boldsymbol{x}_n^T) \right) \boldsymbol{u}. \tag{22}$$

Given the covariance matrix $\boldsymbol{\Lambda} = \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n\boldsymbol{x}_n^T)$, the solution of problem (22) is given by the eigenvector of the covariance matrix $\boldsymbol{\Lambda}$. If we denote the top K eigenvectors of $\boldsymbol{\Lambda}$ by $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_K$ and $K < M$, a dimensionality reduction expression of $\boldsymbol{x}_n$ can be formulated as:

$$\boldsymbol{y}_n = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_K]^T \boldsymbol{x}_n, \tag{23}$$

where $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_K$ are the first K principle components of the training samples.

By contrast, the ICA attempts to find a new basis for representing original samples that are assumed to be a linear weighted superposition of some unknown latent variables. It aims for decomposing multivariate variables into a set of additive subcomponents, which are non-Gaussian variables and are statistically independent from each other. As for the independent components, also termed as the latent variables, they exhibit the maximum possible "statistical independence", which can be commonly characterized by either the minimization of their mutual information quantified in terms of the Kullback-Leibler divergence metric and the maximum entropy criterion, or by the maximization of what is termed in parlance as the non-Gaussianity relying on kurtosis and negentropy, for example.

Let us consider the linear noiseless ICA model in a simple example, where the multivariate training variables are denoted by $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$. Its latent independent component vector is represented by $\boldsymbol{s} = [s_1, s_2, \ldots, s_M]^T$. Each component of $\boldsymbol{x}$ can be generated by a linearly weighted sum of independent components, i.e. we have $x_n = a_{n1}s_1 + a_{n2}s_2 + \cdots + a_{nM}s_M$, where a_{nm} is the weighting coefficient. The vectorial form of $\boldsymbol{x}$ can be expressed as:

$$\boldsymbol{x} = \sum_{m=1}^{M} s_m \boldsymbol{a}_m, \tag{24}$$

where $\boldsymbol{a}_m = [a_{1m}, a_{2m}, \ldots, a_{Nm}]^T$. Furthermore, let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M)$. Then the original multivariate training variables can be rewritten as:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}, \tag{25}$$

where the unknown matrix $\boldsymbol{A}$ is referred to as the mixing matrix. ICA algorithms attempt to estimate both the mixing matrix $\boldsymbol{A}$ and the independent component vector $\boldsymbol{s}$ relying on setting up a cost function, which again, either maximizes the non-Gaussianity or minimizes the mutual information. Thus, we can recover the independent component vector by computing $\boldsymbol{s} = \boldsymbol{A}^{-1}\boldsymbol{x}$, where $\boldsymbol{A}^{-1}$ is termed as the 'unmixing' matrix. Usually, we assume that $N = M$ and that the mixing matrix $\boldsymbol{A}$ is a square-shaped

matrix. Moreover, the *apriori* knowledge of the probability distribution of s is beneficial in terms of formulating the cost function.

2) Applications: As for the application of PCA and ICA in wireless networks, Shi *et al.* [258] utilized PCA to extract the most relevant feature vectors from fine-grained subchannel measurements for improving the localization and tracking accuracy in an indoor location tracking system. Moreover, Morell *et al.* [259] designed an efficient data aggregation method for WSNs based on PCA amalgamated with a non-eigenvector projection basis, while keeping the reconstruction error below a predefined threshold. Quer *et al.* [260] exploited PCA for inferring the spatial and temporal features of a range of signals monitored by a WSN. Based on this they recovered the large original data set from a small observation set.

Additionally, Qiu [261] combined ICA with PCA in a smart grid scenario for recovering smart meter data, which were jointly capable of enhancing the transmission efficiency both by avoiding the channel estimation in each frame and by eliminating wide-band interference or jamming signals. A semi-blind received signal detection method based on ICA was proposed by Lei *et al.* [262], which additionally estimated the channel information of a multicell multiuser massive MIMO system. Moreover, Sarperi *et al.* [263] proposed an ICA based blind receiver structure for MIMO OFDM systems, which approached the performance of its idealized counterpart relying on perfect CSI. ICA was also used for digital self-interference cancellation in a full duplex system [264], which relied on a reference signal used for estimating the leakage into the receiver. More explicitly, in full duplex systems the high-power transmit signal leaks into the receiver through a nonlinear leakage path and drowns out the low-power received signal. Hence its cancellation requires at least 120 dB interference rejection. Furthermore, in [265], the Boolean ICA concept was proposed based on the integration of Boolean functions of binary signals for inferring the activities of the underlying latent signal sources. Specifically, it was shown that given m SUs, the activities of up to $(2m - 1)$ PUs can be determined.

V. REINFORCEMENT LEARNING IN WIRELESS NETWORKS

Reinforcement learning deals with an agent interacting with the environment. Three specific aspects of reinforcement learning, multi-arm bandit problem, Markov decision process (MDP) and temporal-difference (TD) learning can be very useful for wireless networks. Then, we explore further on these algorithms of reinforcement learning to wireless networks.

A. Multi-Armed Bandit and Its Applications

1) Methods: The multi-armed bandit technique, also called K-armed bandit, models a decision making problem, where an agent is faced with a dilemma of K different actions. After each choice, the agent receives a reward relying on a stationary probability distribution that is associated with its decision. The agent attempts to maximize its expected total reward over a series of decision making rounds relying on a balance striking a trade-off between consulting existing knowledge and acquiring new knowledge when optimizing its decisions. The action of referring to existing knowledge to make decisions is termed as "exploitation", while the trial of acquiring new knowledge is referred to as "exploration". Striking a trade-off between exploration and exploitation is also sought by other reinforcement learning algorithms, where exploitation is the plausible action for maximizing the expected reward within the current round, while exploration may produce a greater reward in the long run.

In a K-armed bandit model, K possible actions, $a_1, a_2, \ldots, a_K$, yield different rewards associated with the K unknowns of the problem at hand, which may have different distributions with K mean values of $\mu_1, \mu_2, \ldots, \mu_K$, respectively. The agent iteratively chooses an action A_i at the round i and receives the corresponding reward of R_i. Up to the round i, the expected reward of an action a can be expressed as $Q_i(a) = E[R_i | A_i = a]$. Upon striking a balance between the exploration and the exploitation, we may arrive at a simple bandit algorithm as follows, for example. In each decision-making round, we greedily opt for the action $A = \arg\max_a Q(a)$ relying on the probability of $(1 - \varepsilon)$, whilst riskily embarking on a random action selection based on the probability of ε, where ε is the probability of a brave attempt for exploring new knowledge.

In contrast to the above-mentioned ε-greedy bandit algorithm, there are also more complex bandit algorithms, such as the gradient aided bandit algorithm, associative-search bandit, non-stationary bandit, etc [47]. Moreover, the multi-armed bandit problem can be extended into a multi-play and multi-armed bandit problem [266], where the reward of each agent depends on others' actions, and each agent tries to find its optimal decision by predicting the future actions of the other agents relying on previous decision making strategies.

2) Applications: As mentioned before, multi-armed bandit based techniques are capable of dealing with uncertainties in the context of future wireless networks because of limited prior knowledge and the associated resource-thirsty feedback. Moreover, it is beneficial to model the selfishness and the decision conflicts of/among the users during the decision making process. Hence, the multi-armed bandit based algorithms have become powerful tools for rational decision making in wireless networks both for distributed users and APs as well as for the central control center. Specifically, Maghsudi *et al.* [267] proposed a small cell activation scheme relying on the multi-armed bandit philosophy given only limited information about the available energy of the small cell BS as well as the number of users to be served. The overall heterogeneous network's throughput was improved with the aid of an energy-efficient small cell on-off switching regime controlled by the macro BS, while the inter-interference level was reduced. Another compelling application of the multi-armed bandit regime in the heterogeneous network is

TABLE III: Compelling Applications of Unsupervised Learning in Wireless Networks

Paper	Application	Method	Description
[243]	gateway deployment	K-means	reduce delay and improve network throughput for optical/wireless hybrid networks
[244]	relay selection	K-means	create small cells in an LTE macro cell with low power cluster constraint
[245]	sensor partitioning	K-means	balance the load of storage centers and minimize the total network cost
[246]	anomaly detection	K-means	verify spatio-temporal varying users' genuineness relying on ground truth information
[247]	intrusion detection	K-means	improve intrusion detection accuracy and reduce the learning complexity
[248]	blind transceiver	K-means	not require pilot duration and pilot power for saving energy consumption
[249]	signal detection	K-means	burst-mode data transmission with an unbalanced ratio of bits zero and bits one
[250]	channel estimation	EM algorithm	construct a GMM to estimate channel parameters in both target cell and adjacent cells
[251]	PU detection	EM algorithm	SUs estimate PU's sojourn time and signal strength relying on a HMM model
[252]	channel state detection	EM algorithm	jointly estimate channel frequency responses, noise variance and PU's signal
[253]	symbol detection	EM algorithm	joint symbol detection and channel estimation for MIMO-OFDM systems
[254]	network state detection	EM algorithm	joint estimate the sequential target state and network synchronization state
[255]	active user detection	EM algorithm	detect active user for the low-activity CDMA based M2M communications
[256]	source localization	EM algorithm	formulate localization as a joint sparse signal recovery and parameter estimation problem
[258]	indoor location	PCA	extract relevant feature vectors from fine-grained subchannel measurements
[259]	data aggregation	PCA	limit the reconstruction error based on a non-eigenvector projection basis
[260]	data recovery	PCA	exploit PCA to extract spatial and temporal features of real signals
[261]	data recovery	ICA & PCA	enhance transmission efficiency by avoiding channel estimation and eliminating jamming signals
[262]	channel estimation	ICA	differentiate and decode the received signal, and estimate the channel information
[263]	blind receiver	ICA	yield an ideal performance close to that with perfect CSI
[264]	interference cancellation	ICA	digital interference cancellation based on the reference signal from transmitter power amplifier
[265]	signal detection	ICA	infer the activities of latent signal sources based on the Boolean functions

constituted by the dynamic network selection in the context of uncertain heterogeneous network state information. Wu *et al.* [268] formulated the optimal network selection problem as a continuous-time multi-armed bandit problem considering diverse traffic types. Moreover, the network access cost function and the QoE reward were defined as the metrics of evaluating the proposed network selection schemes. In [269], given the time-varying and user-dependent fading channels of wireless peer-to-peer (P2P) networks, a multi-armed bandit aided optimal distributed transmitter scheduling policy was conceived for multi-source multimedia transmission, which was beneficial of maximizing the data transmission rate and reducing the related power consumption in the light in terms of the realistic energy constraints of wireless mobile devices. In addition to transmitter scheduling, Maghsudi and Stańczak applied the covariate multi-armed bandit regime [270] for solving the relay selection problem in the wireless network, where the geographical location of relay nodes was assumed to be known by the source node, but no knowledge was assumed about the corresponding fading gains. The proposed covariate multi-armed bandit model is capable of dealing with the exploitation-exploration dilemma of the relay selection process. Lee *et al.* [271] proposed a $K\epsilon$-greedy multi-armed bandit based framework for exploiting the gains provided by frequency diversity in Wi-Fi channels. They struck a trade-off between the achievable gain stemming from frequency diversity and the resource consumption imposed by channel estimation and coordination.

Given the open broadcast nature of the wireless channel environment and the access contention mechanism among multi-priority users, multi-armed bandit based techniques have played a special role in cognitive networks [272]–[277]. For example, Zhao *et al.* [272] formulated a multi-armed restless bandit model for opportunistic multi-channel access, which approached the maximum attainable throughput by accurately predicting which is next idle channel likely to become. In [275], a channel selection scheme was investigated which was capable of adapting to the link quality and hence finding the optimal channel for avoiding interferences and deep fading. Moreover, Gwon *et al.* [273] and Zhou *et al.* [277] further considered the choice of access strategy in the presence of both legitimate desired users and jamming cognitive radio nodes, which was resilient to adaptive jamming attacks with different strengths spanning from near no-attack to the full-attack across the entire spectrum. In contrast to only sensing and accessing a single channel, considering the correlated rewards of different arms, a sequential multi-armed bandit regime was conceived by Li *et al.* [274] for identifying multiple channels to be sensed in a carefully coordinated order. Furthermore, Avner and Mannor [276] studied multi-user coordination in cognitive networks, where each user's successful channel selection relies on both the channel state as well as on the decisions of the other users.

3) An Example: Visible light communication (VLC) systems have the compelling benefit of a wide unlicensed communication bandwidth as well as innate security in downlink (DL) transmission scenarios, hence they may find their way into the construction of future wireless networks. However, considering the limited coverage and dense deployment of light-emitting diodes (LED), traditional network association strategies are not readily applicable to VLC networks. Hence by exploiting the power of online learning algorithms, in [278], the authors focused their attention on sophisticated multi-LED access point selection strategies conceived for hybrid indoor LiFi-WiFi communication systems with the aid of a multi-armed bandit model. Explicitly, since light-fidelity (LiFi) VLC transmissions are less suitable for uplink (UL) transmissions, a classic WiFi UL was used in this study.

To elaborate, in the indoor VLC system, the communication between the devices and the backbone network relies on the VLC DL as well as on the RF WiFi UL, which hence can be viewed as a hybrid LiFi-WiFi network. In the system model, it is assumed that there are M low-energy LED lamps in the indoor space considered.

(a) LOS path   (b) Reflected and LOS path
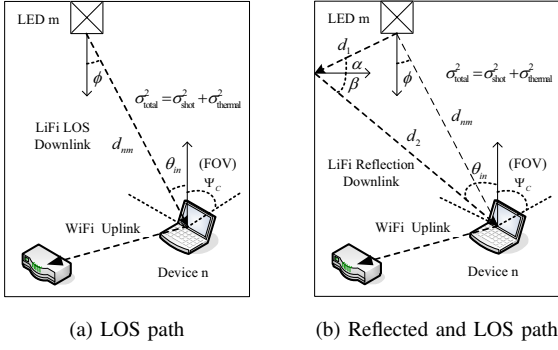
Fig. 15: The model of the indoor VLC link including both the LOS path and the one-hop reflected path, where FOV represents the field of view, while σ^2_{slot} and $\sigma^2_{\text{sthermal}}$ denote the variance of the shot noise as well as the thermal noise, respectively [278] ©IEEE.

Fig. 16: The normalized throughput of the selected VLC link versus K for different LED AP selection schemes. ©IEEE

Fig. 17: The system's normalized throughput versus K for different LED AP selection schemes. ©IEEE

Moreover, regardless of their positions, the N mobile devices are capable of accessing any of the M indoor LED lamps and of downloading packets from the Internet via VLC. When a decision round is due, the access control strategy obeys the decision probability distribution of $P = \{p_1, p_2, ..., p_M\}$. And it has $\sum_{m=1}^{M} p_m = 1$, where p_m denotes the probability of accessing the mth LED lamp. Furthermore, the service time of each LED lamp obeys the negative exponential distribution with a departure rate ς, while the interval between system access requests, in the same way, obeys the negative exponential distribution with an arrival rate λ. The VLC DL channel is characterized by a diffuse link, where the light beam is radiated within a certain angle. Thus, the indoor VLC channel can be modelled by combining the line of sight (LOS) path (Fig. 15 (a)) as well as a single one-hop reflected path (Fig. 15 (b)).

The expectation of the accumulated reward gap function is defined as the metric for characterizing the performance of our AP selection scheme, which represents the difference between the maximum theoretical reward and the actually acquired reward after sequential decision making experiments relying on the system's decision probability distribution, which is formulated as.

$$R_P(K) = \max_{i_1, i_2, ..., i_K} E\left[\sum_{k=1}^{K} Q(i_k, t_k)\right] - E\left[\sum_{k=1}^{K} Q(a_k, t_k)\right],$$
(26)

where $Q(i_k, t_k)$ denotes the user rate associated with the kth decision round in terms of the access decision i_k at the instant t_k, with a_k being the actual access decision.

