Intelligent and Converged Networks ISSN 2708-6240
2021, 2(4): 295−305 DOI: 10.23919/ICN.2021.0021

QSPCA: A two-stage efficient power control approach in D2D communication for 5G networks

Saurabh Chandra, Prateek, Rohit Sharma, Rajeev Arya*, and Korhan Cengiz

Abstract: The existing literature on device-to-device (D2D) architecture suffers from a dearth of analysis under imperfect channel conditions. There is a need for rigorous analyses on the policy improvement and evaluation of network performance. Accordingly, a two-stage transmit power control approach (named QSPCA) is proposed: First, a reinforcement Q-learning based power control technique and; second, a supervised learning based support vector machine (SVM) model. This model replaces the unified communication model of the conventional D2D setup with a distributed one, thereby requiring lower resources, such as D2D throughput, transmit power, and signal-to-interference-plus-noise ratio as compared to existing algorithms. Results confirm that the QSPCA technique is better than existing models by at least 15.31% and 19.5% in terms of throughput as compared to SVM and Q-learning techniques, respectively. The customizability of the QSPCA technique opens up multiple avenues and industrial communication technologies in 5G networks, such as factory automation.

Key words: device-to-device (D2D); interference; Internet of Things (IoT); machine learning; power control; Q-learning; support vector machine (SVM); 5G

1 Introduction

With the enhancement of the usage of smart devices, their performance demands improvement in data rate and power consumption and lesser interference. Increased demand on the performance and data rate by users adversely impacts the 5th-generation (5G) network. Various applications, such as sharing of documents among users, live data, video calls, live conferences, live streaming ultra-high-quality 4K videos, and multimedia content sharing[1, 2], are some of the use-case scenarios in practical applications.

To meet users' expectations, the performance enhancement of wireless network systems is urgently needed. Due to the wide variety of real-time applications, the Internet of Things (IoT) has determined several applications in 5G networks[3]. Quality of service (QoS) is affected by the congestion of bandwidths as the number of users increases. Hence, device-to-device (D2D) communication is taken into consideration because it provides a viable solution for various issues and demands. Devices can directly communicate over D2D channels. Without any involvement of network infrastructure (NI), D2D allows users to share information directly. D2D means communication between devices, such as cell phones and vehicles. Network capacity issues can be solved with the help of 5G; wherein multiple devices can rapidly share data or files in an authentic way. 5G supports numerous applications and fulfills the high data rate demanded by users. The ultimate purpose of D2D communication is the performance improvement of 5G networks in terms of energy efficiency, throughput, latency, and spectrum utilization[4]. Furthermore, the primary purpose of 5G networks is to

• Saurabh Chandra, Prateek, Rohit Sharma, and Rajeev Arya are with the Wireless Sensor Network Lab, the Department of Electronics and Communication Engineering, National Institute of Technology Patna, Ashok Rajpath, Patna 800005, India. E-mail: {saurabh.ec.jrf21, prateek.ec18, rohits.ph21.ec, rajeev.arya}@nitp.ac.in.
• Korhan Cengiz is with the Department of Electrical-Electronics Engineering, Trakya University, Edirne 22030, Turkey. E-mail: korhancengiz@trakya.edu.tr.
* To whom correspondence should be addressed.
Manuscript received: 2021-06-08; revised: 2021-09-08; accepted: 2021-09-29

give access to ubiquitous connectivity in every scenario. Distributed denial of service (DDoS)-type flooding attacks mainly target resources, such as bandwidth, processing power, memory, and accessibility to users. Adaptive agent-based models are used to protect the network application layer against DDoS flooding attacks.

In D2D communication, users can directly communicate with their respective neighboring users without involving the base station (BS). Benefits include low latency, spectrum efficiency, high throughput, and energy efficiency. However, the reuse of resources causes interference among D2D user pairs and interference between the cellular user and D2D transmitter user. Therefore, detracting the interference is the biggest challenge in underlaid D2D communication. The power control technique is one of the most suitable methods to reduce interference between users. Works (e.g., Ref. [5]) have paid attention to mitigating interference in this regard. Due to the random location of D2D users, cellular users inside a cell, where the BS is centrally located, cause serious interference to the D2D communication because D2D users reuse the resources of cellular users. Interference reduction between users is the most important issue in D2D communication. To solve related issues, power control techniques play a significant role in guaranteeing the QoS and maximizing the throughput while maintaining a minimum signal-to-interference-plus-noise ratio (SINR). Two main types of power control approaches have been taken into account. The first type includes the traditional game theory, graph theory, and stochastic process, and the second one is machine learning (ML), particularly reinforcement learning (i.e., Q-learning) and supervised learning (i.e., deep neural network). The application of ML on resource management has attracted wide attention from researchers. Reinforcement learning is an emerging field used to solve the problems of resource and power allocations in 5G networks[6]. Techniques, such as multi-agent-based distributed spectrum sharing[7], to enhance data rate and spectrum capacity and to optimize resource allocation (RA) are active areas of pursuit[8].

