Option2Vec: Learning Temporal-State Abstraction Embeddings on the MDP

Chang Li, Dongjin Song, and Dacheng Tao, *Fellow, IEEE*

Abstract—Temporally extended actions refer to actions that last for multiple time steps. The option framework provides a natural scheme to incorporate extended actions into reinforcement learning. Learning an option, however, requires modeling the initiation set, intra-option policy, and termination condition, respectively, which is sub-optimal for knowledge representation and is computationally expensive. To address these issues, we consider temporal and state abstraction jointly. In particular, we formulate the option framework as a Hidden-Markov-Model-style Probabilistic Graphical Model (PGM), thus enabling each option to be parameterized as a compact embedding vector. To optimize this PGM, we first propose a novel *Markovian Option-Value function*, and prove it is an unbiased estimation of the conventional value function. We then derive a two-stage policy gradient theorem based on the above value function. Finally, we implement this PGM upon the Transformer architecture to encode options into fixed-length embeddings, named Option2Vec. Extensive experiments on challenging *Mujoco* environments demonstrate Option2Vec's efficiency and effectiveness: under widely used configuration, with merely 15.8% parameters, Option2Vec achieves competitive, if not better, performance compared to the state-of-the-art baselines on all finite horizon and transfer learning environments. Moreover, Option2Vec significantly outperforms all baselines on infinite horizon environments while exhibiting smaller variance, faster convergence, and interpretability.

Index Terms—The Option Framework, Hierarchical Reinforcement Learning, Markov Decision Process

1 Introduction

THIS demo file is intended to serve as a "starter file" for IEEE Computer Society journal papers produced under LaTeX using IEEEtran.cls version 1.8b and later. I wish you the best of success.

mds
August 26, 2015

2 Conclusion

The conclusion goes here.

2.1 Subsection Heading Here

Subsection text here.

2.1.1 Subsubsection Heading Here

Subsubsection text here.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgments

The authors would like to thank...

- *Chang Li and Dacheng Tao were with the UBTECH Sydney AI Centre, School of Computer Science, University of Sydney, Sydney, NSW 2000, Australia. E-mail:* chli4934@uni.sydney.edu.au dacheng.tao@sydney.edu.au
- *Dongjin Song University.*

Manuscript received December 19, 2021

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

PLACE PHOTO HERE

Michael Shell Biography text here.

John Doe Biography text here.

Jane Doe Biography text here.