Furthermore, in [278] a pair of multi-armed bandit learning techniques, i.e. the 'exponential weights for exploration and exploitation' (EXP3) as well as the 'exponentially-weighted algorithm with linear programming' (ELP), were advocated for updating the AP-assignment decision probability distribution of each AP at each time instant for the sake of improving the link throughput based on the probability distribution $R_P(K)$ of (26)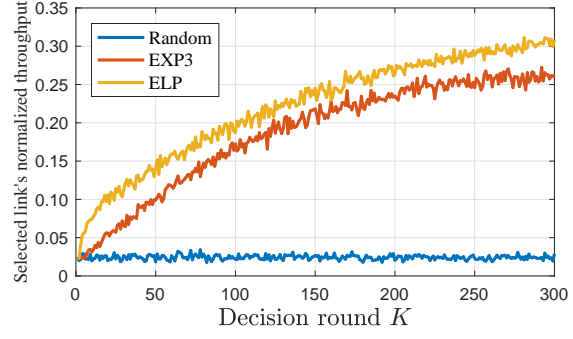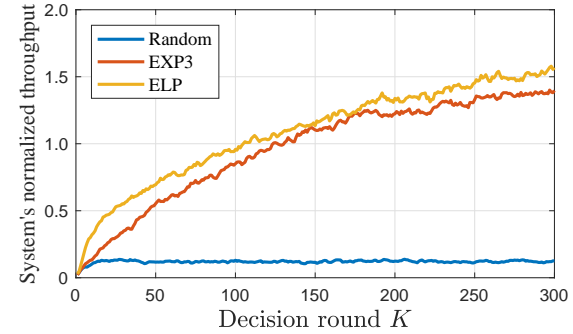. More explicitly, in contrast to the trial-and-error EXP3 algorithm, the ELP based AP selection algorithm was constructed for taking into account both the partially observed conditions of the APs as well as the network topology.

The theoretical upper bound of the expected value of the accumulated reward gap function of the EXP3- and ELP-based multi-armed bandit learning algorithms was also derived in [278]. In Fig. 16 and Fig. 17, the normalized throughput of the selected VLC links and of the whole system relying on the EXP3-based, ELP-based as well as on random LED AP selection schemes was compared. By contrast, the random selection scheme granted an identical decision probability of accessing any of the M LEDs, namely $1/M$, for each lamp at each decision-making time instant. It was assumed that the negative exponential departure probability of each downloading service was $\varsigma = 0.2$. Moreover, the initial state of the number of downloading services supported by each lamp was randomly chosen between $[1, 30]$. Upon increasing the number of decision rounds K, the EXP3- and ELP-based selection schemes had a higher accumulated normalized throughput than random selection. Furthermore, relying on more neighbor observation information as well as by exploiting the connection of the LED lamps, the ELP-based AP-selection scheme was shown to outperform that based on EXP3.

B. MDP & POMDP and Their Applications

1) Methods: The classic Markov decision process (MDP) [279] constitutes a framework of making decisions in the context of a discrete-time stochastic environment of Markov state transitions, which provides the decision maker with the optimal actions to opt for at each state. It has been used in a wide range of disciplines, especially in automatic control [280]. The goal of the decision maker, generally speaking, is to maximize the cumulative reward received over a long run and to find the corresponding optimal policy π^* which represents a mapping from each state to the specific probabilities of choosing each legitimate action.

In an MDP model, the system's state transition follows the Markovian property, where the system's response at time epoch $(t+1)$ depends exclusively on the current state and on the agent's action at time epoch t. Mathematically, at time epoch t, the system is in a certain state S, where the agent selects a legitimate action A that is available in the state S. As a result, the system then acts at the next time epoch $(t + 1)$ by moving into a new state S' relying on the system's state transition probability of $p(S'|S, A)$. At the same time, the decision maker receives the corresponding reward $r(S, A)$. The associated value function is then defined for quantifying how well the agent carries our its action over a long run commencing from the initial state S_0, which can be formulated as:

$$v_\pi = E_\pi \left[\sum_{t=0}^{\infty} \gamma^t r_{t+1} | S_t \right], \qquad (27)$$

where γ represents the discount factor and the mapping $\pi(A|S)$ represents the probability of opting for action A in the state S. Hence, the optimal policy π^* can be formulated by maximizing the value function considered, i.e. we have $\pi^* = \arg\min_\pi v_\pi$. The maximization of the value function can be reformulated as an iterative equation with the aid of Bellman's optimality theorem [281], which is given by:

$$\pi^*(A \mid S) = \arg\max_A v^*(S)$$
$$= \arg\max_A \sum_{S'} p(S' \mid S, A) \left[r(S, A) + \gamma v^*(S') \right]. \qquad (28)$$

By contrast, as an extension of MDP, the partially observable Markov decision process (POMDP) only relies on partial knowledge about the hidden Markov system which is eminently suitable for scenarios, where the agent cannot directly observe the underlying system's state transitions. Hence, the agent has to constitute belief states and the associated belief transition function by relying on a set of observations instead of the real system states. In a nutshell, the POMDP framework can be formulated as a quintuple of $\langle S, b, A, B, r \rangle$, i.e.

- *System's State S:* The system's state S represents the system's legitimate state;
- *Belief State b:* The belief state b benchmarks the degree of the similarity between each of the system's legitimate state S and the state estimated by the agent;

- *Action A:* The action A denotes the specific action that can be selected in the given state;
- *Belief Transition Function B:* The belief transition function $B(b' \mid b, A)$ represents the probability of the belief state traversing from b to b' conditioned on selecting action A;
- *Reward Function r:* The reward function $r(b, A)$ quantifies the immediate reward received by performing the selected action.

Similarly, the optimal policy π^* can be obtained by solving the optimization problem of:

$$\pi^*(A \mid b) = \arg\max_A v^*(b)$$
$$= \arg\max_A \sum_{b'} B(b' \mid b, A) \left[r(b, A) + \gamma v^*(b') \right]. \qquad (29)$$

2) Applications: As another important decision-making tools, which is different from the multi-armed bandit solutions, MDP/POMDP should firstly model the environment relying on either fully or partially observed knowledge. To elaborate a little further, Massey *et al.* [282] proposed an MDP based downlink service scheduling policy for wireless service providers. Considering the time-sensitive nature of wireless tele-traffic patterns, their proposed scheduling policy was capable of maximizing the expected reward for the wireless service provider in the context of a multiplicity of services. In [283], Tang *et al.* resorted to the MDP approach for enhancing a basic node-misconduct detection method, where a novel reward-penalty function was defined as a function of both correct and wrong decisions. The resultant adaptive node-misconduct detector maximized this reward-penalty function in diverse network states. Moreover, Kong *et al.* conceived a discrete-time MDP (DTMDP) aided mechanism [284] for dynamically activating and deactivating certain resources of the B-S in the context of time-varying network traffic. More explicitly, at each decision round, the DTMDP had the option of activating a new resource module, deactivating the currently active resource module and no operation. The proposed switching mechanism reduced the power consumption, i.e. improved the energy efficiency at the BS.

As a further development, relying on the POMDP paradigm, Tseng *et al.* [285] designed a cell selection scheme for improving the network's capacity, where the full cell loading status was not observable. Hence, it predicted the unavailable cell loading information from set of non-serving base stations and then took actions for improving the various performance metrics, including the system's capacity, the handover time as well as the mobility management as a whole. Moreover, the belief state was defined for representing the state uncertainty in terms of the statistical probability of a cell's specific loading state. The simulation results of Tseng *et al.* [285] showed that their solution outperformed the conventional signal-strength aided and load-balancing based methods. In order to save the energy of sensors, Fei *et al.* [286] proposed a POMDP aided K-sensor scheduling policy, which guaranteed the sensors' high-quality coverage and reduced
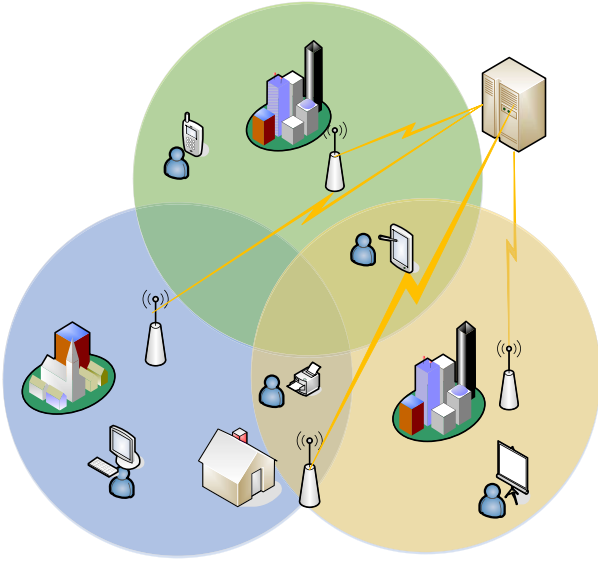
3) An Example: As shown in Fig. 18, the 'super-WiFi' network concept has been originally proposed for nationwide Internet access in the USA. However, the traditional mains power supply is not necessarily ubiquitous in this large-scale wireless network. Furthermore, the non-uniform geographic distribution of both the BSs and of the tele-traffic requires carefully considered user-association. Relying on the rapidly developing energy harvesting techniques, in [292], a POMDP-based access point selection strategies was conceived for an energy harvesting aided super-WiFi network.

It was assumed that both the battery states as well as the user access states were completely observable. However, in practice the solar radiation intensity changes over time in a year, as influenced by the weather conditions. Furthermore, the radiation sensors have a limited sampling rate, which makes it hard to simultaneously record the solar radiation intensity and to accurately estimate the system's battery state. Fortunately, relying on historical solar radiation observation data provided by the University of Queensland, Australia [293], in a short period of time, say, within an hour, the real-time harvested solar power can be modeled as $P_H = \theta + \kappa$, where θ is constant for an hour, while κ is a small perturbation. Moreover, multiple factors, such as the effective irradiation area, the clouds' distribution, the sensors' operating status, etc. may independently affect the harvested power. Relying on the central-limit theorem, the perturbation κ can be regarded as being Gaussian distributed. Hence, the distribution of P_H can be written as $P_H \sim \mathcal{N}(\theta, \sigma^2)$, where θ and σ^2 can be learned from the harvested data set.

Moreover, a queue-based user-association state model as well as a dynamic battery state model was established. Hence, the system's state having K APs is constituted by both the user-association states as well as by the battery states. Let $U = (U_1, U_2, ..., U_K)$ denote the user-association states, while $B = (B_1, B_2, ..., B_K)$ represent the AP battery states, where $U_k \leq U_M$ and $B_k \leq B_M - 1$. Furthermore, the super-WiFi system state can be written as a $2K$-element vector $S = (U_1, B_1, U_2, B_2, ..., U_K, B_K)$, which includes both the K APs' user-association states and the K APs' battery states. Assuming the independence of each AP's two sub-states, the system's state transition probability can be expressed as:

$$P(S'|S) = \prod_{k=1}^{K} \Phi_k(U_k'|U_k, A)\Psi_k(B_k'|B_k, A), \quad (30)$$

where A represents the users' actions in terms of which available APs they request association with.

Since the requesting users only have partial knowledge of the entire super-WiFi system's state, relying on the above definitions and hypotheses, we construct the POMDP decision-making model in terms of a quintuple of $\langle S, b, A, B, r \rangle$ as mentioned above. The POMDP formulation can be reduced to a belief MDP with the aid of the belief state vector. Therefore, the expected reward of the system relying on strategy Π after an infinite number

Fig. 18: The scenario of the 'super-WiFi' network [292]. ©IEEE

the total energy consumption. Similarly, by striking a trade-off between the detection performance and energy consumption, Zois *et al.* [287] designed a POMDP aided sensor node selection scheme for WBANs by maximizing the system's lifetime as well as optimizing the physical state detection accuracy. The main goal of the sensor node selection was to devise a schedule under which the sensors alternated between the active state and the dormant state relying on the specific network activity. Upon relying on the decentralized POMDP (DEC-POMDP), Pajarinen *et al.* [288] proposed a MAC solution, which promptly adapted both to the spatial and temporal opportunities facilitated by the wireless network dynamics, which yielded an increased throughput and reduced latency compared to the traditional carrier-sense multiple access relying on conventional collision avoidance (CSMA/CA) methods. Here, the POMDP tackled the uncertainty both in the environment's evolution and in the associated inaccurate observations. Thanks to the cross layer optimization employed, more information can be gleaned from the lower layers for enhanced network condition estimation. Then, Xie *et al.* [289] used the POMDP model for solving the frame size selection problem of the ubiquitous transmission control protocol (TCP) with the objective of improving the total estimated throughput by striking a tradeoff between the contention probability and back-off time based on the current network condition. Furthermore, Michelusi and Mitra [290] conceived a cross-layer framework for jointly optimizing the spectrum-sensing and access processes of cognitive wireless networks with the objective of maximizing the throughput of the SU under a strict constraint on the maximal performance degradation imposed on the PU. Furthermore, the high complexity of the POMDP formulation was mitigated by a low-dimensional belief representation, which was achieved by minimizing the Kullback-Leibler divergence defined in [291].

of time slots can be written as:

$$V^{\Pi}(S|S^0) = E[\sum_{t=0}^{\infty} \gamma^t r^t(S^t, \Pi(S^t)) b(S^t)|S^0], \quad (31)$$

where S^0 is the initial system state, while $b(S)$ is the belief state vector reflecting the grade of similarity between the current estimated state and the legitimate system state S. Moreover, r is the immediate reward of the system and γ represents the discount rate. Then, the optimal strategy can be constructed by invoking dynamic programming aided iterative algorithms for maximizing the expected reward function.

Bearing in mind the large values of K, U_M and B_M, as well as the users' rapidly fluctuating arrival rate λ and departure rate μ, obtaining the optimal POMDP solution may face the curse of dimension disaster. In order to reduce the computational complexity, a suboptimal algorithm was proposed in [292]. Explicitly, Algorithm 2 of [292] aimed for maximizing the expectation of the system's energy function, which was defined as:

$$H[b(S)] = \sum_{S\in\mathbb{S}} b(S) \sum_{i=1}^{K} \min_{\Pi} \{E_{iR} + E_{iH} - E_{iC}, E_{max}\},$$
$$(32)$$

where E_{iR} represents the residual energy of AP i, while E_{iH} is its energy harvested under the assumption that the harvested power level remains quasi-static during the information transmission interval and E_{iC} denotes the energy consumption. Finally, E_{max} is the capacity of the AP's battery.

The efficiency of the AP selection algorithms proposed in [292] was compared in terms of the system's access efficiency defined as $\xi = N_S/T_S$, where N_S is the total number of successful access attempts during the entire simulation time T_S. In Fig. 19 and Fig. 20, multiple APs ($K = 2$) are considered with the maximum number of admitted users being $U_M = 1$, while having a maximum number of battery states given by $B_M = 3$. Moreover, the departure rate is $\mu = 0.05$. We may conclude from Fig. 19 that a highly loaded system makes the carrier-sense multiple access with collision detection (CSMA/CD) method almost useless, when the users' arrival rate reaches a certain value. As shown in Fig. 20, where $\lambda = 0.4$, the system's access efficiency recorded for all the AP selection algorithms only increases with the solar radiation intensity in a relatively small range. However, the performance of the CSMA/CD, CSMA/CA[7], as well as of the random selection algorithm remains unchanged, regardless of the increase in solar radiation intensity. Moreover, the suboptimal Algorithm 2 of [292] is capable of outperforming the POMDP method at a strong solar radiation intensity, which may be deemed to be the result of the approximations and hypotheses inherent in the POMDP model.

