

Machine Learning-Based Resource Optimization for D2D Communication Underlying Networks

Lingting Zhu, Chonghe Liu, Jiantao Yuan, and Guanding Yu

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China

{ltzhu, liuchonghe, yuanjiantao, yuguanding}@zju.edu.cn

Abstract—Deploying device-to-device (D2D) communication over underlying cellular network can significantly enhance the spectrum utilization. However, co-channel interference will occur when D2D pairs share the same radio resource with cellular users. To mitigate the interference within a reasonable range, a machine learning based resource reuse scheme for D2D and cellular users is proposed in this paper. Specifically, we formulate an optimization problem to maximize the total throughput of D2D pairs and cellular users by optimally allocating subcarrier and power within the limits of the interference threshold. Since the formulated problem is a mixed integer non-linear programming problem, we solve it in two steps. First, we assign the orthogonal subcarriers to different cellular users to maximize the total throughput of all cellular users. Then, D2D pairs are allowed to reuse different subcarriers to further enhance the throughput without affecting the performance of cellular users. The second step is still NP-hard and therefore we propose a low-complexity algorithm based on the pointer network, a specific neural network structure proposed recently. Results reveal that, with remarkably reduced complexity, the proposed scheme outperforms the conventional resource allocation algorithms.

I. INTRODUCTION

Device-to-device (D2D) communication technique is an effective approach to improve the cellular network capacity [1]. In D2D communication, the D2D transmitter (D2DT) can send data directly to the D2D receiver (D2DR) via a direct link without the assistance of the base station (BS). This communication method will bring many benefits, such as increasing data rate, reducing energy consumption, and decreasing end-to-end latency [2].

There are two deployment modes for the D2D devices coexisting with the cellular users (CUEs) [3]: overlaying mode and underlying mode. In the overlaying mode, the CUEs and D2D devices are assigned dedicated radio resources without co-channel interference. However, the spectrum resources are usually insufficient due to the existence of a large number of users in the cellular network [4]. On the other hand, the underlying mode allows the CUEs and D2D devices to share the same radio resources. Therefore, the D2D devices will interfere the CUEs when they use the same radio resources. Generally, the spectrum efficiency obtained by this mode is higher than that in the overlaying mode if appropriate interference management mechanisms are adopted. Nevertheless, the interference management in underlying mode is a very challenging issue [5].

In recent years, many works have investigated the interference management and resource allocation for D2D networks. For instance, a two-phase auction algorithm has been proposed to analyze the interference for D2D communication in [6]. The game-theoretic model has been formulated to investigate the radio resource allocation in underlying D2D communication [7]. The graph-based resource allocation problem has been studied to maximize the performance for D2D underlying cellular networks in [8] and [9]. Energy efficient power control scheme and energy-aware cooperative traffic offloading via D2D cooperations have been studied in [10] and [11]. To maximize the sum rate while meeting the successive interference cancellation decoding constraint, a joint D2D mode selection and resource allocation scheme has been investigated in [12]. The study in [13] aims at maximizing the weighted sum-rate with the consideration of quality-of service (QoS) for both cellular and D2D users.

However, each D2D pair is only allowed to reuse one resource block (RB) with a cellular user in the above-mentioned works. In fact, to improve the performance, a D2D pair can also reuse multiple RBs with multiple CUEs in one frame if the co-channel interference can be properly controlled. To tackle this issue, a resource allocation optimization problem has been formulated for underlying D2D networks in [14], and the maximal independent sets and knapsack algorithm have been adopted to solve the combinatorial problem. The multi-agent deep reinforcement learning has been adopted to decrease the computational complexity and reduce signaling overhead in [15].

In this paper, a subcarrier and power allocation problem to maximize the total throughput of D2D pairs and CUEs is formulated. We consider that the CUEs adopt the orthogonal frequency division multiple access (OFDMA) technology to use the subcarriers [16], and the D2D pairs reuse the CUEs' resources by the underlying mode. To avoid severe interference to both CUEs and D2D links, the received interference on each subcarrier from both sides is strictly controlled within a reasonable demodulation threshold. Since the optimization problem is non-convex, we re-formulate the original problem into a knapsack problem with two restrictions. This problem belongs to multidimensional knapsack problem (MKP) [17]. Even though existing heuristic algorithms can approach the MKP, they need to be designed elaborately and would still suffer from high computational complexity. Therefore, we adopt the pointer network [18] to develop a low-complexity

Machine Learning-Based Resource Optimization for D2D Communication Underlying Networks

