Multi-agent Covering Option Discovery through Kronecker Product of Factor Graphs

Extended Abstract

Jiayu Chen
Purdue University
West Lafayette, United States
chen3686@purdue.edu

Jingdi Chen
The George Washington University
Washington, DC, United States
jingdic@gwu.edu

Tian Lan
The George Washington University
Washington, DC, United States
tlan@gwu.edu

Vaneet Aggarwal
Purdue University
West Lafayette, United States
vaneet@purdue.edu

ABSTRACT

Covering option discovery has been developed to improve the exploration of reinforcement learning in single-agent scenarios with sparse reward signals, through connecting the most distant states in the embedding space provided by the Fiedler vector of the state transition graph. However, these option discovery methods cannot be directly extended to multi-agent scenarios, since the joint state space grows exponentially with the number of agents in the system. Thus, existing researches on adopting options in multi-agent scenarios still rely on single-agent option discovery and fail to directly discover the joint options that can improve the connectivity of the joint state space of agents. In this paper, we show that it is indeed possible to directly compute multi-agent options with collaborative exploratory behaviors among the agents, while still enjoying the ease of decomposition. Our key idea is to approximate the joint state space as a Kronecker graph – the Kronecker product of individual agents' state transition graphs, based on which we can directly estimate the Fiedler vector of the joint state space using the Laplacian spectrum of individual agents' transition graphs. This decomposition enables us to efficiently construct multi-agent joint options by encouraging agents to connect the sub-goal joint states which are corresponding to the minimum or maximum values of the estimated joint Fiedler vector. The evaluation based on multi-agent collaborative tasks shows that the proposed algorithm can successfully identify multi-agent options, and significantly outperforms prior works using single-agent options or no options, in terms of both faster exploration and higher cumulative rewards.

KEYWORDS

Covering Option Discovery; Multi-agent Reinforcement Learning; Kronecker Product

ACM Reference Format:
Jiayu Chen, Jingdi Chen, Tian Lan, and Vaneet Aggarwal. 2022. Multi-agent Covering Option Discovery through Kronecker Product of Factor Graphs: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 INTRODUCTION

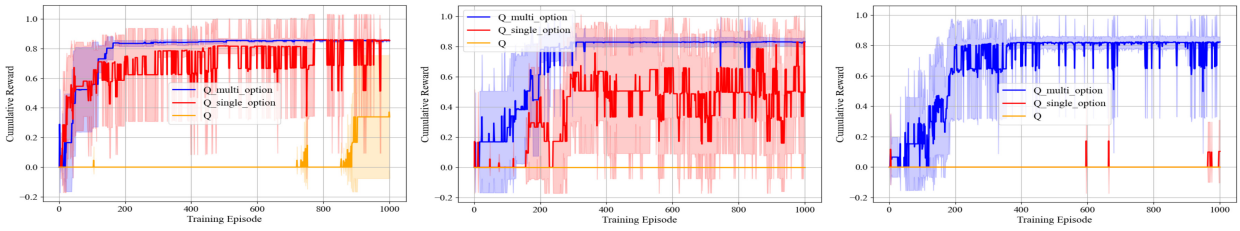
In this paper, we consider the problem of constructing and utilizing covering options in multi-agent reinforcement learning (MARL). Due to the exponentially-large state space in multi-agent scenarios, a commonly-adopted way to solve this problem [1, 2, 4, 8, 9] is to construct the single-agent options as if in a single-agent environment first, and then learn to collectively leverage these individual options to tackle multi-agent tasks. This method fails to consider the coordination among agents in the option discovery process, and thus can suffer from very poor behavior in multi-agent collaborative tasks. To this end, in our work, we propose a framework that makes novel use of Kronecker product of factor graphs to directly construct the multi-agent options in the joint state space, and adopt them to accelerate the joint exploration of agents in MARL. Also, instead of directly adopting the *Covering Option Discovery* to the joint state space since its size grows exponentially with the number of agents, we build multi-agent options based on the individual state transition graphs, making our method much more scalable.

2 BACKGROUND

Kronecker product of graphs [12]: Let $G_1 = (V_{G_1}, E_{G_1})$ and $G_2 = (V_{G_2}, E_{G_2})$ be two state transition graphs, corresponding to the individual state space $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. The Kronecker product of them denoted by $G_1 \otimes G_2$ is a graph defined on the set of vertices $V_{G_1} \times V_{G_2}$, such that: Two vertices of $G_1 \otimes G_2$, namely (g, h) and (g', h'), are adjacent if and only if g and g' are adjacent in G_1 and h and h' are adjacent in G_2. Thus, the Kronecker Product Graph can capture the joint transitions of the agents in their joint state space very well, and we propose to use the Kronecker Product Graph as an effective approximation of the joint state transition graph, so that we can discover the joint options based on the factor graphs.