1.1 Related work

ML-based techniques, such as reinforcement learning, deep learning, supervised learning, and unsupervised learning, have been actively implemented independently to enhance D2D performance. The authors in Ref. [9] proposed a support vector machine (SVM) based approximation, which is a supervised learning-based scheme used to solve the transmission power problem. However, Q-learning and multicast D2D communication remain to be explored in their work to mitigate interference by D2D transmission power control. In Ref. [10], although the authors segregated users with respect to the overall throughput, energy consumption, and fairness, independent learning approaches remain to be analyzed. Furthermore, the implications of the value function, which converges with the throughput function, were not discussed by the authors. In Ref. [11], reinforcement learning based QoS maintainability was proposed using team Q-learning and distributed Q-learning to introduce transmission power control schemes in D2D underlaid in cellular networks. It was also used to enhance the sum throughput and secure QoS[12]. However, it lacks training efficiency, which can be improved using the attention mechanism, followed by the proof of convergence of the solution. In Ref. [13], a trilateration-based cooperative network was proposed to reduce the average latency and enhance the success rate with offload traffic. However, the proposed mechanism is not capable of making a decision and selecting the appropriate mode for resource sharing. In Ref. [2], the authors proposed Q-learning-based schemes to select the relay, distribute power, improve energy efficiency, and mitigate traffic congestion. Multicasting has been considered, but the authors did not analyze the quality of experience (QoE) parameter for enhancing and optimizing the sum rate. Online reinforcement learning based hybrid transmission mode selection[14] was proposed to solve the trade-off factor problem and enhance the utilization of resources. However, the proposal[14] lacks the capacity to handle the complexity of the network and cannot offload traffic data. To address the binary classification problem, SVM-based classifications were proposed by the authors in Ref. [15]. However, they did not consider reinforcement learning based approaches to reduce latency and transmission power control. This limitation may be solved using the Q-learning approach[16]. In Ref. [17], the authors proposed a deep reinforcement learning (DRL) based approach to provide the optimum solution to identified problems, such as shortage of memory, and to store the content and enhance the QoE in the cache system. However, this scheme is not effective for the utilization of resources and energy-efficient D2D underlaid communication. To study and address the problems of traffic delay and lack of energy consumption, the authors in Ref. [18] proposed a DRL-based non-cooperative and real-time approach. However, this proposal may not be effective for optimizing the storage capacity of content by increasing the memory size, and the maintainability of QoS also needs improvement for enhancing the throughput. In most of the previous works, the focus was on several reinforcement learning based approaches to control the transmission power in D2D underlaid communication. Moreover, their primary objective was to achieve the maximum sum rate and maintain the QoS of the overall network.

1.2 Motivation

Surveys[9−11, 13−15, 17, 18] show that RL-based power control schemes are generally used to enhance user performance. The existing literature suffers from the lack of channel information between users due to the inadequate knowledge on the location of users. Hence, there is a need for rigorous analyses on users who are having the worst channel conditions. The main research gaps identified are as follows:

(1) Lack of analysis under imperfect channel conditions

The existing approaches have focused only on users with good channel conditions to allocate resource blocks (RBs), but users with the worst channel conditions and below-the-average data rates were not considered.

(2) Lack of policy evaluation and improvement

Two-stage implementations consisting of policy evaluation and improvement have not been attempted in the domain of D2D communication.

To overcome these gaps, we propose a new and efficient power control technique using ML in the D2D domain communication, where RBs must be equally distributed to users irrespective of the channel conditions to maximize the throughput and ensure QoS. The main objective of the proposed power control approach is to control the transmission power and assign resources using the BS to cellular users based on the knowledge of user location. ML techniques have been employed to control the power by slackening the interference to get the maximal throughput and strengthen the performance of the network.

1.3 Contribution

In this study, we investigated several ML approaches to controlling the transmission power and mitigating interference gain between D2D and cellular users. To achieve a better performance of the system and overall improvement of the throughput by solving the spectrum sharing and power control problem in the most efficient way, we propose an efficient transmit power control approach for D2D users through a two-stage distributed model based QSPCA approach using Q-learning and SVM techniques. The major contributions of this work are summarized as follows:

- A two-stage efficient power control approach is proposed based on ML. Then, we derive a general expression for the calculation of the channel gain, SINR, and throughput for cellular and D2D users.
- To maximize the throughput for the improvement of the network performance, the RA and power control problem are formulated under maximum power constraints and QoS with minimum SINR. The formulated problem is then solved with the help of an efficient power control based approach (named QSPCA), which is performed in two stages.
- In the first stage, the power level is selected on the basis of getting rewards to reduce the impact of co-channel interference between users. A respective reward is set for each D2D user to achieve the target.
- In the second phase of QSPCA, the generated dataset is then used for the classifications in terms of the optimized transmission power and channel gain for

maximizing the throughput while ensuring QoS to improve network performance and maintain minimum SINR requirements.

The outline of this article is categorized as follows: The network model and formulated problem are discussed in Section 2. The proposed power control approach to providing the solution to the problem is elaborated in Section 3. The performance and simulation results are discussed in Section 4. Finally, the article is summarized in Section 5. The notations and symbols used it this paper are listed in Table 1.

2 Network model and problem formulation

In this study, we considered D2D underlaid communication in a single-cell scenario with a single BS located at the center of a cell and where users are distributed uniformly inside a cell. We also considered multiple D2D and cellular users existing within a cell, but the location of the D2D users is at the edge of the cell boundary. If the D2D receiver is in close proximity to the D2D transmitter, then they only established a direct link for data sharing and direct communication by satisfying the threshold or minimum distance between them.

Table 1 List of notations and symbols used in the paper.

