

Article

Leveraging Machine-Learning for D2D Communications in 5G/Beyond 5G Networks

Sherief Hashima ^{1,5*,‡}, Basem M. ElHalawany ^{2,6‡}, Kohei Hatano ^{1,3,‡}, Kaishun Wu ^{2,8,‡}, and Ehab Mahmoud Mohamed ^{4,7‡}

¹ RIKEN Advanced Intelligence Project (AIP), Fukuoka 819-0395 Japan; {sherief.hashima, kohei.hatano}@riken.jp

² Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen University, Shenzhen, 518060, China; {basem.mamdoh, wu}@szu.edu.cn

³ Faculty of Arts and Science, Kyushu University, Fukuoka 819-0395, Japan, and also with RIKEN AIP, Japan; hatano@inf.kyushu-u.ac.jp

⁴ College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Aldwaser 11991, Saudi Arabia; ehab_mahmoud@aswu.edu.eg

⁵ Engineering Dept, Egyptian Atomic Energy Authority, Cairo Inshas 13759, Egypt

⁶ Faculty of Engineering at Shoubra, Benha University, Benha, Egypt

⁷ Electrical Engineering Dept, Faculty of Engineering, Aswan University, Aswan 81542, Egypt

⁸ PCL Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen, China.

* Correspondence: sherief.hashima@riken.jp; Tel: +8109092997501 (Japan)

‡ These authors contributed equally to this work.

Abstract: Device-to-device (D2D) communication is a promising paradigm for the fifth generation (5G) and beyond 5G (B5G) networks. Although D2D communication provides several benefits, including limited interference, energy efficiency, reduced delay, and network overhead, it faces a lot of technical challenges such as network architecture, and neighbor discovery, etc. The complexity of configuring D2D links and managing their interference, especially when using millimeter-wave (mmWave), inspire researchers to leverage different machine-learning (ML) techniques to address these problems towards boosting the performance of D2D networks. In this paper, a comprehensive survey about recent research activities on D2D networks will be explored with putting more emphasis on utilizing mmWave and ML methods. After exploring existing D2D research directions accompanied with their existing conventional solutions, we will show how different ML techniques can be applied to enhance the D2D networks performance over using conventional ways. Then, still open research directions in ML applications on D2D networks will be investigated including their essential needs. A case study of applying multi-armed bandit (MAB) as an efficient online ML tool to enhance the performance of neighbor discovery and selection (NDS) in mmWave D2D networks will be presented. This case study will put emphasis on the high potency of using ML solutions over using the conventional non-ML based methods for highly improving the average throughput performance of mmWave NDS.

Keywords: D2D communication; mmWave; machine-learning applications; 5G; B5G

1. Introduction

Future wireless data traffic keeps growing, especially with recent data-hungry applications such as high definition video, virtual and augmented reality applications. This traffic explosion motivated network operators and designers to race to satisfy the market and customers' needs and expectations. On the other hand, the expected massive connectivity of such applications in B5G and 6G wireless networks presents various challenges in resource allocation (RA), link, and interference management (IM).

Device-to-device (D2D) communication represents one of the main pillars of future networks that facilitates traffic offloading and relaxes the traffic load of the whole system [1]. D2D networks harness

such features by enabling direct communication between wireless nodes without traversing the macro base station (Macro BS) or the core network [2]. Besides traffic offloading, the D2D concept can be exploited to enable communication in disaster situations at which the Macro BS is malfunctioned due to natural disasters such as earthquakes, floods, and typhoons, etc. D2D concept can solve such scenario and either connect them to the closest working ground network or identify their specific locations.

Generally, D2D can be classified into in-band and out-band [1,2] based on the dedicated frequency band. In the in-band D2D networks, D2D communication is overlaid or underlaid the cellular band. However, in the out-band D2D networks, the D2D communication uses the unlicensed frequency bands, i.e., industrial, scientific and medical (ISM) bands. Although out-band D2D has the advantages of high capacity and no-interference with the cellular users (CUs), it suffers from integration/management problems due to the use of different types of interfaces, e.g., LTE and Wi-Fi.

Although D2D communications provide a lot of benefits, it introduces interference to the CUs, especially for in-band schemes. In order to mitigate the interference, many power control and resources reuse algorithms have been proposed in the literature. Another attractive coexistence is the natural symbiosis between the promising millimeter wave (mmWave) band, from 30 up to 300 GHz, and D2D communications [2],[3]. The fact that millimeter waves are characterized by short-range intermittent transmission comes from its fragile channel. It can be sharply directed using antenna beamforming (BF) techniques makes it a perfect candidate to coexist with D2D to create low mutual interference high data rates D2D links. Figure 1 shows some case studies of current critical D2D scenarios in real life, such as disaster management, unmanned aerial vehicles (UAV) communications [4], vehicle to everything (V2X) applications[5], etc.

On the other hand, the recent progress in machine-learning (ML) leads to employing it in various applications in communication networks, including spectrum management, intelligent RA, recovery of channel state information, modulation, detection, and mmWave BF. In this context, researchers have investigated several ML algorithms for enabling efficient D2D communication networks related to technical problems such as neighbor discovery and selection (NDS), optimal D2D pairs matching, RA, and multi-hop D2D relaying for coverage extension. Table 1 summarizes the nomenclature used throughout this paper.

This paper main contributions can be summarized as follow:-

- A comprehensive survey about recent research activities for D2D communication will be explored accompanied with their existing conventional solutions.
- The state-of-the-art ML hypothesis will be presented including its different approaches, i.e. supervised, unsupervised learning, and reinforcement learning (RL). Then, ML solutions for

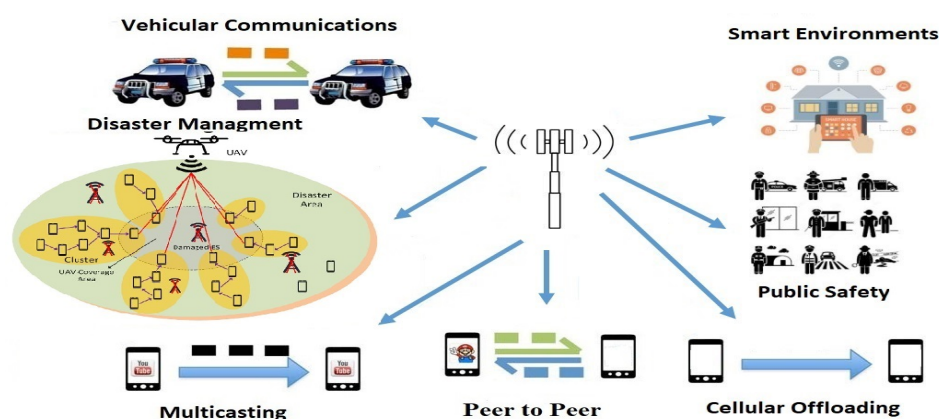


Figure 1. Samples of D2D Applications.

