Ali *et al. EURASIP Journal on Wireless Communications and Networking* (2019) 2019:178
https://doi.org/10.1186/s13638-019-1498-x

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Q-learning-enabled channel access in next-generation dense wireless networks for IoT-based eHealth systems

Rashid Ali[1], Yazdan Ahmad Qadri[1], Yousaf Bin Zikria[1], Tariq Umer[2], Byung-Seo Kim[3] and Sung Won Kim[1*]

Abstract

One of the key applications for the Internet of Things (IoT) is the eHealth service that targets sustaining patient health information in digital environments, such as the Internet cloud with the help of advanced communication technologies. In eHealth systems, wireless networks, such as wireless local area networks (WLAN), wireless body sensor networks (WBSN), and wireless medical sensor networks (WMSNs), are prominent technologies for early diagnosis and effective cures. The next generation of these wireless networks for IoT-based eHealth services is expected to confront densely deployed sensor environments and radically new applications. To satisfy the diverse requirements of such dense IoT-based eHealth systems, WLANs will have to face the challenge of assisting medium access control (MAC) layer channel access in intelligent adaptive learning and decision-making. Machine learning (ML) offers services as a promising machine intelligence tool for wireless-enabled IoT devices. It is anticipated that upcoming IoT-based eHealth systems will independently access the most desired channel resources with the assistance of sophisticated wireless channel condition inference. Therefore, in this study, we briefly review the fundamental models of ML and discuss their employment in the persuasive applications of IoT-based systems. Furthermore, we propose Q-learning (QL) that is one of the reinforcement learning (RL) paradigms as the future ML paradigm for MAC layer channel access in next-generation dense WLANs for IoT-based eHealth systems. Our goal is to contribute to refining the motivation, problem formulation, and methodology of powerful ML algorithms for MAC layer channel access in the framework of future dense WLANs. This paper also presents a case study of next-generation WLAN IEEE 802.11ax that utilizes the QL algorithm for intelligent MAC layer channel access. The proposed QL-based algorithm optimizes the performance of WLAN, especially for densely deployed devices environment.

Keywords: Internet of Things, eHealth systems, Machine learning, Next-generation dense WLANs, MAC layer channel access

1 Introduction

Internet of Things (IoT) technology connects physical objects with the help of sensors and actuators by utilizing the existing infrastructure of communication networks, specifically with the help of unlicensed wireless networks [1]. Therefore, IoT technology uses the existing network infrastructure and communication technologies to ensure its strength. Sensors and actuators play a vital role in connecting the physical world to the digital world [2, 3]. The applications of IoT technology such as smart-cities, smart-industries, smart-metering, smart-grid, and smart-healthcare systems (IoT-based eHealth) are continuously increasing. It is expected that by the end of 2020, wireless-enabled devices will increase to 36.5 billion, and 70% of those would comprise sensor devices [1].

One of the key applications for the IoT is the eHealth service that targets sustaining patient health information in digital environments such as the Internet cloud with the help of advanced communication technologies. The World Health Organization (WHO) conducted a survey in 2013 and highlighted that upcoming decades would face the challenge of shortage of global health workforce, which would reach 12.9 million [4]. The main reasons of

*Correspondence: swon@yu.ac.kr
[1]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea
Full list of author information is available at the end of the article

the decline are decreased interest in young people to pursue this profession, aging of current workforce, and the growing risk of non-infectious diseases such as cancer and heart stroke [4]. However, nowadays, health-related information can be easily monitored and tracked with the help of smart sensors and devices. This IoT-based eHealth enables people to allow emergency services/hospitals, doctors, and relatives to access their health-related data through different applications for immediate and efficient treatment. Handheld devices, such as smartphones and fitness bands, can act as on-body coordinators for personalized health monitoring because they are equipped with a variety of sensors, such as heart rate measurement sensor, blood glucose and pressure sensors, temperature sensors, humidity measurement sensors, accelerometers, magnetometers, and gyroscope (Fig. 1) [5]. There exist several built-in applications in such handheld smartphones, such as S-Health, to keep track of daily body fitness. However, there are always concerns regarding data privacy and security, reliability, and trustworthiness in the extensive usage of wearable smart devices [5].

One of the key issues in IoT-based eHealth systems is the requirement of appropriate communication technologies for efficient information sharing [6, 7]. Particularly, reliable connectivity is essential for real-time health-related information sharing. Wireless communication technologies are flexible and cost-effective for IoT-based information sharing. As shown in Fig. 1, a combination of both short-range wireless communication technologies, such as Bluetooth Low Energy (BLE), ZigBee, and IEEE 802.11 wireless local area network (WLAN), and long-range wireless communication technologies, such as Sub-1 Giga, LoRaWAN, and 4G/5G/LTE cellular systems, are typically considered [8, 9]. Both academic and industrial communities have recognized the significant attention given to future WLANs (IEEE 802.11) for IoT-based eHealth systems. One of their motivating services is the promisingly high throughput to support extensively advanced technologies even in densely deployed devices environment [10, 11]. However, unlicensed WLAN would face huge challenges in the future to access the shared channel resources, especially for highly dense IoT device deployments. The use of small cells and information-centric sensor networks in forthcoming IoT-based system may help to reduce the performance degradation issues [3, 12]. The most popular wireless channel resource utilization technique utilized by the WLAN medium access control (MAC) protocol is known as carrier sense multiple access with collision avoidance (CSMA/CA). To achieve maximum channel resource utilization through fair channel access in the WLANs with the ever-increasing density of contending IoT devices, the CSMA/CA scheme is very important as a part of IoT-based systems. CSMA/CA uses a binary exponential backoff (BEB) as its typical and