Lingting Zhu, Chonghe Liu, Jiantao Yuan, and Guanding Yu

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China

{ltzhu, liuchonghe, yuanyiantao, yuguanding}@zju.edu.cn

Abstract—Deploying device-to-device (D2D) communication over underlying cellular network can significantly enhance the spectrum utilization. However, co-channel interference will occur when D2D pairs share the same radio resource with cellular users. To mitigate the interference within a reasonable range, a machine learning based resource reuse scheme for D2D and cellular users is proposed in this paper. Specifically, we formulate an optimization problem to maximize the total throughput of D2D pairs and cellular users by optimally allocating subcarrier and power within the limits of the interference threshold. Since the formulated problem is a mixed integer non-linear programming problem, we solve it in two steps. First, we assign the orthogonal subcarriers to different cellular users to maximize the total throughput of all cellular users. Then, D2D pairs are allowed to reuse different subcarriers to further enhance the throughput without affecting the performance of cellular users. The second step is still NP-hard and therefore we propose a low-complexity algorithm based on the pointer network, a specific neural network structure proposed recently. Results reveal that, with remarkably reduced complexity, the proposed scheme outperforms the conventional resource allocation algorithms.

I. INTRODUCTION

Device-to-device (D2D) communication technique is an effective approach to improve the cellular network capacity [1]. In D2D communication, the D2D transmitter (D2DT) can send data directly to the D2D receiver (D2DR) via a direct link without the assistance of the base station (BS). This communication method will bring many benefits, such as increasing data rate, reducing energy consumption, and decreasing end-to-end latency [2].

There are two deployment modes for the D2D devices coexisting with the cellular users (CUEs) [3]: overlaying mode and underlying mode. In the overlaying mode, the CUEs and D2D devices are assigned dedicated radio resources without co-channel interference. However, the spectrum resources are usually insufficient due to the existence of a large number of users in the cellular network [4]. On the other hand, the underlying mode allows the CUEs and D2D devices to share the same radio resources. Therefore, the D2D devices will interfere the CUEs when they use the same radio resources. Generally, the spectrum efficiency obtained by this mode is higher than that in the overlaying mode if appropriate interference management mechanisms are adopted. Nevertheless, the interference management in underlying mode is a very challenging issue [5].

In recent years, many works have investigated the interference management and resource allocation for D2D networks. For instance, a two-phase auction algorithm has been proposed to analyze the interference for D2D communication in [6]. The game-theoretic model has been formulated to investigate the radio resource allocation in underlying D2D communication [7]. The graph-based resource allocation problem has been studied to maximize the performance for D2D underlying cellular networks in [8] and [9]. Energy efficient power control scheme and energy-aware cooperative traffic offloading via D2D cooperations have been studied in [10] and [11]. To maximize the sum rate while meeting the successive interference cancellation decoding constraint, a joint D2D mode selection and resource allocation scheme has been investigated in [12]. The study in [13] aims at maximizing the weighted sum-rate with the consideration of quality-of service (QoS) for both cellular and D2D users.

However, each D2D pair is only allowed to reuse one resource block (RB) with a cellular user in the above-mentioned works. In fact, to improve the performance, a D2D pair can also reuse multiple RBs with multiple CUEs in one frame if the co-channel interference can be properly controlled. To tackle this issue, a resource allocation optimization problem has been formulated for underlying D2D networks in [14], and the maximal independent sets and knapsack algorithm have been adopted to solve the combinatorial problem. The multi-agent deep reinforcement learning has been adopted to decrease the computational complexity and reduce signaling overhead in [15].

In this paper, a subcarrier and power allocation problem to maximize the total throughput of D2D pairs and CUEs is formulated. We consider that the CUEs adopt the orthogonal frequency division multiple access (OFDMA) technology to use the subcarriers [16], and the D2D pairs reuse the CUEs' resources by the underlying mode. To avoid severe interference to both CUEs and D2D links, the received interference on each subcarrier from both sides is strictly controlled within a reasonable demodulation threshold. Since the optimization problem is non-convex, we re-formulate the original problem into a knapsack problem with two restrictions. This problem belongs to multidimensional knapsack problem (MKP) [17]. Even though existing heuristic algorithms can approach the MKP, they need to be designed elaborately and would still suffer from high computational complexity. Therefore, we adopt the pointer network [18] to develop a low-complexity

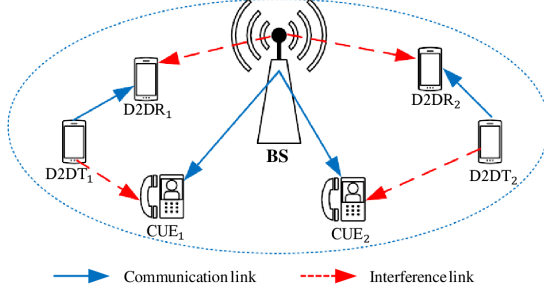


Fig. 1. System model of D2D communication underlying network.

algorithm in this paper. The pointer network is a special kind of neural network proposed recently, which has the potential to solve combinatorial optimization problems efficiently [19]. To the best of our knowledge, this paper is the first attempt to apply the pointer network to solve wireless resource allocation problem.