Covering Option Discovery: As defined in [10], an option ω consists of three components: an intra-option policy $\pi_\omega : \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$, a termination condition $\beta_\omega : \mathcal{S} \rightarrow \{0, 1\}$, and an initiation set $I_\omega \subseteq \mathcal{S}$. An option $< I_\omega, \pi_\omega, \beta_\omega >$ is available in state s if and only if $s \in I_\omega$. If the option ω is taken, actions are selected according to π_ω until ω terminates stochastically according to β_ω (i.e., $\beta_\omega = 1$).

The authors of [7] proposed *Covering Option Discovery* – discovering options by minimizing the upper bound of the expected



(a) four-room with 3 agents (b) four-room with 4 agents (c) four-room with 5 agents

Figure 1: Evaluation on n-agent four-room tasks.

cover time of the state space. First, they compute the Fiedler vector F of the Laplacian matrix L of the state transition graph. Then, they collect the states s_i and s_j with the largest $(F_i - F_j)^2$ (F_i is the i-th element in F), based on which they construct two symmetric options: $\omega_{ij} = < I_{\omega_{ij}} = \{s_i\},\ \pi_{\omega_{ij}},\ \beta_{\omega_{ij}} = \{s_j\} >, \omega_{ji} = < I_{\omega_{ji}} = \{s_j\},\ \pi_{\omega_{ji}},\ \beta_{\omega_{ji}} = \{s_i\} >$ to connect these two subgoal states bidirectionally, where π_ω is defined as the optimal path between the initiation and termination state. This whole process is repeated until they get the required number of options. The intuition of this method is that: $(F_i - F_j)^2$ gives the first order approximation of the increase in $\lambda_2(L)$ (i.e., algebraic connectivity) by connecting (s_i, s_j) [6], and it's empirically proved in [7] that the larger the algebraic connectivity is, the smaller the upper bound of the expected cover time would be and the easier the exploration tends to be.

3 PROPOSED ALGORITHM

Multi-agent Covering Option Discovery: In order to discover the multi-agent options, we need to find the Fiedler vector of the joint state transition graph. Given that the size of the joint state space grows exponentially with the number of agents, we propose to use the Kronecker Product Graph $\otimes_{i=1}^{n} G_i$ as an approximation of the joint state transition graph $\widetilde{G}$ so as to decompose the eigenfunction calculation to single-agent state spaces. Inspired by [3] which proposed an estimation of the Laplacian spectrum of the Kronecker product of two factor graphs, we have the following THEOREM 3.1 (The proof is provided in the full version of our paper [5]).

Theorem 3.1. *For graph $\widetilde{G} = \otimes_{i=1}^{n} G_i$, we can approximate the eigenvalues μ and eigenvectors v of its Laplacian L by:*

$$\mu_{k_1,\ldots,k_n} = \left\{ \left[1 - \prod_{i=1}^{n} (1 - \lambda_{k_i}^{G_i}) \right] \prod_{i=1}^{n} d_{k_i}^{G_i} \right\} \tag{1}$$

$$v_{k_1,\ldots,k_n} = \otimes_{i=1}^{n} v_{k_i}^{G_i} \tag{2}$$

where $\lambda_{k_i}^{G_i}$ and $v_{k_i}^{G_i}$ are the k_i-th smallest eigenvalue and corresponding eigenvector of $\mathcal{L}_{G_i}$ (normalized Laplacian matrix of G_i), and $d_{k_i}^{G_i}$ is the k_i-th smallest diagonal entry of D_{G_i} (degree matrix of G_i).

Through enumerating $(k_1, \cdots, k_n)$, we can collect the eigenvalues of $\otimes_{i=1}^{n} G_i$ by Equation (1) and the corresponding eigenvectors by Equation (2). Then, the eigenvector $v_{\hat{k}_1,\cdots,\hat{k}_n}$ corresponding to the second smallest eigenvalue $\mu_{\hat{k}_1,\cdots,\hat{k}_n}$ is the estimated Fiedler vector of the joint state transition graph, namely $F_{\widetilde{G}}$. Based on it, we can define the joint states corresponding to the maximum or minimum in $F_{\widetilde{G}}$ as the initiation or termination joint states, which can be connected with joint options. Further, consider an MDP with n agents and m states for each agent. To compute the Fiedler vector directly from the joint state transition graph would require time complexity $O(m^{3n})$, since there are in total m^n joint states and the time complexity of eigenvalue decomposition is cubic with the size of the joint state space. While, our solution can significantly reduce the problem complexity from $O(m^{3n})$ to $O(nm^3)$ for multi-agent problems. Also, we would like to point out that for problems with continuous state space (i.e., m is large), our approach could be directly integrated with sample-based techniques for eigenfunction estimation, like [11, 13]. Hence, the bottleneck on computational complexity can be overcome.