Table 1. Nomenclature

| | | | |
|---------|----------------------------------------------|----------|-------------------------------------------|
| D2D | Device to device | 5G | fifth generation |
| BS | base station | macro BS | macro base station |
| CUs | cellular users | mmwave | millimeter-wave |
| ML | machine-learning | IM | Interference management |
| ISM | industrial, scientific and medical | RA | resource allocation |
| BF | beamforming | UAV | unmanned aerial vehicles |
| V2X | vehicle to everything | NDS | neighbor discovery and selection |
| MAB | multi armed bandit | CMAB | combinatorial MAB |
| 3GPP | 3rd generation partnership project | Prose | proximity services |
| NC | network-centric | DC | device-centric |
| LIA | limited interference area | COMP | coordinated multi point |
| FD | full-duplex | SIC | successive interference cancellation |
| DF | decode and forward | AF | amplify and forward |
| HD | half-duplex | NN | neural network |
| FNN | forward neural network | KNN | K-nearest neighbor |
| SVM | support vector machine | DT | decision tree |
| GPU | graphical processing unit | DNN | deep neural network |
| RNN | recurrent neural network | mmwave | millimeter-wave |
| RECOME | RElative COre MErge | GMM | Gaussian mixture model |
| PCA | principal component analysis | RL | Reinforcement-Learning |
| RL-LCDC | RL-based latency controlled D2D connectivity | RRM | radio resource management |
| SE | spectral efficiency | SINR | signal to interference ratio |
| PC | Power control | MAAC | multi-agent actor critic |
| NAAC | neighbor-agent actor critic | RF | radio frequency |
| HT | Hilbert transform | CV | cross-validation |
| FL | Federated learning | DSGD | decentralized stochastic gradient descent |
| UCB | Upper confidence bound | MOSS | Minimax Optimal Stochastic Strategy |
| LOS | line of sight | NLOS | non line of sight |
| UAV | Unmanned Aerial Vehicle | AOA | Angle of Arrival |

D2D challenges will be provided showing their superior performances over the conventional counterparts.

- We discuss the open research challenges and future research directions of ML based D2D networks, especially with novel ML techniques like federated learning (FL).
- A case study on applying MAB as an efficient online ML tool in enhancing the performance of mmWave NDS will be presented. In this case study, different multi armed bandit (MAB) techniques such as upper confidence bound (UCB) and minimax optimal stochastic strategy (MOSS) will be investigated to show the effectiveness of using ML tools in enhancing the average throughput performance of mmWave NDS over the existing traditional solutions, namely direct NDS and random selection. Besides, we show that such performance enhancements come with a sufficient learning convergence rate.

The rest of the paper is organized as follows: Section II overviews the recent research directions for D2D communications. In section III, we summarize different ML techniques that can be generally employed for D2D network solutions. Section IV discusses different applications of ML algorithms in various D2D scenarios. D2D challenges, future research directions, and applications are introduced in section V. A case study of applying MAB on mmWave D2D is presented by section VI. Finally, section VII concludes the work.

2. Recent Research Direction for D2D Communications

Before we dig deep on the applications of ML in D2D communication, we have to highlight the future research directions for D2D communication in both sub 6GHz and mmWaves bands, as summarized in Fig. 2, which are given as follows.

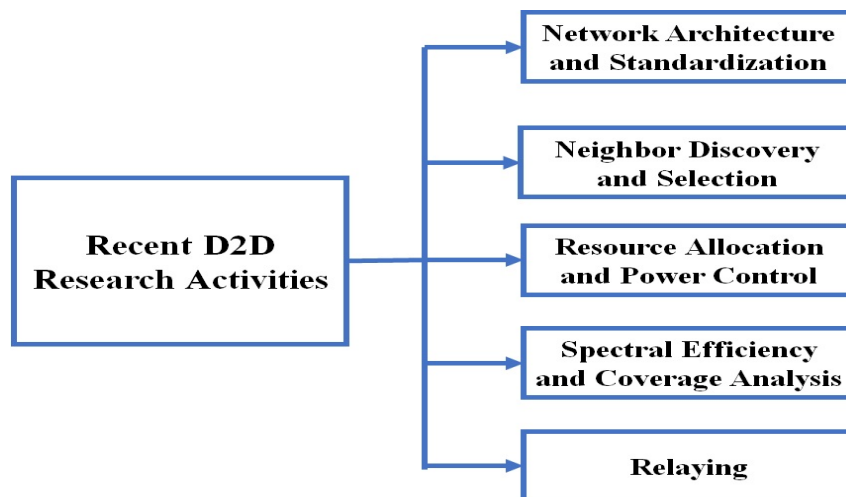


Figure 2. Recent D2D Research Directions

2.1. Network Architectures and Standardization

D2D communication modes have been included in the 3rd generation partnership project (3GPP) by defining the proximity services (ProSe) into the standard [2]. Later, ProSe network architecture has been modified to incorporate mmWaves D2D applications. Towards that, a message exchange entity was added to the ProSe architecture to manage the operations of mmWave links discovery, establishment, and maintenance. Also, a new paradigm of mmWave D2D networks was introduced in [6] based on the interworking between LTE, unlicensed μW and mmWave bands, where wide coverage μW band, e.g., Wi-Fi, was used to overcome the shortcoming of the mmWave transmissions and assists the construction/management of the mmWave D2D links.

2.2. Neighbor Discovery and Selection (NDS)

NDS is a crucial design aspect in D2D networks, in which a device should discover its neighbor devices and select the best one for constructing the D2D link. There are two approaches of D2D ND, namely network-centric (NC) and device-centric (DC). In NC ND, the cellular network itself will discover the nearby devices, while in DC, the devices themselves will do the job. Both NC and DC have their advantages and disadvantages. However, in dense user scenario NC ND performs better than DC ND, and vice versa. Quick ND is preferred due to the limited battery capacity of the devices in addition to reducing the consumed overhead. In mmWave D2D, the problem of ND turns to be more significant due to the use of BT, which consumes high energy and overhead. To overcome this problem, the authors in [6] used out-of-band ND assistance, in which the wide coverage unlicensed μW band is used to assist the discovery of the mmWave devices and hence reducing the overhead and energy consumption.