[7]Strictly speaking, the CSMA/CD and CSMA/CA in this paper are different from the Ethernet's data link layer protocols. Here, both of them represent the access control mechanisms. We use the same acronym CSMA/CD and CSMA/CA for convenience.
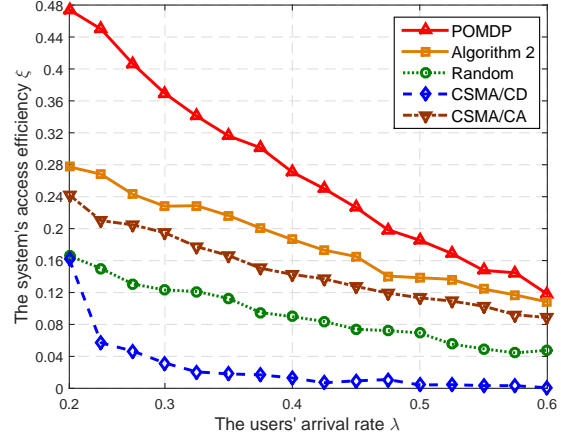

Fig. 19: The system access efficiency versus users' arrival rate for different access schemes ($K = 2$, $U_M = 1$, $B_M = 3$). ©IEEE
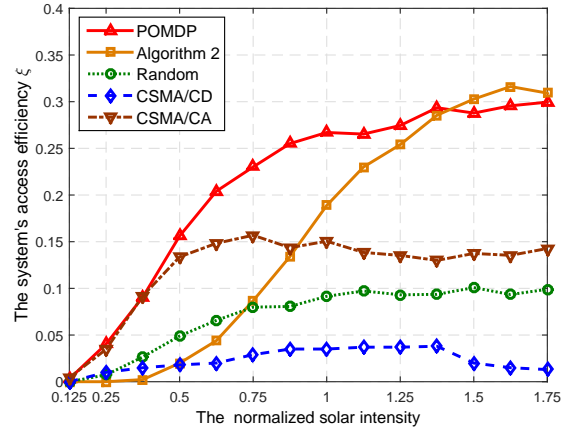

Fig. 20: The system access efficiency versus solar radiation intensity for different access schemes ($K = 2$, $U_M = 1$, $B_M = 3$). ©IEEE

C. Temporal Difference Learning and Its Applications

1) Methods: Temporal-difference (TD) learning is a model-free reinforcement learning method, which is capable of directly gleaning knowledge from raw experience without a model of the environment or receiving delayed reward, which can be typically viewed as a combination of Monte Carlo methods and of dynamic programming. More specifically, it samples the environment like the Monte Carlo methods, and then updates the corresponding parameters relying on current estimates like dynamic programming does. By contrast, TD learning operates in an on-line fashion by relying on the result of a single time step, rather than waiting for the final outcome until the end of an episode of the Monte Carlo method. Moreover, it has an advantage over the dynamic programming methods since it does not require a model of the state transition probabilities as shown in Fig. 21. TD learning can be readily invoked for finding an optimal action policy for any finite MDP associated with an unknown system model. Fig. 21 shows the difference between the MDP, POMDP
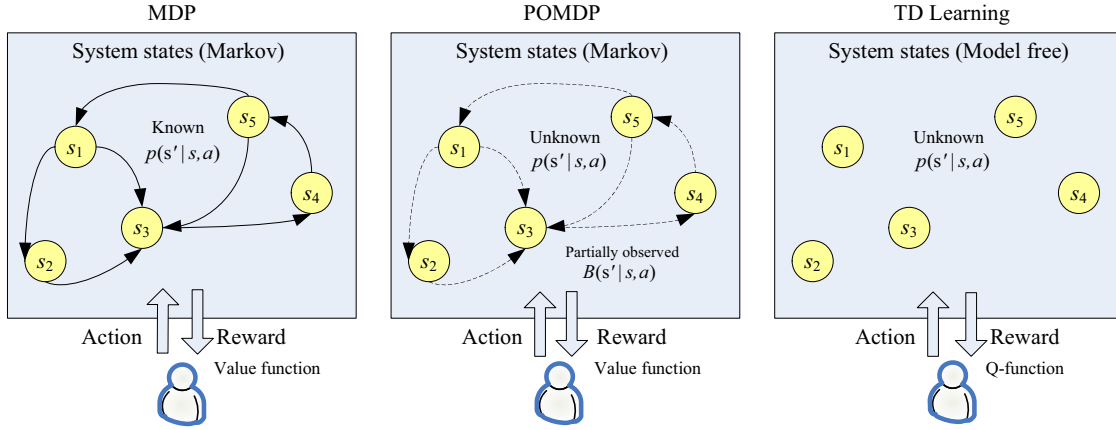
Fig. 21: The comparison between the MDP, POMDP and TD learning.

and TD learning.

A pair of popular representatives of the TD learning family are constituted by the Q-learning and by the "state-action-reward-state-action" (SARSA) technique, which interacts with the environment and updates the state-action value function, namely the Q-function, based on the action it takes. In contrast to SARSA, Q-learning updates the Q-function relying on the *maximum reward* provided by one of its available actions. Specifically, the update of the Q-function in SARSA can be formulated as [55]:

$$Q(S, A) \leftarrow (1 - \alpha)Q(S, A) + \alpha\left[r + \gamma Q(S', A')\right], \quad (33)$$

while in Q-learning, the update of the Q-function can be cast as [88]:

$$Q(S, A) \leftarrow (1 - \alpha)Q(S, A) + \alpha\left[r + \gamma \max_{A \in \mathbb{A}} Q(S', A)\right], \quad (34)$$

where S represents the system's state and A is the action selected by the agent, whilst $\mathbb{A}$ represents the available set of actions. Moreover, α is the update weighting coefficient and γ denotes the discount factor. As for the convergence analysis, SARSA is capable of converging with probability 1 to an optimal policy as well as to an optimal state-action value function, provided that all the state-action pairs are visited a sufficiently high number of times. However, because of the independence of making an action and that of updating the Q-function, Q-learning has no delayed reward as TD-learning, which tends to facilitate an earlier convergence than SARSA [47].

2) Applications: As a benefit of being free from modeling the environment, TD learning is capable of providing competent decisions even in unknown environments. Table IV summarizes a variety of compelling applications found in wireless networks for both SARSA and Q-learning along with their brief description.

VI. Deep Learning in Wireless Networks

A. Deep Artificial Neural Networks and Their Applications

1) Methods: Artificial neural networks [304] constitute a set of algorithms conceived by imitating the interaction amongst neurons in the human brain, which are designed to extract features for clustering and classification tasks.

In a common artificial neural network (ANN) model [305], the input of each artificial neuron is a real-valued signal, where the output of each artificial neuron is subjected to by some non-linear functions, namely the so-called activation functions. This is formulated as:

$$y_k = \psi(u_k + b_k), \quad (35)$$

where

$$u_k = \Sigma_{i=1}^m w_{ki} x_i. \quad (36)$$

Here, x_i and y_k represent the input signal and output signal, respectively, while w_{ki} is the associated weight and b_k is the bias. Moreover, $\psi(\bullet)$ is the activation function. Artificial neurons and their connections typically use a weighting factor for adjusting the "speed" of the learning process. Moreover, artificial neurons are organized in the form of layers. Different layers perform different transformations of their inputs. Basically, the input signals travel from the first layer to the last layer, possibly via multiple hidden layers.

The general deep neural network (DNN) is characterized by multiple hidden layers between the input and output layers as shown in Fig. 22 (a), which is capable of modeling complex relationships of the processed data with the aid of multiple non-linear transformations. In a DNN, the provision of extra layers facilitates the composition of features from lower layers, which is beneficial in terms of more accurately modeling complex data than a 'shallow' network having a single hidden layer. In comparison to 'shallow' networks, DNNs are capable of attaining a performance, despite requiring less training. Furthermore, DNNs may be viewed as a type of feed-forward network, where the processed data flows in the direction from the input layer to the output layer without looping back. The forward propagation algorithms and backward propagation algorithms represent a step-by-step problem solving process. Specifically, the forward propagation algorithms apply the carefully trained weight matrixes and bias vectors for carrying out the associated linear and

TABLE IV: Compelling Applications of the SARSA Algorithm and Q-learning in Wireless Networks

Paper	Method	Scenario	Application & Description
[294]	reduced-state SARSA	cellular network	dynamic channel allocation considering both mobile traffic and call handoffs.
[295]	on-policy SARSA	CR network	distributed multiagent sensing policy relying on local interactions among SU
[296]	on-policy SARSA	MANET	energy-aware reactive routing protocol for maximizing network lifetime
[297]	on-policy SARSA	HetNet	resource management for maximizing resource utilization and guaranteeing QoS
[298]	approximate SARSA	P2P network	energy harvesting aided power allocation policy for maximizing the throughput
[299]	Q-learning	WBAN	power control scheme to mitigate interference and to improve throughput
[300]	Q-learning	OFDM system	adaptive modulation and coding not relying on off-line training from PHY
[301]	Q-learning	cooperative network	efficient relay selection scheme meeting the symbol error rate requirement
[302]	decentralized Q-learning	CR network	aggregated interference control without introducing signaling overhead
[303]	convergent Q-learning	WSN	sensors' sleep scheduling scheme for minimizing the tracking error

activation operations. By contrast, the backward propagation algorithms, which are widely used in the industrial field define a so-called loss function for quantifying the difference between the output produced by the training samples' and the real output. Then, by minimizing the loss or error we can optimize the connection weights of the neural network. Although there are impressive applications of DNN, improving their convergence behavior requires further research.

By contrast, in a recurrent neural network (RNN) a neuron in one layer is capable of connecting to the neurons of previous layers. Therefore, a RNN is capable of exploiting the dynamic temporal information hidden in a time sequence and it beneficially exploits its "memory" inherited from the previous layers for processing the future inputs, as shown in Fig. 22 (b). RNNs have the advantage of requiring reduced training and impose a reduced computational complexity, because fewer tensor operations have to be calculated. Popular algorithms used for training the RNN include the real time recurrent learning technique of [306], the causal recursive backpropagation algorithm of [307] and the backpropagation through time algorithm of [308], just to name a few.

The convolutional neural network (CNN) constitutes a popular class of feed-forward deep artificial neural networks, which only require modest preprocessing.They are capable of reducing the number of parameters step-by-step as we move from the input layer to the hidden layers. As seen in Fig. 22 (c), a basic CNN architecture is composed of an input layer, an output layer as well as multiple hidden layers, which are often referred to as convolutional layers, pooling layers and fully connected layers. More particularly, the convolutional layers carry out the convolution operation, also termed as the cross-correlation operation, which generates a multi-dimensional feature map relying on a number of so-called filters. The CNN has been successfully used both in image and video recognition [309], in natural language processing [310], in recommender systems [311], etc. Fig. 22 contrasts the basic architecture of DNN, RNN and CNN, respectively.

2) Applications: In this subsection, we will consider the benefits of deep artificial neural network algorithms in a variety of wireless networking scenarios. As mentioned before, deep artificial neural networks are capable of capturing the non-linear and often dynamically varying relationship between the inputs and outputs. Hence they have a powerful prediction, inference and data analysis capability by exploiting the vast amount of data generated both by the environment and by the users. As for *learning from the environment*, we are able to harness DNNs trained by the data gleaned over the air for the sake of channel estimation [312], interference identification [313], localization [314]–[318], etc. By contrast, with regard to *learning from the users or devices*, DNN algorithms can also be used for predicting the users' behaviors, such as their content interests [319], mobility patterns [320], etc. in order to beneficially design the dynamic content caching of BSs and to efficiently allocate wireless resources, for example.

Traditional signal processing approaches supported by statistics and information theory in communication systems substantially rely on accurate and tractable mathematical models. Unfortunately, however, practical communication systems may have a range of imperfections and non-linear factors, which are difficult to model mathematically. Given that DNN algorithms do not require a tractable model, they are capable of *remedying the imperfections in the physical layer* by learning both from the environment and from previous inputs relying on a specific hardware configuration. To elaborate, Ye *et al.* [312] proposed a DNN aided channel estimation method for learning the wireless channel characteristics, such as the nonlinear distortion, interference and frequency selectivity. The DNN aided channel estimation method was shown to be more robust than traditional methods, especially in the context of having fewer training pilots, in the absence of cyclic prefix, as well as in the face of nonlinear clipping noise. Apart from estimating the channel characteristics, DNNs can also be used for classifying modulated signals in the physical layer. Rajendran *et al.* [321] conceived a date-driven automatic modulation classification (AMC) scheme hinging on the long short term memory (LSTM) aided RNN, which captured the time domain (TD) amplitude and phase information of modulation schemes carried in the training data without expert knowledge. Their simulations showed that the novel AMC had an average classification accuracy of about 90% in the context of time-varying SNR ranging from 0dB to 20dB. As for signal detection, Farsad and Goldsmith [322] developed a deep learning aided signal detector, where the transmitted signal can be efficiently estimated from its corrupted version observed at the receiver. The detector was trained relying on known transmitted signals, but without any knowledge of the underlying wireless channel model and estimated the likelihood of each symbol, which was beneficial for carrying out soft decision error correction afterwards. In the application of
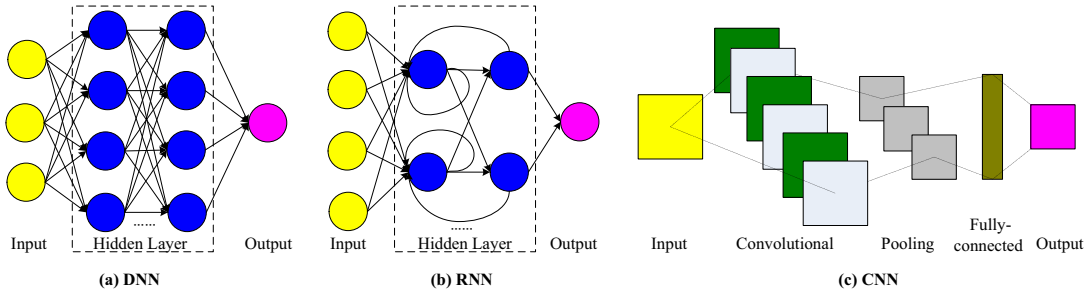
Fig. 22: The basic architecture of DNN, RNN and CNN.

interference identification, Schmidt *et al.* [313] proposed a 64-feature-map assisted CNN based wireless interference identification scheme. The CNN model learned the relevant features through self-optimization during the GPU based training process, which was first designed in [323]. By carefully considering the realistic capability of wireless sensors, the model relied on the time- and frequency-limited sensing snapshots having the duration of 12.8 μs as well as the bandwidth of 10MHz. The proposed CNN based wireless interference identifier was shown to have a higher identification accuracy than the state-of-the-art schemes in the context of low SNRs, such as -5dB, for example.

Furthermore, we can use DNNs for *modelling the entire physical layer of a communication system* without any classic components such as source coding, channel coding, modulation, equalization, etc. In [324], O'Shea *et al.* used a DNN to represent a simple communication system with one transmitter and one receiver that can be trained as a so-called auto-encoder without knowing the accurate channel model. Moreover, a CNN algorithm was conceived for modulation classification based on both sampled radio frequency time-series data and expert knowledge integrated by radio transformer networks (RTN). Additionally, O'Shea *et al.* [325] extended the DNN aided auto-encoder to a single user MIMO communication scenario, where the physical layer encoding and decoding processes were jointly optimized as a single end-to-end self-learning task. Their simulation results showed that the auto-encoder based system outperformed the classic space time block code (STBC) at 15dB SNR. Furthermore, Dörner *et al.* [326] also developed a DNN based prototype system solely composed of two unsynchronized off-the-shelf software-defined radios (SDR). This prototype system was capable of mitigating the current restriction on short block lengths.

DNNs also play a critical role in supporting a variety of compelling upper layer applications, such as traffic prediction [327], packet routing [328] and control [329], traffic offloading [330], resource allocation [331], attack detection [332], just to name a few. For instance, Wang *et al.* [327] presented a hybrid deep learning aided structure for spatio-temporal traffic modeling and prediction in cellular networks by mining information from the China Mobile dataset. It used a novel deep learning aided auto-encoder for modeling the spatial features of wireless traffic, while using LSTM units for temporal modeling. Additionally, Kato *et al.* [328] proposed a supervised DNN aided traffic routing scheme, which outperformed the classic open shortest path first (OSPF) scheme in terms of requiring a lower overhead, whilst maintaining a higher throughput and lower delay. By contrast, a real-time deep CNN based traffic control mechanism learning from previous network anomalies was conceived by Tang *et al.* [329], which substantially reduced the average delay and packet loss rate. Hence, deep learning aided traffic control may indeed constitute a potential candidate for gradually replacing traditionally routing protocols in future wireless networks. Furthermore, Li *et al.* [330] integrated both the DNN structure and the edge computing technique into the multimedia IoT, which was able to improve the efficiency of multimedia processing. Sun *et al.* [331] treated the power control problem in interference-limited wireless networks as a 'black box'. They proposed an 'almost-real-time' power control algorithm relying on a DNN structure trained by simulated data. In comparison to traditional mathematical tools, the approximation error of the DNN aided algorithm is closely related to the depth of the DNN considered. As for network security issues, for example, He *et al.* [332] constructed a conditional deep belief network (CDBN) for the real-time detection of malicious false data injection (FDI) attacks in the smart grid, which was trained by historical measurement data. The simulations conducted using the IEEE 118-bus test system and the IEEE 300-bus test system showed that the CDBN aided FDI detection scheme was resilient to the environmental noise and had a higher detection accuracy than its SVM aided counterparts.

As a successful example of learning from the environment, DNNs are beneficial in terms of extracting electromagnetic fingerprint information from the wireless channel for *indoor localization*. In [314]–[316], Wang *et al.* proposed a DNN having three hidden-layer for training the calibrated CSI phase data, where the fingerprint information was represented by the DNN's weights. Their experimental results showed that the DNN aided localization scheme performed well in different propagation environments, including an empty living room, and a laboratory in the presence of mobile users. In [317], Wang *et al.* proposed a deep learning method for supporting device-free wireless localization and activity recognition relying on learning from the wireless signals around the

target, where a sparse auto-encoder network was used for automatically learning the discriminative features of wireless signals. Furthermore, a softmax-regression-based framework [333] was formulated for the location and activity recognition based on merged features. Moreover, in [318], Zhang *et al.* constructed a four-layer DNN for extracting reliable high level features from massive Wi-Fi data, which was pre-trained by the stacked denoising auto-encoder. Additionally, an HMM aided high-accuracy localization algorithm was proposed for smoothening the estimate variation. Their experimental results showed a substantial localization accuracy improvement in the context of a widely fluctuating wireless signal strength.