Adopting Multi-agent Options in MARL: In order to take advantage of options in the learning process, we adopt a hierarchical algorithm framework: when making decisions, the RL agent first decides on which option ω to use according to the high-level policy, and then decides on the action to take based on the corresponding intra-option policy π_ω. Note that the agent does not decide on a new option with the high-level policy until the current option terminates. The multi-agent options can be adopted either in a decentralized or centralized manner. That is, the agents can either choose their own options independently, or be forced to execute the same multi-agent option simultaneously. The decentralized manner is more flexible but has a larger search space. While, the centralized manner fails to consider all the possible solutions but makes it easier for the agents to visit the sub-goal joint states, since the agents simultaneously select the same joint option which will not terminate until the agents arrive at a sub-goal state.

4 EVALUATION

Please refer to the full version of our paper [5] for the completed evaluation results. We compare our method with: (1) Agents with single-agent options – utilizing single-agent options in MARL, like [1, 2, 4, 8, 9]. (2) Agents without options – directly adopting MARL algorithms. In Figure 1(a)-1(c), we show the results on n-agent four-room tasks (n agents need to reach the goal area simultaneously to complete the task), using Independent Q-learning as the high-level policy. We can observe that the performance improvement brought by our approach are more and more significant as the number of agents increases. When $n = 5$, both the baselines fail to complete the task, while agents with five-agent options can converge within $\sim$ 200 episodes.

REFERENCES

[1] Christopher Amato, George Konidaris, Leslie P Kaelbling, and Jonathan P How. 2019. Modeling and planning with macro-actions in decentralized POMDPs. *Journal of Artificial Intelligence Research* 64 (2019), 817–859.

[2] Christopher Amato, George Dimitri Konidaris, and Leslie Pack Kaelbling. 2014. Planning with macro-actions in decentralized POMDPs. In *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14*. IFAAMAS/ACM, 1273–1280.

[3] Milan Bašić, Branko Arsić, and Zoran Obradović. 2021. Another estimation of Laplacian spectrum of the Kronecker product of graphs. arXiv:2102.02924 [cs.SI]

[4] Jhelum Chakravorty, Patrick Nadeem Ward, Julien Roy, Maxime Chevalier-Boisvert, Sumana Basu, Andrei Lupu, and Doina Precup. 2020. Option-Critic in Cooperative Multi-agent Systems. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '20*. 1792–1794.

[5] Jiayu Chen, Jingdi Chen, Tian Lan, and Vaneet Aggarwal. 2022. Multi-agent Covering Option Discovery based on Kronecker Product of Factor Graphs. arXiv:2201.08227 [cs.MA]

[6] Arpita Ghosh and Stephen Boyd. 2006. Growing Well-connected Graphs.

[7] Yuu Jinnai, Jee Won Park, David Abel, and George Dimitri Konidaris. 2019. Discovering Options for Exploration by Minimizing Cover Time. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 3130–3139. http://proceedings.mlr.press/v97/jinnai19b.html

[8] Youngwoon Lee, Jingyun Yang, and Joseph J. Lim. 2020. Learning to Coordinate Manipulation Skills via Skill Behavior Diversification. In *8th International Conference on Learning Representations, ICLR 2020*. OpenReview.net.

[9] Jing Shen, Guochang Gu, and Haibo Liu. 2006. Multi-agent hierarchical reinforcement learning by integrating options into maxq. In *First international multi-symposiums on computer and computational sciences (IMSCCS'06)*, Vol. 1. IEEE, 676–682.

[10] Richard S. Sutton, Doina Precup, and Satinder P. Singh. 1999. Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning. *Artif. Intell.* 112, 1-2 (1999), 181–211. https://doi.org/10.1016/S0004-3702(99)00052-1

[11] Kaixin Wang, Kuangqi Zhou, Qixin Zhang, Jie Shao, Bryan Hooi, and Jiashi Feng. 2021. Towards Better Laplacian Representation in Reinforcement Learning with Generalized Graph Drawing. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 11003–11012. http://proceedings.mlr.press/v139/wang21ae.html

[12] Paul M Weichsel. 1962. The Kronecker product of graphs. *Proceedings of the American mathematical society* 13, 1 (1962), 47–52.

[13] Yifan Wu, George Tucker, and Ofir Nachum. 2019. The Laplacian in RL: Learning Representations with Efficient Approximations. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. https://openreview.net/forum?id=HJlNpoA5YQ