Simultaneous State-Dependent Updating of multi-options in Flexible Option Learning

Jerome Francis*
Tata Consultancy Services (TCS)
New Delhi, India
francis.jerome999@gmail.com

Abstract

Within intra-option learning, Flexible Option Learning was formulated to enable simultaneous update of all option-policies which was consistent with their primitive action choices. However, updating all options without any goal-related heuristic reduces the degree of diversity of options within the option set, which is a major drawback. We revisit and extend Flexible Option learning to introduce a state-dependent multi-option update method to add more flexibility to the way multi-option updates can be performed in the context of deep reinforcement learning. Our method utilizes the concept of distance as a goal-related heuristic to generalize the multi-option updates in different important transitions like bottleneck situations in environment.

1 Introduction

Temporal abstraction in reinforcement learning (RL) allows better knowledge transfer and explicit reasoning at different temporal situations, which makes it an important part of an intelligent agent. Flexible Option Learning Klissarov and Precup (2021)revisited the intra-option learning to include simultaneous multi-option updates in function approximation setting, which can be naturally incorporated in hierarchical RL frameworks. However, the method promoted a multi-option update in every situation which was controlled by a hyperparameter. This restricts the degree of diversity in an option set and the way we may want to update options or a group of options.

In this method, we propose a state-dependent function to control the multi-option updates in certain situations noticed by a goal-related heuristic; which is based on the simple concept of distance between the goal and the current state. We also study this flexibility over multi-option updates in an tabular setting over the Option-Critic architecture by investigating the state-option pairs where the agent usually promotes multi-option updates, which can denote places of important transition like bottleneck states, initial state with an new and different goal, etc.

2 Method

We follow all the notations and the assumptions for MDPs and options as mentioned in the Flexible Option Learning Klissarov and Precup (2021). In this work, we introduce a state-dependent function $\eta(s)$ which is based on the concept of distance $Distance$ from the current state **s** in an environment till the goal G. Hence, $\eta(s)$ can be written as follows,

$$\eta(s) = \frac{1}{Distance(s, G)} \tag{1}$$

*Portfolio : https://jerryfrancis-97.github.io/

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

where $Distance$ can be Euclidean distance or the Manhattan (Cityblock) distance. Therefore,

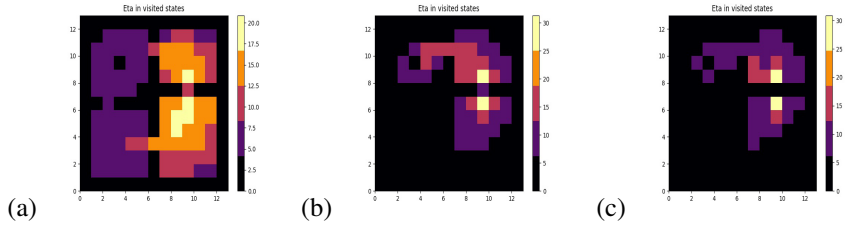
$$Distance = Euclidean(X, Y) = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \tag{2}$$

and

$$Distance = Cityblock(X, Y) = Abs(X_i - X_j) + Abs(Y_i - Y_j) \tag{3}$$

3 Experiments & Discussion

For our experiments of the FourRooms domain we based our implementation on [Bacon et al., 2016] and ran the experiments for 100 episodes that lasted a maximum of 1000 steps with goal located in the right hallway. Here, all the options' parameters are learned from scratch. The result are then averaged over 10 independent runs with use of exponential smoothing $\alpha = 0.8$. Random seeds to augment the randomness of the environment the agent receives.



(a) (b) (c)

Figure 1: **Option Visualization** : Representation of $\eta(s)$ in FourRooms environment. In a) we have $\eta = 0.3$, where a lot of activity takes place in the right hallway, but doesn't traverse to left-side rooms. In contrast to a), b) and c) figures where η is based on Euclidean and Cityblock distance respectively,show some directed exploration to left-side rooms along with choosing important states near right hallway to traverse to other rooms.

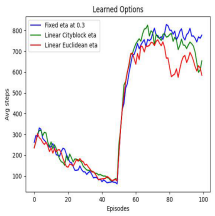


Figure 2: **FourRooms Domain**: In this graph, we can see that after changing the goal, Euclidean based state-dependent function (in red) shows significant improved performance (reduction in steps for completion) in around 25 episodes, when compared with Cityblock based function (in green) or fixed η (in blue).

4 Conclusion

The presented work identifies to generalize the flexibility in which diversity can be changed according to the situation like important state transitions, changes in goal.task, etc. Some future works can be to incorporated other skill diversity methods Eysenbach et al. (2018)in this framework.

References

Eysenbach, B., Gupta, A., Ibarz, J., Levine, S., 2018. Diversity is all you need: Learning skills without a reward function. URL: https://arxiv.org/abs/1802.06070, doi:10.48550/ARXIV.1802.06070.

Klissarov, M., Precup, D., 2021. Flexible option learning. URL: https://arxiv.org/abs/2112.03097, doi:10.48550/ARXIV.2112.03097.