Symbol	Definition
P_n^t	Transmit power of D2D user
P_m^t	Transmit power of cellular user
$S_{m,b}^t$	Transmitted message signal of the m-th cellular user to base station
$S_{n,b}^t$	Transmitted message signal of the n-th D2D user to base station
$h_{m,b}^t$	Channel gain from the m-th cellular user to evolved Node B
$h_{n,b}^t$	Channel gain from the n-th D2D user to evolved Node B
δ_t^m	SINR of cellular user
δ_t^n	SINR of D2D user
$\delta_{t\min}^n$	Minimum SINR of D2D user or threshold SINR
$\delta_{t\min}^m$	Minimum SINR of cellular user
$P_{\mathrm{c}\max}$	Maximum transmission power of cellular user
$P_{\mathrm{d}\max}$	Maximum transmission power of D2D user
$\gamma_{m,n}^t$	Resource block indicator
β	Bandwidth
φ	Noise power
C_{gi}	Channel gain
ζ	Path-loss constant
χ_i	Fast fading gain
μ_i	Slow fading gain
K	Distance between user
α	Path-loss exponent
R_{c}	Throughput of cellular user
R_{d}	Throughput of D2D user
$R_{k,m}$	Learning parameter
μ_D	Lagrange duality function

We denoted BS as the evolved Node B (eNB). Let M be the set of cellular users, i.e., $M = \{1, 2, 3, \ldots, m, \ldots, M\}$, and N be the set of D2D users, i.e., $N = \{1, 2, 3, \ldots, n, \ldots, N\}$. Orthogonal frequency-division multiple access techniques have been used to communicate with cellular users during uplink transmission. Let β be the bandwidth given to RBs. Here, similar to the log-distance path loss model, we assume that the channel suffers from large- and small-scale fading due to multipath propagation.

Here, cellular users communicate with (or transmit signals to) BS. Thus, the received signal at the D2D receiver and eNB over the T-th RB is given by Eq. (1):

$$y_m^{t,c} = \sqrt{P_m^t} h_m^t S_m^t + \sum_{m\in M,\, n\in N}^{t\in T} \gamma_m^t \sqrt{P_n^t} h_{n,b}^t S_{n,b}^t + \varphi \tag{1}$$

where P_m^t is the cellular user transmitter power, P_n^t is the D2D transmitter power over the T-th RB, S_m^t and $S_{n,b}^t$ are the transmitted message signal of the m cellular user and n D2D transmitter to the BS, φ is the signal noise power at the BS over the T-th RB, and $h_{m,b}^t$ and $h_{n,b}^t$ are the channel gain from the m-th cellular user to the eNB and n D2D transmitter to the eNB, respectively. $\gamma_{m,n}^t$ represents the RB allocation indicator, and it is given by Eq. (2):

$$\gamma_{m,n}^t = \begin{cases} 1, & \text{if the number of D2D user gets the } T\text{-th RB;} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Due to the multipath fading and shadowing, we considered fast and slow fading. The overall gain of the channel link can be represented as in Eq. (3):

$$C_{gi} = \zeta \chi_i \mu_i K_i^{-\alpha} \tag{3}$$

where ζ is the path-loss constant, χ_i is the fast fading gain with an exponential distribution with the unit mean, μ_i is the slow fading gain, α is the path-loss exponent, and K is the distance.

In this model, each channel gain coefficient suffers from path loss and fast and slow fading with shadowing.

Based on Eq. (1), the SINR of the cellular user is calculated by the following expression:

$$\delta_t^m = \frac{P_t^m \cdot Cg_{m,t}^0}{\sum\limits_{t=1, j\in\mathbb{R}_s}^{T,n\in N} P_t^{nj} \cdot Cg_{j,t}^{j0} + \varphi} \tag{4}$$

where P_t^{nj} and P_t^m represent the D2D transmitter power and cellular user, respectively, and φ represents the noise power.

Similarly, the SINR of a D2D user on sharing the T-th RB can be expressed as

$$\delta_t^n = \frac{P_t^{ni} \cdot Cg_{ni,t}^{ii}}{\sum\limits_{t=1, j\in\mathbb{R}_s}^{T,n\in N} \alpha_{n,m}^t P_t^{nj} \cdot Cg_{nj,t}^{ji} + P_t^m \cdot Cg_t^{mi} + \varphi} \tag{5}$$

where $Cg_{ni,t}^{ii}$, $Cg_{nj,t}^{ji}$, and Cg_t^{mi} represent the channel link gain of the i-th D2D transmitter, channel link gain of the i-th transmitter of one D2D pair to the j-th receiver of other D2D user pairs, and channel gain of the m-th cellular user to the i-th receiver of the D2D user pair, respectively. α is the binary variable used to indicate resources. When $\alpha = 1$, the i-th D2D transmitter shares the resources of the m-th cellular users; otherwise, $\alpha = 0$.

According to the Shannon capacity theorem, the achievable throughput for the m-th cellular user using Eq. (4) is given as

$$R_{\mathrm{c}} = \beta \log_2(1 + \delta_t^m) \tag{6}$$

Similarly, the achievable throughput for the n-th D2D user using Eq. (5) is given as

$$R_{\mathrm{d}} = \beta \log_2(1 + \delta_t^{ni}) \tag{7}$$

where β is the bandwidth.

The formulated problem that maximizes the overall D2D throughput is modeled in Formula (8):

$$\begin{gathered} \max\Big[\sum_{t=1}^{T}\sum_{i\in\mathbb{R}_t}\gamma_{n,m}^t \cdot (\beta\log_2(1+\delta_t^m) + \beta\log_2(1+\delta_t^{ni}))\Big], \\ \text{s.t.}, \sum_{t=1}^{T}\gamma_{n,m}^t \leqslant 1 \text{ and } \gamma \in \{0,\ 1\}, \\ \delta_t^n \geqslant \delta_{t\min}^n \text{ and } \delta_t^m \geqslant \delta_{t\min}^m, \\ P_{\mathrm{c}} \leqslant P_{\mathrm{c}\max} \text{ and } 0 \leqslant P_{\mathrm{d}} \leqslant P_{\mathrm{d}\max} \end{gathered} \tag{8}$$

where $\delta_{t\min}^n$ and $\delta_{t\min}^m$ are the threshold SINR of the cellular user and D2D user pair, respectively; $P_{\mathrm{c}\max}$ and $P_{\mathrm{d}\max}$ are the maximum transmission power constraints of the cellular and D2D user pairs, respectively.