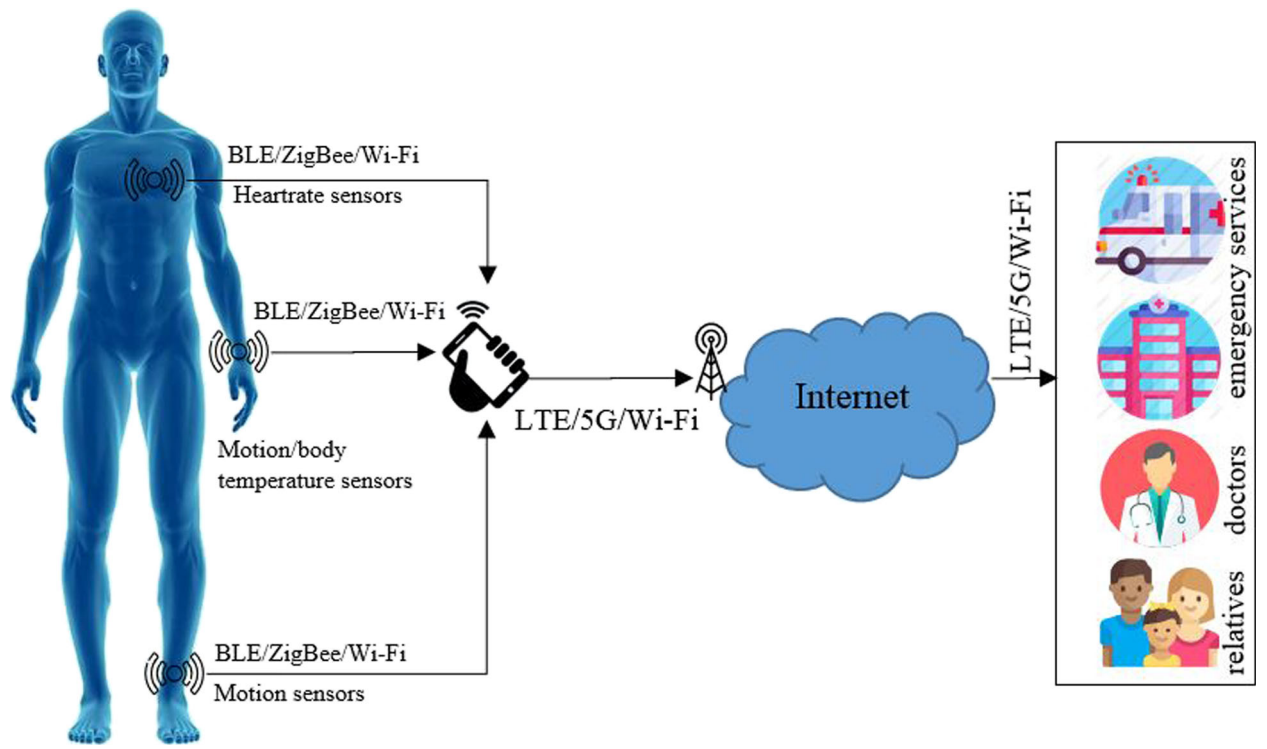

Fig. 1 Role of communication technologies in IoT-based eHealth system

traditional channel contention mechanism [11]. In BEB, a backoff value for contention is generated randomly from a specified contention window (CW). The CW size is exponentially increased for each unsuccessful transmission and reset to its initial size once transmitted successfully. For a network with a heavy load, resetting CW to its minimum size after successful transmission will result in more collisions and poor network performance. Similarly, for fewer contending devices, the blind exponential increase of CW for collision avoidance causes an unnecessary long delay. Besides, this blind increase/decrease of the backoff CW is more inefficient in highly dense networks proposed for IoT-based systems. Thus, the current CSMA/CA mechanism does not allow wireless networks to achieve high efficiency in highly dense environments.

Future dense WLANs are anticipated to infer the diverse and interesting features of both the devices' environments and their behavior to spontaneously optimize the reliability and efficiency of communication. Machine learning (ML), which is one of the prevailing machine intelligence tools, establishes an auspicious paradigm for optimization of the performance of WLANs [13]. As illustrated in Fig. 2, we can imagine an intelligent IoT device that is capable of accessing channel resources with the aid of ML. Therefore, an intelligent device would observe and learn the performance of a specific action with the objective of preserving a specific performance metric. Further, based on this learning, the intelligent device aims to reliably improve its performance while executing future actions by exploiting previous experience. ML algorithms are typically categorized into supervised [14] or unsupervised [15] learning algorithms. The supervised and unsupervised algorithms specify whether there are categorized samples in the available data (usually known as training data). Recently, another class of ML, known as reinforcement learning (RL), has emerged. It is encouraged by behavioral psychology [16, 17]. RL is concerned with a certain form of reward for a learner (such as an intelligent IoT device) that is associated with its environment (such as IoT-based eHealth system) through its observations and actions.

In this study, we briefly assess the fundamental perceptions of ML and propose services in persuasive applications for IoT-based systems based on the supervised, unsupervised, and RL categories. ML can be used extensively for revealing numerous practical problems in the future dense WLANs of the IoT-based application like eHealth systems. Examples include massive multiple-input multiple-output (MIMO), device-to-device (D2D) communications, femto/small cell-based heterogeneous networks, and high contention in dense WLAN environments. Following are the contributions of this paper:

- We briefly present the fundamental insights of ML in persuasive applications for IoT-based systems.
- Furthermore, we propose Q-learning (QL) that is one of the prevailing algorithms of RL as the future ML paradigm for channel access in contention-based dense WLANs for IoT-based systems.

The goal of this paper is to aid readers in refining the enthusiasm for problem devising and the approach to powerful ML algorithms for channel access in the framework of future dense WLANs to tap into previously unexplored applications of IoT-based systems. Table 1 shows list of acronyms used in this paper.

2 Machine learning in WLANs for IoT-based systems

As aforementioned, ML is usually categorized as supervised, unsupervised, and the most recently evolved RL algorithms. In this section, we elaborate the role of these categories in wireless communication networks for IoT-based systems. Figure 3 summarizes the family architecture of ML techniques, models, and their potential applications in dense IoT-based systems.

2.1 Supervised learning

In supervised ML, the learning agent learns from a labeled training dataset supervised by an erudite exterior supervisor. Each labeled training dataset is a depiction of a state comprising a specification, label, particular action,

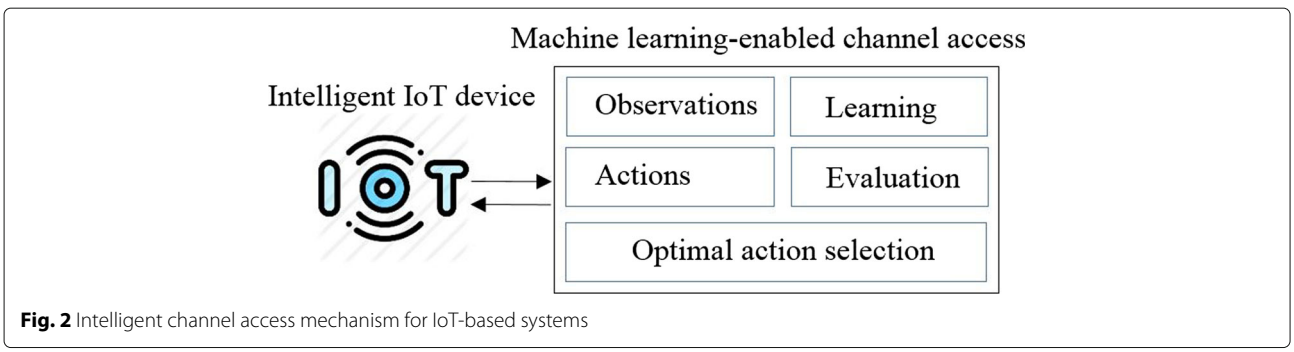

Fig. 2 Intelligent channel access mechanism for IoT-based systems

Table 1 List of acronyms used in this paper

Acronyms	Full description
AP	Access point
BL	Bayesian learning
BLE	Bluetooth Low Energy
CRNs	Cognitive radio networks
CSMA/CA	Carrier sense multiple access/collision avoidance
D2D	Device-to-device
HMM	Hidden Markov model
ICA	Independent component analysis
IoT	Internet of Things
LoRaWAN	Long-range wireless area network
LTE	Long-Term Evolution
MAC	Medium access control
MDP	Markov decision process
MIMO	Multiple-input multiple-output
MIP	Mixed integer programming
ML	Machine learning
PCA	Principle component analysis
POMDP	Partially observed MDP
QL	Q-learning
RA	Regression analysis
RL	Reinforcement learning
SVM	Support vector machine
WBSN	Wireless body sensor network
WHO	World Health Organization
WLAN	Wireless local area network
WMSN	Wireless medical sensor network

and class to which that particular action belongs. The objective of supervised ML is to make the system infer its retorts so that it acts intelligently in states not present in the labeled training dataset [18]. Although supervised ML is a significant type of ML, it is not suitable for a learner to learn the environment without the help of a supervisor and the available training dataset in it. Therefore, for the systems that need to deal interactively, it is often impractical to obtain a sample training dataset of anticipated behavior that is equally precise and descriptive regarding all the states in which the device has to perform actions in the future. In an unexplored environment, wherein ML is expected to be most valuable, a device must be able to learn from its own experience of interaction with the environment [18, 19].

