Power Optimization in Device-to-Device Communications: A Deep Reinforcement Learning Approach With Dynamic Reward

Jiyun Ha and Jamie Shroufe

Table of Contents

Problem Introduction

Device-to-Device (D2D) Communication

⬇

Used to improve system capacity (latency, congestion) and energy efficiency (EE) in cellular networks

Problem Introduction

Downsides of D2D communication:

- Limited network lifetime
- In the absence of base stations -> interference management becomes difficult
- Difficult to optimize energy efficiency while satisfying
  - Throughput constraints
  - Quality of Service (QoS) for D2D pairs and users in underlay D2D network

Problem Summary

How can we improve D2D communication to maximize EE while ensuring the throughput constraints are satisfied?

Previous Literature

Energy-Efficient Power Allocation in OFDMA D2D Communication by Multiobjective Optimization:

- Joint optimization of data rate and EE helps to obtain optimal transit power and ensure high EE in OFDMA based D2D communication

M. R. Mili, P. Tehrani, and M. Bennis, "Energy-efficient power allocation in OFDMA D2D communication by multiobjective optimization," IEEE Wireless. Commun. Lett., vol. 5, no. 6, pp. 668–671, Dec. 2016.

Previous Literature

Coalition Formation Game for Green Resource Management in D2D Communications:

- Joint mode selection and spectrum sharing is modeled as a "coalition formation game"

H. Chen, D. Wu, and Y. Cai, "Coalition formation game for green resource management in D2D communications," IEEE Commun. Lett., vol. 18, no. 8, pp. 1395–1398, Aug. 2014.

Previous Literature

Unrealistic for fast-changing systems as the configuration of D2D users (i.e. position, transmission power, channel gain) are dynamic not static

Previous Literature

- These three papers outlined that deep RL has improved the quality of service (QoS) and minimized improvements for E2E, D2D, and V2V communication
- However, for D2D, the previous approach is not applicable for large-scale network as it lacks tradeoff between EE and throughput -> reduced lifetime of D2D users

Z. Qin, H. Ye, G. Y. Li, and B. F. Juang, "Deep learning in physical layer communications," IEEE Wireless Commun., vol. 26, no. 2, pp. 93–99, Apr. 2019.

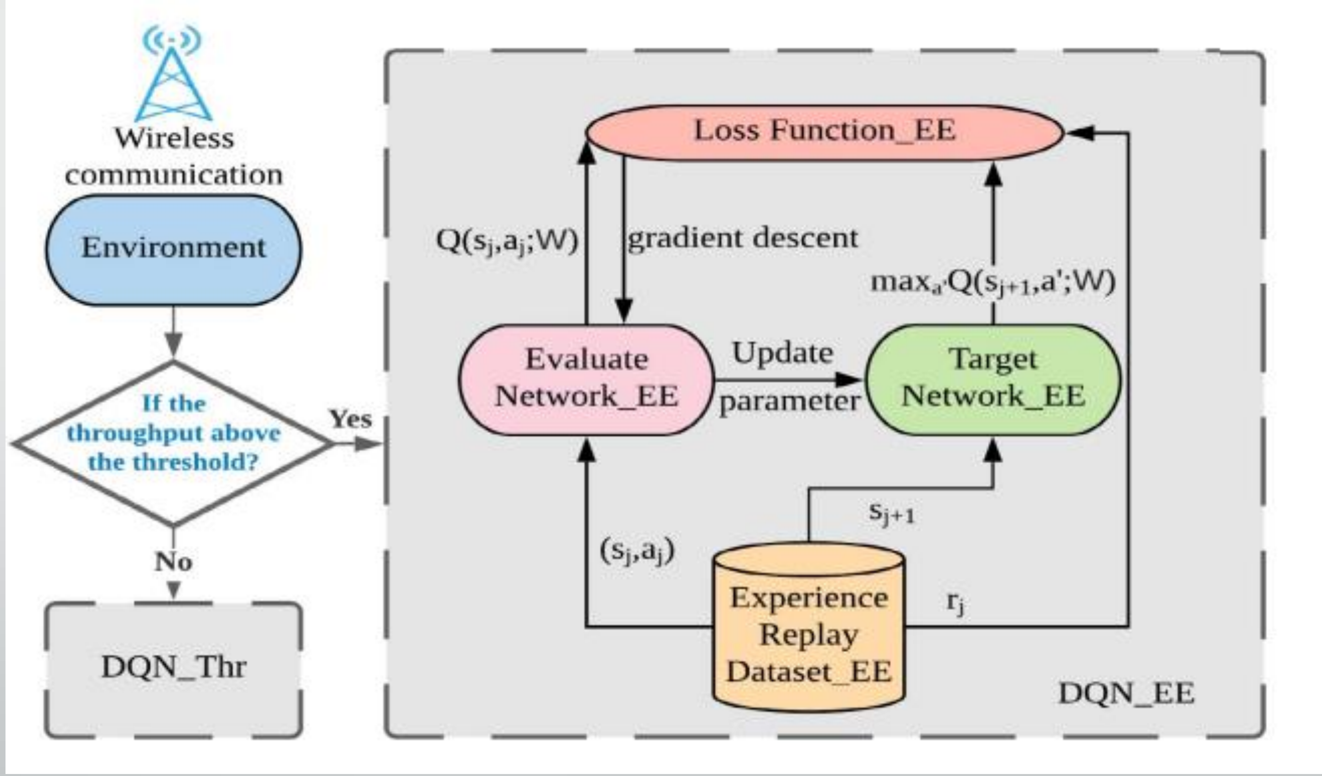
H. Ye, G. Y. Li, and B. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," IEEE Trans. Veh. Technol., vol. 68, no. 4, pp. 3163–3173, Apr. 2019.

