Supplementary Material

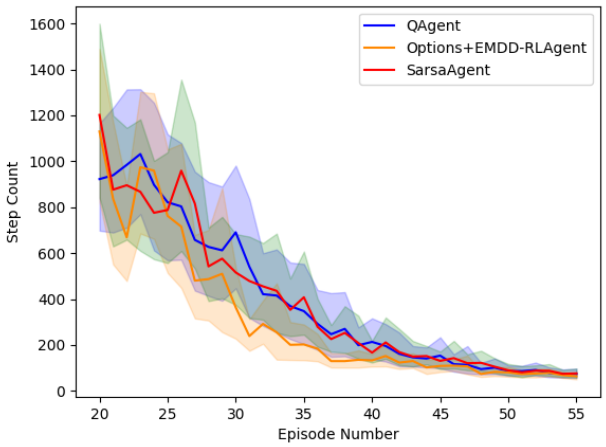

Figure 1: The total number of steps that an agent has taken to reach the goal state per episode is shown for every learning method. The average step value is obtained from 40 experiments in the two-rooms environment. The shaded areas represent the 95% bootstrap confidence interval.

1. Option Creation with EMDD-RL

In order to show the usefulness of subgoals generated by EMDD-RL, we have combined EMDD-RL with the *options* framework and compared their combination with agents employing Q-learning and Sarsa algorithms in the two-rooms environment. We have followed the experimental setup presented by the work [1]. Only one option is permitted for the environment. The agent begins seeking a subgoal after gathering 20 episodes (all have been labeled as positive, and no negative bag is created). Afterward, it updates the average value of detected subgoal candidates during upcoming episodes. If the average value of any candidate exceeds the threshold value (λ, we have set it as 0.6), it is picked as a subgoal. The option is created by inspecting the agent's trajectory history. After option creation, the agent switches from Q-learning to intra-option value learning [2]. The agent starts every episode from the top-left state (state 0), and the states around the starting and goal states have been excluded from EMDD-RL calculations. The same action noise and Q-learning hyperparameter values in the speed experiment are considered for the Q-learning and Sarsa agents.

In Figure 1 (which is very similar to Figure 3 of [1]), the agent creates an option between episodes 27 and 44. The figure shows that the option created helps the agent reach the goal state with fewer steps than the Q and Sarsa agents after the 27^{th} episode till approximately the 50^{th} episode. As a result, EMDD-RL can be employed by combining it with a decomposition method in order to obtain accelerated learning performance in RL problems.

References

[1] McGovern, A., Barto, A.G., 2001. Automatic discovery of subgoals in reinforcement learning using diverse density, in: Proceedings of the Eighteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. p. 361–368.

[2] Sutton, R.S., Precup, D., Singh, S., 1999. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. Artif. Intell. 112, 181–211. URL: https://doi.org/10.1016/S0004-3702(99)00052-1, doi:10.1016/S0004-3702(99)00052-1.