Reward-Respecting Subtasks for Model-Based Reinforcement Learning (Abstract Reprint)

Richard S. Sutton[1,2,3,4], **Marlos C. Machado**[1,2,3,4], **G. Zacharias Holland**[1], **David Szepesvari**[1], **Finbarr Timbers**[1], **Brian Tanner**[1], **Adam White**[1,2,3,4]

[1]DeepMind, Edmonton, Alberta, Canada
[2]University of Alberta, Edmonton, Alberta, Canada
[3]Alberta Machine Intelligence Institute (Amii), Edmonton, Alberta, Canada
[4]Canada CIFAR AI Chair, Canada

Abstract Reprint. This is an abstract reprint of a journal article by Sutton, Machado, Holland, Szepesvari, Timbers, Tanner, and White (2023).

Abstract

To achieve the ambitious goals of artificial intelligence, reinforcement learning must include planning with a model of the world that is abstract in state and time. Deep learning has made progress with state abstraction, but temporal abstraction has rarely been used, despite extensively developed theory based on the options framework. One reason for this is that the space of possible options is immense, and the methods previously proposed for option discovery do not take into account how the option models will be used in planning. Options are typically discovered by posing subsidiary tasks, such as reaching a bottleneck state or maximizing the cumulative sum of a sensory signal other than reward. Each subtask is solved to produce an option, and then a model of the option is learned and made available to the planning process. In most previous work, the subtasks ignore the reward on the original problem, whereas we propose subtasks that use the original reward plus a bonus based on a feature of the state at the time the option terminates. We show that option models obtained from such reward-respecting subtasks are much more likely to be useful in planning than eigenoptions, shortest path options based on bottleneck states, or reward-respecting options generated by the option-critic. Reward respecting subtasks strongly constrain the space of options and thereby also provide a partial solution to the problem of option discovery. Finally, we show how values, policies, options, and models can all be learned online and off-policy using standard algorithms and general value functions.

References

Sutton, R. S.; Machado, M. C.; Holland, G. Z.; Szepesvari, D.; Timbers, F.; Tanner, B.; and White, A. 2023. Reward-respecting subtasks for model-based reinforcement learning. *Artificial Intelligence*, 324: 104001.