Tacit Learning with Adaptive Information Selection for Cooperative Multi-agent Reinforcement Learning

Extended Abstract

Lunjun Liu
College of Electrical and Information Engineering, Hunan University
Changsha, China
Greater Bay Area Institute for Innovation, Hunan University
Guangzhou, China
barryyyliu@hnu.edu.cn

Weilai Jiang
College of Electrical and Information Engineering, Hunan University
Changsha, China
Greater Bay Area Institute for Innovation, Hunan University
Guangzhou, China
jiangweilai@hnu.edu.cn

Yaonan Wang
College of Electrical and Information Engineering, Hunan University
Changsha, China
Greater Bay Area Institute for Innovation, Hunan University
Guangzhou, China
yaonan@hnu.edu.cn

ABSTRACT

In multi-agent reinforcement learning (MARL), the centralized training with decentralized execution (CTDE) framework has gained widespread adoption due to its strong performance. However, the further development of CTDE faces two key challenges. First, agents struggle to autonomously assess the relevance of input information for cooperative tasks, impairing their decision-making abilities. Second, in communication-limited scenarios with partial observability, agents are unable to access global information, restricting their ability to collaborate effectively from a global perspective. To address these challenges, we introduce a novel cooperative MARL framework based on information selection and tacit learning. In this framework, agents gradually develop implicit coordination during training, enabling them to infer the cooperative behavior of others in a discrete space without communication, relying solely on local information. Moreover, we integrate gating and selection mechanisms, allowing agents to adaptively filter information based on environmental changes, thereby enhancing their decision-making capabilities. Experiments on popular MARL benchmarks show that our framework can be seamlessly integrated with state-of-the-art algorithms, leading to significant performance improvements.

KEYWORDS

Multi-agent Reinforcement Learning; Tacit Learning; Adaptive Information Selection

ACM Reference Format:
Lunjun Liu, Weilai Jiang, and Yaonan Wang. 2025. Tacit Learning with Adaptive Information Selection for Cooperative Multi-agent Reinforcement Learning: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Cooperative Multi-Agent Reinforcement Learning (MARL) has emerged as a robust framework for addressing practical challenges across various domains, including autonomous driving [20], gaming [2], swarm robotics [8, 9], and smart grids [10, 11, 17]. Despite its success, learning complex cooperative strategies remains a major challenge. Firstly, neglecting the influence of other agents on the system introduces non-stationarity from the perspective of an individual, potentially leading to environmental instability. Additionally, as the number of agents increases, the observation space for joint actions expands exponentially, which may impede the learning process. To effectively address these challenges, the approach of Centralized Training and Decentralized Execution (CTDE) has been proposed and gained popularity in MARL. CTDE utilizes global information during training while achieving decentralized decision-making based on local information. It serves as the foundation for several prominent methods, including VDN [14], QMIX [12], MADDPG [7], QPLEX [16], COMA [3] and FOP [19].

Constrained by cognitive limitations and individual perspectives, humans exhibit selectivity when receiving information. They process this information based on their knowledge and past experiences, selecting the most relevant details for the present moment. In collaborative settings, individuals often develop a tacit understanding through specific training, enabling them to accurately predict and comprehend their peers' intentions without explicit communication. Inspired by human information processing and cooperation patterns, we propose a novel framework called **S**elective **I**mplicit **C**ollaboration **A**lgorithm (SICA) for multi-agent systems. SICA is built upon the QMIX framework and can be extended to various methods based on CTDE paradigm. The framework comprises three key blocks: the Selection Block, the Communication Block, and the Regeneration Block. During training, the Selection Block assists agents in filtering information relevant to cooperation, which is then shared with other agents through the Communication Block to generate true information. Subsequently, the Regeneration Block utilizes local information to regenerate true information. Through iterative training, SICA gradually reduces reliance on true information, transitioning from a centralized to a decentralized framework.

2 SICA

Selection Block The Selection Block consists of two MLPs and an S6 layer [4]. Since the time intervals of the inputs in the selective tasks are variable, a time-varying model is required, so we integrated the S6 layer into the framework. The two MLPs and

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

other components can be modeled as a Gating Unit (GU) [5], which is responsible for learning long-term dependencies in the input. Therefore, the entire module can be seen as a dual selection mechanism that combines the gating mechanism with the S6 selection mechanism. Additionally, to empower the Selection Block to thoroughly select information, we establish a mini-buffer preceding it. This mini-buffer preserves the preceding b observation-action pairs of a single agent.