The rest of the paper is organized as follows. In Section II, the D2D communication underlying network is introduced. We formulate the optimization problem and develop the low-complexity algorithm in Section III. The performance of the proposed scheme is evaluated via simulation in Section IV. Section V finally concludes this paper.

II. SYSTEM MODEL

In the system model, the downlink data transmission scenario is considered where the BS is placed in the center of the cell. The D2D communication consists of a pair of devices, which can implement direct data transmission with each other without the assistance of the BS. In contrast, the cellular user equipments refer to the mobile terminals which can only communicate through the BS. There are N CUEs and M D2D pairs randomly deployed in the coverage of the BS, as shown in Fig. 1.

The underlying mode is considered where the CUEs and D2D devices coexist on the same radio resources. The OFDMA is adopted by the CUEs and there are F orthogonal available subcarriers in the system. Each subcarrier can only be assigned to at most one CUE to avoid inter-user interference. The D2D pairs reuse the CUEs' radio resources while meeting the interference restriction. The D2D devices are assumed to be already paired, but whether a D2D pair is activated should be determined according to the resource allocation algorithm. We assume that full channel state information (CSI) is available. That is, the BS can acquire all CSI from each CUE and between D2D pairs.

As OFDMA is adopted by the CUEs, interference for a CUE only comes from the D2D devices that reuse the same radio resources. Then, the received interference for the CUE on subcarrier f can be expressed as

$$I_f = \sum_{m=1}^M v_{m,f} p_{m,f} h_{m,0,f}, \quad (1)$$

where $v_{m,f} \in \{0,1\}$ indicates whether the subcarrier f is assigned to the D2D pair m , $p_{m,f}$ stands for the transmit power on subcarrier f for D2DT m , and $h_{m,0,f}$ represents the channel power gain on subcarrier f between the D2DT m and the BS (denoted as 0), including the path loss, shadowing and channel fading.

Similarly, the received interference for D2DR m on subcarrier f is given by

$$I_{m,f} = \sum_{n=1}^N w_{n,f} p_{n,f} h_{n,m,f} + \sum_{m'=1, m' \neq m}^M v_{m',f} p_{m',f} h_{m',m,f}, \quad (2)$$

where $w_{n,f}$ indicates whether the subcarrier f is assigned to CUE n , $p_{n,f}$ stands for the transmit power on subcarrier f for CUE n , $h_{n,m,f}$ and $h_{m',m,f}$ represent the channel power gain on subcarrier f between the D2DR m and the CUE n , and between the D2DR m and the D2DT m' , respectively.

From the expression of Shannon's capacity, the achievable data rate is directly affected by the signal-to-interference-plus-noise ratio (SINR). According to (1) and (2), the SINR expressions for the CUE n and the D2D pair m on subcarrier f can be respectively expressed as

$$\gamma_{n,f} = \frac{p_{n,f} h_{n,0,f}}{I_f + N_{n,f}^C}, \quad (3)$$

$$\gamma_{m,f} = \frac{p_{m,f} h_{m,m,f}}{I_{m,f} + N_{m,f}^D}, \quad (4)$$

where $h_{n,0,f}$ and $h_{m,m,f}$ represent the channel power gain on subcarrier f between the CUE n and the BS, and between the D2DT m and D2DR m , respectively, $N_{n,f}^C$ is the noise power for the CUE n , and $N_{m,f}^D$ is the noise power for D2DR m , both on subcarrier f .

Therefore, the total achievable data rates for all CUEs, and for all D2D pairs, can be respectively expressed as

$$R_C = \sum_{n=1}^N \sum_{f=1}^F w_{n,f} B \log_2 \left(1 + \frac{-1.5 \gamma_{n,f}}{\ln(5p_e^C)} \right), \quad (5)$$

$$R_D = \sum_{m=1}^M \sum_{f=1}^F v_{m,f} B \log_2 \left(1 + \frac{-1.5 \gamma_{m,f}}{\ln(5p_e^D)} \right), \quad (6)$$

where B is the bandwidth of each subcarrier, p_e^C and p_e^D are used to denote the targeted bit error rate (BER) for CUEs and D2D devices, respectively.

III. PROBLEM FORMULATION AND LOW COMPLEXITY ALGORITHM

In this section, an optimization problem is first formulated to maximize the total throughput of all CUEs and D2D pairs. Then, a machine learning based low computational complexity algorithm is designed to solve it.