2.3. Resource Allocation and Power Control

RA and power control (PC) gained considerable attention in the design of the in-band D2D networks, where the D2D users share the same resources with the CUs [7]. A variety of interference mitigation techniques exist in literature to address this problem, which can be divided into interference avoidance, interference coordination and interference cancellation. A limited interference area (LIA) surrounding the D2D users was introduced to prevent any CU from transmitting within the LIA as a way of interference avoidance. For interference coordination, optimal RA and PC were investigated between D2D users and CUs. The genetic algorithm, game theory and the optimization theory were utilized as efficient mathematical tools for coordinating the interference and controlling the power among CUs and D2D users [8],[9]. A new research direction is to use the social activities of the users in

interference mitigation among D2D users and CUs. For interference cancellation, techniques based on successive interference cancellation (SIC), coordinated multi point (CoMP) and full-duplex (FD)-based self-interference cancellation were investigated in the literature to cancel the interference occurs between D2D users and CUs. Fortunately, for the out-of-band D2D networks, including mmWave D2D, the problem of interference between CUs and D2D users does not exist. A low complex and bandwidth efficient transceiver design for superimposed waveform is provided in [10]. A new RA technique based on NOMA called superimposed multi-user shared access (MUSA) that supports much users than conventional techniques was proposed in [11].

2.4. Spectral Efficiency and Coverage Analysis

The main advantages of enabling D2D communications in cellular networks beside relaxing the traffic load on the Macro BS/core network are enhancing the spectral efficiency (SE), reducing outage and increasing the coverage of the D2D users. Several studies were performed to evaluate these metrics in conjunction with D2D networks [1]. Towards that, tools from probability theory and stochastic geometry were extensively used to analyze the performance of D2D networks. Cognitive and energy harvesting D2D networks were modeled and analyzed using a tool from stochastic geometry. Also, the Poisson cluster process was used to model the locations of the devices for coverage analysis of the clustered D2D networks. The coverage probability, the mean number of covered receivers and throughput of the multi-cast D2D transmission were also analyzed using tools from probability theory. Recently, analysis of the underlaid FD D2D network is provided in terms of coverage probabilities and achievable sum-rates for both D2D users and CUs. To study the improvements in mmWave networks coming from enabling D2D links, a tool from stochastic geometry was used to analyze mmWave D2D networks concerning interference, coverage and data rate, especially for mmWave wearable. Also, a fine-grained analysis of mmWave D2D networks using the Poisson bipolar model was given. Moreover, the locations of mmWave devices were modeled by Poisson cluster process to investigate the performance of mmWave clustered D2D network.

2.5. Relaying

The construction of D2D relays was investigated to extend the coverage of the D2D communications, deliver the cellular connection to the out-of-coverage CUs, and route around blockages in case of mmWave D2D. Several D2D relay selection schemes can be found in literature considering different critical parameters when selecting the best relay like the end-to-end data rate, end-to-end delay and the remaining energy of the relayed device. Also, various relaying schemes were considered such as decode and forward (DF), amplify and forward (AF), and demodulate and forward in addition to both half-duplex (HF) and FD transmissions. Optimal resource allocation and power control, in conjunction with D2D relaying were also investigated. Different mathematical tools were used for selecting the relays, such as optimization theory, game theory, fuzzy logic, genetic algorithm, etc. For mmWave D2D, relaying is more critical to not only extend the D2D communication range but also to route around blockages. Another uniqueness of mmWave transmission is the use of BT, which makes the process of relay probing, i.e., exploring the candidate relays, time, and energy consuming. Thus, the research in mmWave D2D relaying is focusing not only on optimizing the relay selection using conventional optimization techniques but also on finding out the optimal number of probed relays considering the trade-off between investigating more relays and maximizing the end-to-end throughput.

3. Overview of Machine-Learning Methods

ML is a branch of artificial intelligence (AI) that allows learning knowledge from examples/data without being explicitly programmed [12]. ML algorithms can find hidden patterns in massive complex data by using different training methodologies, which usually can be categorized as follows:

- **Supervised-Learning:** In this category, the ML model tries to learn a function, $y = f(x)$, that maps an input (x) to an output (y) based on a set of sample pairs (i.e., historical data set), which is used for training the model. There are two sub-categories for supervised-learning, namely the regression and the classification. Regression models like linear and logistic ones that predict real-valued outcomes using linear or sigmoid function approximations [12]. On the other hand, other regression ML models such as neural networks (NNs), random forests, bagging and boosting meta-algorithms are another fundamental regression exploits different techniques [12]. Classification models categorize/classify data samples into one out of several classes. Several classical classification models can be used for D2D applications, including K-nearest neighbor (KNN), support vector machines (SVM), and decision tree (DT) [12]. Additionally, the recent advances in graphical processing units (GPUs) allows artificial deep NNs (DNNs) to be used for large-size datasets. Such DNN have different architectures including the multi-layer feed-forward NN (FNN), convolutional NN (CNN), recurrent NN (RNN), Hopfield Networks, and Boltzmann machine, which are implemented in many new areas in communication networks [12].
- **Unsupervised-Learning:** Unlike supervised-learning, unsupervised-learning models discover and explore hidden patterns and structures of the input data without having data labels [12]. Unsupervised-learning can be sub-categorized into three categories, namely the clustering, density estimation, and dimension reduction. In clustering, the ML algorithm divides and labels data samples into groups/clusters, where the samples in one cluster are similar to each other more than to those samples in different clusters. Representative types of such sub-categories are the K-means and the Relative COre MERge (RECOME) clustering algorithms [12]. On the other hand, the objective of density estimation algorithms is to estimate the distribution density of data samples in the feature space to reveal the high-density regions, which usually show some essential characteristics. The Gaussian mixture model (GMM) is one of the popular algorithms in this sub-category. Finally, dimension reduction techniques, such as principal component analysis (PCA), K-means, and GGMM, transform the data from a high-dimensional space into a low-dimensional space, which reserve the principal structures of the data. Such techniques are widely-utilized in many applications [12].
- **Reinforcement-Learning:** RL is a powerful tool for dealing with real-time control problems at which there are difficulties in using supervised and unsupervised learning techniques. The learning methodology of RL is based on trial-and-error, similar to humans. An RL's agent is rewarded or penalized for the action it took to maximize the long-term rewards. To select a proper action, a recursive environmental feedback is provided to the agent in each step, where the strategy of the agent is to take action is defined as a policy. The most widely-used RL techniques are the Q-learning [12]. On the other hand, MAB is another promising RL based general approach, which is getting more interest specially in communication applications. In its conventional settings, the MAB problem is expressed by a collection of arms or actions, and it takes the exploration-exploitation dilemma for a player. Each time step, the player/learner selects an arm and receives its corresponding reward, which can be modeled as stochastic or non-stochastic. The title bandit means that the player only knows the prize of the chosen arm, while the other arms rewards remain unknown at that specific time. The player wishes to maximize the cumulative reward gained from a sequential selection of the arms. In other words, the player intends to minimize regret compared with the best single arm. MAB is very helpful in sequential decision-making such as network routing [12],[13].