Examples of supervised ML algorithms are regression models [20], k-nearest neighbor (KNN) [21], support vector machine (SVM) [22], and Bayesian learning (BL) [18]. Regression analysis (RA) depends on a statistical method for assessing the relations among input parameters. The objective of RA is to envisage the assessment of one or more continuously valued estimation objectives, given the assessment of a vector of input parameters. The estimation objective is a function of the independent parameters. The KNN and SVM techniques are mostly employed to categorize different objects in the system. In the KNN technique, an agent/device is categorized according to the votes of the neighbor agents. The agent is associated with the category that is most common among its k-nearest neighbors. On the contrary, the SVM algorithm uses non-linear mapping for object classification. First, it converts the original training dataset into a higher measurement, where it befits distinguishability. Later, it explores for the optimized linearly separating hyperplane that is accomplished by distinguishing one category of agents from another [18]. On the contrary, the idea of BL is to estimate a posterior distribution of the target variables, given some inputs and the available training datasets. The hidden Markov model (HMM) is a simple example of reproductive paradigms that can be learned with the help of BL [19]. HMM is a tool for expressing probability distributions of the trail of observations in the system. More specifically, it is a generalization method, where the unseen (hidden) variables of the system are associated with each other through a Markov decision process (MDP) [23]. These hidden variables control the particular constituent to be selected for each observation, while being relatively independent of each other.

These examples of supervised ML paradigms can be used for estimating wireless radio parameters that are related to the quality of service and quality of experience requirements of a particular user/device. Similar to a massive MIMO system of hundreds of radio antennas, the available channel estimation may lead to optimal dimensional search problems, which can be easily learned using any of the abovementioned supervised learning models. The SVM functions are cooperative for data classification problems. A hierarchical SVM (H-SVM), in which each hierarchical level is comprised of a fixed number of SVM classifiers, was proposed in [23]. H-SVM is used to intelligently estimate the Gaussian channel's noise level in a MIMO system by exploiting the training data. KNN and SVM can be pragmatic in finding the optimum handover solutions in wireless networks. Similarly, the BL model can be invoked for wireless channel characteristics learning and estimation in future generation ultra-dense wireless networks. For example, Wen et al. [24] estimated both the radio channel parameters in a specific radio cell and those of the intrusive links of the neighboring radio cells using BL techniques to deal with the pilot contamination problem faced by massive MIMO systems. Another application of BL was proposed in [25], where a Bayesian inference model was proposed for considering and statistically describing a variety of methods that are

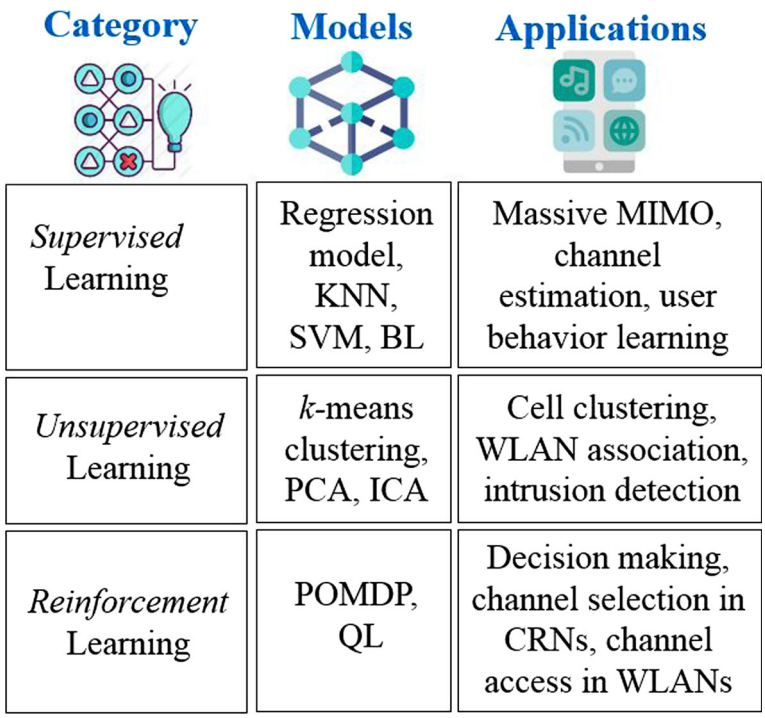

Fig. 3 Machine learning family architecture models and their potential applications in dense IoT systems

proficient at learning the predominant factors for cognitive radio networks (CRNs). Their proposed mechanism covers both the MAC and the network layers of a wireless network.

2.2 Unsupervised learning

Unsupervised ML is usually regarding the verdict structure veiled in a collection of unlabeled training datasets. The terms supervised ML and unsupervised ML would appear to profoundly categorize most ML-based paradigms; however, they are not accurate. The aim of supervised ML is to learn the mapping from an input dataset to an output result where accurate values are provided by a supervisor. On the contrary, in unsupervised learning, there is no external supervisor but only the available input dataset. The objective is to find symmetries in the dataset. There is an edifice of the available dataset space, e.g., that certain patterns occur often, such patterns can help understand the action to be performed in the future for any unknown input. In the statistical context, this is also known as density estimation [18].

Examples of unsupervised ML algorithms are k-means clustering [21], principle component analysis (PCA) [26], and independent component analysis (ICA) [27]. The objective of k-means clustering is to divide user observations into k clusters, where each observation is associated with the adjacent cluster. It uses the center of gravity (centroid) of the cluster, which is the mean value of the observation points within that particular cluster. Continuous iteration of the k-means clustering algorithm keeps assigning an agent to the particular cluster in which the centroid is close to the agent based on a similarity metric. This similarity metric is known as Euclidean distance. Further, the in-cluster differences are also minimized until convergence by iteratively updating the cluster centroid is achieved [18]. PCA is used to transform a set of possibly associated parameters into a set of unassociated parameters that are known as the principal components (PCs). The number of PCs is always less than or equal to the number of original parameters/components. The first PC has the largest possible variance, and each subsequent PC has the utmost variance probable under the limitation that it is unassociated with the prior PCs. Basically, the PCs are orthogonal (unassociated) because they are the eigenvectors of the covariance matrix that is symmetric. Unlike PCA, ICA is a statistical method applied to expose unseen elements that inspire sets of haphazard parameters/components within the system [18].