A. Problem Formulation

As the D2D pairs transmit data opportunistically, a precise power control mechanism will consume a large amount of

control signaling. Then, we assume that a D2D pair would use constant power for data transmission once it is activated, i.e., the transmit power of D2DT m on each subcarrier can be expressed as

$$p_{m,f} = \begin{cases} 0, & \text{if D2D pair } m \text{ is inactive on } f, \\ P, & \text{if D2D pair } m \text{ is active on } f. \end{cases} \quad (7)$$

We denote $S_f = \{m | v_{m,f} = 1\}$ to indicate the set of active D2D pairs on subcarrier f .

To maximize the total throughput of both CUEs and D2D pairs, the optimization problem can be expressed as

$$\text{OP1} : \max_{\{\mathbf{W}, \mathbf{V}, \mathbf{P}\}} \{R_C + R_D\}, \quad (8)$$

subject to

$$\sum_{n=1}^N w_{n,f} \leq 1, \quad \forall f, \quad (8a)$$

$$w_{n,f} \in \{0, 1\}, \quad \forall f, n, \quad (8b)$$

$$v_{m,f} \in \{0, 1\}, \quad \forall f, m, \quad (8c)$$

$$0 \leq \sum_{n=1}^N \sum_{f=1}^F p_{n,f} \leq P_{\max}, \quad (8d)$$

$$\gamma_{n,f} \geq \gamma_{th,C}, \quad \forall f, n, \quad (8e)$$

$$\gamma_{m,f} \geq \gamma_{th,D}, \quad \forall f, \forall m \in S_f, \quad (8f)$$

where $\mathbf{W} = [w_{n,f}]_{N \times F}$ is the N by F subcarrier allocation matrix for CUEs, $\mathbf{V} = [v_{m,f}]_{M \times F}$ is the M by F subcarrier allocation matrix for D2D pairs, and $\mathbf{P} = [p_{n,f}]_{N \times F}$ is the matrix of transmit powers of CUEs on all subcarriers. Constraints (8a) and (8b) ensure that each subcarrier can be allocated to at most one CUE. Constraint (8d) is used to restrict the transmit power of the BS. Constraints (8e) and (8f) represent the SINR requirements for the CUEs and the D2D pairs, respectively. Here, $\gamma_{th,C}$ and $\gamma_{th,D}$ are the minimum required demodulation threshold to ensure the targeted QoS for CUEs and D2D pairs, respectively.

B. The Proposed Algorithm

We can see that OP1 is a mixed integer non-linear programming optimization problem, thus it is NP-hard in general. Therefore, we decouple the original problem (OP1) into two subproblems (i.e., OP2 and OP3), and develop an algorithm with low computational complexity to solve it. The decoupled resource allocation algorithm includes the allocation for CUEs under the constraints of the BS's transmit power and orthogonal subcarrier firstly, and then the allocation for D2D pairs within interference threshold.

1) *Resource Allocation for CUEs*: The first subproblem (OP2) aims to maximize the total throughput of all CUEs by assuming orthogonal subcarriers with no interference coming from D2D pairs. That is,

$$\text{OP2} : \max_{\{\mathbf{W}, \mathbf{P}\}} R_C, \quad (9)$$

subject to (8a), (8b), and (8d).

To obtain the maximum throughput of all CUEs, we can maximize the transmission rate of CUEs on each subcarrier.

Therefore, the maximum carrier-to-interference ratio (Max C/I) algorithm is adopted for CUEs. That is, the CUE with the best channel gain is permitted to transmit data on each subcarrier.

Moreover, the transmit power of CUE n^* is controlled by the open loop power control mechanism to reduce interference on the D2D receivers, i.e.

$$p_{n^*,f} \geq \frac{\gamma_{th,C}(I_{\max} + N_{n^*,f}^C)}{h_{n^*,0,f}}, \quad (10)$$

where I_{\max} is the maximum allowed interference on each subcarrier.

2) *Resource Allocation for D2D pairs*: After determining the resource allocation for all CUEs, the subcarrier assignment for D2D pairs can be formulated as

$$\text{OP3} : \max_{\mathbf{V}} R_D, \quad (11)$$

subject to (8c), (8e), and (8f).