4. Applications of ML in D2D communications

In this section, we highlight some vital ML algorithms utilized for D2D in real-life scenarios emphasizing on their pros over traditional solutions. ML techniques can address variety of D2D

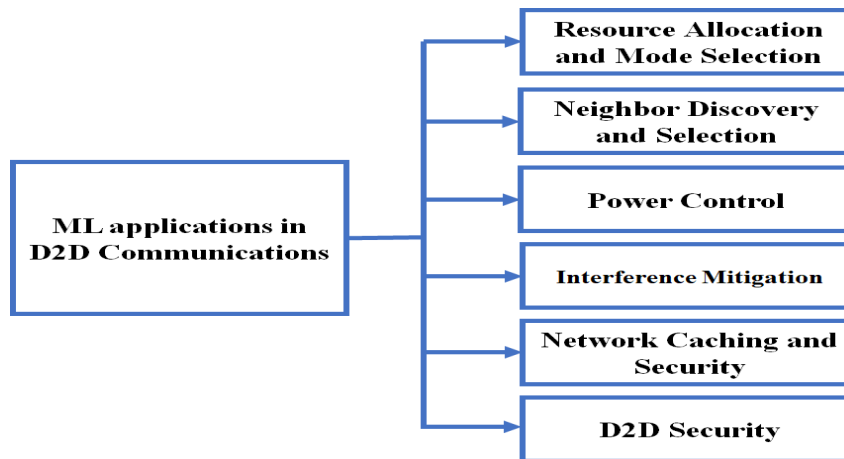


Figure 3. ML applications in D2D networks

communications' challenges as given in Fig. 3 including the traditional ones given in Fig. 2. Furthermore, Table 2 summarizes different ML applications in different D2D scenarios.

4.1. RA and Mode Selection (MS)

The RA and MS processes in D2D communications show high complexity in highly dense networks. Leveraging ML methods for both RA and MS can promote more powerful flexibility to cope with network dynamics. The authors of [14] proposed a DL based algorithm for transmission in D2D networks. They formulated a CNN that selects D2D linkages to transmit data, where 90% of the selected sub-optimal data are utilized to train the CNN based algorithm and the remaining 10% for validation. Their numerical results showed that they obtained 85-95% accuracies from the neural network. In [9], an autoencoder is trained using a supervised learning approach to pair underlay D2D transmitters to reuse the spectrum of CUs, where the optimal pairing in a dataset is obtained using the conventional Hungarian algorithm that minimizes the total cost of selection through combinatorial optimization formed as bipartite graph. The main idea is to use a Deep Learning (DL) approach that is capable of mapping the cost matrices of matching different D2Ds-CUs to the corresponding optimal solutions defined by the assignment matrix obtained from the Hungarian algorithm with less complexity and time. In [8], the authors modeled the joint radio resource management (RRM)-MS problem as two stage online learning combinatorial MAB (CMAB). Their Combinatorial bandit Learning for MS and RA (CBMOS) algorithm achieves fast learning speed (132%) and higher performance (142%) for high channel dynamics. In [7], a RL-based latency controlled D2D connectivity (RL-LCDC) for indoor D2D was proposed. RL-LCDC intelligently finds the neighbors, determines the D2D connection, and adaptively manages the communication area for greatest network connectivity. Distributed Q-learning algorithm can automatically allocate the spectrum to control interference in D2D enabled multi-tier HetNets. The performance of the proposed RL based schemes was optimal compared to other techniques in terms of throughput, SE, signal to interference ratio (SINR), and network coverage. Additionally, clustering algorithms can be exploited for improving RA in single and multi-cell D2D network to obtain better SE and system performance. The authors in [15] utilized k-means clustering algorithm for improving RA in single cell D2D network yielding SE and system performance improvements.

4.2. ML for NDS

Single and multi-hop neighbor probing is a severe problem in D2D communications that can be efficiently solved using ML. In [16], the authors proposed a DL based peer discovery technique that applies information about social network relationships to reject ill-disposed devices. Also the authors of [17] and [18], formulated the problem of mmWave D2D NDS as a stochastic MAB to gain maximum

long term throughput. In [4], [19], a multiplayer MAB algorithms were leveraged for surrounding gateway UAV selection by access UAVs in a disaster area scenario. This results in not only reducing the probability of encountering malicious devices but also enhancing the efficiency of peer discovery.

4.3. ML for Power Control

PC is an important interference management related topic that ML can handle in D2D. In [20], two PC algorithms based on supervised and unsupervised learning were proposed in D2D scenarios. The authors proved the importance of ML in D2D communication by comparing two ML algorithms with conventional PC methods in terms of computational complexity, throughput, and energy efficiency. In [21], the problem of D2D PC is addressed in the case of known channel gains between two D2D users and BSs, while the channel gains among the two users are unknown. A complete automatic power allocation method for internet of things (IoT)-D2D communication based on DL are proposed in [22]. They designed a distributed DL structure that trains the devices as a group but each device works independently and attained near optimal performance. The authors of [23] suggested a mean-field multiagent deep RL model that permits the devices to learn online PC strategies in a fully distributed manner, i.e., selfish strategy each node operates independently.

4.4. ML for Interference Mitigation

D2D transmission may be the source of severe interference to the other D2D links and the CUs. A survey of different popular and AI-based interference mitigation and RA approaches developed in D2D communications is provided in [24]. Additionally, the multi-agent actor critic (MAAC) is a newly proposed algorithm in [25] to mitigate interference by efficiently distributing the spectrum allocation. Moreover, the same paper proposes the neighbor-agent actor critic (NAAC) that uses neighbor users' historical information for centralized training leading to outage probability reduction and sum rate improvement for D2D links. Another RL based method for the RA problem was introduced in [26] based on the K-nearest neighbor algorithm that is utilized to choose the task offloading platform. Moreover, in [15], the concept of a limited D2D communication area is proposed based on ML.

4.5. ML for Network Caching

Former research on D2D caching strategies implies perfect knowledge of the content distribution. However, ML-based caching policy that makes use of the demand history is not only highly promising but also recommended to save time and complexity. A comprehensive survey articles for different ML and DL techniques for caching are provided in [27], [28], respectively. In [29], the D2D caching problem was formulated as multi-agent MAB to maximize the total predicted caching reward. Q-learning was leveraged to learn how to manage caching choices.

