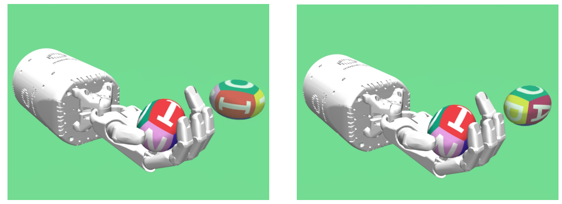
Improving Exploration Efficiency of Single-goal In-hand Manipulation Reinforcement Learning by Progressive Goal Generation

Yingyi Kuang[1], George Vogiatzis[1], and Diego R. Faria[1]

***Abstract*— In recent years, goal-based reinforcement learning framework has become popular for solving robot manipulation tasks, especially when it comes to high dimensional dexterous in-hand manipulation. Hindsight Experience Replay (HER) has shown outstanding performance in solving multi-goal robot manipulation reinforcement learning problems when the reward is provided sparsely. In this paper, we investigate how HER performs when used in single-goal tasks. In addition, we propose a Progressive Goal Generation (PGG) that improves the exploration efficiency when using HER. Several experiments of robot in-hand manipulation have been carried out in simulation. Reported results show that PGG considerably improves the exploration efficiency for in-hand manipulation tasks.**

I. INTRODUCTION

Solving robot manipulation problems utilizing reinforcement learning (RL) has been popular for years [1]. However, dexterous in-hand manipulation using human-like robot hands still remains challenging, mainly due to the high degree of freedom[2]. To assist the research for robot manipulation RL problems, goal-based RL framework has been proposed to reconstruct the observation representation by using a function to extract goals from the state space. Several goal-based robot manipulation environments [3] are provided in simulation platform (e.g. MuJoCo [4]) to help researchers studying such RL problems. In many RL for robot manipulation scenarios, the reward is provided sparsely. Traditional RL algorithms such as DDPG[5] are proved to be inefficient on learning such tasks [3]. Hindsight Experience Replay (HER) method [6] is proposed to improve the sample efficiency for these RL problems. Instead of learning to reach the original goals which could be difficult for the agent, HER changes to learn more from the achieved goals which have been reached previously. It reconstructs the experiences to replace their original goals in the states with the achieved goals and recalculate the rewards. These hindsight experiences largely increase the percentage of positive rewards that agent used for learning, thus accelerating the learning process. HER is originally proposed to help learning multi-goal RL problems where the agent is required to learn goals from a certain range (in goal space). Many approaches have been proposed to improve learning efficiency of HER in multi-goal scenarios [7][8]. In this paper, we investigate how HER performs on single-goal tasks, i.e. the agent only requires to learn a certain target goal. In addition, we propose an automatic goal generation method to improve the exploration efficiency of HER for single-goal tasks. We evaluate our method with two single-goal in-hand object re-orientation tasks in a simulation platform, as shown in Figure 1.



(a) HandEggRotateEasy (b) HandEggRotateHard

Fig. 1: Two single-goal RL tasks based on OpenAI Gym robotics hand environment [3]. The goal is to rotate an object (egg shape) within the hand to reach a certain orientation (template displayed at the right side of the image, near the hand).

II. PROPOSED APPROACH

When combining HER with other off-policy algorithms (such as DDOG) for single-goal RL tasks, we found that the learning efficiency is improved comparing with only using the off-policy algorithm. However, there are single-tasks that are still challenging, especially when the state space has high dimension (i.e. in-hand manipulation) and the target goal state is highly different from the initial state. The reason could be lacking of guidance during learning. During the earlier stage of the learning process, since the agent has no knowledge of how to reach the target goal, it keeps randomly exploring around its initial states for a long period of time before it explores further states. The hindsight goals are randomly selected from the past trajectories by HER and are distributed evenly around the initial states. They do not progress towards the original target goal.

In this work, we considered the design of a curriculum learning approach that proposes goals for the agent, which allow it to learn the target goal step-by-step. It first learns how to reach easier goals, then gradually approaches towards more complex goals for finally being able to achieve the target goal. The goals proposed by the curriculum need to be chosen between the initial state and the target goal state. They also need to have a proper level of difficulty for the current agent. Inspired by [9], we propose a Progressive Goal Generation (PGG) used for single-goal RL problems that generates goals for the agent from its previous experiences.

All authors are with [1]Aston Robotics, Vision and Intelligent Systems Lab (ARVIS), Department of Computer Science, School of Engineering and Applied Science, Aston University, Birmingham, UK. Emails: {kuan1801; g.vogiatzis; d.faria}@aston.ac.uk.

This work is supported by EPSRC-UK InDex project (EU CHIST-ERA programme), with reference EP/S032355/1.

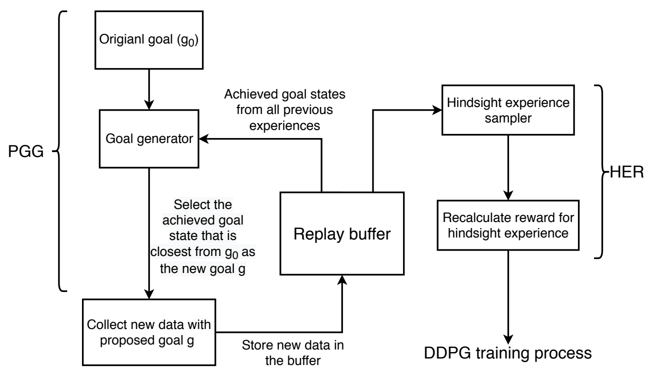


Fig. 2: The proposed Progressive Goal Generation (PGG) approach with HER and DDPG