From (1) and (3), we can rewrite (8e) as

$$\sum_{m=1}^M v_{m,f} p_{m,f} h_{m,0,f} \leq \min_n \left\{ \frac{p_{n,f} h_{n,0,f}}{\gamma_{th,C}} - N_{n,f}^C \right\}. \quad (12)$$

Similarly, (8f) can be rewritten as

$$\sum_{m'=1}^M v_{m',f} p_{m',f} h_{m',m,f} \leq \min_{m \in S_f} \left\{ \frac{P h_{m,m,f}}{\gamma_{th,D}} - N_{m,f}^D - \sum_{n=1}^N w_{n,f} p_{n,f} h_{n,m,f} + v_{m,f} P h_{m,m,f} \right\}. \quad (13)$$

To make the restriction conveyed in (8f) have the same form as (12), we first rewrite it as (13). Meanwhile, since S_f is related to the values of $v_{m,f}$ and is hard to be determined before resource allocation, we relax the right part of (13) using the set of all D2D pairs. By swapping the index m' with m , we can rewrite (13) as

$$\sum_{m=1}^M v_{m,f} p_{m,f} \max_{m'} \{h_{m,m',f}\} \leq \min_{m'} \left\{ \frac{P h_{m',m',f}}{\gamma_{th,D}} - N_{m',f}^D - \sum_{n=1}^N w_{n,f} p_{n,f} h_{n,m',f} \right\}. \quad (14)$$

After the transformation of (8e) and (8f), (OP3) can be transformed into a two-dimensional knapsack problem, i.e., the weights of the entities of the knapsack problem have two dimensions. This kind of problem belongs to multidimensional knapsack problem (MKP) [17] and only heuristic algorithms exist to solve such an NP-hard problem. But effective heuristic algorithms still suffer from high time complexity, making it infeasible for practical application. In this paper, the pointer network, which has the potential for solving combinatorial optimization problems, is adopted to effectively solve it.

3) *Algorithm Implementation*: As mentioned above, the reuse problem for each subcarrier f can be seen as a two-dimensional knapsack problem. Given the allocation status of CUEs, we represent each D2D pair as a 3D attributes

Algorithm 1 The machine learning based algorithm.

- 1: **Phase 1: Resource allocation for CUEs.**
- 2: Set the total number of CUEs N , the total number of D2D pairs M , and total number of simulations times;
- 3: Initialize $R_C = 0$, $R_D = 0$, and $p_{n,f} = P_{\max}/N$;
- 4: **for** each $f \in [1, F]$ **do**
- 5: Obtain n^* with the best channel gain on subcarrier f , update the transmit power of CUE n^* by (10), and update $p_{n,f}$, R_C , and the required arguments in (12) and (14);
- 6: **end for**
- 7: **Phase 2: Resource allocation for D2D pairs.**
- 8: Set the training set S , number of training steps T , batch size B , pointer network parameter θ ;
- 9: **for** each $t \in [1, T]$ **do**
- 10: Select a batch of sample s_i for $i \in \{1, 2, \dots, B\}$;
- 11: Sample solution o_i based on $p_\theta(\cdot|s_i)$ for $i \in \{1, 2, \dots, B\}$;
- 12: Compute value $V(o_i|s_i)$;
- 13: Let $g_\theta = \frac{1}{B} \sum_{i=1}^B (V(o_i|s_i) - b(s_i)) \nabla_\theta \log p_\theta(o_i|s_i)$;
- 14: Update $\theta = \text{ADAM}(\theta, g_\theta)$;
- 15: Update baseline $b(s_i) = b(s_i) + \alpha(V(o_i|s_i) - b(s_i))$ for $i \in \{1, 2, \dots, B\}$;
- 16: **end for**
- 17: **for** each $f \in [1, F]$ **do**
- 18: Utilize pointer network to compute the $v_{m,f}$ for $m \in \{1, 2, \dots, M\}$ and update R_D ;
- 19: **end for**

vector (v, x, y) , where v is the equivalent value of the entity from (6), x and y are two weights in the restrictions of the two-dimensional knapsack problem and can be derived from (12) and (14), respectively. As the pointer network is based on the sequence-to-sequence model, the input should be a sequence which compromises the 3D attributes vectors mentioned above. The output is also a sequence and can be derived by the pointing mechanism of the pointer network, which re-orders the input. The output corresponds to one set of valid entities that satisfy the constraint. To be specific, we traverse the output sequence and stop when the found entities exceed the restrictions in (12) and (14). The found entities are the solution to the knapsack problem. We call it solution o and denote $V(o)$ as the total value of the corresponding set of the entities.

Now we introduce the architecture of pointer network, which is shown in Fig. 2. As in the figure, the pointer network compromises two recurrent neural network (RNN) modules depicted by different colors to distinguish the encoder and the decoder, respectively. The two RNN modules both consist of long short-term memory (LSTM) [20] cells. We encode each two-dimensional knapsack instance as a sequence of 3D vectors. Specifically, the sequence can be represented as $s = \{(v_i, x_i, y_i)\}_{i=1}^M$. Here, we use sequence s as the input and the embedding of s_i is the input of the encoder at time step i . The d -dimensional embedding is obtained by linear

transformation of s_i shared across all input steps. Attention mechanism is applied as a pointer to select a member of the input sequence at each time step [18]. And the embedding of the selected one is used as the input of the decoder at the next time step. At the first time step, the input of the decoder is a d -dimensional vector which is trainable. Fig. 2 shows the detailed pointing mechanism, including the red pointer in the encoder and the input layer in the decoder.