4.6. ML for D2D Security and commercial availability

There are still ongoing discussions on D2D commercial pricing and network security, where ML can handle such problems [1]. ML can help in addressing new D2D security challenges and threats related to device and user authentication to prohibit unauthorized access and attacks on the complete network. As an example of such usage, a radio frequency (RF) fingerprint based identification method of D2D device is proposed in [30]. Firstly, Hilbert transform (HT) and PCA are utilized to create the RF fingerprint of D2D device. Then, cross-validation (CV)-SVM is used as the classifier. Moreover the development of ML solution will make the issues of related commercial products close to appear in the market by large mobile companies like Huawei and Samsung with reasonable prices.

5. Challenges and Future Research Directions

Although most researchers leveraged ML techniques in wireless communication and D2D networks, it is crucial to identify and address different problems and challenges in practical D2D

Table 2. ML based D2D Communications

| Reference & D2D Use-case | Objective | ML Technique | Main Outcome | Shortcomings |
|-------------------------------|--------------------------------------------------------------------|--------------------------------|-----------------------------------------------------------------------|------------------------------------------------------------------------------------|
| [1] RA | Pairing D2D Transmitters to reuse CUs Spectrum | DL/Autocovector | near optimal accuracy with less complexity and time | Performance degradation at large number of paired devices |
| [8] RA | Efficiently solve RBM and MS problem | RL, CMAB | Proposed scheme achieve fast learning speed and higher performance. | Needs work on tuning of the learning parameters according to the network dynamics. |
| [7] Indoor D2D RA | Find neighbor, determine connection and manage communication area. | RL, Q-learning | Optimal connectivity with small delay. | Single-agent RL method |
| [14] RA | SI improvement and cross-interference cancellation. | DL (CNN) | Large accuracy and near identical scheduling results | Improper for practical spatial and mobile models. |
| [13] RA and FC | FD system performance improvement and interference mitigation | K-means clustering | Improved SE and system fairness | Single cell considered not multiple ones. |
| [20] RA | Next generation vehicular communications improvement | KNN | Latency cost and average system cost reduction with small delay | Proposed algorithm needs to be optimized. |
| [18] NDS | Efficient peer discovery | DI | Connection with trusted devices only. | Concentrates on social network information only. |
| [17][18] NDS | mmWave D2D neighbor discovery | MAB | prolong network life time with good mmWave linkage | Unsuitable for multiple devices |
| [20] FC | Interference Control | Q-learning, CART decision tree | Better FC performance than traditional techniques. | Non-cooperative Multi-agent |
| [21] FC | Interference mitigation and capacity maximization | Supervised ML | Increased capacity without information exchange between the two users | Improper for multi channels between D2D pairs |
| [22] FC | Complete automatic power allocation for IoT-D2D communication | DI | Distributed DI architecture with near optimal cell throughput | Neglects small scale fading and unsuitable for centralized scenario |
| [23] FC | Online FC for Large energy harvesting networks | Multiagent deep RL | High FC efficiency and near ideal performance | Unsuitable for distributed scenarios |
| [28] Network caching | Efficient content caching. | RL | Large average downloading latency and high caching rate. | Optimal transmission range and number of neighbors are not considered. |
| [19] Security | Device authentication and recognition | SVM, CV-SVM | 90% recognition rate, efficient D2D recognition technique. | No investigation on detecting different attacks. |
| [14][19] UAV in disaster area | Gateway UAV selection problem | MAB | Access UAV chooses proper gateway UAV with perfect mmWave linkage | Selfish policy cannot mitigate collisions |

networks. In this section we present the following exciting and challenging future research directions that worth further studies.

5.1. Fast Learning Process in Highly Dynamic D2D Networks

The implementation of ML-based algorithms requires fast speed learning, especially for fast-moving devices. Additionally, suitable models for high channel dynamicity should be learned and updated. However, it is not always achievable to model in practice, which forms a bottleneck in D2D communication networks, especially for high-speed trains (HST) and vehicles. In mmWave D2D, due to the inherent complexity of the BT algorithms, proposing an optimal/sub-optimal ML-based algorithm with fast inference time is a challenging task, especially if we take into consideration different D2D use cases in Fig.1.

5.2. Adaptive ML for Easy/Adversarial Environment

So far, stochastic MAB algorithms are applied in several wireless communications applications. These algorithms are designed for stochastic stationary environments, which are not suitable for adversarial /dynamic environments like D2D communications. On the other hand, some of ML techniques (e.g., online learning algorithms) have theoretical performance guarantees in the worst-case scenarios. These techniques, however, do not perform better in practice since the environments are not always so adversarial and they do not fully take advantage of such easiness in the settings. Easy data approach in ML like in [31], attempt to develop algorithms that perform adaptively in both the best and worst cases simultaneously. This approach would be helpful in solving different D2D problems, where the environments sometimes stationary and some other times are dynamic.

5.3. Full Duplex D2D

Although FD communications double the network capacity, FD D2D communications result in severe interference and complex resource management problems. Further FD-D2D research is required for real future implementation.

5.4. Future ML Algorithms

Some problems in D2D applications are inherently combinatorial. For example, in multi-hop communication tasks, the transmitter chooses a path (a sequence of intermediate devices) to the receiver to send the message. Therefore, the underlying problem can be formulated as an optimization problem over paths in the graph formed by intermediate devices. Appropriate problem formulations and combinatorial/adaptive ML algorithms could overcome problems such as clustering, NDS, and multi-hop relay probing. Also, appropriate ML algorithms for the centralized and decentralized peer to peer network setting have to be addressed. Moreover the symbiosis relation between ML and communication communities have to help each other to solve future D2D problems via distributed learning.

5.5. MmWave Environment

Recent research directions that use ML algorithms for tackling small and separated problems can not efficiently address the highly dynamic and ultra-sense features in B5G and 6G networks, especially

in mmWave based environment. The existing ML algorithms are applied without considering the adversarial mmWave environment, such as path blocking and spatial transmissions coming from BF. Moreover, DL-based solutions utilize continuous optimization that increases the system overhead. Also, these solutions require an incredible offline-learning phase, making it unsuitable for future mmWave B5G/6G applications. However, online/ discrete adaptive optimization techniques will be more suitable to cope with the mmWave nature, especially for multi-hop transmission scenarios.

5.6. Distributed learning information

Distributed AI can control the future D2D generations. The mutual information between ML and communications communities are essential and strongly required for promising unique solutions. The D2D community designers should provide ML designers with sufficient information about the devices/ locations/ speeds/ environments so as to invent suitable algorithms that succeed to perform distributive learning.

5.7. Energy Harvesting and Cognitive Radio

Saving energy in D2D networks is an essential requirement for prolonging network lifetime. ML-based energy harvesting and stochastic optimization schemes are urgently required to mitigate the harvested energy outage. The concept of CR can be intelligently implemented with the aid of ML.

