Beam-Selection MiniGrid Environment for 5G and 6G Networks Applications

Rebecca Aben-Athar, Carnot Braun, Ingrid Nascimento, João Borges and Aldebaro Klautau

***Abstract*— Applications of Reinforcement Learning (RL) are increasingly relevant in the context of future networks. However, although there is a great availability of environments for other research areas such as computer vision, there is a lack of good ones for telecommunication purposes. This paper proposes a Beam-Selection MiniGrid, based on Minimalist Gridworld environment.**

***Keywords*— Reinforcement learning, 5G, 6G, beam-selection.**

I. INTRODUCTION

Applications of Reinforcement Learning (RL) are increasingly relevant in the context of the 5th Generation (5G) and 6th Generation (6G) of mobile networks, being used in several areas, such as resource allocation [1], network slicing [2], congestion control [3] and the 5G physical layer (PHY) [4]. However, although the frameworks to apply this method are numerous, to the best of the authors' knowledge, all lack complete coverage of potentially useful features and environments. In this work, we focus our investigation on the Beam-Selection MiniGrid environment, which is based on a version of RL Minimalist Gridworld [5] scenario.

To understand this area and how RL can contribute, we must first look to the radio of 5G communications systems, which leverage the weakly explored spectrum range of the millimeter waves (mmWaves) to serve its users. As mmWave propagation is more prone to fading and blockage, it depends on massive Multiple Input Multiple Output (MIMO) techniques, such as *beamforming*, to produce directional beams, to propagate the antenna reach more efficiently. The 5G Base Station (BS) task then is twofold, to keep track of the valid beams and to choose the most appropriate one in a process called *beam-selection*, that consists of selecting a beam-pair for BS and User Equipment (UE) to achieve the best communication channel in any given time.

For that, the BS needs to be aware of its surroundings and the user locations, which can be hard to achieve, specially given that 5G networks will enable not only traditional ground users, such as pedestrians, but also aerial vehicles as well. So, in this context, the site covered by the BS becomes increasingly dynamic, and the process of beam-selection gets proportionally more complex. Traditional ways of gathering information about the environment relies on the transmission of pilot tones, but this can lead to an increase in the network overhead, as the BS needs to use the available radio resources to transmit the tones. So, the nature of the task, in which the BS is an agent that must take optimal actions, in other words, choose the best beams, in an complex and dynamic environment, suggests that RL is indeed a relevant method to be investigated, and so, having meaningful environments to help the researcher apply this method is of great importance and the main contribution of this paper is to present the methodology for Beam-Selection Minigrid environment based on the Minimalist Gridworld scenario.

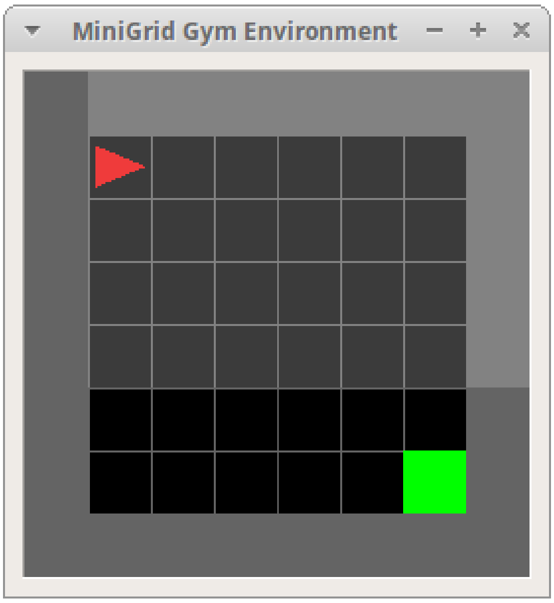
This work was supported by the Innovation Center, CNPq, Capes Foundation, Brazil. Emails: (rebecca.athar, carnot.filho, ingrid.nascimento, joao.tavares.borges) @itec.ufpa.br, aldebaro@ufpa.br

II. ENVIRONMENTS

A. MiniGrid environment

The Minimalist Gridworld Environment (MiniGrid) [5], Figure 1, is a OpenAI Gym based environment, designed to represent the classic Gridworld scenario. Originally, the state space is represented by the position of RL agent in the grid and action space represents 4 directions the agent can move inside the grid. Each movement made by the agent is rewarded with 0 or a negative value depending on how it is implemented. When agent finds the final stage it is rewarded with maximum value, this value is decrease by the number of steps the agent takes to find the terminal stage. The episode ends when final stage is encountered.