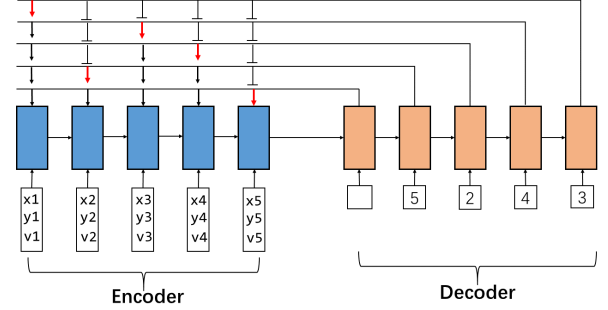


Fig. 2. Architecture of pointer network.

4) *Training Process*: The training process is combined with the reinforcement learning method, which provides an excellent paradigm for training process for combinatorial optimization problems [19]. We use model-free policy-based reinforcement learning to optimize the parameters. The value $V(o)$ is set as the training objective which is called reward in our training process. To implement the training process conveniently, policy gradient method is used. In particular, we apply the well-known REINFORCE algorithm [21] to obtain the gradient and use the gradient to update the parameters of the network by Adam [22]. With Monte Carlo sampling, we utilize the REINFORCE algorithm in a more feasible way. After setting the batch size B , by generating B independent and identically distributed (i.i.d.) sample knapsack problems, the gradient can be expressed in a stochastically average form, as

$$g_\theta = \frac{1}{B} \sum_{i=1}^B (V(o_i|s_i) - b(s_i)) \nabla_\theta \log p_\theta(o_i|s_i), \quad (15)$$

$$\theta = \text{ADAM}(\theta, g_\theta). \quad (16)$$

In the above, $b(s)$ denotes the baseline function in training process which does not depend on the permutation of the sequence in the network and estimates the expected value to reduce the variance of the gradients. Here, we use an exponential moving average of the reward obtained by the network over time. The detailed principle can be found in [19].

C. Computational Complexity Analysis

We now analyze the computational complexity of the proposed algorithm. Regarding the Phase 1 in Algorithm 1, the computational complexity is $O(N)$. The basic modules in the pointer network in Phase 2 are long short-term memory

TABLE I
SIMULATION PARAMETERS

Parameters	Value
Center frequency	2.3GHz
Subcarrier bandwidth	15kHz
The maximum transmission power of BS, P_{\max}	20W
The maximum coverage radius of BS	200m
Distance between the D2DT and D2DR	15m
The targeted BER for CUEs, p_e^C	0.1
The targeted BER for D2D pairs, p_e^D	0.1
The minimum required SINR to ensure the targeted QoS for CUEs, $\gamma_{th,C}$	4dB
The minimum required SINR to ensure the targeted QoS for D2D pairs, $\gamma_{th,D}$	8dB

cells with attention. Since the computation of attention vector performs M times at each output time [18], the computational complexity is $O(M^2)$ for any given fine-tuned pointer network. Therefore, the total computational complexity of the algorithm is $O(N + M^2)$.

IV. SIMULATION RESULTS

In this section, we present the simulation results for the proposed algorithm. In the simulation setup, we consider an NR-based network with one BS, $N = 10, 20$ CUEs and $M = 10, 20, 30, 40, 50$ D2D pairs. In addition, the CUEs and D2D pairs are randomly located in the service coverage of the BS. We assume that the channel fading on both CUEs and D2D pairs follows the Rayleigh distribution with uniform variance. In detail, the major parameters in the simulation are listed in Table I.

As for the pointer network, we train our network for each given condition. For each experiment, we use 8,000 training samples and 2,000 testing samples. We set the batch size as 128 and utilize LSTM cells with 128 hidden units.

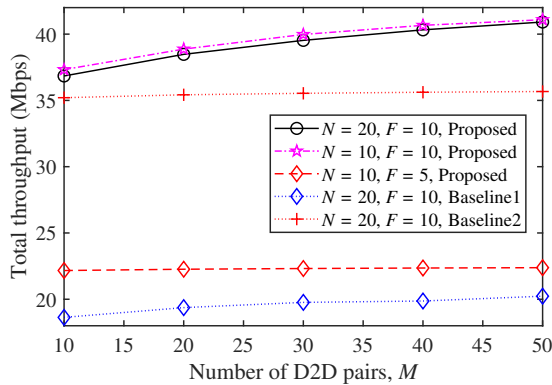


Fig. 3. Total throughput of CUEs and D2D pairs.

