Integrating GAN-Generated Goals into Hindsight Experience Replay in Multi-Goal RL

Berkay Alp Cakal, Tomas Ruiz

May 2020

Objective: A hard problem for Reinforcement Learning (RL) agents in environments with sparse rewards is that the agent initially never reaches its goal with random exploration, and therefore, does not receive any learning signal. One solution to tackle this problem is Hindsight Experience Replay (HER) [And+17] where target goals are substituted by relatively 'easier' virtual goals. However, HER does not measure nor select the most useful goals in terms of its difficulty to the agent. In their work, [Flo+17a] generate goals of the right difficulty using a GAN. Our project will implement their goal generation scheme and integrate it into the HER learning workflow. This is promising because it could potentially improve the sample efficiency and reduce the experience required to train the RL policy.

Related Work: HER substitutes the target goal with a virtual goal that is chosen in a heuristic random manner. This process does not tell apart easy goals from difficult ones. In the mentioned work by [Flo+17a], they generate goals that are just difficult enough for the agent to be useful training goals using a GAN. They formalize the concept of difficulty and estimate it based on the agent's rate of success trying to reach the goal. For every episode, the GAN generates a difficult-enough-goal to train the policy on. The performance of the policy helps labeling the generated goals in terms of difficulty. The GAN is updated at the end of the episode to take into account this difficulty information. Other approaches to select or generate useful goals are e.g. G-HER [Bai+19], where they use a pre-trained RNN network to generate intermediate goals. In [Ren+19], they use a Wasserstein metric to bias the intermediate goals towards the target goal they seek to achieve. Yet another different approach, by [Flo+17b] is to generate starting states close to the goal, instead of goals close to the starting state.

Technical Outline We will apply the idea that goals should be generated optimally with just the right difficulty from [Flo+17a] to HER framework proposed by [And+17] for increasing the learning performance of the current agent. This application is proposed as a future work in [Flo+17a]. Concretely, all target goals will be generated by the GAN, and their difficulty will be assessed and tuned dynamically. The policy training machinery of [Flo+17a] (on-policy TRPO) will replaced with that of HER (off-policy DDPG) and the hinsight mechanism: In the replay buffer, we will include trajectories that substitute virtual goals for the GAN-generated goals. Let's call this combination HER+GAN. We expect to see an improvement in sample efficiency compared to HER, since we are tuning the difficulty of the target goals, as well as compared to the work of [Flo+17a], because the policy will learn from failed attempts as well. We will perform a benchmark of HER+GAN against ablations using HER only and the GAN-generated goals only.

References

[And+17] Marcin Andrychowicz et al. *Hindsight Experience Replay*. 2017. arXiv: 1707.01495 [cs.LG].

[Flo+17a] Carlos Florensa et al. *Automatic Goal Generation for Reinforcement Learning Agents*. 2017. arXiv: 1705.06366 [cs.LG].

[Flo+17b] Carlos Florensa et al. *Reverse Curriculum Generation for Reinforcement Learning*. 2017. arXiv: 1707.05300 [cs.AI].

[Bai+19] Chenjia Bai et al. "Guided goal generation for hindsight multi-goal reinforcement learning". In: *Neurocomputing* 359 (2019), pp. 353–367. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2019.06.022. URL: http://www.sciencedirect.com/science/article/pii/S0925231219308495.

[Ren+19] Zhizhou Ren et al. *Exploration via Hindsight Goal Generation*. 2019. arXiv: 1906.04279 [cs.LG].