CtoD-MAT: Bridging Centralized and Decentralized Execution in Multi-Agent Reinforcement Learning (Student Abstract)

Shota Takayama[1], **Katsuhide Fujita**[2]

[1]Graduate School of Engineering, Tokyo University of Agriculture and Technology
[2]Institute of Global Innovation Research, Tokyo University of Agriculture and Technology
2-24-16 Naka-cho, Koganei-shi, Tokyo, 184-8588, Japan
takayama@katfuji.lab.tuat.ac.jp, katfuji@cc.tuat.ac.jp

Abstract

Although centralized training with centralized execution (CTCE) excels at multi-agent coordination, its reliance on global information limits its use in the real world. Conversely, the practical decentralized execution (CTDE) paradigm often struggles with complex coordination. This paper bridges this critical gap by introducing the Centralized-to-Decentralized (CtoD) learning concept: a novel framework for transferring the knowledge of a powerful centralized policy into a robust, practical decentralized policy. Our method, CtoD-MAT, realizes this transition through a curriculum that gradually shifts agents from centralized to decentralized control. A key innovation is our dynamic scheduling mechanism, featuring a mediator module, which ensures a robust and effective knowledge transfer. Using challenging SMAC benchmarks, we demonstrate that CtoD-MAT successfully produces competitive decentralized policies, notably solving complex coordination tasks that are difficult for standard CTDE methods.

Introduction

Multi-agent reinforcement learning (MARL) has primarily evolved into two paradigms. Centralized training with decentralized execution (CTDE) (Yang et al. 2020) is crucial for real-world applications because its policies rely solely on local observations, rendering them scalable, robust, and communication-efficient. However, this decentralization often constrains team coordination. Conversely, centralized training with centralized execution (CTCE) achieves superior coordination by formulating MARL as a sequence modeling problem (Wen et al. 2022). Despite its effectiveness, this approach requires global information at execution time, which confines it to the CTCE setting.

This paper addresses the critical gap between CTDE's practicality and CTCE's performance and asks the following research question: can the profound coordination knowledge from a CTCE policy be effectively transferred to create a practical, yet highly coordinated, decentralized policy?

To answer this question, we make two core contributions:

- We introduce **centralized training with hybrid execution (CTHE)** as a novel paradigm to formally bridge the two extremes. This paradigm enables the **Centralized-to-Decentralized (CtoD)** learning concept, a new form of knowledge transfer in MARL.
- We design and implement **CtoD-MAT**, a concrete framework that realizes the CtoD transition. By leveraging curriculum learning (Bengio et al. 2009), CtoD-MAT provides the first empirical demonstration of a viable pathway from a CTCE to a CTDE policy.

Our framework validates this conceptual bridge by producing decentralized policies that are competitive with strong MARL baselines.

CtoD-MAT Framework

Our central contribution is the CTHE paradigm. As illustrated in Figure 1, CTHE dynamically partitions the team of agents into two subgroups: a centralized group ($\mathcal{N}^{CE}$) that acts sequentially by conditioning on joint information, and a decentralized group ($\mathcal{N}^{DE}$) that acts in parallel using only local observations. This hybrid model provides the key mechanism for a smooth and principled transition between the fully centralized and decentralized extremes.

Our implementation, CtoD-MAT, realizes this paradigm through a curriculum-guided scheduling mechanism. The key insight is that the first agent in a sequential process and a decentralized agent share a structurally identical input: their respective local observation. This crucial similarity allows a single, shared decoder to support both execution modes. The centralized agents leverage latent representations from a shared encoder, while decentralized agents use a distinct MLP path, facilitating a seamless knowledge transfer.

This transfer process is governed by a two-part curriculum. First, a **Scheduler** determines the *size* of the centralized group ($|\mathcal{N}^{CE}|$). It employs a sigmoid-based "shift function" that reduces the number of centralized agents as training progresses, thereby controlling the difficulty of the curriculum. Second, a **Mediator** determines *which* specific agents form this group. To ensure robust training, the initial agent is chosen randomly before the mediator selects the subsequent agents, requiring tight coordination. Architecturally, the mediator is a Transformer decoder that predicts the next agent to act based on the current context. This dynamic scheduling of both the size and composition of the coordinated group is vital for solidifying knowledge transfer.

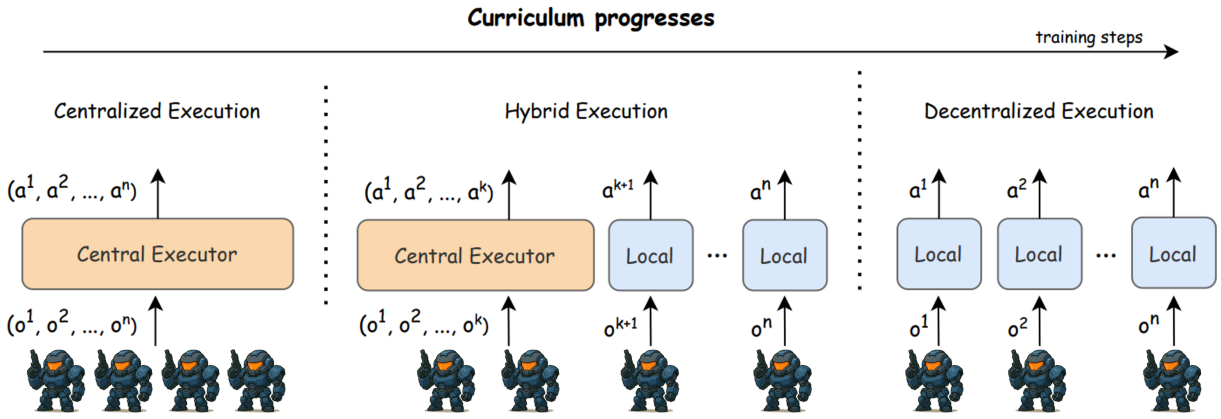

Figure 1: The curriculum-guided hybrid execution framework. The system transitions from fully centralized execution (left), through a hybrid phase, in which some agents act centrally whereas other agents act locally (middle), to fully decentralized execution (right) as training progresses.