The total data rate for both CUEs and D2D pairs, $R_C + R_D$, is plotted in Fig. 3. The transmit power of D2D pairs is fixed as $P = 100$ mW. We compare the proposed algorithm with two algorithms: Baseline 1 and Baseline 2. In Baseline 1, the best D2D pair with the most contribution to the overall throughput can reuse the resource block of each CUE. In Baseline 2, the best two D2D pairs can reuse the resource block. In both baseline algorithms, the resource is randomly allocated to each CUE in advance. So the total throughput of the CUEs is irrespective to the number of D2D pairs.

From Fig. 3, the proposed algorithm can always achieve better performance than the baseline algorithms because of the multi-user diversity brought by the SNR scheduling. In addition, the larger bandwidth is adopted, the higher throughput will be obtained. On the other hand, the total throughput obtained by the proposed algorithm increases gradually as the number of D2D pairs M increases. However, the increment is not significant due to the interfere restriction in the underlying mode.

To measure the performance of the proposed algorithm, the comparison of heuristic algorithm is necessary and of great importance. In Fig. 4, we compare the total throughput of D2D pairs in the proposed algorithm with that of one classical algorithm, Tabu search. The numbers of CUEs and D2D pairs are set to be $N = 20$ and $M = 50$ and the transmit power of each D2D pair is set to be 100 mW. Tabu search [23] uses a local or neighborhood search procedure to search from one potential solution to another iteratively. An effective implementation of MKP based on Tabu search can be found in [24] and interested readers can refer to [23] for detailed information. The computational complexity of Tabu search is $O(I(M^2 + M + L))$, where I is the maximum number of iterations, L is the length of Tabu list consisting of solutions which have changed by the process of moving from one solution to another, and M is the scale of the MKP problem, i.e., the number of D2D pairs. In our experiment, we set I as 300 and L as 200. Besides, the computation of neural network can be accelerated by parallel computing via graphics processing unit (GPU). Therefore, the required time for Tabu search is much larger than that for our proposed algorithm.

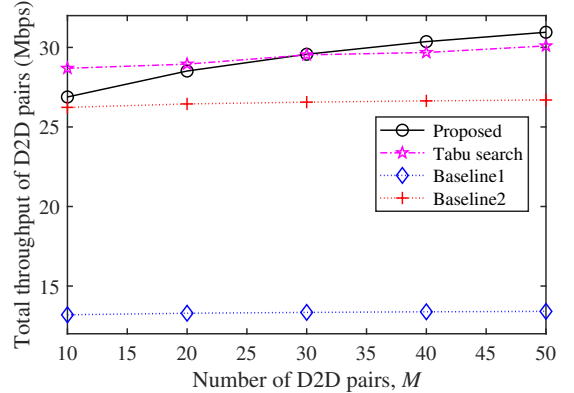


Fig. 4. Total throughput of D2D pairs.

From Fig. 4, it can be seen that the proposed algorithm has a close performance to the Tabu search algorithm. When the number of D2D pairs is small, Tabu search has a better performance. But as the number gets larger, the proposed algorithm has a better one. The results indicate that with remarkably reduced computational complexity, the proposed algorithm based on machine learning achieves a satisfactory performance. However, Tabu search is strictly limited to the size of the problem and it costs much time to search the solution.

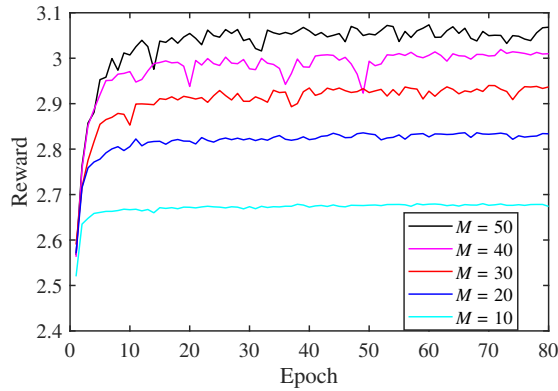


Fig. 5. Convergence of the proposed algorithm.

Fig. 5 depicts the convergence of the proposed algorithm. We plot the convergence curves for different numbers of D2D pairs with the epoch number. From Fig. 5, we can observe that the training process of the proposed machine learning based algorithm converges very fast. As the problem size, M , increases, the convergence rate becomes slow. This is rather intuitive since the model space of pointer networks becomes more complicated as the problem size increases, consuming more time for model convergence.