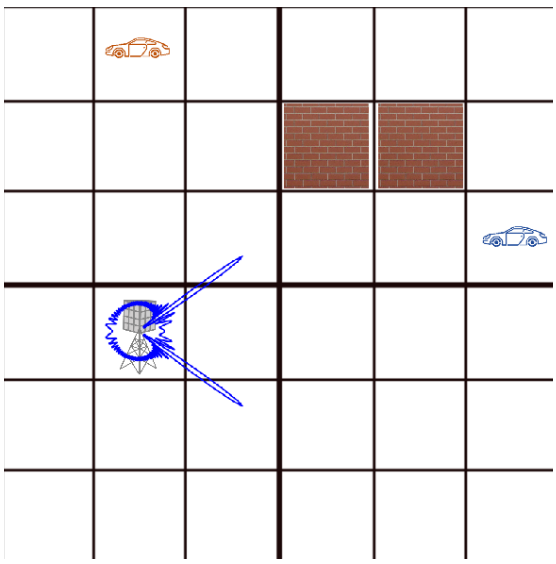
Fig. 1
MINIGRID 5X5 EMPTY ROOM



B. Beam-Selection MiniGrid environment

We could pose the beam selection problem as a minigrid task, in which the RL agent is executed at a BS with an antenna array and serves single-antenna users on downlink using an analog MIMO architecture with N_b beam vector indices. Figure 2 shows an implementation of such problem, where the BS and UEs live in a $M \times M$ Gridworld in which there are M^2 invariant channels depending only on position. An episode lasts N_e time slots, and for each episode, a user moves left, right, up or down. The states would be defined as the users' positions and a list of $N_a - 1$ indices of the previously scheduled users, where N_a represents the number of possible actions. All in a state space dimension of $M^{2N_u} \times (N_u)^{N_a - 1}$, where N_u is the number of users. The action would be the scheduling of one user among the N_u users and the choice if one beam, among the N_b beams, to serve it. The action space dimension would be $N_u \times N_b$, with N_b being the number of possible beams. Finally, in the end of the episode, the reward would be the normalized throughput. In other words, before the episode end, it is -100 if a user is not allocated for N_a consecutive slots, and zero otherwise. The final return would be the sum of rewards.

Fig. 2
EXAMPLE OF THE PROPOSED MIMO BEAM-SELECTION MINIGRID ENVIRONMENT



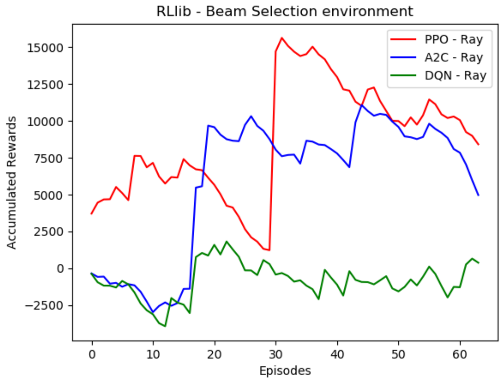
III. EXPERIMENTS ON BEAM-SELECTION MINIGRID USING RL ALGORITHMS

The Beam-Selection MiniGrid represents a more complex environment formatted as a 6x6 grid with a base station that has 32 antennas to attend 2 users in this experiment. Also, all users must be allocated at least once in an episode duration of 30 timesteps. In general, the size of action space is 64 and the number of possible states is 5184 due to the combination of users positions inside the grid as explained in previous section.

In this secction we present an more general experiment using the RL library RLlib [6], in order to show the performance per agents in the environment.

RLlib agents faced challenges for convergence in this environment as can be seen in Figure 3. This analysis shows a narrow difference between PPO, DQN and A2C algorithms in the successful choice of the beam index and allocation of users from half of the simulation to the end.

Fig. 3
RLLIB - ACCUMULATED REWARD X EPISODES



IV. CONCLUSION

The presented results indicate that the MiniGrid environment can be used as the baseline for the implementation of 5G and 6G applications, such as the Beam-Selection MiniGrid environment. In addition, we compare the performances of state-of-art algorithmos implementation, using the library RLlib, where the agent is represented by the BS.

REFERENCES

[1] X. Wang and T. Zhang, "Reinforcement learning based resource allocation for network slicing in 5G C-RAN," in *2019 Computing, Communications and IoT Applications (ComComAp)*, 2019, pp. 106–111.

[2] Y. Kim and H. Lim, "Multi-agent reinforcement learning-based resource management for end-to-end network slicing," *IEEE Access*, vol. 9, pp. 56 178–56 190, 2021.

[3] I. Nascimento, R. Souza, S. Lins, A. Silva, and A. Klautau, "Deep Reinforcement Learning Applied to Congestion Control in Fronthaul Networks," *Proceedings - 2019 IEEE Latin-American Conference on Communications, LATINCOM 2019*, pp. 1–6, 2019.

[4] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, and R. W. Heath, "5g mimo data for machine learning: Application to beam-selection using deep learning," in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–9.

[5] M. Chevalier-Boisvert, L. Willems, and S. Pal, "Minimalistic gridworld environment for openai gym," https://github.com/maximecb/gym-minigrid, 2018.

[6] "Raylib is a Simple and Easy-to-Use Library to Enjoy Videogames Programming." https://www.raylib.com/, accessed: 2021-09-08.