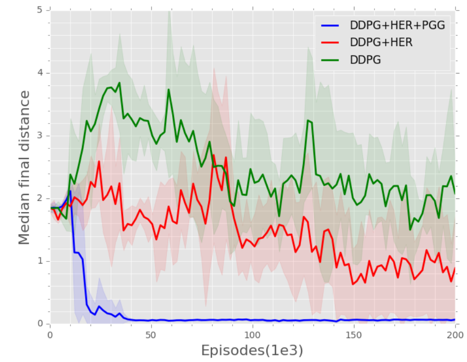
We use the normalized difference of the two goal vectors to represent the distance between two goals. Among all achieved goals of the agent's past trajectories, PGG simply selects the achieved goal that has the minimal distance to the original target goal, which will be the new target goal for the next data collection. By applying this approach, the chosen target goals are slowly moving from the initial state towards the final target goal during the learning process, which always lie in the inter-mediate range between the initial state and the target goal. The whole process of PGG+HER is shown in Figure 2.

III. EXPERIMENTAL RESULTS

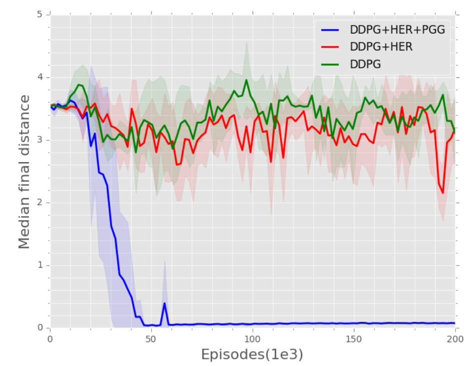
Our experiment focuses on two single-goal in-hand re-orientation tasks, *HandEggRotateFixedTaskEasy* and *HandEggRotateFixedTaskHard*. These tasks are created based on the OpenAI Gym shadow hand RL environments. The goals of these two tasks are to control the joints of a human-like shadow robot hand to rotate an egg shaped object within hand to reach a certain orientation. The initial states of these two tasks are the same. For the task *HandEggRotateFixedTaskEasy*, the orientation of the target goal is around $\pi/2$ degree difference relative to the initial orientation, while the *HandEggRotateFixedTaskHard* the difference is around π degree. Sparse rewards are used in these tasks.

Three different RL approaches are implemented and compared. The pure DDPG method is employed as the baseline. HER is combined with DDPG to verify whether it improves the learning rate. Finally, PGG is combined with HER and DDPG to verify whether it can further accelerate the learning process. Here we use the future strategy with $k = 4$ for HER. The final distance between the original goal and the final achieved goal after the episode ends is calculated during evaluation to measure the performance.

The experiments are conducted with 5 different random seeds. The average final distance with standard deviation for three methods are shown in Figure 3. The task can be considered as learned when the distance is less than 0.1. With DDPG used alone, both two tasks are not learned within the experimental time. When it is combined with HER, the easy rotation task converges faster than DDPG only. However, for the hard rotation task, it does not show any sign of convergence. Both tasks are finally learned after combining with the proposed PGG, and the learning process converges more quickly. Results reveals that DDPG+HER+PGG outperforms the other two learning methods, and PGG significantly improves the learning efficiency for these tasks.



(a) HandEggRotateEasy



(b) HandEggRotateHard

Fig. 3: Comparison of three RL methods on two single-goal in-hand manipulation tasks with sparse reward.

IV. CONCLUSION

In this work we discussed the RL problem for single-goal in-hand manipulation tasks with sparse rewards. We evaluate the traditional DDPG algorithm as well as the recent HER algorithm. We proposed a curriculum-based learning approach called Progressive Goal Generation (PGG) to improve the exploration efficiency for such a problem. Experimental results show that PGG significantly improves the learning efficiency when combining with DDPG and HER. Further work will be carried out to evaluate PGG on other robot manipulation tasks beyond in-hand manipulation. We also intend to deeply investigate the combination of DDPG+PGG without HER. Moreover, it can be explored to use the proposed distance as the dense reward for learning and confirm whether PGG benefits the learning process on dense reward cases.

REFERENCES

[1] H. Nguyen and H. La, “Review of deep reinforcement learning for robot manipulation,” in *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE, 2019, pp. 590–595.

[2] M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray *et al.*, “Learning dexterous in-hand manipulation,” *arXiv preprint arXiv:1808.00177*, 2018.

[3] M. Plappert, M. Andrychowicz, A. Ray, B. McGrew, B. Baker, G. Powell, J. Schneider, J. Tobin, M. Chociej, P. Welinder *et al.*, “Multi-goal reinforcement learning: Challenging robotics environments and request for research,” *arXiv preprint arXiv:1802.09464*, 2018.

[4] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5026–5033.

[5] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *2016 International Conference on Learning Representations (ICLR)*, 2016.

[6] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. P. Abbeel, and W. Zaremba, “Hindsight experience replay,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5048–5058.

[7] R. Zhao, X. Sun, and V. Tresp, “Maximum entropy-regularized multi-goal reinforcement learning,” in *International Conference on Machine Learning*, 2019, pp. 7553–7562.

[8] Y. Kuang, A. I. Weinberg, G. Vogiatzis, and D. R. Faria, “Goal density-based hindsight experience prioritization for multi-goal robot manipulation reinforcement learning,” in *2020 29th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020.

[9] Z. Ren, K. Dong, Y. Zhou, Q. Liu, and J. Peng, “Exploration via hindsight goal generation,” in *Advances in Neural Information Processing Systems*, 2019, pp. 13 464–13 474.