

Artificial Intelligence Project Proposal

Hierarchical Reinforcement Learning for Sparse Rewards

Zhihao Jiang (2016011273), Yiheng Lin (2016012392)

1 Problem

Sparse reward is a fundamental challenging problem for RL. Hierarchical exploration approaches learn to select subgoals and how to achieve subgoals, which seems helpful for reinforcement learning with sparse rewards. There are several papers showing that hierarchical exploration allows more quickly to explore regions far away than basic ϵ -greedy exploration.

Some problems are remained. For example, why hierarchical exploration performs better than ϵ -greedy exploration? Under what conditions can we guarantee their privileges in performance? Is there any other exploration method performing better?

2 Existing Works

In lecture, we studied DRL[2, 3]. It performs not good in the environments with sparse feedbacks, such as Montezuma's revenge.

(Kulkarni et al.)[1] proposed a two layer hierarchical framework for reinforcement learning in the environments where the feedbacks are sparse. The higher hierarchy (Meta Controller) is responsible for defining goals for the lower hierarchy (Controller). The lower hierarchy try to realize the goals use RL algorithms. The paper also uses deep learning methods (convolution) to train the controllers to realize their goals.

(Vezhnevets et al.)[5] generalized federal reinforcement learning framework so that it can be used under more settings. He also proposed a new method (transition policy gradient) to train the Manager model. The manager model (higher hierarchy) use extrinsic rewards as input to generate intrinsic rewards for worker model (lower hierarchy). And dilated LSTM is used when training Manager Model since long term memory is preferred for the Manager Model.

(Nachum et al.)[4] pointed out that the prior works depended too much on specific settings. For example, in [1], we need to build a custom object detector that provides plausible object candidates manually. So they aim to provide a more general model that can fit in more complex settings. In contrast to previous works that needs to set the goal based on the model, the authors use states as goals directly, such that the lower hierarchy can be trained faster and the model can fit in different settings more easily. The author also establishes the concept of off-policy.

3 Limitation

The main limitation of these works on HRL are they lack theoretical analysis on why the proposed framework works better than ϵ -greedy RL. They provide a lot of settings under which their HRL methods can outperform other models. But the question they still need to answer is under what specific conditions can we gaurantee that HRL is better than traditional ϵ -greedy RL.

4 Objectives

We hope to achieve one or more goals below.

- Reimplement or design a hierarchical exploration algorithm which performs good under specific environment.
- Improve hierarchical exploration algorithm mentions in papers.
- Explain why hierarchical exploration performs better than ϵ -greedy exploration under some conditions.
- Design a learning algorithm in environments with sparse feedbacks, which performs better than previous works of hierarchical exploration.

5 Approach

We will try to reimplement the result of the papers to test the stability and advantage of proposed frameworks under some specific settings. Then we will try to figure out the shared features of settings under which HRL can outperform ϵ -greedy RL and make theoretical proofs. After that, we will use experiments to test the conclusions we get.

References

- [1] Tejas D. Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 3675–3683, 2016.
- [2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.
- [3] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin A. Riedmiller, Andreas Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- [4] Ofir Nachum, Shixiang Gu, Honglak Lee, and Sergey Levine. Data-efficient hierarchical reinforcement learning. *CoRR*, abs/1805.08296, 2018.
- [5] Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. Feudal networks for hierarchical reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 3540–3549, 2017.