5.8. Peer to Peer Internet

Employing ML with future D2D networks might help in realizing newly decentralized peer to peer internet. Decentralized multiplayer MAB techniques can help on solving such problem. Also federated and distributive learning techniques will be effective solutions. Future 6G systems will be definitely depend on real time/responsible AI.

5.9. D2D Networks for Decentralized Federated Learning

FL is a type of decentralized ML-based technique used to train networks by exploiting local models training and client-server communication [32]. This type of decentralized model is suitable for networks where the training data is distributed over a large number of devices with a fraction of the data. At the same time, those devices exchange their locally-trained models instead of exchanging their private data. Specifically, FL enables a joint ML training over distributed data sets with limited disclosure of local data. In [32], the authors provided an implementation of decentralized stochastic gradient descent (DSGD) technique for large-scale wireless D2D networks. However, exploiting such FL techniques for joint scheduling and resource allocation in D2D networks is a challenging task, especially under channel uncertainty and connection availability in each iteration.

6. Case Study: MAB based mmWave D2D Scenario

This section demonstrates the effectiveness of ML-based methods over conventional solutions of mmWave D2D NDS problem. Fig. 4 illustrates the simulated mmWave D2D network, where a mmWave device is located at the center of a micro-BS area of $125 \times 125m^2$, and it desires to establish a D2D link with one of its neighbor devices. Conventionally, the center device should exhaustively search over all its surrounding devices using BT and select the best one maximizing the achievable data rate of the D2D linkage. This will highly decrease the link throughput because of the incredible training overhead. Instead, the mmWave NDS problem is modeled as a stochastic MAB, where the center device will act as the player aiming to maximize its long term reward, which is the achievable data rate. This is done via playing over the surrounding devices, serving as the arms of the bandit. Through proactive online learning, the center device will reach up at the device, maximizing its achievable data rate while examining one nearby device at a time, which highly reduces the NDS process's training overhead and increases the throughput consequently. UCB and MOSS [33] algorithms are utilized to

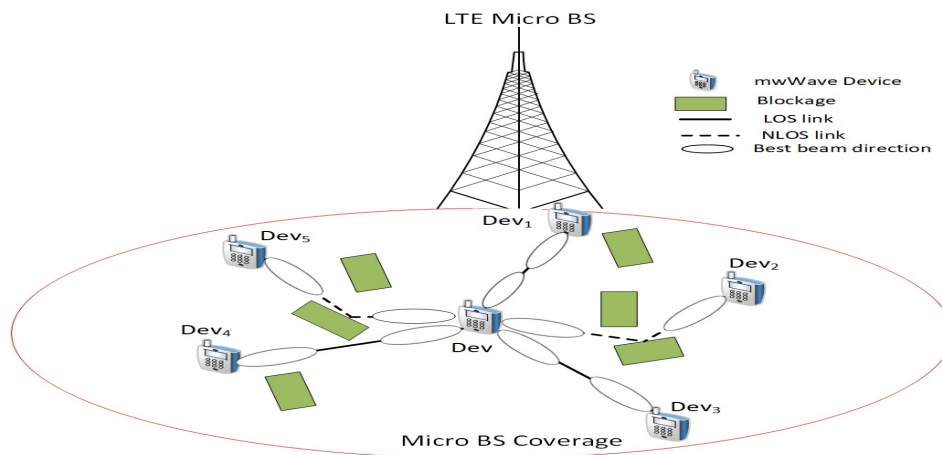


Figure 4. Utilized mmWave D2D System Model

Table 3. Case Study Simulation Parameters

| Parameter | Value |
|---------------------------------------------------------|------------------------------|
| Bandwidth, Data transmission time | 2.16 GHz, 20msec |
| Beamform training time, and Data length | 0.28 msec, 1 Gbit |
| Transmitted power, Half power beam width for AOA | 10 dBm and 20 |
| LOS and NLOS pathloss exponents | 2.22 and 3.88 |
| LOS and NLOS shadowing's standard deviation | 10.3, 14.6 |
| Blocking object thinning factor, Blocking object radius | 1 and uniform [0.3 – 0.6] m |
| Noise | $-174 + 10\log_{10}(B) + 10$ |

prove the effectiveness of the MAB based mmWave NDS over the conventional NDS, and random selection [13]. UCB attempts to improve the action selection's confidence every round by reducing the uncertainty, while MOSS is appropriate for both stochastic and adversarial MAB settings. Hence, both are suitable for the mmWave D2D NDS problem. In random selection, a random nearby device is selected every round for establishing the D2D link. Although it highly relaxes the NDS overhead, it results in a poor achievable data rate and low throughput. The UCB and MOSS algorithms are modified to select the best nearby device with maximum long-term data rate for constructing the D2D link. The mmWave channel model plus blocking formulations are given in details in [17]. Table 3 summarizes the simulation parameters values utilized in this case study.

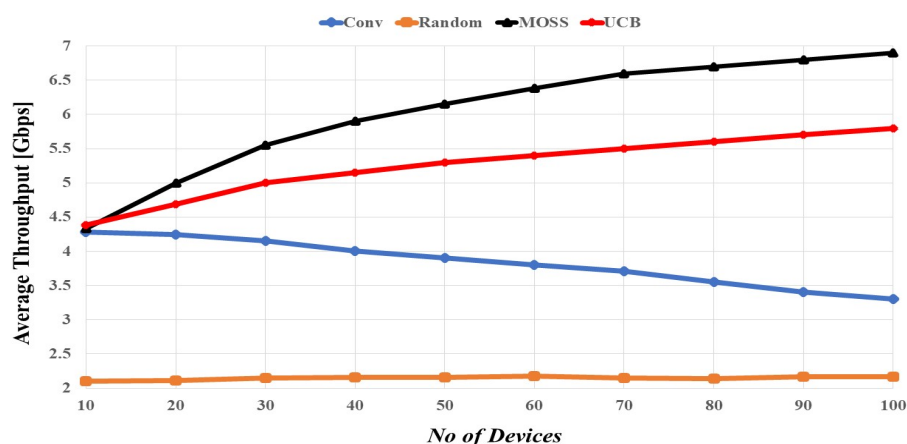


Figure 5. Average Throughput comparisons at no blockage

Figure 5 shows the average throughput comparisons using a separate number of distributed devices at no blocking and all the paths are line of sight (LOS). The proposed MAB based schemes

