PORTAL: Automatic Curricula Generation for Multiagent Reinforcement Learning

Extended Abstract

Jizhou Wu
College of Intelligence and Computing, Tianjin University
Tianjin, China
research5@tju.edu.cn

Tianpei Yang*
University of Alberta and Alberta Machine Intelligence Institute
Edmonton, Canada
tpyang@tju.edu.cn

Xiaotian Hao
College of Intelligence and Computing, Tianjin University
Tianjin, China
xiaotianhao@tju.edu.cn

Jianye Hao*
College of Intelligence and Computing, Tianjin University
Tianjin, China
jianye.hao@tju.edu.cn

Yan Zheng
College of Intelligence and Computing, Tianjin University
Tianjin, China
yanzheng@tju.edu.cn

Weixun Wang
College of Intelligence and Computing, Tianjin University
Tianjin, China
wxwang@tju.edu.cn

Matthew E.Taylor
University of Alberta and Alberta Machine Intelligence Institute
Edmonton, Canada
matthew.e.taylor@ualberta.ca

ABSTRACT

Despite many breakthroughs in recent years, it is still hard for MultiAgent Reinforcement Learning (MARL) algorithms to directly solve complex tasks in MultiAgent Systems (MASs) from scratch. In this work, we study how to use Automatic Curriculum Learning (ACL) to reduce the number of environmental interactions required to learn a good policy. In order to solve a difficult task, ACL methods automatically select a sequence of tasks (i.e., curricula). The idea is to obtain maximum learning progress towards the final task by continuously learning on tasks that match the current capabilities of the learners. The key question is how to measure the learning progress of the learner for better curriculum selection. We propose a novel ACL framework, ***PrOgRessive mulTiagent Automatic curricuLum*** (**PORTAL**), for MASs. PORTAL selects curricula according to two criteria: 1) How difficult is a task, relative to the learners' current abilities? 2) How similar is a task, relative to the final task? By learning a shared feature space between tasks, PORTAL is able to characterize different tasks based on the distribution of features and select those that are similar to the final task. Also, the shared feature space can effectively facilitate the policy transfer between curricula. Experimental results show that PORTAL can train agents to master extremely hard cooperative tasks, which can not be achieved with previous state-of-the-art MARL algorithms.

* Corresponding author.

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

KEYWORDS

Multiagent Reinforcement Learning; Automatic Curriculum Learning; Transfer Learning

ACM Reference Format:
Jizhou Wu, Tianpei Yang*, Xiaotian Hao, Jianye Hao*, Yan Zheng, Weixun Wang, and Matthew E.Taylor. 2023. PORTAL: Automatic Curricula Generation for Multiagent Reinforcement Learning: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023,* IFAAMAS, 3 pages.

1 INTRODUCTION

Although reinforcement learning (RL) agents can learn sophisticated behaviors by continuously interacting with the environment [5], they suffer from the notorious sample inefficiency problem [6, 8]. In MultiAgent Systems (MASs), this problem is more severe since the agents need to learn under partially observable and non-stationary environments, which makes it difficult for agents to achieve cooperate and even leads to algorithmic failures [1, 9, 10]. One potential way to address difficult MARL problems is to use curriculum learning (CL) [3] to construct a sequence of tasks from easy to hard to improve the agents' learning process [3]. The key idea is to obtain maximum learning progress towards the (final) target task by continuously learning on tasks that match the current capabilities of the agents.

In this paper, we propose a novel ACL framework ***PrOgRessive mulTiagent Automatic curricuLum*** (**PORTAL**) to facilitate MARL algorithms. We study 1) how to measure the learning progress of learners for better curriculum selection and 2) how to design efficient transfer mechanism for better curriculum transfer, two critical issues in multiagent ACL. The main contributions are:

(1) We propose a novel curriculum selection criterion that considers both the difficulty of tasks for the learners and the relevance of tasks to the final task.
(2) To facilitate curriculum transfer, we propose a shared semantic feature space to align observations of different tasks, enabling efficient policy transfer between tasks.
(3) Experimental results show that PORTAL outperforms other MARL curriculum methods and can master extremely hard (cooperative) tasks, which can not be achieved with prior state-of-the-art (SOTA) MARL algorithms.