L. Liang, H. Ye, G. Yu, and G. Y. Li, "Deep-learning-based wireless resource allocation with application to vehicular networks," Proc. IEEE, vol. 108, no. 2, pp. 341–356, Feb. 2020.

Proposed Solution

- A deep RL algorithm with two parallel DQNs (deep Q network), one for throughput and one for EE
- Chooses between the two according to their current state and which would optimize throughput and EE (a literal comparison tradeoff)
- Algorithm provides flexibility in different scenarios as it converges to optimal EE under different constraints (SINR, max power limits, QoS)

Proposed Solution



Proposed Solution

Key Algorithm Components:

- **State (S)**: Current values of SINR and transmission power of D2D users and BS
- **Action (A)**: Adjusts the power levels for D2D users, modifying the transmission power by a small step each time
- **Reward (R)**:
  - **EE-Oriented Reward**: When throughput is above the threshold, focus on maximizing EE
  - **Throughput-Oriented Reward**: When throughput is below the threshold focus on increasing throughput
  - Violation of Constraints -> Penalize

Proposed Solution

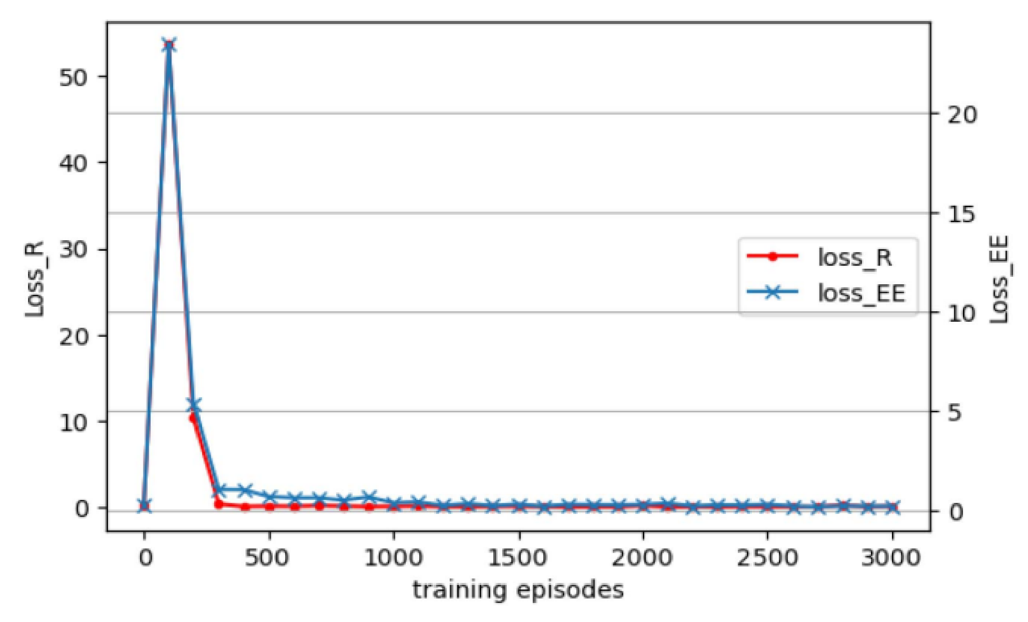
Dynamic Decision Making:

- **DQN_EE**: Focuses on maximizing EE
- **DQN_Thr**: Focuses on maintaining throughput

Training Process:

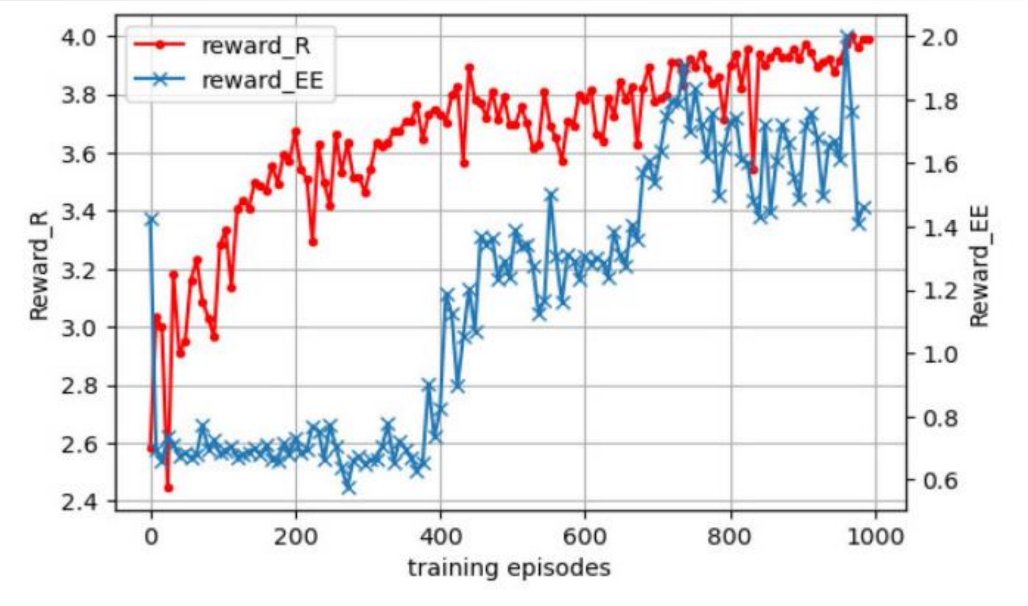
1. Observe the current state
2. Select an action adjusting the power levels
3. Given new state, a reward is provided
4. Stores the transitions in a replay buffer

Training Loss



- Initial rapid decrease
- Convergence: rapid decline suggests the algorithm learns optimal policy quickly
- Stabilization: loss remains close to zero

Training Reward



- **Throughput Reward**: Rapidly increases, gradually improves, stabilizes around ~3.9-4.0 at optimal level
- **EE Reward**: Slow increase, rapid improvement, stabilizes around ~1.8-2.0 at optimal balance
- Both stabilize by the end of training indicating the algorithm has learned the policy that balances throughput and EE

Recommendations for Future Work

- Considering more QoS metrics such as application performance/quality and service reliability (accounting for more user satisfaction)
- Extending beyond 5G systems to even more dense networks
- Comparing how this algorithm performs in urban vs. more rural areas
- Real-world testing

Conclusion

- **Problem**: How can we improve D2D communication to maximize EE while ensuring the throughput constraints are satisfied?
- **Solution**: A deep RL algorithm with parallel DQNs with dynamic reward mechanisms balancing EE and system throughputs
- **Results**: Demonstrated optimal EE and effective handling of the tradeoff between EE and throughput

Thank You.