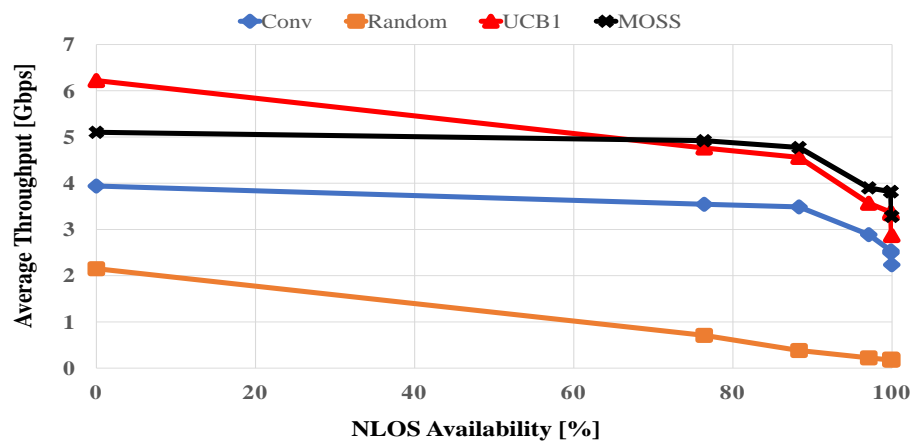


Figure 6. Average Throughput vs NLOS availability for 60 Devices

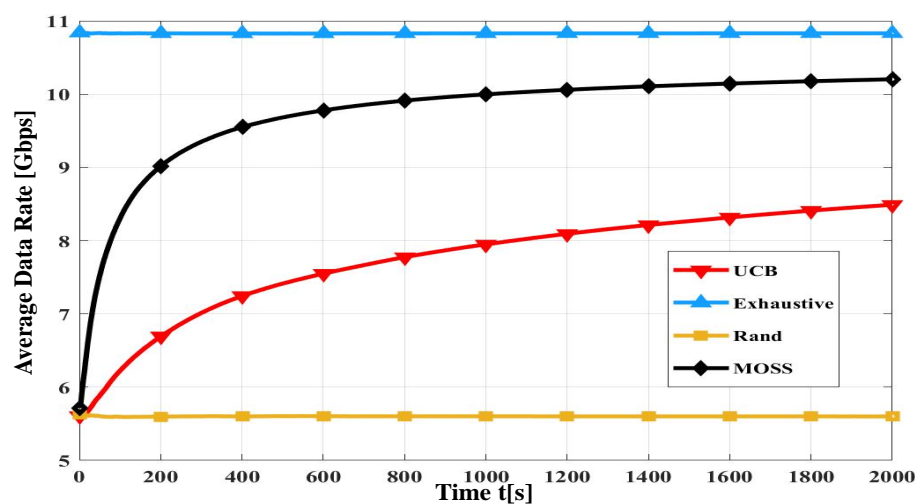


Figure 7. UCB and MOSS convergence vs Horizon

show superior performance over either conventional (Conv) or random NDS methods. The average throughput is increased as we increase the number of devices due to the pros of MAB based algorithms that reduce the overhead, unlike traditional solutions. The conventional NDS method reduces the average throughput as the number of devices increases. Figure 6 presents the average throughput against the percentage of NLOS availability for the compared algorithms. It is worth noting that MAB based solutions have superior performance even at high LOS blockage. The figure confirms the ML-based algorithms' advantage for solving the NDS problem in mmWave D2D considering harsh LOS blockage environment. One of the main challenges for MAB solutions is the convergence of the algorithm. In Fig. 7, we study the convergence rate of the utilized MAB algorithms against the horizon where the optimal solution is added as an upper limit. It is clearly shown that the proposed MAB algorithms achieve a high convergence rate towards the optimal data rate obtained through exhaustively searching all available nearby devices. At $t = 400$, MOSS and UCB converge to 86%, 70% of the ideal average throughput, respectively.

7. Conclusion

This paper has presented a general overview of the applicability of ML algorithms in the area of D2D networks. The above investigation has identified difficulties and challenges to be addressed by the community to establish practical ML-based solutions that support D2D in B5G and 6G systems. The scope of future research when ML meets D2D is broad. Hence, we introduced a few exciting and challenging research issues that worth additional investigations. Furthermore, we give a case study to

emphasize the effectiveness of the MAB based techniques to solve the NDS problem in mmWave D2D communications over conventional solutions.

Author Contributions: All authors contributed equally in this paper.

Funding: This research was funded by RIKRN-AIP Computational Learning Theory Team, Japan <https://www.riken.jp/en/research/labs/aip/>

Conflicts of Interest: The authors declare no conflict of interest.

8. References

1. Ansari, R.I.; Chrysostomou, C.; Hassan, S.A.; Guizani, M.; Mumtaz, S.; Rodriguez, J.; Rodrigues, J.J.P.C. 5G D2D Networks: Techniques, Challenges, and Future Prospects. *IEEE Systems Journal* **2018**, *12*, 3970–3984. doi:10.1109/JSYST.2017.2773633.
2. Venugopal, K.; Valenti, M.C.; Heath, R.W. Device-to-Device Millimeter Wave Communications: Interference, Coverage, Rate, and Finite Topologies. *IEEE Transactions on Wireless Communications* **2016**, *15*, 6175–6188. doi:10.1109/TWC.2016.2580510.
3. Xiao, M.; Mumtaz, S.; Huang, Y.; Dai, L.; Li, Y.; Matthaiou, M.; Karagiannidis, G.K.; Björnson, E.; Yang, K.; I, C.; Ghosh, A. Millimeter Wave Communications for Future Mobile Networks. *IEEE Journal on Selected Areas in Communications* **2017**, *35*, 1909–1935.
4. E. Mahmoud Mohamed, S. Hashima, A.K.H.; Abdelghany, M. Gateway Selection in Millimeter Wave UAV Wireless Networks Using Multi-Player Multi-Armed Bandit. *MPDI Sensors (Basel)* **2020**, *20*, 3938–3947.
5. Lien, S.; Deng, D.; Lin, C.; Tsai, H.; Chen, T.; Guo, C.; Cheng, S. 3GPP NR Sidelink Transmissions Toward 5G V2X. *IEEE Access* **2020**, *8*, 35368–35382.
6. Mohamed, E.M.; Abdelghany, M.A.; Zareei, M. An Efficient Paradigm for Multiband WiGig D2D Networks. *IEEE Access* **2019**, *7*, 70032–70045. doi:10.1109/ACCESS.2019.2918583.
7. Sreedevi, A.; Rama Rao, T. Reinforcement learning algorithm for 5G indoor Device-to-Device communications. *Transactions on Emerging Telecommunications Technologies* **2019**, *30*, e3670. doi:10.1002/ett.3670.
8. Ortiz, A.; Asadi, A.; Engelhardt, M.; Klein, A.; Hollick, M. CBMoS: Combinatorial Bandit Learning for Mode Selection and Resource Allocation in D2D Systems. *IEEE Journal on Selected Areas in Communications* **2019**, *37*, 2225–2238. doi:10.1109/JSAC.2019.2933764.
9. Zaky, A.B.; Huang, J.Z.; Wu, K.; ElHalawany, B.M. Generative neural network based spectrum sharing using linear sum assignment problems. *China Communications* **2020**, *17*, 14–29.
10. Çatak, E.; Durak-Ata, L. An efficient transceiver design for superimposed waveforms with orthogonal polynomials. 2017 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 2017, pp. 1–5. doi:10.1109/BlackSeaCom.2017.8277657.
11. Çatak, E.; Tekçe, F.; Dizdar, O.; Durak-Ata, L. Multi-user shared access in massive machine-type communication systems via superimposed waveforms. *Physical Communication* **2019**, *37*, 100896. doi:https://doi.org/10.1016/j.phycom.2019.100896.
12. He, R.; Ding, Z. *Applications of Machine Learning in Wireless Communications*; IET, London, UK., 2019.
13. Burtini, G.; Loeppky, J.L.; Lawrence, R. A Survey of Online Experiment Design with the Stochastic Multi-Armed Bandit. *ArXiv* **2015**, *abs/1510.00757*.
14. Ban, T.W.; Lee, W. A Deep Learning Based Transmission Algorithm for Mobile Device-to-Device Networks. *Electronics* **2019**, *8*. doi:10.3390/electronics8111361.
15. Huang, X.; Zeng, M.; Fan, J.; Fan, X.; Tang, X. A Full Duplex D2D Clustering Resource Allocation Scheme Based on a K-Means Algorithm. *Wireless Communications and Mobile Computing* **2018**, *2018*, 1–8. doi:10.1155/2018/1843083.
16. Long, Y.; Yamamoto, R.; Yamazaki, T.; Tanaka, Y. A Deep Learning Based Social-aware D2D Peer Discovery Mechanism. 2019 21st International Conference on Advanced Communication Technology (ICACT), 2019, pp. 91–97. doi:10.23919/ICACT.2019.8701911.