Clustering is one of the common problems in densely deployed wireless networks of IoT-based systems, especially in heterogeneous network environments with diverse cell sizes. In such cases, small cells have to be wisely grouped to avoid interference using coordinated multi-point transmission, whereas the mobile devices are grouped to follow an optimum offloading strategy. The devices are grouped in device-to-device (D2D) wireless

networks to attain high energy efficiency, and the WLAN users are grouped to uphold an optimum access point (AP) association. Xia et. al. [28] proposed a hybrid scenario to diminish inclusive wireless traffic by encouraging the exploitation of a high-capacity optical infrastructure. They formulated a mixed-integer programming (MIP) problem to cooperatively optimize both network gateway splitting and the virtual radio channel provision based on typical *k*-means clustering. Both PCA and ICA are formulated to recover statistically autonomous source signals from their linear combinations using powerful statistical signal processing techniques. One of their key applications is in the area of intrusion detection in wireless networks, which depends on traffic monitoring. Besides, similar issues may also be resolved in the dense wireless communications technologies of IoT-based systems. PCA and ICA can also be invoked to classify user behavior in CRNs. In [29], the authors applied PCA and ICA in a smart grid scenario of IoT systems to improve the concurrent wireless transmissions of smart devices set up in the smart home. The statistical possessions of the received signals were oppressed to blindly isolate them using ICA. Their proposed mechanism enhances transmission capability by evading radio channel assessment and data security by excluding any wideband intrusion.

2.3 Reinforcement learning

Reinforcement learning (RL) is motivated by behaviorist sensibility and a control philosophy, where an agent can achieve its objective by interacting with and learning from its surroundings. In RL, the agent does not have clear information whether it is close to its target objective. However, the agent can observe the environment to augment the aggregate reward in an MDP [30]. RL is an ML technique that learns about the environment, what to do, and how to outline circumstances to current actions to maximize a numerical reward signal. Mostly, the agent is not informed about which actions to perform, and it has to learn which actions will produce maximum reward. In some exciting and inspiring situations, it is possible that actions will affect not only the instant reward but also the following state, and consequently, all succeeding rewards. MDPs offer a precise framework for modeling decision-making in particular circumstances, where the consequences are comparatively haphazard, and the decision-maker partially governs the consequences.

Partially observable MDP (POMDP) [31] and QL [17] are the examples of RL. POMDP might be seen as speculation with MDP, where the agent is inadequate to perceive the original state transitions in a straightforward manner; therefore, it only has constrained information. The agent has to retain the trajectory of the probability distribution of the appropriate states based on a set of annotations, and the probability distribution of both the observation probabilities and the original MDP [32]. QL might be conjured up to discover an optimum strategy for performing action from any finite MDP, particularly when the environment is unknown [18].

The uses of POMDP paradigms create vital tools for supportive decision-making in IoT-based systems, where the IoT devices may be considered agents and the wireless network constitutes the environment. In a POMDP problem, the technique first postulates the environment's state space and the agent's action space. Additionally, it endorses the Markov property among the states. Secondly, it constructs the state transition probabilities formulated as the probability of navigating from one state to another under a specific action. The third and final step is to enumerate both the agent's instant reward and its long-term reward via Bellman's equation [17]. Later, a wisely constructed iterative algorithm may be considered to classify the optimum action in each state. The applications of POMDP comprise the network selection problems of heterogeneous networks, channel sensing, and user access in CRNs. In [32], the authors proposed a mechanism for transmission power control problems of energy-harvesting systems, which were scrutinized with the help of the POMDP model. In their proposed investigation, the battery, channel, data transmission, and data reception states are defined as the state space, and an action by the agent is related to transmitting a packet at a certain transmission power. QL, usually in aggregation with the MDP models, has also been used in applications of heterogeneous networks. Alnwaimi et al. [33] presented a heterogeneous, fully distributed, multi-objective strategy for the optimization of femtocells based on a QL model [33]. Their proposed model solves both the channel resource allocation and interference coordination issues in the downlink of heterogeneous femtocell networks. Their proposed model acquires channel distribution awareness and classifies the accessibility of vacant radio channel slots for the establishment of opportunistic access. Further, it helps choose sub-channels from the vacant spectrum pool.

3 Q-learning-enabled channel access for dense WLANs

As described in the previous section, QL has already been extensively applied in heterogeneous wireless networks [14]. In such a case, the QL paradigm also covers a set of states where an agent can make a decision on an action from a set of available actions. By performing an action in a particular state, the agent collects a reward with the objective of exploiting its collective rewards. A collective reward is illustrated as a Q-function and is updated in an iterative approach after the agent performs an action and attains the subsequent reward [18]. The trade-off between exploration and exploitation is one of

the challenges arising in QL but not in other types of ML techniques. To achieve considerable rewards, it is obligatory for a QL agent to choose those actions that it has tried before, and found them to be effective in constructing the reward (exploitation). However, to learn more about the environment, an agent has to try actions that it has not selected before (exploration). In exploitation, the agent has to attain what it has already experienced to optimize the process, and additionally, it must explore the environment to maximize the aggregated reward to make better selections in the future. The quandary is that neither exploration nor exploitation can be pursued exclusively without failing in the other process. The agent must try a diversity of actions and gradually favor those that appear to be the best. It is not possible to both explore and exploit with a particular action selection; therefore, we frequently refer to the "tussle" between these two.

3.1 Q-learning prototype

As aforementioned, QL algorithm utilizes a form of RL to solve MDPs without possessing complete information. In addition to the agent and the environment, a QL system has four main sub-elements: a policy, a reward, a Q-value function, and sometimes a model of the environment as an optional entity [17], as shown in Fig. 4.

3.1.1 Policy

The learning agent's manner of behaving at a particular time is defined as a policy. A policy can be a modest utility or a lookup table; however, it may comprise extensive computations such an exploration process. A policy is fundamental for a QL agent because it alone is adequate to determine the behavior of an agent. Generally, policies might be stochastic. A policy decides which action to perform in which state [17].

3.1.2 Reward

In each iteration, the QL agent receives a particular quantity from the environment known as the reward. The main objective of a QL algorithm is to collect as much reward as possible. An agent's exclusive goal is to exploit the accumulated reward collected over the long run. The reward describes the pleasant and unpleasant events for the agent. Reward signals are the instant and crucial topographies of the problem faced by the agent. The agent decides to change its policy based on the reward. For example, if the current action of the policy is followed by a low reward, then an agent may decide to select other actions in the future [17].

3.1.3 Q-value function

Although the reward specifies what is good at one instant, a Q-value function stipulates what is good in the end. Therefore, the Q-value of a state is the accumulated amount of reward that an agent gains at this state to presume in the future [17]. For example, although a state may continuously produce a low instant reward, it may have a high Q-value owing to being repeatedly trailed by other states that produce high rewards. In a WLAN environment, rewards are similar to a high channel collision probability (unpleased) and a low channel collision probability (pleased), whereas Q-values resemble a more sophisticated and prophetic verdict of how pleased or unpleased the agent is in a particular state (e.g., the back-off stage). If there is no reward, then there will be no Q-value, and the only purpose of estimating the Q-value is to attain additional rewards. An agent is most anxious about the Q-value while giving and assessing verdicts. An agent selects optimum actions based on Q-value findings. It seeks actions that carry states of a maximum Q-value and not a maximum reward because these actions attain the highest amount from the rewards for the agent over the long run.