With regard to learning from users or devices, Ouyang *et al.* [320] conceived a CNN-aided online learning architecture for understanding human mobility patterns relying on analyzing continuous mobile data streams. Al-Molegi *et al.* [334] integrated both the spatial features gleaned from GPS data and the temporal features extracted from the associated time stamps for predicting human mobility based on a RNN. Moreover, Song *et al.* [335] proposed an intelligent deep LSTM RNN based system for predicting both human mobility and the specific transportation mode in a large-scale transportation network, which was beneficial in terms of providing accurate traffic control for intelligent transportation systems (ITS). Additionally, a mobility prediction technique relying on a complex extreme learning machine (CELM) was developed by Ghouti *et al.* [336] in order to jointly optimize both the bandwidth and the power MANETs. In [337], both the multi-layer perception and RNN models were employed by Agarwal *et al.* for characterizing the activity of primary users in CR networks, where three different traffic distributions, namely Poisson traffic, interrupted Poisson traffic and self-similar traffic were used for training the related models.

Table V lists a range of typical applications of DNNs along with a brief description.

B. Deep Reinforcement Learning and Its Applications

1) Methods: The deep reinforcement learning technique is constituted by the integration of the aforementioned DNNs and reinforcement learning. Explicitly, in deep reinforcement learning methods, DNNs are used for approximating certain components of reinforcement learning, including the state transition function, reward function, value function and the policy. These components can be viewed as a function of the weights in these DNNs, which can be updated with the aid of the classic stochastic gradient descent.

In particular, the deep Q-Network (DQN) constitutes the first deep reinforcement learning solution, which was proposed by Mnih *et al.* in 2015 [58], which avoids the instability of the reinforcement learning algorithm, which may even become divergent when its action-value function is approximated relying on a non-linear function. To elaborate a little further, DQN stabilizes the training process of the action-value function approximation by relying on experience replay. Furthermore, DQN only requires modest domain knowledge. The deep Q-learning

algorithm in DQN is a variant of the classical Q-learning algorithms, which is integrated with the deep CNN model, where the convolutional filters seen in Fig. 22 (c) are used for representing the effects of receptive fields. One of the outputs of the deep CNN involved yields the specific value of the Q-function for a possible action. Beyond the DQN, substantial efforts have also been invested in improving the performance and stability, as exemplified by the double DQN [342] and the dueling DQN [343]. Thanks to the powerful feature representation capability of DNNs and of the reinforcement learning algorithms, DQN performs well in a range of compelling applications as exemplified by the AlphaGo, which is the first super-human program to defeat a professional human chess-player.

2) Applications: Deep reinforcement learning is eminently suitable for supporting the interaction in autonomous systems in terms of a higher level understanding of the visual world, which can be readily applied to a diverse analytically intractable problems in future wireless networks.

Given the intrinsic advantages of the reinforcement learning in environment in interactive decision making, it may play a significant role in the field of control decision [344], [345]. Specifically, Zhang *et al.* [344] proposed a model-free UAV trajectory control scheme relying on deep reinforcement learning for data collection in smart cities, where a powerful deep CNN was used for extracting the necessary features, while a DQN model was used for decision making. Given the sensing region and the related tasks, this algorithm supported efficient route planning for both the UAVs and mobile charging stations involved. In [345], a deep reinforcement learning aided communication-based train control system was conceived by Zhu *et al.* which jointly optimized the communication handoff strategy and the control functions, while reducing the energy consumption. Real channel measurements and real-time train position information were used for training the DQN model, which resulted in optimal communication and control decisions.

Furthermore, the resource allocation problems of wireless networks, such as energy scheduling, traffic scheduling, caching decisions, user association, etc. can be efficiently solved by deep reinforcement learning at a low computational complexity [153], [346]–[351]. For example, Zhang *et al.* [153] proposed a deep Q-learning model for system's dynamic energy scheduling, which relied on the amalgamated stacked auto-encoder and Q-learning model. More specifically, the stacked auto-encoder was used for learning the state-action value function of each strategy in any of the available system states. Moreover, Xu *et al.* [346] proposed a deep reinforcement learning framework for power-efficient resource allocation in CRANs, which optimized the expected and cumulative long term power consumption, including the transmit power consumption, the sleep/active transition power consumption as well as the RRU's power consumption. A two-step deep reinforcement learning aided decision making scheme was conceived, where the learning agent first decides on activating/deactivating the sleeping mode of

TABLE V: Compelling Applications of Deep Artificial Neural Networks in Wireless Networks

Paper	Application	Method	Description
[312]	channel estimation	DNN	learn nonlinear distortion, interference and frequency selectivity of wireless channels
[321]	modulation classification	RNN	capture amplitude and phase information without expert knowledge
[322]	signal detection	DNN	transmit signal detection from noisy and corrupted signals without underlying CSI
[313]	interference identification	CNN	learn features through self-optimization during the GPU based training process
[324]	PHY representation	DNN	represent simple system having one transmitter and receiver without accurate CSI
[325]	PHY representation	DNN	represent single user MIMO system relying on DNN aided auto-encoder
[326]	software-defined radio	DNN	be capable of easing the current restriction on short block lengths
[327]	traffic prediction	DNN	deep auto-encoder and LSTM for modeling spatial and temporal features
[328]	packet routing	DNN	traffic routing scheme with little signal overhead, large throughput and small delay
[329]	traffic control	CNN	consider previous network abnormalities, lower average delay and packet loss rate
[330]	traffic offloading	DNN	integrate both DNN structure and edge computing technique into multimedia IoT
[331]	power control	DNN	an almost-real-time power control algorithm in interference-limited wireless networks
[332]	network security	DBN	a real-time detection of malicious false data injection attack in smart grid
[314]–[318]	indoor localization	DNN	device-free wireless localization and recognition by learning from ambient wireless signals
[320]	mobility prediction	CNN	learn human mobility pattern relying on analyzing continuous mobile data stream
[334]	mobility prediction	RNN	integrate spatial feature from GPS and temporal feature from associated time stamps
[335]	transportation mode	RNN	predict both human mobility and transportation mode for large-scale transport networks
[337]	activity prediction	RNN	characterize primary users' activity in CR with different traffic distribution
[338]	traffic prediction	CNN	traffic modeling and prediction based on a convolutional long short-term memory network
[339]	resource allocation	DNN	universal DNN framework for a number of common wireless resource allocation problems
[340]	PHY representation	CNN	modulation, equalization, and demodulation with a convolutional autoencoder
[341]	link scheduling	CNN	optimal scheduling of interfering links in dense wireless networks

each RRU, and then determines the optimal beamformer's power allocation. As for traffic scheduling, Zhu *et al.* [347] designed a stacked auto-encoder assisted deep learning model for packet transmission planning in the face of multiple contending channels in cognitive IoT networks, which aimed for maximizing the system's throughput. In this architecture, MDP was used for modelling the system states. Given the large state-action space of the system, the stacked auto-encoder was used for constructing the mapping between the state and the action for accelerating the process of optimization. Furthermore, a deep Q-learning algorithm was conceived for designing both the cache allocation and the transmission rate in content-centric IoT networks for the sake of maximizing the long-term QoE [348], where He *et al.* considered both the networking cost as well as the users' mean opinion score. In [349] and [350], He *et al.* proposed a DQN based user scheduling scheme for a cache-enabled opportunistic interference alignment (IA) assisted wireless network in the context of realistic time-varying channels formulated as a finite-state Markov model. More specifically, the DQN was constructed by relying on a sophisticated action-value function for the sake of reducing the computational complexity. Their simulation results demonstrated that the DQN aided IA assisted user scheduling was beneficial in terms of substantially improving the network's throughput vs energy efficiency trade-off. To elaborate a little further, He *et al.* [351] utilized deep reinforcement learning for constructing their resource allocation policy relying on a joint optimization problem, which considered the programmable networking, information-centric caching as well as mobile edge computing in the context of connected vehicular scenarios. Moreover, the ε-greedy policy was utilized for striking an attractive trade-off between exploration and exploitation.

In order to curb the potentially excessive computational complexity resulting from having a large state space and to deal with its partial observability in cognitive radio networks, Naparstek and Cohen developed a distributed dynamic spectrum access scheme relying on deep multi-user reinforcement learning, where each user maps his/her current state to spectrum access actions with the aid of a DQN for the sake of maximizing the network's utility which was achieved without any message exchanges [352]. Additionally, Wang *et al.* [353] proposed an adaptive DQN algorithm for dynamic multichannel access, which was capable of achieving a near-optimal performance outperforming the Myopic policy [354][8] and the Whittle's Index-based heuristic algorithms [355][9] in complex scenarios.

Table VI lists some typical applications of deep reinforcement learning in wireless networks.

VII. FUTURE RESEARCH AND CONCLUSIONS

A. Future Machine Learning Aided Network Applications

It is anticipated that the next-generation network infrastructure will be migrating to a service-oriented architecture by taking advantage of both NFV and SDN for supporting eMBB, mMTC, and uRLLC along with attractive social media and augmented reality (AR) aided entertainment, smart manufacturing, smart energy, e-health and autonomous driving.

Given that ML can be readily executed both by agents, as well as by edge- and cloud-computing, it is eminently suitable for supporting sophisticated networking functionalities. We can categorize ML-based networking into:

- *ML-Aware Networking*: The networking entities are aware of the availability of ML functionalities and optionally might exploit advantages of ML, for example, when the user load escalates.
- *On-/Off-line ML-Aided Networking*: The network functionalities must resort to invoking ML, for example, because the search-space of joint scheduling

[8]The Myopic policy is one that simply optimizes the average immediate reward. It is called 'myopic' in the sense that it merely considers the single criterion, while it has the advantage of being easy to implement.

[9]The Whittle's Index-based algorithm is one of index heuristic algorithms which is designed to solve a problem in a more efficient way than traditional methods often used for solving NP-hard problems. More explicitly, the Whittle's Index policy is a low-complexity heuristic policy.

TABLE VI: Compelling Applications of Deep Reinforcement Learning in Wireless Networks

Paper	Scenario	Application	Description
[344]	UAV network	trajectory control	a model-free UAV trajectory control scheme in smart cities relying on DQN
[345]	ITS	train control	jointly optimize the communication handoff strategy and control performances
[153]	energy-aware network	energy scheduling	associate the stacked auto-encoder and the deep Q-learning model
[346]	CRAN	power allocation	decide RRU's sleeping mode and the optimal beamformer's power allocation
[347]	cognitive IoT	traffic scheduling	construct the mapping between states and actions relying on stacked auto-encoder
[348]	content-centric IoT	cache allocation	jointly design cache allocation and transmission rate for maximizing long-term QoE
[350]	IA network	user scheduling	obtain the action-value function relying on DQN for lowering complexity
[351]	vehicular network	resource allocation	consider programmable SDN, information-centric caching and mobile edge computing
[352]	CR network	spectrum access	distributed spectrum access for maximizing network utility without message exchanges
[353]	CR network	multichannel access	adaptive DQN aided multichannel access yielding a near-optimal performance
[356]	MEC network	resource allocation	DQN assisted cache, computation and power resources joint association
[357]	CR network	energy scheduling	energy efficiency optimization for distributed cooperative spectrum sensing
[358]	vehicular network	resource allocation	edge computing and caching resources dynamic association for IoV

and resource-allocation is excessive for full search. Depending on the specific application, this may rely either on off-line ML or on on-line ML.

- *On-line ML-Aided Networking*: The network functionalities must resort to on-line ML, for example, anticipatory mobility management requires on-line ML.

To facilitate prompt ML-aided network control, new tele-traffic samples may potentially be included for near-instantaneous learning according to the following use cases:

- In contrast to traditional networking, near-instantaneous mobility trajectory inference relies on on-line datasets for training the ML module.
- Context information relying on location information inference or on user experience as a function of latency, jitter, packet dropping probability can be taken into account.
- Planning and policy relying on ML functionalities for improving a certain work flow, an agent's specific policy and/or rewards, along with the extracted features or learning models may be exploited, where the agent may be constituted either by a network entity or a smart UE.

To expound a little further, ML can be used for diverse application scenarios, such as:

- *Channel State Estimation*: Having accurate CSI is critical for all air-interface techniques, which has been inferred/estimated with the aid of deep learning.
- *User Behavior Inference*: User behavior may be characterized for example in terms of their mobility patterns, while supporting high-quality network management and mobility management with the aid of big data analysis relying on ML.
- *Traffic Prediction*: Network intelligence and user mobility patterns can be used for predicting the wired/wireless network traffic in support of efficient network/radio resource allocation.
- *Network Security*: ML is also capable of enhancing the network security, whilst detecting attacks and intrusions.
- *Anticipatory Networking Mechanism*: Online learning can be used to develop new predictive networking mechanism. For example, anticipatory mobility management (AMM) using Naive Bayesian and recursive

belief update was conceived for ultra-low latency wireless networking [359], [360].

Based on the aforementioned future network architectural features, we list a range of research ideas on promising applications of ML in future wireless networks.

- *UAV-aided networking*: Given the agility of UAV nodes as well as the bursty and often unpredictable nature of terrestrial wireless traffic, ML models can be used both for predicting the traffic demand and for adaptively adjusting the UAVs' location.
- *mMTC and uRLLC network*: While wireless networks have primarily served communications among individuals, in the era of the IoT, wireless networking also supports myriads of machines and intelligent devices. In this era a pair of 5G operational modes - namely mMTC and uRLLC - are expected to play key roles [361]. ML is capable of enhancing conventional networks designed for mMTC, for example by invoking reinforcement learning to appropriately select the access points of MTC [362]. The uRLLC mode of operation constitutes a rather young technological territory, which can be jointly designed with mMTC [363]. To reduce the network's latency from hundreds of milliseconds as experienced in state-of-the-art mobile communications to the desirable range of just a few milliseconds, ML is capable of supporting so-called anticipatory mobility management, which integrates the naive Bayesian classification of the previously used APs and geographical regression for the predictive analysis of data. Another example of disruptive technical trend is to investigate how wireless networking impacts on the smart agents operated by ML.
- *Narrow band IoT (NB-IoT)*: NB-IoT allows a large number of low-power devices to connect to the cellular network, where the devices require long-term Internet access and dense wireless coverage. ML algorithms are capable of supporting intelligent resource allocation, optimal AP deployment and efficient access.
- *Socially-aware wireless networking*: The operation of socially-aware wireless networks relies on a variety of social attributes, where ML schemes are beneficial in terms of facilitating feature extraction, social group-of-interest formation, classification and prediction of

these social attributes, such as human mobility, social relations, behavior preference, etc.

- *Wireless Virtual reality (VR) networks*: VR networks facilitate for the users to experience and interact with immersive environments, which requires flawless audio and video data processing capability. ML algorithms have the potential of circumventing the conception of complex joint source- and channel-coding schemes by further developing the auto-encoder principles.

- *Network integration, representation and design*: ML may provide an alternative for network representation, where we can integrate each classic communication-theoretic blocks including source- and channel-encoding, modulation, demodulation, decoding, etc. into a "black box". By simply learning from and processing previous input and output signals, the receivers become capable of adaptively understanding the operational mechanism of the "black box" considered.

- *Wireless network tomography*: State-of-the-art wireless networks support a vast number of nodes, such as those of the IoT, where the provision of global information is practically impossible for each node. Hence a new class of problems arises in the context of distributed wireless networks, which is related to the acquisition of network-related information. Classic network tomography [364] defines the problem as: $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{X}$ is a J-dimensional vector of the network's dynamically fluctuating parameters, such as the link delay or traffic activity, $\mathbf{Y}$ is the I-dimensional vector of measurements and $\mathbf{A}$ is known as the routing matrix that is likely to define a rank-deficient scenario in network tomography. Such inverse matrix problems of inference may be solved by EM or Markov Chain Monte Carlo methods, and they are mathematically similar to the regression problems of ML. Network tomography can be applied to the inference of a wireless network's status and the subsequent optimization of the network's operation [365]. Further generalization using statistical inference and singular value decomposition for statistically characterizing and enhancing the radio resource utilization relies on accurate spectrum sensing in cognitive radio networks [366]. Another interesting application is to infer activities beyond sight (i.e. see-through walls) by applying variance-based radio tomography, Tikhonov least-squares regularization, and Kalman tracking [367].