Based on the network model presented above, the problems are identified and formulated. The objective function is defined in terms of the throughput to be maximized, which is dependent on QoS constraints and maximum power constraints. D2D users reuse the spectrum of cellular users, which causes co-channel interference and even the worst channel condition. To detract from the interference among users, the scenario needs an algorithm that can handle the identified problems and can be solved efficiently. To address these issues, an efficient power control based approach is proposed, which will be discussed in the next section.

3 Proposed power control algorithm based on ML techniques

In this section, an efficient power control approach (QSPCA) is proposed in two stages. In the first stage, the dataset is constructed offline while considering the RB for multiple D2D and cellular users. For generating the datasets, the feature vectors are as follows: Cellular user, D2D user, RB, location of D2D, and cellular user. Training datasets are constructed offline, so the location of the users based on the knowledge of the channel state information is available with the resources assigned to the cellular use by the BS and D2D users reusing the spectrum. In the second phase of the implementation, the generated dataset is used for training and testing the samples taken from the classified dataset. Using Eqs. (3)−(5), various system parameters, such as channel gain, interference, and SINR, are calculated.

3.1 QSPCA Stage 1

In the first stage of QSPCA, we consider the set of users sharing the K-th RB $R_{k,n}$, which is the learning parameter, where $R_{k,n} = \{1, 2, \ldots, n\}$. Each D2D transmitter i in $R_{k,m}$ tries to adjust their actions to get the rewards in terms of the throughput. In Q-learning[19], we define different elements, such as agents, actions, states, and reward function as follows:

Agent: The D2D user pair is the agent of each D2D transmitter.

Action: The action is the set of power transmission levels of a D2D user pair, and it is given by Eq. (9):

$$p_k = (p_1, p_2, ..., p_{k_m}) \tag{9}$$

State: The set of states for the D2D user on the RB is defined as in Eq. (10):

$$S_k = \{T_k\} \tag{10}$$

where $T_k \in \{0, 1\}$ denotes the standard or stage of interference, which is defined by Eq. (11):

$$T_k = \begin{cases} 1, & \delta_k^m \geqslant \delta_{k\min}; \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where $\delta_{k\min}$ is the threshold SINR requirement to maintain the QoS over all users. Here, we make an assumption that cellular users report to the BS regarding the SINR and D2D users will know this information.

Strategy: There are multiple ways to take the best actions from the present Q-value. In this study, the exploration strategy is called the ε-greedy strategy. Let a random action be p_{rand}. During agent exploitation, this strategy will pick the best action based on the estimated present Q-value, and while exploring, it will randomly pick an action and aim for a high reward, which can be represented as

$$\pi_p^s = \begin{cases} \text{select } p_{\text{rand}}: \ E(p_{\text{rand}}) = \varepsilon, & \forall\, \varepsilon \in [0, 1]; \\ \text{select } \arg\max_{p_k \in P_k} Q(s_k, p_k): \ E(p) = 1-\varepsilon, & \forall\, \varepsilon \in [0, 1] \end{cases} \tag{12}$$

The agent may explore or exploit in every step, which can be decided by the ε-greedy strategy, and it uses $0 < \varepsilon < 1$.

Reward: The reward for the D2D user is the throughput, which is given in Eq. (13):

$$R_k = \begin{cases} \beta \log_2(1+\delta_t^n), & \delta_t^m \geqslant \delta_{t\min}^m; \\ -1, & \text{otherwise} \end{cases} \tag{13}$$

The actions are selected relentlessly during the training process. If and only if the forthcoming Q-value is higher than the current Q-value, then the Q-value tries to improve their values. The updated Q-value expression is given in Eq. (14):

$$Q_{k+1}(s_{k+1}, p_{k+1}, k) = \begin{cases} \max(Q_k(s_k, p_k, k), \\ R_{k+1} + \eta \cdot \max(Q_k(s_{k+1}, p_{k+1}, k))), & \text{if } s = s_k \text{ and } p = p_k; \\ Q_k(s_k, p_k, k), & \text{otherwise} \end{cases} \tag{14}$$

where $Q_{k+1}(s_{k+1}, p_{k+1})$ is the Q-value, s_{k+1} is the state, p_{k+1} is the action, and η is the discount factor.

In sum, the transmit power of the D2D user is optimized in the first stage. The generated dataset will be used for training and testing in the next stage.

3.2 QSPCA Stage 2

The generated dataset is used for classification purposes using the SVM classifier. The kernel technique is used to transform the data and get an optimal boundary between the achievable outputs on the basis of the transformations. In Stage 2, performance is dependent on the kernel function, whose objective is to find a mapping of training samples in terms of the transmit power of the D2D user and channel gain for the calculation of system parameters.

SVM is used to search for an optimal hyperplane that divides data input into two groups[20]. Each occurrence of the input denoted by X_s represents a pair (X_s, Z_s), where Z_s is the binary class label. The characteristics of the classification are given as a positive class that belongs to the predicted power devices and a negative class that belongs to the actual power. We can write the hyperplane as in Eq. (15):

$$W \cdot X_s + \sigma = 0 \tag{15}$$

where the classifier can be defined by Eq. (16):

$$\phi(X_s) = \begin{cases} X_s W + \sigma, & Z_s < -1; \\ X_s W + \sigma, & Z_s \geqslant 1 \end{cases} \tag{16}$$

where W is the weight, i.e., $W \in \mathbb{R}_s$, and σ is the bias factor.

The following problem defines the decision boundary. Now, we need to determine W and σ with the help of Formula (17) that can minimize $||W||$, such that for all data vectors,

$$Z_s \phi(X_s) \geqslant 1 \tag{17}$$

Then, the support vector X_s on the boundary is $Z_s \phi(X_s) = 1$.

The problem is commonly given as a similar problem of minimizing weight $||W||$. This is nothing but a quadratic programming problem[21].