3.1.4 Environment model

Environment model is an optional element of QL, which imitates the performance of the system to some extent. Typically, it allows drawing inferences to be made about how the environment will perform [17]. For example, given a state and an action, the model might envision the

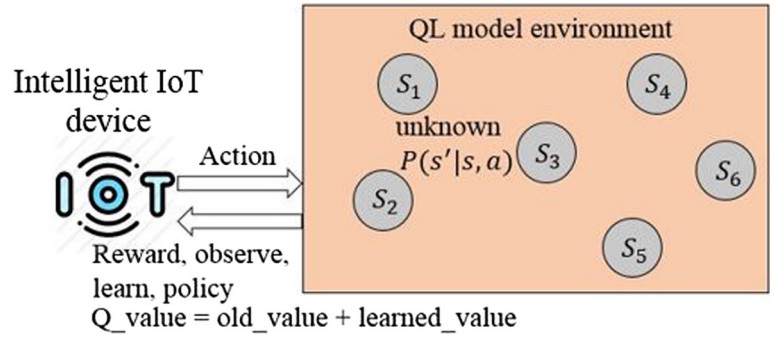

Fig. 4 Q-learning system environment with its sub-elements

subsequent state and the next reward. Environment models are used for planning a method to decide on a sequence of actions by considering latent future situations. In an example of a WLAN system, a device would like to plan its future decisions based on the given state (e.g., the backoff stage) and action, along with its rewards (e.g., channel collision probability).

3.2 Q-learning algorithm

Let S represent a finite set of conceivable states of an environment and A represent a finite set of allowable actions to be performed. At time t, a learner (IoT device) observes the current state (s) of the environment and performs an action (a), i.e., $a_t = a \in A$, based on both the apparent state and its previous experience. The action a_t changes the environmental state from s_t to $s_{t+1} = s^* \in S$; consequently, the agent receives the reward (r) at time t, r_t for the specific action: a_t. The QL algorithm finds an optimal policy for state s that optimizes the rewards over a long period of time. In the QL algorithm, a Q-value function, $Q(s,a)$, estimates the reward as the cumulative discounted reward. An optimal Q-value, i.e., $Q^{\text{opt}}(s,a)$, is determined using the Q-values. The QL algorithm finds the optimal Q-value in a greedy manner. The Q-value is updated as:

$$Q(s,a) = (1-\alpha) \times Q(s,a) + \alpha \times \Delta Q(s,a), \tag{1}$$

where α is the learning rate and takes values such as $0 \leq \alpha \leq 1$. When α is minimum, i.e., zero, the agent does not learn from the environment; therefore, the Q-value is not updated. When α is maximum, i.e., 1, the agent always learns; therefore, learning occurs quickly as seen in the following equation:

$$\Delta Q(s,a) = \left\{ r(s,a) + \beta \times \max_{a'} Q\left(s', a'\right) \right\} - Q(s,a), \tag{2}$$

where β ($0 \leq \beta \leq 1$) weighs the immediate rewards more heavily than future rewards, and is known as the discount factor. Over a considerable period of time, $Q(s,a)$ converges into $Q^{\text{opt}}(s,a)$. The simplest policy for action selection is to choose one of the actions with the maximum measured Q-value (i.e., exploitation). If there are more than one greedy actions, then a choice is randomly made among them. This greedy action selection method can be written as:

$$a^{\text{opt}} = \text{argmax}_a Q(s,a), \tag{3}$$

where argmax_a signifies the action a, for which the expression that follows it is exploited. An agent continuously exploits current knowledge to maximize the instant reward. A simple substitute is to perform greedily in most cases; however, sometimes (e.g., with a small probability ε) the agent can randomly select from all the equal probability actions, independent of the Q-value. The method using this greedy and non-greedy action selection rule is known as the ε-greedy method [17]. An advantage of such a technique is that as the number of iterations increases, every action will guarantee that $Q(s,a)$ converges to $Q^{\text{opt}}(s,a)$. This leads to the inference that the probability of choosing the optimum action converges to a value that is larger than $1-\varepsilon$, i.e., to adjacent certainty. In WLANs, for dense IoT-based systems, an agent would choose greedy actions from high-value actions (exploitation) to improve the throughput performance, and would perform a non-greedy action (exploration) to know the dynamicity of the network environment.

3.3 Case study: DCF-based backoff mechanism

The QL-based channel access scheme can be used to guide densely deployed IoT devices and allocate radio resources more efficiently. When an IoT device is deployed in a new environment, usually, no data are available on historical scenarios. Therefore, QL algorithms are the best choice to observe and learn the environment for optimal policy selection. For example, we consider the case study of DCF-based backoff mechanism of dense WLANs in IoT-based systems. In a densely deployed WLAN, channel collision is the most vital issue causing performance degradation. To tackle collision issues at the MAC layer, we propose adopting the QL algorithm. QL finds solutions through interacting and learning with an environment; therefore, we propose using the QL algorithm to model the optimal contention window (CW) in a channel observation-based

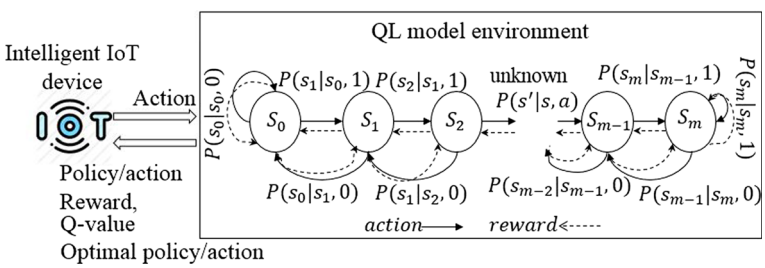

Fig. 5 QL model environment with its elements in the case study of DCF-based backoff mechanism

scaled backoff (COSB) mechanism [34] for dense wireless networks of IoT-based systems. In other words, a station (STA; a WLAN-enabled IoT device referred to as an STA) controls the CW selection intelligently with the aid of the QL-based algorithm.

In COSB [34] protocol, STAs select a random backoff value from the initial CW (CW_{min}) to contend for the wireless medium after observing the channel in an idle state for a distributed inter-frame space (DIFS) period. The period after DIFS is divided into B_{obs} discrete observation time slots. The duration of each discrete time slot is either a constant idle slot time (σ) or a variable busy slot time (owing to successful or collided transmission). In COSB, each STA proficiently measures the channel observation-based collision probability (p_{obs}) as:

$$p_{obs} = 1/B_{obs} \times \sum_{(k=0)}^{(B_{obs}-1)} S_k, \tag{4}$$

where $S_k = 0$ if B_{obs} is observed as idle or if the transmission is successful, whereas $S_k = 1$ if B_{obs} is observed as busy or the transmission has collided [34].