V. CONCLUSIONS

In this paper, a resource allocation algorithm is proposed for D2D communication underlying networks. To better utilize the limited spectrum resources, the D2D pairs are allowed to reuse multiple CUEs resources as long as the interference is below the interference threshold. We formulate a mixed-integer non-linear programming optimization problem and develop a low-complexity algorithm based on the pointer network. Detailed network structure design and training process are provided. Simulation results reveal that the proposed algorithm can achieve satisfactory performance with low computational complexity.

REFERENCES

- [1] D. Feng, L. Lu, Y. Wu, G. Li, S. Li, and G. Feng, "Device-to-device communications in cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 49-55, Apr. 2014.
- [2] N. Saxena, F. H. Kumbhar, and A. Roy, "Exploiting social relationships for trustworthy d2d relay in 5g cellular networks," *IEEE Commun. Mag.*, vol. 58, no. 2, pp. 48-53, Feb. 2020.

- [3] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surv. & Tut.*, vol. 16, pp. 1801-1819, Fourthquarter, 2014.
- [4] G. Yu, L. Xu, D. Feng, R. Yin, and G. Li, "Joint mode selection and resource allocation for device-to-device communications," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3814-3824, Nov. 2014.
- [5] P. Gandotra, R. Kumar, and S. Jain, "A survey on device-to-device (D2D) communications: architecture and security issues," *J. Network & Computer*, vol. 78, pp. 9-29, Jan. 2017.
- [6] Y. Hassan, F. Hussain, S. Hossen, S. Choudhury, and M. M. Alam, "Interference minimization in d2d communication underlying cellular networks," *IEEE Access*, vol. 5, pp. 22471-22484, Oct. 2017.
- [7] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 136-144, Jun. 2014.
- [8] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference graph-based resource allocation (ingra) for d2d communications underlying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3844-3850, Aug. 2015.
- [9] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Resource allocation for d2d communication underlaid cellular networks using graph-based approach," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 7099-7113, Oct. 2016.
- [10] Y. Wu, J. Wang, L. Qian, and R. Schober, "Optimal power control for energy efficient D2D communication and its distributed implementation," *IEEE Communications Letters*, vol. 19, no. 5, pp. 815-818, Feb. 2015.
- [11] Y. Wu, J. Chen, L. Qian, J. Huang, and X. Shen, "Energy-aware cooperative traffic offloading via device-to-device cooperations: an analytical approach," *IEEE Transactions on Mobile Computing*, vol. 16, no. 1, pp. 97-114, Jan. 2017.
- [12] R. Yin, C. Zhong, G. Yu, Z. Zhang, K. Wong, and X. Chen, "Joint spectrum and power allocation for D2D communications underlying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2182-2195, Apr. 2016.
- [13] R. Wang, J. Zhang, S. H. Song, and K. B. Letaief, "Optimal qos-aware channel assignment in D2D communications with partial CSI," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7594-7609, Nov. 2016.
- [14] A. Kse and B. Zbek, "Resource allocation for underlying device-to-device communications using maximal independent sets and knapsack algorithm," in *Proc. IEEE 29th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Bologna, Italy, pp. 1-5, Sep. 2018.
- [15] Z. Li and C. Guo, "Multi-agent deep reinforcement learning based spectrum allocation for d2d underlay communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1828-1840, Feb. 2020.
- [16] I. G. Fraimis and S. A. Kotsopoulos, "Qos-based proportional fair allocation algorithm for ofdma wireless cellular systems," *IEEE Commun. Lett.*, vol. 15, no. 10, pp. 1091-1093, Oct. 2011.
- [17] J. Puchinger, G. Raidl, and U. Pferschy, "The multidimensional knapsack problem: structure and algorithms," *INFORMS Journal on Computing*, vol. 22, no. 2, pp. 250-265, Apr. 2010.
- [18] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer networks," *Advances in Neural Information Processing Systems*, pp. 2692-2700, Dec. 2015.
- [19] I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, "Neural combinatorial optimization with reinforcement learning," *arXiv preprint arXiv:1611.09940*.
- [20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computations*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
- [21] R. Williams, "Simple statistical gradient following algorithms for connectionist reinforcement learning," *Machine Learning*, pp. 229-256, May 1992.
- [22] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *3rd International Conference for Learning Representations, ICLR 2015*, San Diego, CA, USA, May 2015.
- [23] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Computers and Operations Research*, vol. 13, no. 5, pp. 533-549, May 1986.
- [24] S. Hanafi and A. Freville, "An efficient tabu search approach for the 0-1 multidimensional knapsack problem," *European Journal of Operational Research*, vol. 106, no. 2-3, pp. 659-675, Apr. 1998.