Agent i's observation-action pairs from the current time step and the previous b time steps are integrated and passed through a GU before being fed to the S6 layer. The computation process for the entire module is outlined as follows:

$$z_i^t = \text{MLP}(x_{i1}^t)\sigma(\text{MLP}(x_{i2}^t)), \{x_{i1}^t, x_{i2}^t\} = \text{Split}(x_i^t) \quad (1)$$
$$h_i^t = \bar{A}h_i^{t-1} + \bar{B}z_i^t$$

Communication Block The Communication Block enhances the Selection Block by allowing agents to integrate global information into their decision-making process using attention-weighted mechanisms [15]. Given the hidden states of agents i and j as input, we define two learnable matrices: the self-query matrix $q_i^t = W_q h_i^t$ and the cognition matrix $k_j^t = W_k h_j^t$, where W_q , W_k are both learnable linear transformations. The calculation of attention weights proceeds as follows:

$$c_{i,j}^t = \frac{(q_i^t)^T k_j^t}{\sqrt{d_h}} \quad (2)$$

$$w_{i,j}^t = \frac{\exp(c_{i,j}^t)}{\sum_{k=1}^{N} \exp(c_{i,k}^t)} \quad (3)$$

Here, d_h represents the dimension of the hidden state. Then the true information v_i^t can be calculated as $v_i^t = \sum_{i \neq j} w_{i,j} h_j^t$.

Regeneration Block To ensure that decision-making relies solely on local information, we introduce the Regeneration Block, comprising the Selection Block and an MLP. The Regeneration Block allows us to derive the regenerated information $\hat{v}_i$, which continuously approximates the true information v_i.

At timestep t, we utilize exponential weighted averaging to ensure that the regenerated information $\hat{v}_i^t$ converges towards the true information v_i^t. Then, we compute the cross-information $\bar{v}_i^t$:

$$\bar{v}_i^t = (1 - \alpha(t))\hat{v}^t + \alpha(t)v_i^t \quad (4)$$

Where $\alpha(t)$ is dynamic, and to ensure a smoother transition within the framework, we update it using a method similar to cosine annealing[6]:

$$\alpha(t) = \alpha_{start} + (\alpha_{final} - \alpha_{start}) \cos(\frac{t}{t_{\max}}\pi). \quad (5)$$

We set α_{start} to 1 and α_{final} to 0.

Learning Objective The overall learning objective of our method is divided into two parts: the TD loss function[12] and the minimization of the regeneration information error:

$$\mathcal{L}_{tot}(\boldsymbol{\tau}, \boldsymbol{u}, s, h_i, \boldsymbol{h}; \theta) = \mathcal{L}_{TD} + \sigma(t)\mathcal{L}_{Align} \quad (6)$$

Here, $\mathcal{L}_{Align}$ represents the MSE[1] loss between v_i and $\hat{v}_i$, $\sigma(t)$ is a threshold function, which is a hyperparameter.

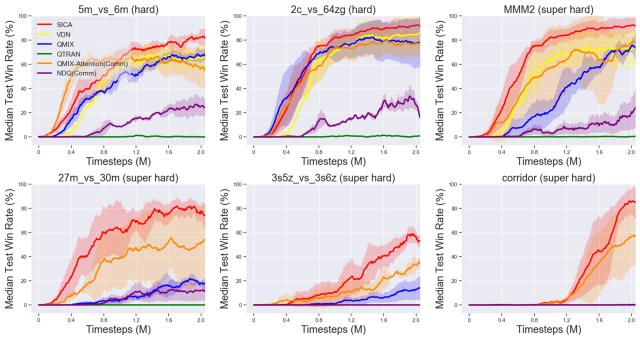

Figure 1: Performance comparison between SICA and baselines on SMAC.

3 EXPERIMENTS

We conducted experiments using the StarCraft II Multi-Agent Challenge (SMAC) benchmark, where the objective is to control a team of allied units against an enemy team governed by built-in policies.

The median win rates across different maps are shown in Figure 1. SICA consistently outperforms the baselines across all maps, even surpassing explicit communication methods. This highlights the effectiveness of SICA in information processing and highlights the robustness of its information regeneration capability. Across all methods, there is a noticeable decline in win rates as we transition from hard to super hard maps, which aligns with expectations given the heightened complexity of the latter scenarios. QTRAN[13] exhibits suboptimal performance across all maps, potentially attributable to challenges in credit assignment resulting in the development of passive agents. Meanwhile, NDQ[18] demonstrates efficacy solely on select maps, potentially stemming from instability in its message passing methodology.

4 CONCLUSION

In this paper, we introduced a novel MARL architecture named SICA, designed to enhance agents' information handling capabilities and improve the framework's generality. By integrating information selection with communication mechanisms, SICA empowers agents to autonomously choose relevant information while incorporating information from other agents. To accommodate to communication-limited environments, SICA gradually learns the tacit understanding between agents, eventually transitioning to a fully decentralized framework. Experimental results illustrate SICA's effectiveness in regenerating global information and significantly enhancing performance in challenging multi-agent tasks through information selection.

ACKNOWLEDGMENTS

This paper is supported by the National Natural Science Foundation of China under Grant 62473138, the Project of Natural Science Foundation Youth Enhancement Program of Guangdong Province under Grant 2024A1515030184, the Project of Guangzhou City Zengcheng District Key Research and Development under Grant 2024ZCKJ01, and the General Project of Natural Science Foundation of Hunan Province under Grant 2022JJ30162.