The optimized solutions (W', σ') validate the classification vector X_s and is given in Eq. (18):

$$C = \text{sign}(xW' + \sigma') = \text{sign}(\phi'(x)) \tag{18}$$

where $\phi'(x)$ is the classification score and x represents the distance from the decision boundary.

We have considered positive Lagrange multipliers ψ_s, which can be multiplied by each constraint and subtracted from the objective function, which is given in Eq. (19).

The dataset is trained using an SVM classifier and support vector set. After the classification, the problem is solved using a Lagrange dual function, which is given as

$$\mu_{\text{D}} = \sum_k \psi_s - \sum_k \sum_l \psi_s \psi_l Z_k Z_l X_k{}' X_l \tag{19}$$

where μ_{D} is the Lagrange duality function. The pseudo code of the proposed QSPCA algorithm is given as Algorithm 1.

Algorithm 1 QSPCA algorithm with implementation steps

1: Initialize $Q_k(s_k, p_k)$ against all state $s_k \in S_k$ and action $p_k \in P_k$ with zero;
2: Initialize and evaluate the state S_k of the agent where s_k belongs to P_k in the current state;
3: **for**
  if $r < \varepsilon$
    current action selected from data sample;
  else
    current action will be taken from that index which is max of Q-value;
4: Repeat above step for each iteration;
5: Choose p_k from P_k i.e., action using the derived policy from $Q_{k+1}(s_k, p_k)$;
6: Take action p_k and observe the next state s_{k+1} and earn the rewards R_k;
7: Update the Q-table and Q-value as well as state and action in each iteration;
8: Now determine the next state and it will be $s_k \rightarrow s_{k+1}$ and take the current action as $p_k \rightarrow p_{k+1}$;
9: Now the system parameters with optimized actions will be used for training dataset which is done by SVM classifier;
10: Train the dataset, to obtain corresponding SVM classifiers and support vector sets X_s;
11: Combine all datasets and train them to get the final output by using Lagrangian dual function L_{D};
12: Observe the data to be classified;
13: Recognize the suitable action p of the suitable state s.

4 Simulation results and discussion

In this section, we discuss various ML techniques by observing the simulated results, including the analysis of the simulation parameter. We utilize different ML-based approaches to provide a solution for the optimized problem. The simulation results signify that the proposed algorithm maximizes the throughput by 15.31% and 19.5% as compared to Q-learning and SVM techniques to control the D2D power by mitigating the co-channel interference. Table 2 summarizes the basic system parameters used in this study.

In Fig. 1, the plot between the D2D throughput with respect to the SINR of the D2D user is investigated. Here, we compare the results of the different ML-based

Table 2 System parameters.

Parameter	Value
Radius of cell (m)	500
Noise power (dBm)	−116
Number of BS	1
Number of D2D user	100
Maximum transmit power (dBm)	23
Bandwidth (MHz)	10
Multipath fading	Exponential distribution with mean equals to 1
Standard deviation due to shadowing (dB)	8
Maximum D2D distance (m)	50
Pathloss exponent	4
Pathloss constant	10^{-2}
Pathloss model for D2D links (dB)	$22.7 + 40\log_{10}d$
Pathloss model for cellular links (dB)	$15.3 + 37.6\log_{10}d$

Note: The unit of d is km.

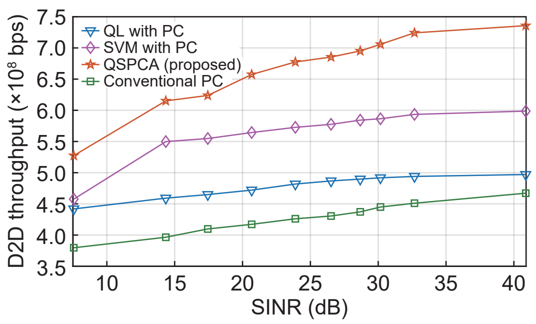


Fig. 1 D2D throughput versus SINR.

power control approaches. In the case of the proposed QSPCA algorithm, when the value of the SINR increases, the enhancement in the throughput is observed, and the performance is improved as a greater number of D2D users can get a higher data rate. The relationship between the D2D throughput and SINR of the D2D user is analyzed by considering the variant number of the D2D user. As we can see, the increment in the SINR leads to a strengthened throughput.

In Fig. 2, the graph between the D2D throughput and bit error rate (BER) is plotted. Here, the throughput decreases with the increase in the BER. However, the throughput slowly degrades with the respective BER value. To reduce the BER, the SINR must be increased. Therefore, the BER parameter should be taken into consideration when the performance of the network is analyzed.

In Fig. 3, a graph is plotted between the transmit power of the D2D user and that of multiple D2D users, where the number of D2D users n is 100 and the maximum power $P_{\max}$ is 23 dBm. Here, we can see that the proposed algorithm minimizes the power transmission of the D2D user as compared to the other power control techniques by the variant number of the D2D user.

Figure 4 shows the plotted graph of the D2D throughput with respect to the D2D user. If the number of D2D users increases, then the throughput of the D2D user keeps increasing. Accordingly, the proposed algorithm provides a better throughput as compared to the other power control mechanisms. Moreover, it has no impact on the performance of the D2D users because it keeps increasing the throughput as per the demand of users even in the crowded region.

Figure 5 shows the plotted graph between the transmit power of the D2D users and D2D throughput. The D2D throughput is relatively better as the D2D transmit power is limited to $P_{\max}$ = 23 dBm for performance improvement. Thus, the proposed algorithm maximizes the D2D throughput as compared to the existing ML techniques, even if the transmit power is bounded.

Figure 6 shows the graph between the SINR of the D2D users and the number of D2D users. The graph shows that the value of the SINR keeps improving for a large number of D2D users to achieve better performance and throughput. Moreover, we estimate that interference has been reduced between users.