We assume backoff stages of COSB as a set of m states, i.e., $S = \{0, 1, 2, ..., m\}$, where an intelligent IoT device performs an action a from a finite set of permissible actions $A = \{0, 1\}$, where 0 indicates decrement and 1 indicates increment. This is because in COSB, there are two possible actions: increase or decrease the CW size [34]. At time t, the STA collects reward r_t in the response to an action a_t following policy π in a particular state s_t; i.e., $r_t(s_t, a_t)$ with the objective to exploit collective reward $Q(s_t, a_t)$, which is a Q-value function defined in Eqs. (1) and (2). Figure 5 depicts the proposed QL model environment with its elements in a DCF-based backoff mechanism for channel access in WLANs.

The selection of optimal action following π^{opt} is known as a greedy action $\left(a^{\pi^{opt}}\right)$ selection policy that is defined in Eq. (3). A naivest policy can be to exploit in most cases; however, sometimes, the STA explores according to the default policy π, independent of $a^{\pi^{opt}}$. The exploration with probability (ε) and exploitation with probability $(1 - \varepsilon)$ is called ε-greedy method [17]. The ε-greedy technique guarantees the convergence of learning estimate $\Delta Q(s, a)$ with the increase of episodes (instances). In a dense WLAN environment, exploitation can be used to improve throughput performance by an IoT device, and exploration can be used to know the dynamicity of the WLAN environment.

In COSB [34], an STA conducts p_{obs} at every transmission attempt. Therefore, we express p_{obs} as the reward of the action at any specific state. Therefore, reward r_t produced by action a_t taken in state s_t at time t can be described as:

Table 2 MAC layer and PHY layer simulation parameters

Parameter type	Value
Frequency	5 GHz
Channel bandwidth	160/20 MHz
Data rate	1201/11 Mbps
Payload size	1472 bytes
Transmission range	10 m
Simulation time	100/500 s
Propagation loss model	Log distance
Mobility model	Constant position
Rate adaptation models	Constant rate/minsttrel

$$r_t(s_t, a_t) = 1 - p_{obs}. \tag{5}$$

The above equation indicates how pleased the STA was with its action a_t in state s_t.

4 Experimental results and discussion

We used ns3.28 [35] simulator to perform experiments of the proposed *i*QRA mechanism. Some important PHY layer and MAC layer simulation parameters are shown in Table 2. The results in Fig. 6a and b indicate that a small value of α and a large value of β make ΔQ (learning estimate) converge faster. The convergence of ΔQ clearly indicates that there exist optimal values that can be learned and exploited in the future. The throughput performance optimization of COSB using proposed *i*QRA is depicted in Fig. 7a. The performance of *i*QRA may degrade in small networks (i.e., for < 10 contending STAs as shown in Fig. 7a owing to low and irregular rewards).

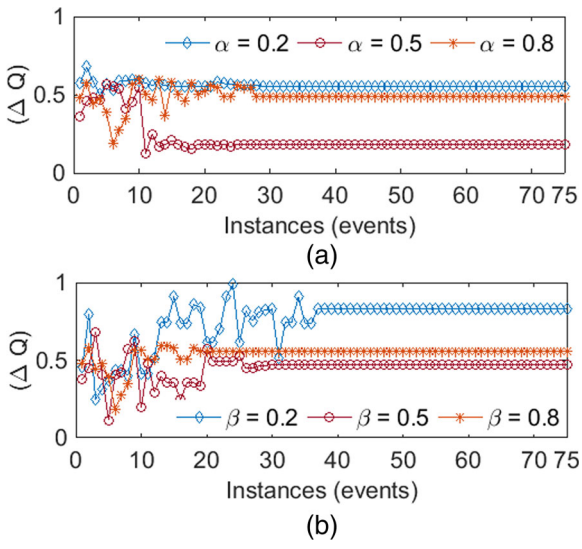

Fig. 6 Learning estimate (ΔQ) convergence with varying **a** learning rate α ($\varepsilon = 0.5$) and **b** discount factor β ($\varepsilon = 0.5$)

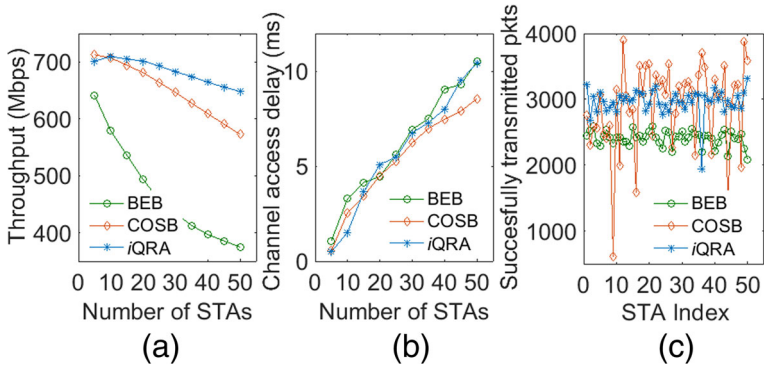

Fig. 7 Comparison of BEB, COSB, and *i*QRA for **a** throughput (Mbps), **b** channel access delay (ms), and **c** successfully transmitted packets (fairness)

Additionally, the channel access delay is also increased for *i*QRA as compared to COSB; this is obvious owing to the environment inference characteristics; however, it remains lower than the conventional binary exponential backoff (BEB) [11] mechanism shown in Fig. 7b. Figure 7c portrays that the proposed *i*QRA also improves the fairness of COSB. The optimized performance of COSB using *i*QRA clearly stipulates that the QL-based proposed mechanism is effective in learning the network environment. Additionally, *i*QRA is essentially intended to intelligently adjust its learning parameters according to the dynamics of the WLAN. Therefore, we simulated a dynamic network environment by increasing the number of contenders by 5 after every 50 s until the number of STAs reached 50. Figure 8 depicts the properties of network dynamics on ΔQ. The figure shows simulation of a 500 s period with 1500 learning instances of a tagged STA. As shown in the figure, with the network dynamics, a tagged STA observes fluctuation in its learning estimate ΔQ, thereby indicating the inference of change in the network. We see that the throughput performance of *i*QRA eventually reaches a steady state in a dynamic network environment, as shown in Fig. 9a. To evaluate the performance of the proposed *i*QRA for moving devices in the network, we simulated a distance-based rate adaptation model. This model changes the transmission rate of the sender device according to the distance between the sender and receiver to achieve the best possible performance. IEEE 802.11a (11 Mbps) WLAN with 10 contending STAs is simulated for distance-based rate adaptation performance evaluation, as shown in Fig. 9b. Contending STAs are placed randomly around the access point (AP) within a distance of 25 m. A tagged STA starts moving away from the AP that is initially placed at a 1-m distance. As the distance from the AP increases, performance of a tagged STA degrades for all the three compared algorithms (BEB, COSB, and *i*QRA), as shown in Fig. 9b. It is observed that the throughput of the BEB algorithm approaches close to zero after the STA reaches a distance of 60 m, and it finally becomes zero after reaching a distance of 80 m. Owing to the observation-based nature of COSB, it achieves higher throughput even after a 60-m distance, compared to BEB. However, the proposed *i*QRA performs optimally, even if the distance reaches 80 m owing to its network inference capability.