- *Large-scale Pareto-optimization*: In the above-mentioned optimization problems the designer is routinely faced by having to strike a trade-off amongst conflicting factors, as exemplified by the ubiquitous bandwidth- vs. power-efficiency trade-off in wireless communications. However, there are numerous other trade-offs to be taken into account, such as for example the data-security vs. integrity trade-off or the BER vs. delay trade-off, as well as the BER vs. complexity trade-off, just to mention a few. Hence it is desirable

to find all the Pareto-optimal operating points of future wireless systems, which jointly constitute the optimal Pareto-front of multi-component optimization problems [134]. Again, none of the solutions on the optimal Pareto-front may be improved without degrading at least one of the parameters. Naturally, as we mentioned, the optimization search-space is extended every time, when a new parameter is incorporated. Hence ML constitutes a promising tool for the fertile ground of multi-component Pareto-optimization of a plethora of systems.

B. Future Wireless Networking Techniques for Multi-Agent Systems Using Machine Learning

In addition to applying ML in wireless networking, the question of how to design a wireless network for example for multi-robot systems or for multi-agent systems arises.

In this context, Legg and Hutter [368] gave an informal definition of machine intelligence: "Intelligence measures an agent's ability to achieve goals in a wide range of environments." Distributed artificial intelligence has been brought into the lime-light of artificial intelligence research over three decades ago [369], which has a pair of popular sub-disciplines: distributed problem solving and multi-agent systems. Distributed problem solving typically decomposes tasks into several not completely independent sub-problems that can be executed on different processors and then synthesizes a solution. On the other hand, multi-agent systems are typically constituted by robots, having certain goals and actions in challenging operating environments. Although the communication supporting the actions of agents has been studied for a while, under idealized simplifying assumptions, the features of realistic wireless communications and networking have not been taken into consideration [370]–[372].

An interesting study disseminated in [373] has considered the collective behavior of autonomous vehicles moving across Manhattan in New York city. Each autonomous vehicle acted as an intelligent agent using reinforcement learning. In contrast to human-to-human personal communication, the reward map and policy of another autonomous vehicle within the interaction range would be useful information to exchange. In this wireless networking scenario, the freshness of such information is critical hence ultra-low-latency wireless networking is preferred. In fact, for collaborative robots relying on their own machine learning algorithms while aiming for achieving a common goal, ultra-low-latency wireless networking is of paramount importance. Given the plethora of open research problems in networked multi-agent systems, such as their network topology, this is a compelling research area.

C. Conclusions

In summary, we have reviewed the thirty-year history of ML. We surveyed a range of typical algorithms and models conceived for supervised learning, unsupervised learning,

reinforcement learning as well as deep learning, respectively. Furthermore, we have highlighted the development tendency of wireless network techniques and a variety of representative scenarios for future wireless networks as seen in Fig. 5 and Fig. 8. We also have provided a case-by-case description of numerous compelling applications relying on ML algorithms in wireless networks as shown in Table VII, followed by a pair of detailed application examples relying on our recent research results. In comparison with state-of-the-art survey papers seen in Fig. 1, our paper overviews all the four popular kinds of learning schemes and their applications in future wireless networks, which has a full scope of how ML algorithms bear fruits in the past decades in wireless networks. Indeed, powerful ML methods are poised to occupy an important position in tackling intractable and hitherto uncharted scenarios and applications in wireless networks.

Appendix

The applications of ML in wireless networks are summarized in Table VII.

References

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, Feb. 2016.

[2] P. Rawat, K. D. Singh, H. Chaouchi, and J. M. Bonnin, "Wireless sensor networks: A survey on recent developments and potential synergies," *The Journal of Supercomputing*, vol. 68, no. 1, pp. 1–48, Apr. 2014.

[3] D.-J. Deng, Y.-P. Lin, X. Yang, J. Zhu, Y.-B. Li, J. Luo, and K.-C. Chen, "IEEE 802.11 ax: highly efficient WLANs for intelligent information infrastructure," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 52–59, Dec. 2017.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] M. Kalil, A. Al-Dweik, M. F. A. Sharkh, A. Shami, and A. Refaey, "A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks," *IEEE Access*, vol. 5, pp. 20 814–20 827, 2017.

[6] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.

[7] E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.

[8] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[9] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015.

[10] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks*, vol. 10, no. 5, pp. 988–999, May 1999.

[11] V. Vapnik, "Principles of risk minimization for learning theory," in *Advances in Neural Information Processing Systems*, Denver, Colorado, Nov. 1992, pp. 831–838.

[12] I. Guyon, V. Vapnik, B. Boser, L. Bottou, and S. A. Solla, "Structural risk minimization for character recognition," in *Advances in Neural Information Processing Systems*, Denver, Colorado, Nov. 1992, pp. 471–479.

[13] N. Sebe, I. Cohen, A. Garg, and T. S. Huang, *Machine learning in computer vision*. Springer Science & Business Media, 2005, vol. 29.

[14] K. Fu, "Learning control systems and intelligent control systems: An intersection of artificial intelligence and automatic control," *IEEE Transactions on Automatic Control*, vol. 16, no. 1, pp. 70–72, Feb. 1971.

[15] P. Baldi and S. Brunak, *Bioinformatics: The machine learning approach*. MIT Press, 2001.

[16] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[17] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 67, May 2016.

[18] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, May 2014.

[19] Cisco, "Cisco visual networking index: global mobile data traffic forecast update, 2014–2019," *Cisco White Paper*, Feb. 2015.

[20] X. Cao, L. Liu, Y. Cheng, and X. Shen, "Towards energy-efficient wireless networking in the big data era: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 303–332, Nov. 2017.

[21] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*. Springer, 2011, pp. 760–766.

[22] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: A survey," in *Artifical Intelligence for Human Computing*. Springer, 2007, pp. 47–71.

[23] A. Pentland and A. Liu, "Modeling and prediction of human behavior," *Neural computation*, vol. 11, no. 1, pp. 229–242, Jan. 1999.

[24] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1996–2018, Apr. 2014.

[25] R. V. Kulkarni, A. Forster, and G. K. Venayagamoorthy, "Computational intelligence in wireless sensor networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 1, pp. 68–96, May 2011.

[26] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1136–1159, Oct. 2013.

[27] L. Gavrilovska, V. Atanasovski, I. Macaluso, and L. A. DaSilva, "Learning and reasoning in cognitive radio networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1761–1777, Mar. 2013.

[28] A. He, K. K. Bae, T. R. Newman, J. Gaeddert, K. Kim, R. Menon, L. Morales-Tirado, Y. Zhao, J. H. Reed, W. H. Tranter *et al.*, "A survey of artificial intelligence for cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1578–1592, May 2010.

[29] T. Park, N. Abuzainab, and W. Saad, "Learning how to communicate in the Internet of Things: Finite resources and heterogeneity," *IEEE Access*, vol. 4, pp. 7063–7073, Nov. 2016.

[30] A. Forster, "Machine learning techniques applied to wireless ad-hoc networks: Guide and survey," in *IEEE 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, Melbourne, Australia, Apr. 2007, pp. 365–370.

[31] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2392–2431, Jul. 2017.

[32] H. A. Al-Rawi, M. A. Ng, and K.-L. A. Yau, "Application of reinforcement learning to routing in distributed wireless networks: A review," *Artificial Intelligence Review*, vol. 43, no. 3, pp. 381–416, Jan. 2015.

[33] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2432–2455, May 2017.

[34] T. T. Nguyen and G. Armitage, "A survey of techniques for Internet traffic classification using machine learning," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 56–76, Oct. 2008.

[35] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, Oct. 2016.

TABLE VII: Applications of ML in Wireless Networks

Scenario	Application	Learning Technique	Specific Method	Rreference
Ad hoc network	reactive routing	reinforcement learning	on-policy SARSA	[296]
body area network	power control	reinforcement learning	Q-learning	[299]
cellular network	dynamic channel allocation	reinforcement learning	reduced-state SARSA	[294]
cellular network	traffic prediction	deep learning	DNN	[327]
cellular network	traffic prediction	deep learning	CNN	[338]
cellular network	relay selection	unsupervised learning	K-means	[244]
cellular network	interference alignment	deep learning	DQN	[350]
centralized CR network	channel selection	supervised learning	SVM	[227]
cognitive femtocell network	interference control	reinforcement learning	decentralized Q-learning	[302]
CR ad hoc network	signal detection	unsupervised learning	ICA	[265]
CR ad hoc network	multi-agent sensing	reinforcement learning	on-policy SARSA	[295]
CR network	data recovery	unsupervised learning	ICA & PCA	[261]
CR network	relay selection	reinforcement learning	Q-learning	[301]
CR network	activity prediction	deep learning	RNN, CNN	[337]
CR network	spectrum access	deep learning	deep reinforcement learning	[352]
CR network	channel access	deep learning	DQN	[353]
CR network	energy scheduling	deep learning	deep reinforcement learning	[357]
CR network	PHY representation	deep learning	DNN	[324]
CRAN	power allocation	deep learning	deep reinforcement learning	[346]
dense cellular network	link scheduling	deep learning	CNN	[341]
energy-aware network	energy scheduling	deep learning	DQN	[153]
heterogeneous network	resource management	reinforcement learning	on-policy SARSA	[297]
heterogeneous network	packet routing	deep learning	DNN	[328]
heterogeneous network	data estimation	supervised learning	H-SVM, SVM	[221]
heterogeneous network	gateway deployment	unsupervised learning	K-means	[243]
heterogeneous network	data prediction	supervised learning	SVM	[224]
IoT network	traffic scheduling	deep learning	deep Q-learning	[347]
IoT network	cache allocation	deep learning	deep reinforcement learning	[348]
IoT network	traffic estimation	supervised learning	regression	[199]
IoT network	traffic offloading	deep learning	DNN	[330]
M2M network	user detection	unsupervised learning	EM algorithm, Bayes method	[255]
MANET	behavior learning	supervised learning	SVM	[225]
MANET	anomaly detection	unsupervised learning	K-means	[246]
MEC network	resource allocation	deep learning	DQN	[356]
MIMO	blind transceiver	unsupervised learning	K-means	[248]
MIMO	channel estimation	unsupervised learning	EM algorithm, Bayes method	[250]
MIMO	symbol detection	unsupervised learning	EM algorithm, Bayes method	[253]
MIMO	channel estimation	unsupervised learning	ICA	[262]
MIMO	blind receiver	unsupervised learning	ICA	[263]
MIMO	PHY representation	deep learning	DNN	[325]
MIMO	modulation and coding	reinforcement learning	Q-learning	[300]
mobile computing network	user location	supervised learning	SVM	[223]
multiantenna CR network	primary user detection	unsupervised learning	EM algorithm	[251]
multiantenna CR network	channel state detection	unsupervised learning	EM algorithm	[252]
OFDM network	channel estimation	deep learning	DNN	[312]
P2P network	power allocation	reinforcement learning	SARSA	[298]
smart grid network	network security	deep learning	DBN	[332]
software-defined network	resource allocation	deep learning	DNN	[326]
UAV network	trajectory control	deep learning	DQN	[344]
vehicular network	traffic control	deep learning	deep reinforcement learning	[345]
vehicular network	traffic prediction	supervised learning	KNN	[203]
vehicular network	resource allocation	deep learning	deep reinforcement learning	[351]
vehicular network	resource allocation	deep learning	deep reinforcement learning	[358]
Wi-Fi network	interference elimination	supervised learning	KNN	[208]
WI-Fi network	indoor localization	deep learning	DNN	[314]
Wi-Fi network	attacker counting	supervised learning	SVM	[228]
Wi-Fi network	indoor location	unsupervised learning	PCA	[258]
Wi-Fi network	wireless coexistence	supervised learning	logistic regression	[197]
wireless communication	signal detection	unsupervised learning	K-means	[249]
wireless communication	modulation classification	supervised learning	KNN	[207]
wireless communication	interference cancellation	unsupervised learning	ICA	[264]
wireless communication	modulation classification	deep learning	RNN, CNN	[321]
wireless communication	signal detection	deep learning	DNN	[322]
wireless communication	interference identification	deep learning	CNN	[313]
wireless communication	mobility prediction	deep learning	CNN	[320]
wireless communication	mobility prediction	deep learning	RNN	[334]
wireless communication	transportation mode	deep learning	RNN, DNN	[335]
wireless communication	PHY representation	deep learning	CNN	[340]
wireless mesh networks	traffic control	deep learning	CNN	[329]
wireless network	antenna selection	supervised learning	Bayes method, SVM	[232]
wireless network	QoE prediction	supervised learning	Bayes, SVM	[236]
wireless network	intrusion detection	unsupervised learning	K-means	[247]
wireless network	power control	deep learning	DNN	[331]
wireless network	resource allocation	deep learning	DNN	[339]
wireless sensor network	sensor partitioning	unsupervised learning	*K-mean*	[245]
wireless sensor network	interference estimate	supervised learning	regression	[195]
wireless sensor network	spectrum sensing	supervised learning	regression	[196]
wireless sensor network	PHY authentication	supervised learning	logistic regression	[198]
wireless sensor network	map reconstruction	supervised learning	regression	[200]
wireless sensor network	wireless localization	supervised learning	regression, SVM	[201]
wireless sensor network	anomaly detection	supervised learning	KNN	[204]
wireless sensor network	missing data estimation	supervised learning	KNN	[206]
wireless sensor network	localization estimation	supervised learning	SVM	[222]
wireless sensor network	signal classification	supervised learning	SVM,DNN	[226]
wireless sensor network	network association	supervised learning	Bayes	[233]
wireless sensor network	network state detection	unsupervised learning	EM algorithm	[254]
wireless sensor network	source localization	unsupervised learning	EM algorithm	[256]
wireless sensor network	data aggregation	unsupervised learning	PCA	[259]
wireless sensor network	data recovery	unsupervised learning	PCA	[260]
wireless sensor network	sleeping scheduling	reinforcement learning	Q-learning	[303]
WLAN	indoor location	supervised learning	Bayes method, SVM	[235]
WLAN	anomaly detection	supervised learning	naive Bayes method	[234]

[36] J. Xie, F. R. Yu, T. Huang, R. Xie, and Y. Liu, "A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 393–430, Jan. 2019.

[37] F. Pacheco, E. Exposito, M. Gineste, C. Baudoin, and J. Aguilar, "Towards the deployment of machine learning solutions in network traffic classification: A systematic survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1988–2014, Apr. 2018.

[38] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Communications Surveys & Tutorials, DOI:10.1109/COMST.2019.2924243*, vol. PP, no. 99, pp. 1–1, 2019.

[39] M. Usama, J. Qadir, A. Raza, H. Arif, K.-L. A. Yau, Y. Elkhatib, A. Hussain, and A. Al-Fuqaha, "Unsupervised machine learning for networking: Techniques, applications and research challenges," *arXiv preprint, arXiv:1709.06599*, 2017.

[40] K.-L. A. Yau, P. Komisarczuk, and P. D. Teal, "Reinforcement learning for context awareness and intelligence in wireless networks: Review, new features and open issues," *Journal of Network and Computer Applications*, vol. 35, no. 1, pp. 253–267, Jan. 2012.

[41] M. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," *IEEE network*, vol. 30, no. 3, pp. 22–29, May 2016.

[42] K. Ota, M. S. Dao, V. Mezaris, and F. G. De Natale, "Deep learning for mobile multimedia: A survey," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 13, no. 3s, p. 34, Aug. 2017.

[43] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2595–2621, Oct. 2018.

[44] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials, DOI:10.1109/COMST.2019.2926625*, vol. PP, no. 99, pp. 1–1, 2019.

[45] S. Suthaharan, "Supervised learning algorithms," in *Machine Learning Models and Algorithms for Big Data Classification*. Springer, 2016, pp. 183–206.

[46] H. B. Barlow, "Unsupervised learning," *Neural Computation*, vol. 1, no. 3, pp. 295–311, 1989.

[47] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998.

[48] T. M. Mitchell, "Does machine learning really work?" *AI magazine*, vol. 18, no. 3, pp. 11–20, 1997.

[49] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 924–937, Aug. 2003.

[50] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. MIT press, 2012.

[51] M. B. Christopher, *Pattern Recognition and Machine Learning*. Springer, 2016.

[52] T. Hastie, R. Tibshirani, and J. Friedman, "Unsupervised learning," in *The Elements of Statistical Learning*. Springer, 2009, pp. 485–585.

[53] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.

[54] I. Guyon and A. Elisseeff, "An introduction to feature extraction," in *Feature extraction*. Springer, 2006, pp. 1–25.

[55] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, May 1996.

[56] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[57] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015.

[58] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[59] G. E. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[60] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *arXiv preprint arXiv:1312.6026*, 2013.

[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, Dec. 2012, pp. 1097–1105.

[62] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[63] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, no. 4, pp. 500–544, 1952.

[64] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, p. 386, 1958.

[65] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Science*, vol. 9, no. 1, pp. 147–169, 1985.

[66] J. J. Hopfield and D. W. Tank, ""neural" computation of decisions in optimization problems," *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152, 1985.