2 MULTIAGENT CURRICULUM LEARNING

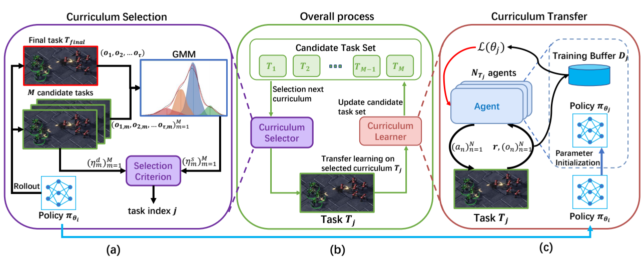

Figure 1: PORTAL Framework: (a) Curriculum Selection: at the time we finished training on current task, we collect data on all candidate tasks and calculate the curriculum selection criterion. (b) Overall process: we select task T_j as next task and then train the agent and update the candidate task set. (c) Curriculum Transfer: We reload previous policy π_{θ_i} as an initialization for the policy π_{θ_j} on new task. In figure, a black arrow denotes data and a red arrow denotes gradients.

The multiagent curriculum learning problem is formally defined as follows: given an easy initial task $T_{initial}$, a difficult final task T_{final}, and a set of candidate tasks $\mathcal{T} = \{T_m\}_{m=1}^{M}$, the goal is to learn a policy π_θ that maximizes the accumulated return $R_t = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$ on the final task T_{final}. To achieve that, the curriculum learning algorithm needs to select tasks from $\mathcal{T}$ to form a curriculum sequence $Seq = \{T_{\bar{m}}\}_{\bar{m}=1}^{\bar{M}}$. At any time, the policy is trained on one task T_i, expressed as current task $t = T_i$, and the policy denoted as π_{θ_i} will be trained on T_i using any MARL algorithm until it converges. Figure 1 shows a single step example, which includes three parts:

(1) Overall process: In Figure 1.b, when we finish the previous task, select next task from the set of candidate tasks, and then learn on the newly selected task.
(2) Curriculum Selection: In Figure 1.a, the policy π_{θ_i} trained on T_i, generates trajectories $\{\langle \boldsymbol{o}_i, \boldsymbol{a}_i, r_i, \boldsymbol{o}_{i+1}\rangle\}$ and $\left\{\langle \boldsymbol{o}_{i,m}, \boldsymbol{a}_{i,m}, r_{i,m}, \boldsymbol{o}_{i+1,m}\rangle\right\}_{m=1}^{M}$ on both final task T_{final} and M candidate tasks. The task similarity criterion η^s is calculated using $\{\langle \boldsymbol{o}_i \rangle\}$ and $\left\{\langle \boldsymbol{o}_{i,m}\rangle\right\}$, which reflects the difference on the state visitation distribution between the final task and candidate tasks. The task difficulty criterion η^d is calculated using $\left\{\langle r_{i,m}\rangle\right\}$, which reflects the difficulty of a candidate task to the current policy.
(3) Curriculum Transfer: In Figure 1.c, transfer learning allows an agent to learn the current task T_j by starting from the previous policy π_{θ_i}.

With $t = T_j$ and $\pi = \pi_{\theta_j}$, we repeat the above procedure multiple times until learning $\pi = \pi_{\theta_{final}}$ on task $t = T_{final}$.

3 EXPERIMENTS

We focus on Starcraft MultiAgent Challenge (SMAC) [4], a widely-used MARL benchmark. The experiments are carried out on three different task series: Marines, Stalkers & Zealots (S & Z) and Medivac & Marauders & Marines (MMM). The final tasks are all extremely hard tasks: 7m_vs_9m, 3s5z_vs_4s8z, MMM10.

We compare PORTAL with the SOTA non-curriculum algorithm HPN-QMIX [2] and the curriculum algorithm DYMA [7]. PORTAL generates the curriculum sequences [5m, 5m_vs_6m, 8m_vs _10m, 7m_vs_9m] for Marines, [2s3z, 3s5z_vs_3s6z, 3s5z_vs_4s7z, 3s5z_vs_4s8z] for S & Z, and [MMM, MMM4, MMM7, MMM10] for MMM. Figure 2 shows the test win rate on the final task for all three series. We can see that PORTAL achieves the best performance over other baselines on the final task.