Figure 7 shows the plotted graph between the cumulative density function (CDF) and SINR. It also

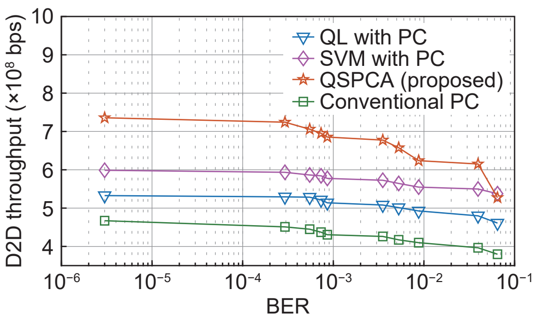


Fig. 2 D2D throughput versus BER.

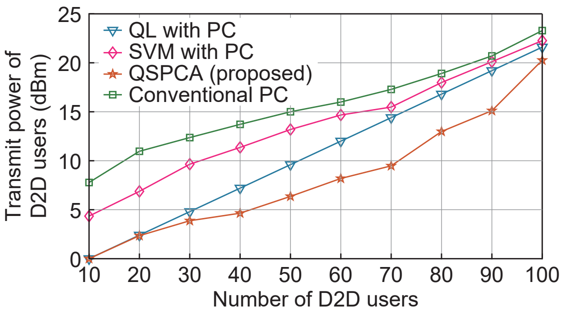


Fig. 3 Transmit power of D2D users versus the number of D2D users.

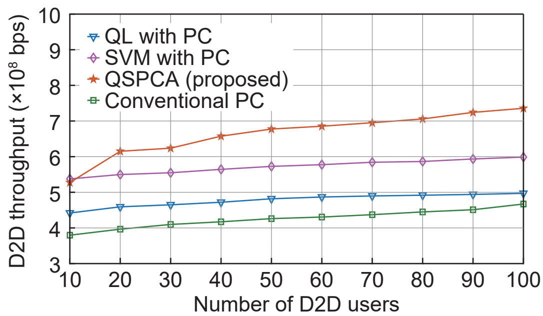


Fig. 4 D2D throughput versus the number of D2D users.

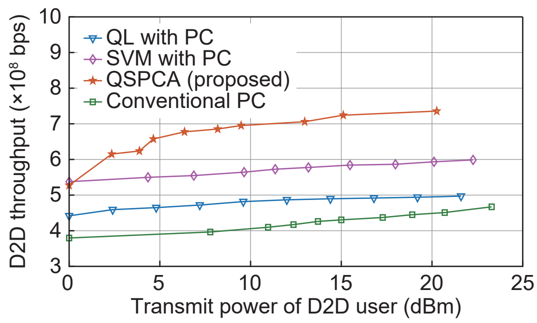


Fig. 5 D2D throughput versus transmit power of D2D users.

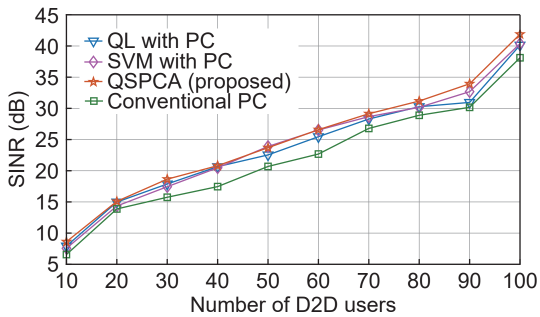


Fig. 6 SINR of the D2D users versus the number of D2D users.

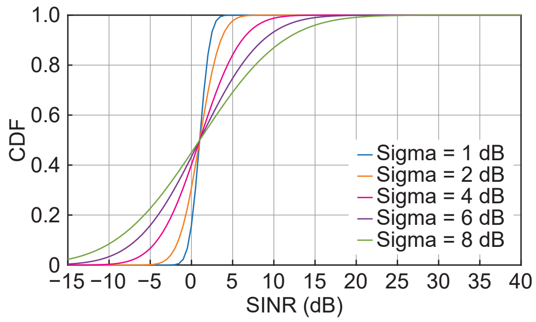


Fig. 7 CDF versus SINR.

depicts the CDF of the SINR of the receiving D2D throughput. Here, the SINR ranges from −15 dB to 40 dB. We can also see the behavior of the CDF for the different standard deviation values. In this study, for the standard deviation sigma $\sigma = 8$ dB, the SINR is improved with the increases in deviation by 1, 2, 4, 6, and 8 dB.

Figure 8 shows the plotted graph between the Q-value and Q-learning iteration. Here, the convergence behavior of the Q-learning algorithm corresponds to the Q-learning iteration, which shows the change in the Q-value over the iteration in a learning process of D2D users.

Figure 9 shows the simulated classification process

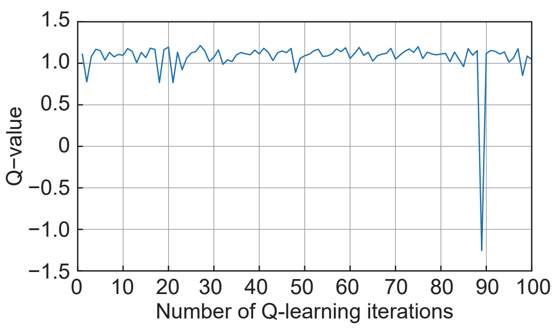


Fig. 8 Q-value versus the number of Q-learning iterations.