[67] D. E. Rumelhart, G. E. Hinton, R. J. Williams *et al.*, "Learning representations by back-propagating errors," *Cognitive Modeling*, vol. 5, no. 3, p. 1, 1988.

[68] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," Colorado University at Boulder Department of Computer Science, Tech. Rep., 1986.

[69] D. S. Broomhead and D. Lowe, "Radial basis functions, multivariable functional interpolation and adaptive networks," Royal Signals and Radar Establishment Malvern (United Kingdom), Tech. Rep., 1988.

[70] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[71] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[72] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[73] L. Breiman, *Classification and regression trees*. Routledge, 2017.

[74] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[75] ——, *C4.5: Programs for machine learning*. Elsevier, 2014.

[76] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[77] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[78] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B*, vol. 20, no. 2, pp. 215–232, 1958.

[79] T. M. Cover, P. E. Hart *et al.*, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[80] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[81] A. McCallum, K. Nigam *et al.*, "A comparison of event models for naive bayes text classification," in *AAAI Workshop on Learning for Text Categorization*, vol. 752, no. 1, Madison, WI, 1998, pp. 41–48.

[82] A. Gammerman, V. Vovk, and V. Vapnik, "Learning by transduction," in *The Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, Jul. 1998, pp. 148–155.

[83] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *The Eleventh Annual Conference on Computational Learning Theory*, Madison, WI, Jul. 1998, pp. 92–100.

[84] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," Tech. Rep., 2001.

[85] X. J. Zhu, "Semi-supervised learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep., 2005.

[86] R. R. Bush and F. Mosteller, "Stochastic models for learning," Tech. Rep., 1955.

[87] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine Learning*, vol. 3, no. 1, pp. 9–44, 1988.

[88] C. J. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, May 1992.

[89] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*. University of Cambridge, 1994, vol. 37.

[90] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[91] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, p. 354, 2017.

[92] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," in *The Twelfth International Conference on Artificial Intelligence and Statistics*, Clearwater Beach, FL, 2009, pp. 448–455.

[93] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013, pp. 6645–6649.

[94] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *The 28th Annual Conference on Neural Information Processing Systems*, Montreal, Canada, Dec. 2014, pp. 2672–2680.

[95] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[96] G. E. Hinton and R. S. Zemel, "Autoencoders, minimum description length and helmholtz free energy," in *Advances in neural information processing systems*, 1994, pp. 3–10.

[97] R. Agrawal, R. Srikant *et al.*, "Fast algorithms for mining association rules," in *The 20th International Conference on Very Large Data Bases*, vol. 1215, Santiago, Chile, Sep. 1994, pp. 487–499.

[98] M. J. Zaki, "Scalable algorithms for association mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, May 2000.

[99] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," vol. 29, no. 2, pp. 1–12, 2000.

[100] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[101] R. Sibson, "SLINK: An optimally efficient algorithm for the single-link cluster method," *The Computer Journal*, vol. 16, no. 1, pp. 30–34, 1973.

[102] D. Defays, "An efficient algorithm for a complete link method," *The Computer Journal*, vol. 20, no. 4, pp. 364–366, 1977.

[103] R. Michalski, "Knowledge acquisition through conceptual clustering: A theoretical framework and algorithm for partitioning data into conjunctive concepts," *International Journal of Policy Analysis and Information Systems*, vol. 4, pp. 219–243, 1980.

[104] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *The Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 14, Oakland, CA, 1967, pp. 281–297.

[105] E. H. Ruspini, "A new approach to clustering," *Information and Control*, vol. 15, no. 1, pp. 22–32, 1969.

[106] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–22, 1977.

[107] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[108] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Departmental Papers (CIS)*, p. 107, 2000.

[109] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," vol. 25, no. 2, pp. 103–114, 1996.

[110] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *The Second International Conference on Knowledge Discovery and Data Mining (KDD)*, vol. 96, no. 34, Portland, OR, 1996, pp. 226–231.

[111] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering points to identify the clustering structure," vol. 28, no. 2, pp. 49–60, 1999.

[112] K. Pearson, "On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[113] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[114] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[115] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, Nov. 2008.

[116] R. Stratonovich, "Conditional Markov processes," *Theory of Probability and its Applications*, vol. 5, no. 2, p. 156, 1960.

[117] J. Moussouris, "Gibbs and Markov random systems with constraints," *Journal of Statistical Physics*, vol. 10, no. 1, pp. 11–33, 1974.

[118] J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artificial Intelligence*, vol. 29, no. 3, pp. 241–288, 1986.

[119] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," *Departmental Papers (CIS)*, 2001.

[120] D. N. Perkins, G. Salomon *et al.*, "Transfer of learning," *International Encyclopedia of Education*, vol. 2, pp. 6452–6457, 1992.

[121] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[122] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and Computation*, vol. 121, no. 2, pp. 256–285, 1995.

[123] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[124] I. Barandiaran, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[125] J. A. Gutierrez, E. H. Callaway, and R. L. Barrett, *Low-rate wireless personal area networks: Enabling wireless sensors with IEEE 802.15. 4*. IEEE Press, 2004.

[126] B. P. Crow, I. Widjaja, J. G. Kim, and P. T. Sakai, "IEEE 802.11 wireless local area networks," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 116–126, Sep. 1997.

[127] C. Eklund, R. B. Marks, S. Ponnuswamy, K. L. Stanwood, and N. J. Van Waes, *WirelessMAN: Inside the IEEE 802.16 standard for wireless metropolitan area networks*. IEEE Press, 2006.

[128] R. A. Hansen and P. T. Rawles, "Wide area networks," in *Encyclopedia of Information Technology Curriculum Integration*. IGI Global, 2008, pp. 971–978.

[129] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Jul. 2015.

[130] P. Baronti, P. Pillai, V. W. Chook, S. Chessa, A. Gotta, and Y. F. Hu, "Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards," *Computer Communications*, vol. 30, no. 7, pp. 1655–1695, May 2007.

[131] M. Ilyas, *The Handbook of Ad Hoc Wireless Networks*. CRC press, 2017.

[132] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour, "Wireless body area networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1658–1686, Jan. 2014.

[133] N. Abramson, "Development of the ALOHANET," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 119–123, Mar. 1985.

[134] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, "A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 550–586, 2017.

[135] Y. Censor, "Pareto optimality in multiobjective problems," *Applied Mathematics and Optimization*, vol. 4, no. 1, pp. 41–59, Mar. 1977.

[136] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[137] J. Zhang, T. Chen, S. Zhong, J. Wang, W. Zhang, X. Zuo, R. G. Maunder, and L. Hanzo, "Aeronautical ad hoc networking for the internet-above-the-clouds," *Proceedings of the IEEE*, vol. 107, no. 5, pp. 868–911, 2019.

[138] P. Pan, Y. Zhang, X. Ju, and L.-L. Yang, "Capacity of generalised network multiple-input–multiple-output systems with multicell cooperation," *IET Communications*, vol. 7, no. 17, pp. 1925–1937, Jul. 2013.

[139] L. Liu, R. Chen, S. Geirhofer, K. Sayana, Z. Shi, and Y. Zhou, "Downlink MIMO in LTE-advanced: SU-MIMO vs. MU-MIMO," *IEEE Communications Magazine*, vol. 50, no. 2, Feb. 2012.

[140] H. Wang, P. Pan, L. Shen, and Z. Zhao, "On the pair-wise error probability of a multi-cell MIMO uplink system with pilot

contamination," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5797–5811, Oct. 2014.

[141] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[142] P. Pan, H. Wang, Z. Zhao, and W. Zhang, "How many antenna arrays are dense enough in massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3042–3053, Apr. 2018.

[143] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, "On design of non-orthogonal pilot signals for a multi-cell massive MIMO system," *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 129–132, Apr. 2015.

[144] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, May 2003.

[145] J. Wang, C. Jiang, Z. Bie, T. Q. Quek, and Y. Ren, "Mobile data transactions in device-to-device communication networks: Pricing and auction," *IEEE Wireless Communications Letters*, vol. 5, no. 3, pp. 300–303, Jun. 2016.

[146] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3541–3551, Aug. 2013.

[147] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, Apr. 2011.

[148] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: Architectures, standards, and QoS improvement," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, Apr. 2011.

[149] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, Apr. 2011.

[150] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, Jun. 2015.

[151] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, Jun. 2016.

[152] K. Kaur, S. Garg, G. S. Aujla, N. Kumar, J. J. Rodrigues, and M. Guizani, "Edge computing in the industrial internet of things environment: Software-defined-networks-based edge-cloud interplay," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 44–51, Feb. 2018.

[153] Q. Zhang, M. Lin, L. T. Yang, Z. Chen, and P. Li, "Energy-efficient scheduling for real-time systems based on deep Q-learning model," *IEEE Transactions on Sustainable Computing, DOI: 10.1109/TSUSC.2017.2743704*, Aug. 2017.

[154] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wang, and T. Q. Quek, "Resource allocation for cognitive small cell networks: A cooperative bargaining game theoretic approach," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3481–3493, Jun. 2015.

[155] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Communications Magazine*, vol. 54, no. 10, pp. 184–190, Oct. 2016.

[156] J. Hoadley and P. Maveddat, "Enabling small cell deployment with HetNet," *IEEE Wireless Communications*, vol. 19, no. 2, pp. 4–5, Feb. 2012.

[157] H. Zhang, H. Liu, C. Jiang, X. Chu, A. Nallanathan, and X. Wen, "A practical semidynamic clustering scheme using affinity propagation in cooperative picocells," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4372–4377, Sep. 2015.

[158] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wen, and M. Tao, "Resource allocation in spectrum-sharing OFDMA femtocells with heterogeneous services," *IEEE Transactions on Communications*, vol. 62, no. 7, pp. 2366–2377, Jul. 2014.

[159] C. Jiang, Y. Chen, K. R. Liu, and Y. Ren, "Optimal pricing strategy for operators in cognitive femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 5288–5301, Sep. 2014.

[160] H. Zhang, C. Jiang, R. Q. Hu, and Y. Qian, "Self-organization in disaster-resilient heterogeneous small cell networks," *IEEE Network*, vol. 30, no. 2, pp. 116–121, Mar. 2016.

[161] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): A primer," *IEEE Network*, vol. 29, no. 1, pp. 35–41, Jan. 2015.

[162] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks - A technology overview," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 405–426, Sept. 2015.

[163] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on software-defined networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 27–51, Jun. 2015.

[164] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 358–380, Aug. 2015.

[165] H. Kim and N. Feamster, "Improving network management with software defined networking," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 114–119, Feb. 2013.

[166] R. Jain and S. Paul, "Network virtualization and software defined networking for cloud computing: A survey," *IEEE Communications Magazine*, vol. 51, no. 11, pp. 24–31, Nov. 2013.

[167] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," *IEEE Communications Magazine*, vol. 53, no. 2, pp. 90–97, Feb. 2015.

[168] Y. Li and M. Chen, "Software-defined network function virtualization: A survey," *IEEE Access*, vol. 3, pp. 2542–2553, Dec. 2015.

[169] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, Jul. 2011.

[170] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, Mar. 2002.

[171] S. Saleh, M. Ahmed, B. M. Ali, M. F. A. Rasid, and A. Ismail, "A survey on energy awareness mechanisms in routing protocols for wireless sensor networks using optimization methods," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 12, pp. 1184–1207, Dec. 2014.

[172] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[173] C. Jiang, Y. Chen, Y.-H. Yang, C.-Y. Wang, and K. R. Liu, "Dynamic chinese restaurant game: Theory and application to cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1960–1973, Apr. 2014.

[174] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130.

[175] C. Jiang, Y. Chen, Y. Gao, and K. R. Liu, "Joint spectrum sensing and access evolutionary game in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2470–2483, May 2013.

[176] C. Jiang, Y. Chen, K. R. Liu, and Y. Ren, "Renewal-theoretical dynamic spectrum access in cognitive radio network with unknown primary behavior," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 406–416, Mar. 2013.

[177] R. W. Thomas, D. H. Friend, L. A. Dasilva, and A. B. Mackenzie, "Cognitive networks: Adaptation and learning to achieve end-to-end performance objectives," *IEEE Communications Magazine*, vol. 44, no. 12, pp. 51–57, Dec. 2006.

[178] B. Manoj, R. R. Rao, and M. Zorzi, "Cognet: A cognitive complete knowledge network system," *IEEE Wireless Communications*, vol. 15, no. 6, pp. 81–88, Jun. 2008.

[179] C. Jiang, H. Zhang, Y. Ren, and H.-H. Chen, "Energy-efficient non-cooperative cognitive radio networks: Micro, meso, and macro views," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 14–20, Jul. 2014.

[180] M. Turkboylari and G. L. Stuber, "An efficient algorithm for estimating the signal-to-interference ratio in TDMA cellular systems," *IEEE Transactions on Communications*, vol. 46, no. 6, pp. 728–731, Jun. 1998.

[181] M. Ma, X. Huang, and Y. J. Guo, "An interference self-cancellation technique for SC-FDMA systems," *IEEE Communications Letters*, vol. 14, no. 6, Jun. 2010.

[182] M. Kountouris and J. G. Andrews, "Downlink SDMA with limited feedback in interference-limited wireless networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2730–2741, Aug. 2012.

[183] H. Zhang, C. Jiang, X. Mao, and H.-H. Chen, "Interference-limited resource optimization in cognitive femtocells with fairness and imperfect spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1761–1771, Mar. 2016.

[184] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.

[185] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to Gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[186] A. Attar, V. Krishnamurthy, and O. N. Gharehshiran, "Interference management using cognitive base-stations for UMTS LTE," *IEEE Communications Magazine*, vol. 49, no. 8, Aug. 2011.

[187] S. Deb and P. Monogioudis, "Learning-based uplink interference management in 4G LTE cellular systems," *IEEE/ACM Transactions on Networking*, vol. 23, no. 2, pp. 398–411, Feb. 2015.

[188] F. Bernardo, R. Agustí, J. Pérez-Romero, and O. Sallent, "Intercell interference management in OFDA networks: A decentralized approach based onreinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 41, no. 6, pp. 968–976, Jun. 2011.

[189] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 329.

[190] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.

[191] T. A. Max and H. E. Burkhart, "Segmented polynomial regression applied to taper equations," *Forest Science*, vol. 22, no. 3, pp. 283–289, Sep. 1976.

[192] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Apr. 1970.

[193] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[194] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, Mar. 2005.

[195] X. Chang, J. Huang, S. Liu, G. Xing, H. Zhang, J. Wang, L. Huang, and Y. Zhuang, "Accuracy-aware interference modeling and measurement in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 2, pp. 278–291, Feb. 2016.

[196] K. Umebayashi, M. Kobayashi, and M. López-Benítez, "Efficient time domain deterministic-stochastic model of spectrum usage," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1518–1527, Mar. 2018.

[197] M. O. Al Kalaa, S. J. Seidman, and H. H. Refai, "Estimating the likelihood of wireless coexistence using logistic regression: Emphasis on medical devices," *IEEE Transactions on Electromagnetic Compatibility, DOI:10.1109/TEMC.2017.2777179*, 2017.

[198] L. Xiao, X. Wan, and Z. Han, "PHY-layer authentication with multiple landmarks with reduced overhead," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1676–1687, Mar. 2018.

[199] T.-C. Chang, C.-H. Lin, K. C.-J. Lin, and W.-T. Chen, "Traffic-aware sensor grouping for IEEE 802.11 ah networks: Regression based analysis and design," *IEEE Transactions on Mobile Computing, DOI:10.1109/TMC.2018.2840692*, 2018.

[200] J. Chen, U. Yatnalli, and D. Gesbert, "Learning radio maps for UAV-aided wireless networks: A segmented regression approach," in *IEEE International Conference on Communications (ICC)*, Paris, France, May 2017, pp. 1–6.

[201] Q. Lei, H. Zhang, H. Sun, and L. Tang, "Fingerprint-based device-free localization in changing environments using enhanced channel selection and logistic regression," *IEEE Access*, vol. 6, pp. 2569–2577, Dec. 2018.

[202] T. Sherwood, E. Perelman, G. Hamerly, and B. Calder, "Automatically characterizing large scale program behavior," *ACM SIGARCH Computer Architecture News*, vol. 30, no. 5, pp. 45–57, Oct. 2002.

[203] Z. Feng, X. Li, Q. Zhang, and W. Li, "Proactive radio resource optimization with margin prediction: A data mining approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9050–9060, Oct. 2017.

[204] M. Xie, J. Hu, S. Han, and H.-H. Chen, "Scalable hypergrid k-NN-based online anomaly detection in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 8, pp. 1661–1670, Oct. 2013.

[205] O. Onireti, A. Zoha, J. Moysen, A. Imran, L. Giupponi, M. A. Imran, and A. Abu-Dayya, "A cell outage management framework for dense heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2097–2113, Apr. 2016.

