ADLR Project Proposal Draft

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. Our objective is to investigate how recent papers define, measure and take advantage of the concept of **difficulty** and **biasing** of generated virtual goals to speed up the learning in these multi-goal RL settings. Then we will choose an extension in one of these papers to be combined with HER, and run a training in an environment.

Related Work: In Hindsight Experience Replay (HER) [And+17], they substitute the target goal with a virtual goal, such that the agent receives a learning signal. In [Flo+17a], they formalize the concept of difficulty for goals and generate goals for the agent tuned at the right difficulty for its current performance using GAN network. In G-HER [Bai+19], they used a pre-trained RNN network to generate intermediate goals. In [Ren+19], they use a Wasserstein metric to **bias** new intermediate goals towards the target goal they seek to achieve. A different approach, by [Flo+17b] is to generate starting states close to the goal, instead of goals close to the starting state. This is a reversed but also symmetric formulation of the initial problem.

Technical Outline Multi-task settings in Reinforcement Learning (RL) utilize sparse rewards which leads to the requirement of prohibitive amounts of exploration to receive learning signal. [And+17] states that this problem increases exponentially with respect to environment dimension, and suggests Hindsight Experience Replay (HER) to exploit previous replays by using imaginary goals to tackle the challenge of sparse reward signal. The idea is that after performing an episode, the transitions are stored in a replay buffer and another RL optimization is performed with a mini batch from a replay buffer but this time with new virtual goals. The virtual goals are chosen in a heuristic manner, and needs to be generated optimally with just the right **difficulty** for the performance of the current agents. [Flo+17a] utilizes GAN model for this purpose. The idea here is that for every episode, GAN generates a difficult-enough-goal and the policy is updated with this generated goal. The performance of the policy helps labeling the generated goals, and the GAN is updated at the end of the episode so that more difficult goals can be generated in the next episode. In this way, as more episodes are performed, the generated goals help the agent move closer to the actual target goal. However they do not integrate GAN-based goal generation into the HER framework, which is stated as a future work in their paper. One possibility for us to combine these two ideas and implement a training with one of the environments experimented in these papers to make concrete comparisons.

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].