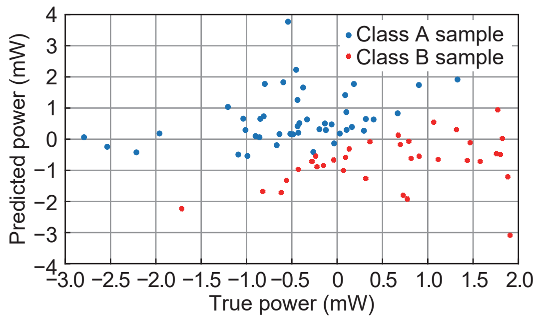


Fig. 9 SVM data classification of the positive and negative classes.

and classification results of our proposed power control algorithm. The observed graph includes a linear hyperplane and support vectors labeled with all datasets related to the SVM model.

5 Conclusion

The scope of the present work was performed while considering N D2D users and M cellular users. The experiment was limited to 100 iterations. Further scalability of the simulation parameter could be performed at later stages. For brevity, the convex analysis could be performed in the future. Geometry was not evaluated because it involves the analysis of topological parameters, which is beyond the scope of the present discussion.

In sum, this article presents a first-hand view of a two-stage transmit power control approach for D2D communication. The effectiveness of the proposed QSPCA algorithm was compared to state-of-the-art techniques. The transmit power of D2D users was first controlled by reducing the co-channel interference, and then the throughput was maximized. The computational results indicate improvements in the overall throughput of the D2D user and minimum SINR requirements while maintaining QoS over all the users. The simulation results imply that the proposed schemes improve the throughput of D2D users by ensuring that the controlled transmission power does not cross the limits as compared to the existing ML approach. The results also confirm that the QSPCA technique is better than existing models by at least 15.31% and 19.5% in terms of the throughput as compared to the SVM and Q-learning techniques, respectively. The control of the D2D transmit power with a lower latency could be utilized in 5G networks for industrial IoT applications, such as factory automation, production, and mining. Research and development in domains, such as time-sensitive networking and enhanced ultra-reliable low-latency communication, could also find immense usage of the proposed technique.

Acknowledgment

This work was supported by the Science and Engineering Research Board (SERB-DST), Govt. of India (No. EEQ/2019/000010).

References

[1] I. Budhiraja, N. Kumar, and S. Tyagi, Deep-reinforcement-learning-based proportional fair scheduling control scheme for underlay D2D communication, *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3143–3156, 2021.

[2] X. Wang, T. Jin, L. S. Hu, and Z. H. Qian, Energy-efficient power allocation and Q-learning-based relay selection for relay-aided D2D communication, *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6452–6462, 2020.

[3] A. Lakhan, M. A. Mohammed, A. N. Rashid, S. Kadry, T. Panityakul, K. H. Abdulkareem, and O. Thinnukool, Smart-contract aware ethereum and client-fog-cloud healthcare system, *Sensors*, vol. 21, no. 12, p. 4093, 2021.

[4] P. Gandotra and R. K. Jha, Device-to-device communication in cellular networks: A survey, *J. Netw. Comput. Appl.*, vol. 71, pp. 99–117, 2016.

[5] N. Lee, X. Q. Lin, J. G. Andrews, and R. W. Heath, Power control for D2D underlaid cellular networks: Modeling, algorithms, and analysis, *IEEE J. Sel. Areas Commun.*, vol. 33, no. 1, pp. 1–13, 2015.

[6] A. Hammoodi, L. Audah, M. A. Taher, M. A. Mohammed, M. S. Aljumaily, A. Salh, and S. A. Hamzah, Novel universal windowing multicarrier waveform for 5G systems, *Comput. Mater. Contin.*, vol. 67, no. 2, pp. 1523–1536, 2021.

[7] K. Zia, N. Javed, M. N. Sial, S. Ahmed, A. A. Pirzada, and F. Pervez, A distributed multi-agent RL-based autonomous spectrum allocation scheme in D2D enabled multi-tier HetNets, *IEEE Access*, vol. 7, pp. 6733–6745, 2019.

[8] M. Lee, G. D. Yu, and G. Y. Li, Accelerating resource allocation for D2D communications using imitation learning, presented at 2019 IEEE 90th Vehicular Technology Conf. (VTC2019-Fall), Honolulu, HI, USA, 2019, pp. 1–5.

[9] M. Lin, J. Ouyang, and W. P. Zhu, Joint beamforming and power control for device-to-device communications underlaying cellular networks, *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 138–150, 2016.

[10] S. Aslam, F. Alam, S. F. Hasan, and M. A. Rashid, A machine learning approach to enhance the performance of D2D-enabled clustered networks, *IEEE Access*, vol. 9, pp. 16114–16132, 2021.

[11] S. W. Nie, Z. Q. Fan, M. Zhao, X. Y. Gu, and L. Zhang, Q-learning based power control algorithm for D2D communication, presented at 2016 IEEE 27th Annual Int. Symp. Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, 2016, pp. 1–6.

[12] B. A. Khalaf, S. A. Mostafa, A. Mustapha, M. A. Mohammed, M. A. Mahmoud, B. A. S. Al-Rimy, S. A. Razak, M. Elhoseny, and A. Marks, An adaptive protection of flooding attacks model for complex network environments, *Secur. Commun. Netw.*, vol. 2021, p. 5542919, 2021.

[13] X. F. Wang, R. B. Li, C. Y. Wang, X. H. Li, T. Taleb, and V. C. M. Leung, Attention-weighted federated deep reinforcement learning for device-to-device assisted heterogeneous collaborative edge caching, *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 154–169, 2021.

[14] L. Feng, Z. X. Yang, Y. Yang, X. Y. Que, and K. Zhang, Smart mode selection using online reinforcement learning for VR broadband broadcasting in D2D assisted 5G HetNets, *IEEE Trans. Broadcast.*, vol. 66, no. 2, pp. 600–611, 2020.

[15] D. Q. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Q. Li, Device-to-device communications underlaying cellular networks, *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3541–3551, 2013.