5 Conclusion

In this study, we investigated the benefits of ML-based intelligent dense wireless networks for IoT-enabled eHealth systems. We presented the key families of ML algorithms and deliberated their application in the context of dense IoT systems including next-generation wireless networks with massive MIMO; heterogeneous IoT

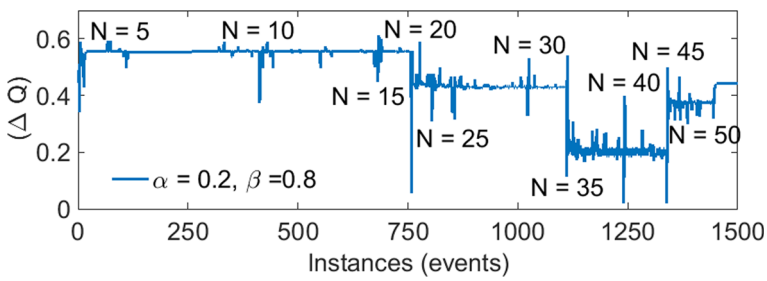

Fig. 8 Convergence of learning estimate (ΔQ) in a dynamic network environment (increasing the number of contenders after every 50 s)

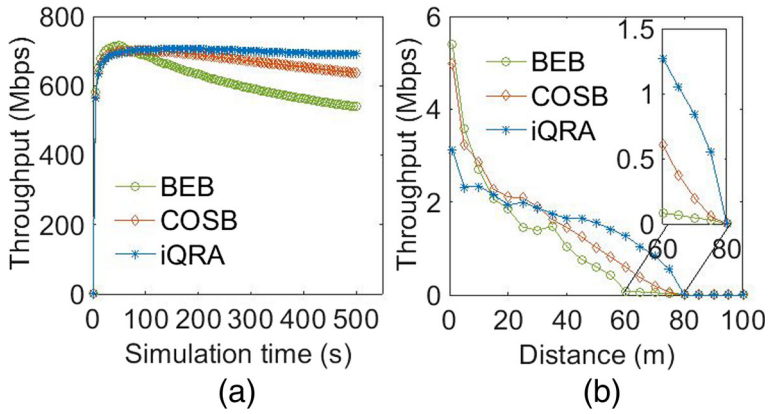

Fig. 9 Throughput comparison of BEB, COSB, and *i*QRA in **a** dynamic network environment with increasing number of contenders after every 50 s and **b** distance-based rate adaptation network environment

networks based on small cells; smart applications, such as the smart grid and smart city; and intelligent cognitive radio. The three well-known categories of ML, supervised learning, unsupervised learning, and RL algorithms, are scrutinized in addition to a consistent sculpting methodology and possible future applications in dense IoT systems. Furthermore, we proposed Q-learning as a promising ML paradigm for MAC layer channel access in dense IoT systems. The proposed paradigm is implemented on a case study of DCF-based backoff mechanism in dense WLANs. We proposed an intelligent Q-learning-based resource allocation (*i*QRA) mechanism to optimize the performance of an existing (COSB) mechanism. The proposed *i*QRA mechanism infers unknown wireless network conditions and exploits rapidly unexpected changes to learn dynamicity in dense WLANs. The experimental results show that *i*QRA significantly enhances the performance of COSB in terms of throughput and fairness. Results reveal the ability of the Q-learning scheme to determine dense wireless network environments in IoT-based systems. In conclusion, ML is a promising area for self-scrutinized intelligence-aided dense wireless network research for IoT-enabled eHealth systems.

In the future, we aim to further investigate the applications of our proposed mechanism in various IoT-based systems such as smart city, smart home, smart grid, and smart industry.

Abbreviations

BLE: Bluetooth Low Energy; CSMA/CA: Carrier sense multiple access with collision avoidance; D2D: Device-to-device; IoT: Internet of Things; KNN: *k*-nearest neighbor; MAC: Medium access control; MIMO: Massive multiple-input multiple-output; ML: Machine learning; QL: Q-learning; RL: Reinforcement learning; SVM: Support vector machine; WBSN: Wireless body sensor network; WHO: World Health Organization; WLAN: Wireless local area network; WMSN: Wireless medical sensor networks

Acknowledgements

This work was supported by the 2019 Yeungnam University Research Grant.

Authors' contributions

RA and YAQ conceived the main idea, designed the algorithm, and proposed the intelligent framework for IoT-based eHealth systems. RA, YBZ, and TU performed the implementation of the proposition in the NS3 simulator. BK and SWK contributed to the structuring, reviewing, and finalizing of the manuscript. All authors read and approved the final manuscript.

Funding

The funding source is the same as described in the acknowledgements.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea. [2]Department of Computer Science, COMSATS University Islamabad, Wah, Pakistan. [3]Department of Computer and Information Communication Engineering, Hongik University, Seoul, 04066, Republic of Korea.

Received: 3 April 2019 Accepted: 24 June 2019
Published online: 08 July 2019

References

1. M. Pasha, W. SMShah, Framework for e-health systems in IoT-based environments. Wirel. Commun. Mob. Comput. **2018**(6183732), 1–11 (2018). https://doi.org/10.1155/2018/6183732
2. G. T. Singh, F. Al-Turjman, Learning data delivery paths in QoI-aware information-centric sensor networks. IEEE Internet Things J. **3**(4), 572–580 (2016). https://doi.org/10.1109/JIOT.2015.2504487
3. M. Z. Hasan, F. Al-Turjman, Evaluation of a duty-cycled asynchronous X-MAC protocol for vehicular sensor networks. EURASIP J. Wirel. Commun. Netw. **2017**(1), 95 (2017). https://doi.org/10.1186/s13638-017-0882-7
4. World Health Organization, Global health workforce shortage to reach 12.9 million in coming decades. http://www.who.int/mediacentre/news/releases/2013/health--workforce-shortage/en/. Accessed 10 Dec 2018
5. H. M Alam, M. I. Malik, T. Khan, A. Pardy, Y. L. Kuusik, A. Moullec, Survey on the roles of communication technologies in IoT-based personalized healthcare applications. IEEE Access. **6**, 36611–36631 (2018). https://doi.org/10.1109/ACCESS.2018.2853148
6. M. Faheem, M. Zahid Abbas, G. Tuna, V. C. Gungor, EDHRP: Energy efficient event driven hybrid routing protocol for densely deployed wireless sensor