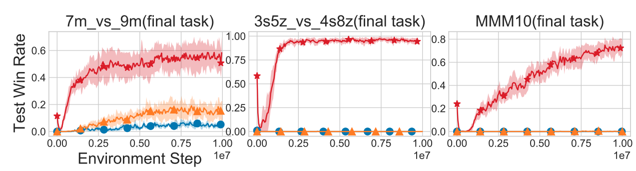

Figure 2: The learning curve uses the environment steps as X-Axis and test win rate as Y-Axis. The darker lines are the means and the lighter areas are the variances, using 95% confidence intervals. Each runs with 3 seeds.

4 CONCLUSION AND FUTURE WORK

In this paper, we propose a novel MARL ACL framework to solve tasks that is hard to learn from scratch using existing MARL algorithms. For automatic curriculum selection, we study how to measure the learning progress of the agents and propose a curriculum selection criterion from two perspectives: learning difficulty and learning relevance. In practice, we calculate observation distributions to measure task similarity, which serves as a proxy for learning relevance. We also learn a semantic feature space shared across tasks to facilitate policy transfer between curricula. With the automatic curriculum selection and transfer mechanism, our approach significantly outperforms existing MARL algorithms. We experimentally verify this on three sets of battle scenarios of SMAC.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Grant No.62106172), the “New Generation of Artificial Intelligence” Major Project of Science & Technology 2030 (Grant No.2022ZD0116402), and the Science and Technology on Information Systems Engineering Laboratory (Grant No.WDZC20235250409, No.WDZC20205250407).

Part of this work has taken place in the Intelligent Robot Learning (IRL) Lab at the University of Alberta, which is supported in part by research grants from the Alberta Machine Intelligence Institute (Amii); a Canada CIFAR AI Chair, Amii; Compute Canada; Huawei; Mitacs; and NSERC.

REFERENCES

[1] Caroline Claus and Craig Boutilier. 1998. The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference.* AAAI Press / The MIT Press, 746–752.

[2] Xiaotian Hao, Weixun Wang, Hangyu Mao, Yaodong Yang, Dong Li, Yan Zheng, Zhen Wang, and Jianye Hao. 2022. API: Boosting Multi-Agent Reinforcement Learning via Agent-Permutation-Invariant Networks. *CoRR* abs/2203.05285 (2022).

[3] Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E. Taylor, and Peter Stone. 2020. Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey. *CoRR* abs/2003.04960 (2020).

[4] Mikayel Samvelyan, Tabish Rashid, Christian Schröder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob N. Foerster, and Shimon Whiteson. 2019. The StarCraft Multi-Agent Challenge. *CoRR* abs/1902.04043 (2019).

[5] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction.* MIT press.

[6] Matthew E. Taylor and Peter Stone. 2009. Transfer Learning for Reinforcement Learning Domains: A Survey. *J. Mach. Learn. Res.* 10 (2009), 1633–1685.

[7] Weixun Wang, Tianpei Yang, Yong Liu, Jianye Hao, Xiaotian Hao, Yujing Hu, Yingfeng Chen, Changjie Fan, and Yang Gao. 2020. From Few to More: Large-Scale Dynamic Multiagent Curriculum Learning. In *Proceedings of The Thirty-Fourth AAAI Conference on Artificial Intelligence.* AAAI Press, 7293–7300.

[8] Tianpei Yang, Jianye Hao, Zhaopeng Meng, Zongzhang Zhang, Yujing Hu, Yingfeng Chen, Changjie Fan, Weixun Wang, Wulong Liu, Zhaodong Wang, and Jiajie Peng. 2020. Efficient Deep Reinforcement Learning via Adaptive Policy Transfer. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence.*

[9] Tianpei Yang, Hongyao Tang, Chenjia Bai, Jinyi Liu, Jianye Hao, Zhaopeng Meng, and Peng Liu. 2021. Exploration in Deep Reinforcement Learning: A Comprehensive Survey. *CoRR* abs/2109.06668 (2021).

[10] Tianpei Yang, Weixun Wang, Hongyao Tang, Jianye Hao, Zhaopeng Meng, Hangyu Mao, Dong Li, Wulong Liu, Yingfeng Chen, Yujing Hu, Changjie Fan, and Chengwei Zhang. 2021. An Efficient Transfer Learning Framework for Multiagent Reinforcement Learning. In *Advances in Neural Information Processing Systems 34.*