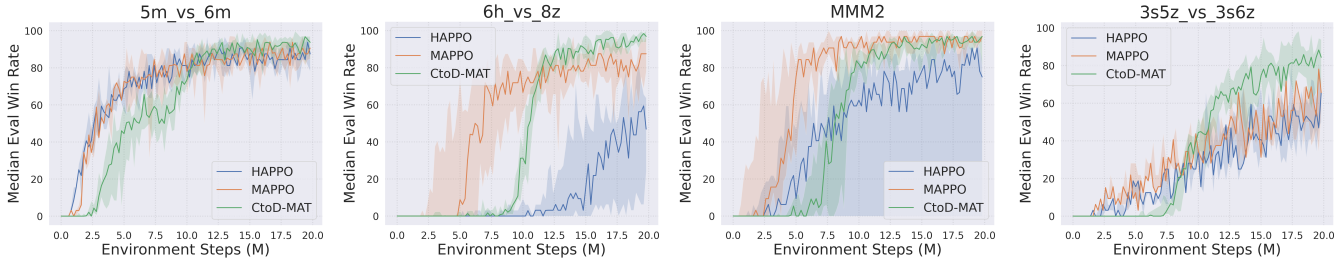

Figure 2: Median win rate (%) in the SMAC task among CtoD-MAT, MAPPO, and HAPPO in the evaluation phase.

Experiments

We validated our framework on four challenging StarCraft Multi-Agent Challenge (SMAC) tasks (Samvelyan et al. 2019). These maps were selected to represent diverse and challenging coordination problems, including homogeneous and heterogeneous team compositions, all under numerical disadvantage. We compared CtoD-MAT against strong CTDE baselines, Heterogeneous-Agent Proximal Policy Optimization (HAPPO) (Kuba et al. 2022) and Multi-Agent Proximal Policy Optimization (MAPPO) (Yu et al. 2022). Our goal is to demonstrate that the proposed CtoD transition is viable, producing a competitive decentralized policy.

Figure 2 and Table 1 show that CtoD-MAT successfully learns effective policies, achieving performance on par with or surpassing those of the strong CTDE baselines across all tested maps. The results on the `6h_vs_8z` map are particularly noteworthy. This map is notoriously difficult for CTDE methods to learn directly; however, high-coordination CTCE models, such as MAT, can solve it. Our framework's strong performance is validated by its successful transfer of the critical coordination knowledge learned in a centralized setting to a fully decentralized policy. This validates our core thesis: the CtoD paradigm provides a viable pathway from a powerful but impractical CTCE model to a practical and competitive CTDE policy, with the performance serving as its validation.

SMAC Map	CtoD-MAT	HAPPO	MAPPO
5m vs 6m	**96.8%**(6.6)	93.8%(6.3)	93.9%(3.5)
MMM2	**96.9%**(3.3)	90.6%(49.6)	**96.9%**(1.7)
6h vs 8z	**98.4%**(2.0)	59.4%(24.2)	87.5%(6.4)
3s5z vs 3s6z	**88.3%**(8.5)	65.6%(9.2)	78.1%(19.0)

Table 1: Median win rates (%). Standard deviation is in parentheses. Bold indicates the highest score.

Conclusion

This work establishes a new perspective on the design of MARL paradigms by introducing the CtoD concept as a formal bridge between CTCE and CTDE. Our framework, CtoD-MAT, provides the first empirical demonstration that this transition is viable, yielding competitive and robust decentralized policies and thus establishing a foundation for a new research area in MARL centered on inter-paradigm knowledge transfer. This contribution moves the field beyond pure performance optimization toward more flexible and practically deployable AI systems.

Future work will involve a detailed ablation study of the contributions of the mediator and the curriculum schedule, as well as a qualitative analysis of the learned behaviors to further characterize the knowledge transfer process.

References

Bengio, Y.; Louradour, J.; Collobert, R.; and Weston, J. 2009. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, 41–48. New York, NY, USA: Association for Computing Machinery. ISBN 9781605585161.

Kuba, J. G.; Feng, X.; Ding, S.; Dong, H.; Wang, J.; and Yang, Y. 2022. Heterogeneous-agent mirror learning: A continuum of solutions to cooperative marl. *arXiv preprint arXiv:2208.01682*.

Samvelyan, M.; Rashid, T.; de Witt, C. S.; Farquhar, G.; Nardelli, N.; Rudner, T. G. J.; Hung, C.-M.; Torr, P. H. S.; Foerster, J.; and Whiteson, S. 2019. The starcraft multi-agent challenge. *arXiv preprint arXiv:1902.04043*.

Wen, M.; Kuba, J.; Lin, R.; Zhang, W.; Wen, Y.; Wang, J.; and Yang, Y. 2022. Multi-agent reinforcement learning is a sequence modeling problem. *Advances in Neural Information Processing Systems*, 35: 16509–16521.

Yang, Y.; Wen, Y.; Wang, J.; Chen, L.; Shao, K.; Mguni, D.; and Zhang, W. 2020. Multi-agent determinantal q-learning. In *International Conference on Machine Learning*, 10757–10766. PMLR.

Yu, C.; Velu, A.; Vinitsky, E.; Gao, J.; Wang, Y.; Bayen, A.; and Wu, Y. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35: 24611–24624.