REFERENCES

[1] Eric Bauer and Ron Kohavi. 1999. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning* 36 (1999), 105–139. https://api.semanticscholar.org/CorpusID:1088806

[2] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemyslaw Debiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Christopher Hesse, Rafal Józefowicz, Scott Gray, Catherine Olsson, Jakub W. Pachocki, Michael Petrov, Henrique Pondé de Oliveira Pinto, Jonathan Raiman, Tim Salimans, Jeremy Schlatter, Jonas Schneider, Szymon Sidor, Ilya Sutskever, Jie Tang, Filip Wolski, and Susan Zhang. 2019. Dota 2 with Large Scale Deep Reinforcement Learning. *ArXiv* abs/1912.06680 (2019).

[3] Jakob N. Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2017. Counterfactual Multi-Agent Policy Gradients. In *AAAI Conference on Artificial Intelligence*.

[4] Albert Gu and Tri Dao. 2023. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *ArXiv* abs/2312.00752 (2023).

[5] Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. 2021. Pay attention to mlps. *Advances in neural information processing systems* 34 (2021), 9204–9215.

[6] Ilya Loshchilov and Frank Hutter. 2016. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv: Learning* (2016). https://api.semanticscholar.org/CorpusID:14337532

[7] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, P. Abbeel, and Igor Mordatch. 2017. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. *ArXiv* abs/1706.02275 (2017).

[8] David Henry Mguni, Joel Jennings, and Enrique Munoz de Cote. 2018. Decentralised Learning in Systems with Many, Many Strategic Agents. *ArXiv* abs/1803.05028 (2018).

[9] David Henry Mguni, Joel Jennings, Sergio Valcarcel Macua, Emilio Sison, Sofia Ceppi, and Enrique Munoz de Cote. 2019. Coordinating the Crowd: Inducing Desirable Equilibria in Non-Cooperative Systems. In *Adaptive Agents and Multi-Agent Systems*.

[10] Dawei Qiu, Jianhong Wang, Zihang Dong, Yi Wang, and Goran Strbac. 2023. Mean-Field Multi-Agent Reinforcement Learning for Peer-to-Peer Multi-Energy Trading. *IEEE Transactions on Power Systems* 38 (2023), 4853–4866.

[11] Dawei Qiu, Jianhong Wang, Junkai Wang, and Goran Strbac. 2021. Multi-Agent Reinforcement Learning for Automated Peer-to-Peer Energy Trading in Double-Side Auction Market. In *International Joint Conference on Artificial Intelligence*.

[12] Tabish Rashid, Mikayel Samvelyan, C. S. D. Witt, Gregory Farquhar, Jakob N. Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. *ArXiv* abs/1803.11485 (2018).

[13] Kyunghwan Son, Daewoo Kim, Wan Ju Kang, David Earl Hostallero, and Yung Yi. 2019. QTRAN: Learning to Factorize with Transformation for Cooperative Multi-Agent Reinforcement Learning. *ArXiv* abs/1905.05408 (2019).

[14] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech M. Czarnecki, Vinícius Flores Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z. Leibo, Karl Tuyls, and Thore Graepel. 2017. Value-Decomposition Networks For Cooperative Multi-Agent Learning. *ArXiv* abs/1706.05296 (2017).

[15] Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Neural Information Processing Systems*.

[16] Jianhao Wang, Zhizhou Ren, Terry Liu, Yang Yu, and Chongjie Zhang. 2020. QPLEX: Duplex Dueling Multi-Agent Q-Learning. *ArXiv* abs/2008.01062 (2020).

[17] Jianhong Wang, Wangkun Xu, Yunjie Gu, Wenbin Song, and Tim C. Green. 2021. Multi-Agent Reinforcement Learning for Active Voltage Control on Power Distribution Networks. *ArXiv* abs/2110.14300 (2021).

[18] Tonghan Wang, Jianhao Wang, Chongyi Zheng, and Chongjie Zhang. 2019. Learning nearly decomposable value functions via communication minimization. *arXiv preprint arXiv:1910.05366* (2019).

[19] Tianhao Zhang, Yueheng Li, Chen Wang, Guangming Xie, and Zongqing Lu. 2021. FOP: Factorizing Optimal Joint Policy of Maximum-Entropy Multi-Agent Reinforcement Learning. In *International Conference on Machine Learning*.

[20] Ming Zhou, Jun Luo, Julian Villela, Yaodong Yang, David Rusu, Jiayu Miao, Weinan Zhang, Montgomery Alban, Iman Fadakar, Zheng Chen, Aurora Chongxi Huang, Ying Wen, Kimia Hassanzadeh, Daniel Graves, Dong Chen, Zhengbang Zhu, Nhat M. Nguyen, Mohamed Elsayed, Kun Shao, Sanjeevan Ahilan, Baokuan Zhang, Jiannan Wu, Zhengang Fu, Kasra Rezaee, Peyman Yadmellat, Mohsen Rohani, Nicolas Perez Nieves, Yihan Ni, Seyedershad Banijamali, Alexander Cowen Rivers, Zheng Tian, Daniel Palenicek, Haitham Ammar, Hongbo Zhang, Wulong Liu, Jianye Hao, and Jun Wang. 2020. SMARTS: Scalable Multi-Agent Reinforcement Learning Training School for Autonomous Driving. *ArXiv* abs/2010.09776 (2020).