networks. J. Netw. Comput. Appl. **58**, 309–326 (2015). https://doi.org/10.1016/j.jnca.2015.08.002
7. M. Faheem, V. C. Gungor, Energy efficient and QoS-aware routing protocol for wireless sensor network-based smart grid applications in the context of industry 4.0. Appl. Soft Comput. **68**, 910–922 (2018). https://doi.org/10.1016/j.asoc.2017.07.045
8. M. Faheem, R. A. Butt, B. Raza, M. W. Ashraf, S. Begum, Md. A. Ngadi, V. C. Gungor, in *Transactions on Emerging Telecommunications Technologies*. Bio-inspired routing protocol for WSN-based smart grid applications in the context of Industry 4.0, (2018). https://doi.org/10.1002/ett.3503
9. M. Faheem, V. C. Gungor, Capacity and spectrum-aware communication framework for wireless sensor network-based smart grid applications. Comput. Stand. Interfaces. **53**, 48–58 (2017). https://doi.org/10.1016/j.csi.2017.03.003
10. S. Demir, F. Al-Turjman, Energy scavenging methods for WBAN applications: a review. IEEE Sensors J. **18**(16), 6477–6488 (2018). https://doi.org/10.1109/JSEN.2018.2851187
11. R. Ali, S. W. Kim, B. Kim, Y. Park, Design of MAC layer resource allocation schemes for IEEE 802.11ax: future directions. IETE Tech. Rev. **35**(1), 28–52 (2018). https://doi.org/10.1080/02564602.2016.1242387
12. F. Al-Turjman, E. Ever, H. Zahmatkesh, Small cells in the forthcoming 5G/IoT: traffic modelling and deployment overview. IEEE Commun. Surv. Tutor. **21**(1), 28–65 (2019). https://doi.org/10.1109/COMST.2018.2864779
13. R. Ali, N. Shahin, Y. B. Zikria, B. Kim, S. W. Kim, Deep reinforcement learning paradigm for performance optimization of channel observation-based MAC protocols in dense WLANs. IEEE Access. **7**, 3500–3511 (2019). https://doi.org/10.1109/ACCESS.2018.2886216
14. C. Zhang, P. Patras, H. Haddadi, Deep learning in mobile and wireless networking: a survey. IEEE Commun. Surv. Tutor. Early Access (2019). https://doi.org/10.1109/COMST.2019.2904897
15. Y. Sun, M. Peng, Y. Zhou, Y. Huang, S. Mao, Application of machine learning in wireless networks: key techniques and open issues. ArXiv e-prints (2018). https://arxiv.org/abs/1809.08707
16. E. M. Joo, *Theory and novel applications of machine learning. 12–16.* (IntechOpen, London, 2009). https://doi.org/10.5772/56681
17. R. S. Sutton, A. G. Barto, *Reinforcement learning: an introduction*, Second ed. (MIT Press, Cambridge, 1998). isbn:0262193981
18. E. Alpaydin, *Introduction to machine learning*, Third ed. (MIT Press, Cambridge, 2014). isbn:978-0-262-028189
19. R. Ali, N. Shahin, R. Bajracharya, B. S. Kim, S. W. Kim, A self-scrutinized backoff mechanism for IEEE 802.11ax in 5G unlicensed networks. Sustainability. **10**, 1201 (2018). https://doi.org/10.3390/su10041201
20. Q. H. Abbasi, S. Liaqat, L. Ali, A. Alomainy, in *2013 First International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare (Ubi-HealthTech)*. An improved radio channel characterisation for ultra wideband on-body communications using regression method, (Jinhua, 2013), pp. 1–4. https://doi.org/10.1109/Ubi-HealthTech.2013.6708063
21. Y. Xu, T. Y. Fu, W. C. Lee, J. Winter, Processing k nearest neighbor queries in location-aware sensor networks. Signal Process. **87**(12), 2861–2881 (2007). https://doi.org/10.1016/j.sigpro.2007.05.013
22. Z. Dong, Y. Zhao, Z. Chen, in *IEEE MTT-S International Wireless Symposium (IWS)*. Support vector machine for channel prediction in high-speed railway communication systems, vol. 2018, (Chengdu, 2018), pp. 1–3. https://doi.org/10.1109/IEEE-IWS.2018.8400912
23. V. S. Feng, S. Y. Chang, Determination of wireless networks parameters through parallel hierarchical support vector machines. IEEE Trans. Parallel Distrib. Syst. **23**(3), 505–12 (2012). https://doi.org/10.1109/TPDS.2011.156
24. C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, P. Ting, Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning. IEEE Trans. Wirel. Commun. **14**(3), 1356–68 (2015). https://doi.org/10.1109/TWC.2014.2365813
25. C.-K. Yu, K.-C. Chen, S.-M. Cheng, Cognitive radio network tomography. IEEE Trans. Veh. Technol. **59**(4), 1980–97 (2010). https://doi.org/10.1109/TVT.2010.2044906
26. M. C. Raja, M. M. A. Rabbani, in *2016 International Conference on Communication and Electronics Systems (ICCES)*. Combined analysis of support vector machine and principle component analysis for IDS, (Coimbatore, 2016), pp. 1–5. https://doi.org/10.1109/CESYS.2016.7889868
27. Z. Luo, C. Li, L. Zhu, Full-duplex cognitive radio using guided independent component analysis and cumulant criterion. IEEE Access. **7**, 27065–27074 (2019). https://doi.org/10.1109/ACCESS.2019.2901815
28. M. Xia, Y. Owada, M. Inoue, H. Harai, Optical and wireless hybrid access networks: design and optimization. IEEE/OSA J. Opt. Commun. Netw. **4**(10), 749–59 (2012). https://doi.org/10.1364/JOCN.4.000749
29. R. C. Qiu, Z. Hu, Z. Chen, N. Guo, R. Ranganathan, S. Hou, G. Zheng, Cognitive radio network for the smart grid: experimental system architecture, control algorithms, security, and micro grid testbed. IEEE Trans. Smart Grid. **2**(4), 724–40 (2011). https://doi.org/10.1109/TSG.2011.2160101
30. R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, H. Zhang, Intelligent 5G: when cellular networks meet artificial intelligence. IEEE Wirel. Commun. **24**(5), 175–183 (2017). https://doi.org/10.1109/MWC.2017.1600304WC
31. Y. Li, B. Yin, H. Xi, Partially observable Markov decision processes and performance sensitivity analysis. IEEE Trans. Syst Man Cybern. Part B Cybern. **38**(6), 1645–1651 (2008). https://doi.org/10.1109/TSMCB.2008.927711
32. A. Aprem, C. R. Murthy, N. B. Mehta, Transmit power control policies for energy harvesting sensors with retransmissions. IEEE J. Sel. Top. Signal Process. **7**(5), 895–906 (2013). https://doi.org/10.1109/JSTSP.2013.2258656
33. G. Alnwaimi, S. Vahid, K. Moessner, Dynamic heterogeneous learning games for opportunistic access in LTE-based macro/femtocell deployments. IEEE Trans. Wirel. Commun. **14**(4), 2294–2308 (2015). https://doi.org/10.1109/TWC.2014.2384510
34. R. Ali, N. Shahin, Y. T. Kim, B. S. Kim, S. W. Kim, Channel observation-based scaled backoff mechanism for high-efficiency WLANs. Electron. Lett. **54**(10), 663–665 (2018). https://doi.org/10.1049/el.2018.0617
35. The network simulator-ns-3. https://www.nsnam.org/. Accessed 01 Sept 2018

Publisher's Note