[206] L. Pan and J. Li, "K-nearest neighbor based missing data estimation algorithm in wireless sensor networks," *Wireless Sensor Network*, vol. 2, no. 2, pp. 115–122, Feb. 2010.

[207] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.

[208] F. Yu, M. Jiang, J. Liang, X. Qin, M. Hu, T. Peng, and X. Hu, "5G WiFi signal-based indoor localization system using cluster K-nearest neighbor algorithm," *International Journal of Distributed Sensor Networks*, vol. 10, no. 12, p. 247525, Dec. 2014.

[209] G. Bontempi, M. Birattari, and H. Bersini, "Lazy learning for local modelling and control design," *International Journal of Control*, vol. 72, no. 7-8, pp. 643–658, Nov. 1999.

[210] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[211] S. Bergman, *The kernel function and conformal mapping*. American Mathematical Soc., 1950, no. 5.

[212] B. Scholkopf and A. J. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press, 2001.

[213] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *The Annals of Statistics*, vol. 36, no. 3, pp. 1171–1220, Jun. 2008.

[214] E. Parzen, "Extraction and detection problems and reproducing kernel Hilbert spaces," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, vol. 1, no. 1, pp. 35–62, Jun. 1962.

[215] K. Fukumizu, F. R. Bach, and M. I. Jordan, "Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces," *Journal of Machine Learning Research*, vol. 5, no. Jan., pp. 73–99, 2004.

[216] J. Kivinen, A. J. Smola, and R. C. Williamson, "Online learning with kernels," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2165–2176, Aug. 2004.

[217] A. Chhabra, *Principles of Communication Engineering*. S. Chand Publishing, 2006.

[218] T. Kailath, "RKHS approach to detection and estimation problems–I: Deterministic signals in Gaussian noise," *IEEE Transactions on Information Theory*, vol. 17, no. 5, pp. 530–549, Sep. 1971.

[219] K. Fukunaga and W. L. Koontz, "Application of the Karhunen-Loeve expansion to feature selection and ordering," *IEEE Transactions on Computers*, vol. 100, no. 4, pp. 311–318, Apr. 1970.

[220] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2230–2231, Oct. 1998.

[221] V.-S. Feng and S. Y. Chang, "Determination of wireless networks parameters through parallel hierarchical support vector machines," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 3, pp. 505–512, Mar. 2012.

[222] D. A. Tran and T. Nguyen, "Localization in wireless sensor networks based on support vector machines," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 7, pp. 981–994, Jul. 2008.

[223] G. Sun and W. Guo, "Robust mobile geo-location algorithm based on LS-SVM," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 3, pp. 1037–1041, May 2005.

[224] B. K. Donohoo, C. Ohlsen, S. Pasricha, Y. Xiang, and C. Anderson, "Context-aware energy enhancements for smart mobile devices," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1720–1732, Aug. 2014.

[225] J. F. C. Joseph, B.-S. Lee, A. Das, and B.-C. Seet, "Cross-layer detection of sinking behavior in wireless ad hoc networks using SVM and FDA," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 2, pp. 233–245, Mar. 2011.

[226] F. Pianegiani, M. Hu, A. Boni, and D. Petri, "Energy-efficient signal classification in ad hoc wireless sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 1, pp. 190–196, Jan. 2008.

[227] K. G. M. Thilina, E. Hossain, and D. I. Kim, "DCCC-MAC: a dynamic common-control-channel-based MAC protocol for cellular cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3597–3613, May 2016.

[228] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and localization of multiple spoofing attackers in wireless networks," *IEEE Transactions on Parallel and Distributed systems*, vol. 24, no. 1, pp. 44–58, Jan. 2013.

[229] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Anomaly detection in wireless sensor networks," *IEEE Wireless Communications*, vol. 15, no. 4, pp. 34–40, Aug. 2008.

[230] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Procedia Computer Science*, vol. 60, pp. 708–713, 2015.

[231] W. Feng, J. Sun, L. Zhang, C. Cao, and Q. Yang, "A support vector machine based naive Bayes algorithm for spam filtering," in *IEEE 35th International Performance Computing and Communications Conference (IPCCC)*, Las Vegas, NV, Dec. 2016, pp. 1–8.

[232] D. He, C. Liu, T. Q. Quek, and H. Wang, "Transmit antenna selection in MIMO wiretap channels: A machine learning approach," *IEEE Wireless Communications Letters, DOI: 10.1109/L-WC.2018.2805902*, Feb. 2018.

[233] P. Abouzar, K. Shafiee, D. G. Michelson, and V. C. Leung, "Action-based scheduling technique for 802.15. 4/ZigBee wireless body area networks," in *IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Toronto, Canada, Sept. 2011, pp. 2188–2192.

[234] M. Klassen and N. Yang, "Anomaly based intrusion detection in wireless networks using Bayesian classifier," in *IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI)*, Nanjing, China, Oct. 2012, pp. 257–264.

[235] R. W. Ouyang, A. K.-S. Wong, C.-T. Lea, and M. Chiang, "Indoor location estimation with reduced calibration exploiting unlabeled data via hybrid generative/discriminative learning," *IEEE Transactions on Mobile Computing*, vol. 11, no. 11, pp. 1613–1626, Nov. 2012.

[236] P. Charonyktakis, M. Plakia, I. Tsamardinos, and M. Papadopouli, "On user-centric modular QoE prediction for VoIP based on machine-learning algorithms," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1443–1456, Jun. 2016.

[237] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[238] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[239] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, Aug. 1987.

[240] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[241] A. Paz and S. Moran, "Non deterministic polynomial optimization problems and their approximations," *Theoretical Computer Science*, vol. 15, no. 3, pp. 251–277, 1981.

[242] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 7, pp. 881–892, Jul. 2002.

[243] M. Xia, Y. Owada, M. Inoue, and H. Harai, "Optical and wireless hybrid access networks: Design and optimization," *Journal of Optical Communications and Networking*, vol. 4, no. 10, pp. 749–759.

[244] M. Hajjar, G. Aldabbagh, N. Dimitriou, and M. Z. Win, "Hybrid clustering scheme for relaying in multi-cell LTE high user density networks," *IEEE Access*, vol. 5, pp. 4431–4438, Mar. 2017.

[245] I. Cabria and I. Gondra, "Potential k-means for load balancing and cost minimization in mobile recycling network," *IEEE Systems Journal*, vol. 11, no. 1, pp. 242–249, Mar. 2017.

[246] M. S. Parwez, D. B. Rawat, and M. Garuba, "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2058–2065, Aug. 2017.

[247] K. El-Khatib, "Impact of feature reduction on the efficiency of wireless intrusion detection systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 8, pp. 1143–1149, Aug. 2010.

[248] H.-W. Liang, W.-H. Chung, and S.-Y. Kuo, "Coding-aided K-means clustering blind transceiver for space shift keying MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 103–115, 2016.

[249] T. Zhao, A. Nehorai, and B. Porat, "K-means clustering-based data detection and symbol-timing recovery for burst-mode optical receiver," *IEEE Transactions on Communications*, vol. 54, no. 8, pp. 1492–1501, Aug. 2006.

[250] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.

[251] K. W. Choi and E. Hossain, "Estimation of primary user parameters in cognitive radio systems via hidden Markov model," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 782–795, Feb. 2013.

[252] A. Assra, J. Yang, and B. Champagne, "An EM approach for cooperative spectrum sensing in multiantenna CR networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1229–1243, Mar. 2016.

[253] X.-Y. Zhang, D.-g. Wang, and J.-B. Wei, "Joint symbol detection and channel estimation for MIMO-OFDM systems via the variational bayesian EM algorithm," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Las Vegas, USA, Mar. 2008, pp. 13–17.

[254] J. Li and A. Nehorai, "Joint sequential target estimation and clock synchronization in wireless sensor networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 74–88, Jun. 2015.

[255] X. Zhang, Y.-C. Liang, and J. Fang, "Novel Bayesian inference algorithms for multiuser detection in M2M communications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7833–7848, Sep. 2017.

[256] W. Meng, W. Xiao, and L. Xie, "An efficient EM algorithm for energy-based multisource localization in wireless sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 3, pp. 1017–1027, Mar. 2011.

[257] B. Sun, Y. Guo, N. Li, and D. Fang, "Multiple target counting and localization using variational Bayesian EM algorithm in wireless sensor networks," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2985–2998, Jul. 2017.

[258] S. Shi, S. Sigg, L. Chen, Y. Ji *et al.*, "Accurate location tracking from CSI-based passive device-free probabilistic fingerprinting," *IEEE Transactions on Vehicular Technology, DOI: 10.1109/TVT.2018.2810307*, pp. 1–14, 2018.

[259] A. Morell, A. Correa, M. Barceló, and J. L. Vicario, "Data aggregation and principal component analysis in WSNs," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 3908–3919, Jun. 2016.

[260] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3447–3461, Oct. 2012.

[261] R. C. Qiu, Z. Hu, Z. Chen, N. Guo, R. Ranganathan, S. Hou, and G. Zheng, "Cognitive radio network for the smart grid: Experimental system architecture, control algorithms, security, and microgrid testbed," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 724–740, Dec. 2011.

[262] L. Shen, Y.-D. Yao, H. Wang, and H. Wang, "ICA based semi-blind decoding method for a multicell multiuser massive MIMO uplink system in Rician/Rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7501–7511, Nov. 2017.

[263] L. Sarperi, X. Zhu, and A. K. Nandi, "Blind OFDM receiver based on independent component analysis for multiple-input multiple-output systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 11, Nov.

[264] J. Li, H. Zhang, and M. Fan, "Digital self-interference cancellation based on independent component analysis for co-time co-frequency full-duplex communication systems," *IEEE Access*, vol. 5, pp. 10 222–10 231, 2017.

[265] H. Nguyen, G. Zheng, R. Zheng, and Z. Han, "Binary inference for primary user separation in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1532–1542, Apr. 2013.

[266] D. P. Zhou and C. J. Tomlin, "Budget-constrained multi-armed bandits with multiple plays," *arXiv preprint arXiv:1711.05928*, 2017.

[267] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5G small cells," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 64–73, Jun. 2016.

[268] Q. Wu, Z. Du, P. Yang, Y.-D. Yao, and J. Wang, "Traffic-aware online network selection in heterogeneous wireless networks,"

IEEE Transactions on Vehicular Technology, vol. 65, no. 1, pp. 381–397, Jan. 2016.

[269] P. Si, F. R. Yu, H. Ji, and V. C. Leung, "Distributed sender scheduling for multimedia transmission in wireless peer-to-peer networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, New Orleans, LO, Dec. 2008, pp. 1–5.

[270] S. Maghsudi and S. Stanczak, "Dynamic bandit with covariates: Strategic solutions with application to wireless resource allocation," in *IEEE International Conference on Communications (ICC)*, Budapest, Hungary, Jun. 2013, pp. 5898–5902.

[271] S. Lee, J. Choi, J. Yoo, and C.-K. Kim, "Frequency diversity-aware Wi-Fi using OFDM-based Bloom filters," *IEEE Transactions on Mobile Computing*, vol. 14, no. 3, pp. 525–537, Mar. 2015.

[272] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5431–5440, Dec.

[273] Y. Gwon, S. Dastangoo, and H. Kung, "Optimizing media access strategy for competing cognitive radio networks," in *IEEE Global Communications Conference (GLOBECOM)*, Atlanta, GA, Dec. 2013, pp. 1215–1220.

[274] B. Li, P. Yang, J. Wang, Q. Wu, S. Tang, X.-Y. Li, and Y. Liu, "Almost optimal dynamically-ordered channel sensing and accessing for cognitive networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2215–2228, Oct. 2014.

[275] P. Li, T. Vermeulen, H. Liy, and S. Pollin, "An adaptive channel selection scheme for reliable TSCH-based communication," in *International Symposium on Wireless Communication Systems (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 511–515.

[276] O. Avner and S. Mannor, "Multi-user lax communications: A multi-armed bandit approach," in *The 35th Annual IEEE International Conference on Computer Communications (INFOCOM)*, San Francisco, CA, Apr. 2016, pp. 1–9.

[277] P. Zhou and T. Jiang, "Toward optimal adaptive wireless communications in unknown environments," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3655–3667, May 2016.

[278] J. Wang, C. Jiang, H. Zhang, X. Zhang, V. C. Leung, and L. Hanzo, "Learning-aided network association for hybrid indoor LiFi-WiFi systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3561–3574, Apr. 2018.

[279] M. L. Puterman, "Markov decision processes: Discrete stochastic dynamic programming," *Technometrics*, vol. 37, no. 3, pp. 353–353, Mar. 1994.

[280] X. R. Cao, "Event-based optimization of Markov systems," *IEEE Transactions on Automatic Control*, vol. 53, no. 4, pp. 1076–1082, May 2008.

[281] L. G. Crespo and J. Q. Sun, "Stochastic optimal control via Bellman's principle," *Automatica*, vol. 39, no. 12, pp. 2109–2114, 2003.

[282] W. A. Massey, K. Ramakrisluian, M. Aravamudan, and G. Pai, "Scheduling algorithms for downlink services in wireless networks: A Markov decision process approach," in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 6, Dallas, TX, Dec. 2004, pp. 4038–4042.

[283] J. Tang and Y. Cheng, "Selfish misbehavior detection in 802.11 based wireless networks: An adaptive approach based on Markov decision process," in *IEEE International Conference on Computer Communications (INFOCOM)*, Turin, Italy, Apr. 2013, pp. 1357–1365.

[284] P.-Y. Kong, "Optimal probabilistic policy for dynamic resource activation using Markov decision process in green wireless networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2357–2368, Oct. 2014.

[285] P.-H. Tseng, K.-T. Feng, and C.-H. Huang, "POMDP-based cell selection schemes for wireless networks," *IEEE Communications Letters*, vol. 18, no. 5, pp. 797–800, May 2014.

[286] X. Fei, A. Boukerche, and R. Yu, "A POMDP based k-coverage dynamic scheduling protocol for wireless sensor networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Miami, FL, Dec. 2010, pp. 1–5.

[287] D.-S. Zois, M. Levorato, and U. Mitra, "A POMDP framework for heterogeneous sensor selection in wireless body area networks," in *IEEE International Conference on Computer Communications (INFOCOM)*, Orlando, FL, Mar. 2012, pp. 2611–2615.

[288] J. Pajarinen, A. Hottinen, and J. Peltonen, "Optimizing spatial and temporal reuse in wireless networks by decentralized partially observable Markov decision processes," *IEEE Transactions on Mobile Computing*, vol. 13, no. 4, pp. 866–879, Apr. 2014.

[289] H. Xie, R. W. Pazzi, and A. Boukerche, "A novel cross layer TCP optimization protocol over wireless networks by Markov decision process," in *IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, 2012, pp. 5723–5728.

[290] N. Michelusi and U. Mitra, "Cross-layer estimation and control for cognitive radio: Exploiting sparse network dynamics," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 1, pp. 128–145, Mar. 2015.

[291] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951.

[292] J. Wang, C. Jiang, Z. Han, Y. Ren, and L. Hanzo, "Network association strategies for an energy harvesting aided super-WiFi network relying on measured solar activity," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3785–3797, Dec. 2016.

[293] "Database of solar radiation," http://solar.uq.edu.au/user/reportPower.php.

[294] N. Lilith and K. Dogancay, "Dynamic channel allocation for mobile cellular traffic using reduced-state reinforcement learning," in *IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 4, Atlanta, GA, Apr. 2004, pp. 2195–2200.

[295] J. Lundén, V. Koivunen, S. R. Kulkarni, and H. V. Poor, "Reinforcement learning based distributed multiagent sensing policy for cognitive radio networks," in *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*. Aachen, Germany: IEEE, May 2011, pp. 642–646.

[296] S. Chettibi and S. Chikhi, "An adaptive energy-aware routing protocol for MANETs using the SARSA reinforcement learning algorithm," in *IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, Madrid, Spain, May 2012, pp. 84–89.

[297] J. Suga and R. Tafazolli, "Joint resource management with reinforcement learning in heterogeneous networks," in *IEEE 78th Vehicular Technology Conference (VTC Fall)*, Las Vegas, NV, Sep. 2013, pp. 1–5.

[298] A. Ortiz, H. Al-Shatri, X. Li, T. Weber, and A. Klein, "Reinforcement learning for energy harvesting point-to-point communications," in *IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[299] R. Kazemi, R. Vesilo, E. Dutkiewicz, and R. Liu, "Dynamic power control in wireless body area networks using reinforcement learning with approximation," in *IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Toronto, Canada, Sep. 2011, pp. 2203–2208.