17. Hashima, S.; Hatano, K.; Takimoto, E.; Mahmoud Mohamed, E. Neighbor Discovery and Selection in Millimeter Wave D2D Networks Using Stochastic MAB. *IEEE Communications Letters* **2020**, *24*, 1840–1844.
18. Hashima, S.; Mohamed, E.; Hatano, K. Minimax Optimal Stochastic Strategy (MOSS) for neighbor discovery and selection in Millimeter Wave D2D Networks. The 23 rd International Symposium on Wireless Personal Multimedia Communications", 2020.
19. Hashima, S.; Hatano, K.; Mohammed, E. Multiagent Multi-Armed Bandit Schemes for Gateway Selection in UAV Networks. 2020 IEEE Globecom Workshops (GC Wkshps), 2020, pp. 1–6.
20. Fan, Z.; Gu, X.; Nie, S.; Chen, M. D2D power control based on supervised and unsupervised learning. 2017 3rd IEEE International Conference on Computer and Communications (ICCC), 2017, pp. 558–563. doi:10.1109/CompComm.2017.8322607.
21. Najla, M.; Gesbert, D.; Becvar, Z.; Mach, P. Machine Learning for Power Control in D2D Communication Based on Cellular Channel Gains. 2019 IEEE Globecom Workshops (GC Wkshps), 2019, pp. 1–6. doi:10.1109/GCWkshps45667.2019.9024549.
22. Kim, J.; Park, J.; Noh, J.; Cho, S. Autonomous Power Allocation Based on Distributed Deep Learning for Device-to-Device Communication Underlying Cellular Network. *IEEE Access* **2020**, *8*, 107853–107864. doi:10.1109/ACCESS.2020.3000350.
23. Sharma, M.K.; Zappone, A.; Debbah, M.; Assaad, M. Multi-Agent Deep Reinforcement Learning based Power Control for Large Energy Harvesting Networks. 2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT), 2019, pp. 1–7. doi:10.23919/WiOPT47501.2019.9144098.
24. Zia, K.; Javed, N.; Sial, M.N.; Ahmed, S.; Iram, H.; Pirzada, A.A. A Survey of Conventional and Artificial Intelligence / Learning based Resource Allocation and Interference Mitigation Schemes in D2D Enabled Networks. *ArXiv* **2018**, *abs/1809.08748*.
25. Li, Z.; Guo, C. Multi-Agent Deep Reinforcement Learning based Spectrum Allocation for D2D Underlay Communications, 2019.
26. Cui, Y.; Liang, Y.; Wang, R. Resource Allocation Algorithm With Multi-Platform Intelligent Offloading in D2D-Enabled Vehicular Networks. *IEEE Access* **2019**, *7*, 21246–21253. doi:10.1109/ACCESS.2018.2882000.
27. Shuja, J.; Bilal, K.; Alanazi, E.A.; Alasmay, W.; Alashaikh, A. Applying Machine Learning Techniques for Caching in Edge Networks: A Comprehensive Survey. *ArXiv* **2020**, *abs/2006.16864*.
28. Wang, Y.; Friderikos, V. A Survey of Deep Learning for Data Caching in Edge Network. *ArXiv* **2020**, *abs/2008.07235*.
29. Jiang, W.; Feng, G.; Qin, S.; Yum, T.S.P.; Cao, G. Multi-Agent Reinforcement Learning for Efficient Content Caching in Mobile D2D Networks. *IEEE Transactions on Wireless Communications* **2019**, *18*, 1610–1622. doi:10.1109/TWC.2019.2894403.
30. Zhang, Z.; Guo, X.; Lin, Y. Trust Management Method of D2D Communication Based on RF Fingerprint Identification. *IEEE Access* **2018**, *6*, 66082–66087. doi:10.1109/ACCESS.2018.2878595.
31. Koolen, W.M.; Erven, T.V. Second-order Quantile Methods for Experts and Combinatorial Games. Proceedings of The 28th Conference on Learning Theory; Grünwald, P.; Hazan, E.; Kale, S., Eds. PMLR, 2015, Vol. 40, pp. 1155–1175.
32. Xing, H.; Simeone, O.; Bi, S. Decentralized Federated Learning via SGD over Wireless D2D Networks, 2020, [arXiv:cs.IT/2002.12507].
33. Audibert, J.Y.; Bubeck, S. Minimax Policies for Adversarial and Stochastic Bandits. Proceedings of the 22nd Annual Conference on Learning Theory (COLT), Proceedings of the 22nd Annual Conference on Learning Theory (COLT) ed., 2009.