[16] S. A. Mostafa, S. S. Gunasekaran, A. Mustapha, M. A. Mohammed, and W. M. Abduallah, Modelling an adjustable autonomous multi-agent internet of things system for elderly smart home, in *Proc. AHFE 2019 Int. Conf. Neuroergonomics and Cognitive Engineering, and the AHFE Int. Conf. Industrial Cognitive Ergonomics and Engineering Psychology*, H. Ayaz, ed. Washington, DC, USA: Springer, 2020, pp. 301–311.

[17] L. X. Li, Y. Xu, J. Y. Yin, W. Liang, X. Li, W. Chen, and Z. Han, Deep reinforcement learning approaches for content caching in cache-enabled D2D networks, *IEEE Internet Things J.*, vol. 7, no. 1, pp. 544–557, 2020.

[18] K. K. Nguyen, T. Q. Duong, N. A. Vien, N. A. Le-Khac, and M. N. Nguyen, Non-cooperative energy efficient power allocation game in D2D communication: A multi-agent deep reinforcement learning approach, *IEEE Access*, vol. 7, pp. 100480–100490, 2019.

[19] G. Künzel, L. S. Indrusiak, and C. E. Pereira, Latency and lifetime enhancements in industrial wireless sensor networks: A Q-learning approach for graph routing, *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5617–5625, 2020.

[20] S. Maldonado, J. Merigó, and J. Miranda, IOWA-SVM: A density-based weighting strategy for SVM classification via OWA operators, *IEEE Trans. Fuzzy Syst.*, vol. 28,

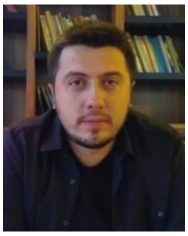
no. 9, pp. 2143–2150, 2020.
[21] S. Zidi, T. Moulahi, and B. Alaya, Fault detection in wireless sensor networks through SVM classifier, *IEEE Sens. J.*, vol. 18, no. 1, pp. 340–347, 2018.

Saurabh Chandra received the BEng degree in electronics & telecommunication engineering from Chhattisgarh Swami Vivekanand Technical University, Bhilai, India, in 2014 and the MTech degree in communication system engineering from Kalinga Institute of Industrial Technology, Bhubaneswar, India, in 2017. He is currently a research fellow in the Department of Electronics and Communication Engineering at National Institute of Technology Patna, Patna, India. His research interests include wireless communication and device to device communication.

Prateek received the BTech degree from Sir Padampat Singhania University, Udaipur, India, in 2014 and the MTech degree in communication system engineering from Kalinga Institute of Industrial Technology, Bhubaneswar, India, in 2017. He is currently pursuing the PhD degree in the Department of Electronics and Communication Engineering at National Institute of Technology Patna, Patna, India. His research interests include wireless communication and soft computing techniques.

Rohit Sharma received the BTech degree in electronics & communication engineering from Uttar Pradesh Technical University, Lucknow, India, in 2009 and the MTech degree in communication engineering from Shobhit University, Meerut, India, in 2011. He is currently pursuing the PhD degree in the Department of Electronics and Communication Engineering at National Institute of Technology Patna, Patna, India. His research interests include wireless communication and IoT.

Rajeev Arya received the BEng degree in electronics and communication engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya University, Bhopal, India, in 2008 and the MTech degree of technology in electronics and communication engineering from Indian Institute of Technology (ISM), Dhanbad, India, in 2012. He received the PhD degree in communication engineering from Indian Institute of Technology, Roorkee, India, in 2016. He has received the Ministry of Human Resource Development Scholarship, MHRD, India, when he was pursing the MTech and PhD degree. He is currently an assistant professor with the Department of Electronics and Communication Engineering at National Institute of Technology Patna, Patna, India. His current research interests are communication systems and wireless communication.



Korhan Cengiz received the BS degree in electronics and communication engineering from Kocaeli University, Umuttepe, Turkey and business administration from Anadolu University, Eskişehir, Turkey, in 2008 and 2009, respectively; he received the MS degree in electronics and communication engineering from Namik Kemal University, Tekirdağ, Turkey, in 2011; and the PhD degree in electronics engineering from Kadir Has University, Istanbul, Turkey, in 2016. He has been an assistant professor since 2018, and he has been the department vice chair of the Department of Electrical-Electronics Engineering, Trakya University, Edirne, Turkey since 2021. He is the author of over 40 articles including *IEEE Internet of Things Journal*, *IEEE Access*, *Expert Systems with Applications*, *Knowledge Based Systems* and *ACM Transactions on Sensor Networks*, 4 international patents, 4 book chapters, and 1 book in Turkish. He is a handling editor of *Microprocessors and Microsystems*, Elsevier, an associate editor of *IET Electronics Letters*, *IET Networks*, and an editor of AEÜ - *International Journal of Electronics and Communications*, Elsevier. He has guest editorial positions in *IEEE Internet of Things Magazine*, CMC-*Computers,* and *Materials&Continua*. He serves several reviewer positions for *IEEE Internet of Things Journal*, *IEEE Sensors Journal*, and *IEEE Access*. He serves several book editorial positions in IEEE, Springer, Elsevier, Wiley, and CRC. He presented more than 20 keynote talks in reputed IEEE and Springer conferences about WSNs, IoT, and 5G. He is the senior member of IEEE and professional member of ACM. His awards and honors include the Tubitak Priority Areas PhD Scholarship, the Kadir Has University PhD Student Scholarship, Best Presentation Award in ICAT 2016 Conference, and Best Paper Award in ICAT 2018 Conference. His research interests include wireless sensor networks, wireless communications, statistical signal processing, indoor positioning systems, internet of things, power electronics, and 5G.