[300] J. P. Leite, P. H. P. de Carvalho, and R. D. Vieira, "A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, Apr. 2012, pp. 809–814.

[301] H. Jung, K. Kim, J. Kim, O.-S. Shin, and Y. Shin, "A relay selection scheme using Q-learning algorithm in cooperative wireless communications," in *IEEE 18th Asia-Pacific Conference on Communications (APCC)*, Jeju Island, South Korea, Oct. 2012, pp. 7–11.

[302] A. Galindo-Serrano and L. Giupponi, "Distributed Q-learning for aggregated interference control in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, May 2010.

[303] L. Prashanth, A. Chatterjee, and S. Bhatnagar, "Two timescale convergent Q-learning for sleep-scheduling in wireless sensor networks," *Wireless Networks*, vol. 20, no. 8, pp. 2589–2604, Nov. 2014.

[304] J. M. Zurada, *Introduction to artificial neural systems*. West publishing Company St. Paul, 1992, vol. 8.

[305] S. Haykin, *Neural networks: A comprehensive foundation*. Prentice Hall PTR, 1994.

[306] R. J. Williams and D. Zipser, "Experimental analysis of the real-time recurrent learning algorithm," *Connection Science*, vol. 1, no. 1, pp. 87–111, Jan. 1989.

[307] P. Campolucci, A. Uncini, F. Piazza, and B. D. Rao, "On-line learning algorithms for locally recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 253–271, Mar. 1999.

[308] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[309] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[310] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational IntelligenCe Magazine*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[311] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative deep learning for recommender systems," in *The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, Australia, Aug. 2015, pp. 1235–1244.

[312] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[313] M. Schmidt, D. Block, and U. Meier, "Wireless interference identification with convolutional neural networks," *arXiv preprint arXiv:1703.00737*, 2017.

[314] X. Wang, L. Gao, and S. Mao, "PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach," in *IEEE Global Communications Conference (GLOBECOM)*. San Diego, CA: IEEE, Dec. 2015, pp. 1–6.

[315] ——, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016.

[316] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[317] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6258–6267, Jul. 2017.

[318] W. Zhang, K. Liu, W. Zhang, Y. Zhang, and J. Gu, "Deep neural networks for wireless localization in indoor and outdoor environments," *Neurocomputing*, vol. 194, pp. 279–287, Jun. 2016.

[319] Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, "Learn to cache: Machine learning for network edge caching in the big data era," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 28–35, Jun. 2018.

[320] X. Ouyang, C. Zhang, P. Zhou, and H. Jiang, "Deepspace: An online deep learning framework for mobile big data to understand human mobility patterns," *arXiv preprint arXiv:1610.07009*, 2016.

[321] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Distributed deep learning models for wireless signal classification with low-cost spectrum sensors," *arXiv preprint arXiv:1707.08908*, 2017.

[322] N. Farsad and A. Goldsmith, "Detection algorithms for communication systems using deep learning," *arXiv preprint arXiv:1705.08044*, 2017.

[323] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International Conference on Engineering Applications of Neural Networks*, Athens, Greece, Aug. 2016, pp. 213–226.

[324] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[325] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Deep learning based MIMO communications," *arXiv preprint arXiv:1707.07980*, 2017.

[326] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, Feb. 2018.

[327] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *IEEE Conference on Computer Communications (INFOCOM)*. Atlanta, USA: IEEE, May 2017, pp. 1–9.

[328] N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 146–153, Jun. 2017.

[329] F. Tang, B. Mao, Z. M. Fadlullah, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 154–160, Feb. 2018.

[330] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of things with edge computing," *IEEE Network*, vol. 32, no. 1, pp. 96–101, Jan. 2018.

[331] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," *arXiv preprint arXiv:1705.09412*, 2017.

[332] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[333] D. Heckerman and C. Meek, "Models and selection criteria for regression and classification," in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. Providence, Rhode Island: Morgan Kaufmann Publishers Inc., Aug. 1997, pp. 223–228.

[334] A. Al-Molegi, M. Jabreel, and B. Ghaleb, "STF-RNN: Space time features-based recurrent neural network for predicting people next location," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, Athens, Greece, Dec. 2016, pp. 1–7.

[335] X. Song, H. Kanasugi, and R. Shibasaki, "DeepTransport: Prediction and simulation of human mobility and transportation mode at a citywide level." in *The 25th International Joint Conference on Artificial Intelligence (IJCAI)*, New York City, NY, Jul. 2016, pp. 2618–2624.

[336] L. Ghouti, "Mobility prediction using fully-complex extreme learning machines," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, Apr. 2014.

[337] A. Agarwal, S. Dubey, M. A. Khan, R. Gangopadhyay, and S. Debnath, "Learning based primary user activity prediction in cognitive radio networks for efficient dynamic spectrum access," in *IEEE International Conference on Signal Processing and Communications (SPCOM)*, Bangalore, India, Nov. 2016, pp. 1–5.

[338] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1389–1401, Mar. 2019.

[339] M. Eisen, C. Zhang, L. F. Chamon, D. D. Lee, and A. Ribeiro, "Learning optimal resource allocations in wireless systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2775–2790, Apr. 2019.

[340] B. Zhu, J. Wang, L. He, and J. Song, "Joint transceiver optimization for wireless communication PHY using neural network," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1364–1373, Mar. 2019.

[341] W. Cui, K. Shen, and W. Yu, "Spatial deep learning for wireless scheduling," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1248–1261, Mar. 2019.

[342] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning." in *The Thirtieth AAAI Conference on Artificial Intelligence*, vol. 16, Phoenix, AZ, Feb. 2016, pp. 2094–2100.

[343] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, "Dueling network architectures for deep reinforcement learning," *arXiv preprint arXiv:1511.06581*, 2015.

[344] B. Zhang, C. H. Liu, J. Tang, Z. Xu, J. Ma, and W. Wang, "Learning-based energy-efficient data collection by unmanned vehicles in smart cities," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1666–1676, Dec. 2017.

[345] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10 705–10 717, Dec.

[346] Z. Xu, Y. Wang, J. Tang, J. Wang, and M. C. Gursoy, "A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs," in *IEEE International Conference on Communications (ICC)*, Paris, France, 2017, pp. 1–6.

[347] J. Zhu, Y. Song, D. Jiang, and H. Song, "A new deep-Q-learning-based transmission scheduling mechanism for the cognitive Internet of things," *IEEE Internet of Things Journal, DOI: 10.1109/JIOT.2017.2759728*, Oct. 2017.

[348] X. He, K. Wang, H. Huang, T. Miyazaki, Y. Wang, and S. Guo, "Green resource allocation based on deep reinforcement learning in content-centric IoT," *IEEE Transactions on Emerging Topics in Computing, DOI: 10.1109/TETC.2018.2805718*, Feb. 2018.

[349] Y. He, Z. Zhang, F. R. Yu, N. Zhao, H. Yin, V. C. Leung, and Y. Zhang, "Deep-reinforcement-learning-based optimization for cache-enabled opportunistic interference alignment wireless

networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 433–10 445, Nov. 2017.

[350] Y. He, C. Liang, F. R. Yu, N. Zhao, and H. Yin, "Optimization of cache-enabled opportunistic interference alignment wireless networks: A big data deep reinforcement learning approach," in *IEEE International Conference on Communications (ICC)*, Paris, France, May 2017, pp. 1–6.

[351] Y. He, N. Zhao, and H. Yin, "Integrated networking, caching, and computing for connected vehicles: A deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 1, pp. 44–55, Jan. 2018.

[352] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for dynamic spectrum access in multichannel wireless networks," *arXiv preprint arXiv:1704.02613*, 2017.

[353] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking, DOI: 10.1109/TCCN.2018.2809722*, Feb. 2018.

[354] S. Çetinkaya and M. Parlar, "Optimal myopic policy for a stochastic inventory problem with fixed and proportional backorder costs," *European Journal of Operational Research*, vol. 110, no. 1, pp. 20–41, Oct. 1998.

[355] P. Ansell, K. D. Glazebrook, J. Niño-Mora, and M. O'Keeffe, "Whittle's index policy for a multi-class queueing system with convex holding costs," *Mathematical Methods of Operations Research*, vol. 57, no. 1, pp. 21–39, Sep. 2003.

[356] Z. Yang, Y. Liu, Y. Chen, and G. Tyson, "Deep reinforcement learning in cache-aided MEC networks," in *IEEE International Conference on Communications (ICC)*, Shanghai, China, May 2019, pp. 1–6.

[357] H. He and H. Jiang, "Deep learning based energy efficiency optimization for distributed cooperative spectrum sensing," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 32–39, Jul. 2019.

[358] Y. Dai, D. Xu, S. Maharjan, G. Qiao, and Y. Zhang, "Artificial intelligence empowered edge computing and caching for Internet of vehicles," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 12–18, Jul. 2019.

[359] C.-Y. Lin, K.-C. Chen, D. Wickramasuriya, S.-Y. Lien, and R. D. Gitlin, "Anticipatory mobility management by big data analytics for ultra-low latency mobile networking," in *IEEE International Conference on Communications (ICC)*, Kansas City, MO, May 2018, pp. 1–7.

[360] K.-C. Chen, T. Zhang, R. D. Gitlin, and G. Fettweis, "Ultra-low latency mobile networking," *IEEE Network*, vol. 33, no. 2, pp. 181–187, Mar. 2018.

[361] D. Soldani, Y. J. Guo, B. Barani, P. Mogensen, I. Chih-Lin, and S. K. Das, "5G for ultra-reliable low-latency communications," *IEEE Network*, vol. 32, no. 2, pp. 6–7, Feb. 2018.

[362] Y.-J. Liu, S.-M. Cheng, and Y.-L. Hsueh, "eNB selection for machine type communications using reinforcement learning based Markov decision process," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11 330–11 338, Dec.

[363] S.-Y. Lien, S.-C. Hung, D.-J. Deng, and Y. J. Wang, "Efficient ultra-reliable and low latency communications and massive machine-type communications in 5G new radio," in *IEEE Global Communications Conference*, Singapore, Dec. 2017, pp. 1–7.

[364] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu, "Network tomography: Recent developments," *Statistical Science*, vol. 19, no. 3, pp. 499–517, Aug. 2004.

[365] Y. E. Sagduyu, Y. Shi, A. Fanous, and J. H. Li, "Wireless network inference and optimization: Algorithm design and implementation," *IEEE Transactions on Mobile Computing*, vol. 16, no. 1, pp. 257–267, Jan. 2017.

[366] C.-K. Yu, K.-C. Chen, and S.-M. Cheng, "Cognitive radio network tomography," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1980–1997, Apr. 2010.

[367] J. Wilson and N. Patwari, "See-through walls: Motion tracking using variance-based radio tomography networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 5, pp. 612–621, May 2011.

[368] S. Legg and M. Hutter, "Universal intelligence: A definition of machine intelligence," *Minds and machines*, vol. 17, no. 4, pp. 391–444, Dec. 2007.

[369] A. H. Bond and L. Gasser, *Readings in distributed artificial intelligence*. Morgan Kaufmann, 2014.

[370] S. Stein, S. A. Williamson, and N. R. Jennings, "Decentralised channel allocation and information sharing for teams of cooperative agents," in *The 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. Valencia, Spain: International Foundation for Autonomous Agents and Multiagent Systems, Jun. 2012, pp. 231–238.

[371] X. Ge and Q.-L. Han, "Distributed formation control of networked multi-agent systems using a dynamic event-triggered communication mechanism," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 10, pp. 8118–8127, Oct. 2017.

[372] L. Zhou, P. Yang, C. Chen, and Y. Gao, "Multiagent reinforcement learning with sparse interactions by negotiation and knowledge transfer," *IEEE transactions on cybernetics*, vol. 47, no. 5, pp. 1238–1250, May 2016.

[373] E. Ko and K.-C. Chen, "Wireless communications meets artificial intelligence: An illustration by autonomous vehicles on manhattan streets," in *IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–7.

Jingjing Wang (S'14-M'19) received his B.S. degree in Electronic Information Engineering from Dalian University of Technology, Liaoning, China in 2014 and the Ph.D. degree in Information and Communication Engineering from Tsinghua University, Beijing, China in 2019, both with the highest honors. From 2017 to 2018, he visited the Next Generation Wireless Group chaired by Prof. Lajos Hanzo, University of Southampton, UK. Dr. Wang is currently a postdoc researcher at Department of Electronic Engineering, Tsinghua University. His research interests include resource allocation and network association, learning theory aided modeling, analysis and signal processing, as well as information diffusion theory for mobile wireless networks. Dr. Wang was a recipient of the Best Journal Paper Award of IEEE ComSoc Technical Committee on Green Communications & Computing in 2018, the Best Paper Award from IEEE ICC and IWCMC in 2019.

Chunxiao Jiang (S'09-M'13-SM'15) received the B.S. degree from Beihang University, Beijing in 2008 and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing in 2013, both with the highest honors. From 2011 to 2012, he visited University of Maryland, College Park as a joint PhD supported by China Scholarship Council. From 2013 to 2016, he was a postdoc with Tsinghua University, during which he visited University of Maryland, College Park and University of Southampton. Since July 2016, he became an assistant professor in Tsinghua Space Center, Tsinghua University. His research interests include space networks and heterogeneous networks. Dr. Jiang is the recipient of the Best Paper Award from IEEE GLOBECOM in 2013, the Best Student Paper Award from IEEE GlobalSIP in 2015, IEEE Communications Society Young Author Best Paper Award in 2017, the Best Paper Award IWCMC in 2017, the Best Journal Paper Award of IEEE ComSoc Technical Committee on Communications Systems Integration and Modeling in 2018.

Haijun Zhang (M'13, SM'17) is currently a Full Professor in University of Science and Technology Beijing, China. He serves as an Editor of IEEE Transactions on Communications, IEEE Transactions on Green Communications Networking, and IEEE Communications Letters. He received the IEEE CSIM Technical Committee Best Journal Paper Award in 2018, IEEE ComSoc Young Author Best Paper Award in 2017, and IEEE ComSoc Asia-Pacific Best Young Researcher Award in 2019.

Yong Ren (M'11-SM'16) received his B.S, M.S and Ph.D. degrees in electronic engineering from Harbin Institute of Technology, China, in 1984, 1987, and 1994, respectively. He worked as a post doctor at Department of Electronics Engineering, Tsinghua University, China from 1995 to 1997. Now he is a professor of Department of Electronics Engineering and the director of the Complexity Engineered Systems Lab in Tsinghua University. He holds 60 patents, and has authored or co-authored more than 300 technical papers in the behavior of computer network, P2P network and cognitive networks. He has serves as a reviewer of IEICE Transactions on Communications, Digital Signal Processing, Chinese Physics Letters, Chinese Journal of Electronics, Chinese Journal of Computer Science and Technology, Chinese Journal of Aeronautics and so on. His current research interests include complex systems theory and its applications to the optimization and information sharing of the Internet, Internet of Things and ubiquitous network, cognitive networks and Cyber-Physical Systems.

Kwang-Cheng Chen (F'07) received the B.S. degree from the National Taiwan University in 1983, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 1987 and 1989, all in electrical engineering. From 1987 to 1998, he was with SSE, COMSAT, IBM Thomas J. Watson Research Center, and National Tsing Hua University, working on mobile communications and networks. From 1998 to 2016, he was a Distinguished Professor with the National Taiwan University, Taipei, Taiwan. He also served as the Director of the Graduate Institute of Communication Engineering, the Director of the Communication Research Center, and the Associate Dean for Academic Affairs with the College of Electrical Engineering and Computer Science, from 2009 to 2015. Since 2016, he has been a Professor of electrical engineering with the University of South Florida, Tampa, FL, USA. His recent research interests include wireless networks, artificial intelligence and machine learning, IoT and CPS, social networks, and cybersecurity. He has been actively involved in the organization of various IEEE conferences as the General/TPC Chair/Co-Chair, and has served in editorships with a few IEEE journals. He also actively participates in and has contributed essential technology to various IEEE 802, Bluetooth, LTE and LTE-A, 5G-NR, and ITU-T FG ML5G wireless standards. He has received a number of awards, including the 2011 IEEE COMSOC WTC Recognition Award, the 2014 IEEE Jack Neubauer Memorial Award, and the 2014 IEEE COMSOC AP Outstanding Paper Award.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk, https://en.wikipedia.org/wiki/Lajos_Hanzo) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his Master degree in electronics in 1976 and his doctorate in 1983. He holds an honorary doctorate by the Technical University of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served as Governor of both IEEE ComSoc and of VTS. He has published 1900+ contributions at IEEE Xplore, 18 Wiley-IEEE Press books and has helped the fast-track career of 119 PhD students.