Efficient Teammate Adaptation with Language-assisted Progressive Intention Alignment

Extended Abstract

Zhichao Wu
Nanjing University
Nanjing, China
wuzc@lamda.nju.edu.cn

Ruiqi Xue
Nanjing University
Nanjing, China
xuerq@lamda.nju.edu.cn

Yichen Li
Nanjing University
Nanjing, China
liyc@lamda.nju.edu.cn

Cong Guan
Nanjing University
Nanjing, China
guanc@lamda.nju.edu.cn

Jingwen Yang
Tencent
Shenzhen, China
jingwenyang@tencent.com

Lei Yuan
Nanjing University
Nanjing, China
yuanl@lamda.nju.edu.cn

Yang Yu
Nanjing University
Nanjing, China
yuy@nju.edu.cn

ABSTRACT

Enabling agents to collaborate effectively with diverse and previously unseen teammates remains a core challenge in multi-agent reinforcement learning (MARL), particularly in open environments. While existing research has made significant strides in adapting to diverse teammate behaviors under a fixed shared reward, the challenge of collaborating with partners who pursue distinct and unobserved personal rewards (intentions) remains unexplored. Moreover, existing teammate modeling relies primarily on low-level behavioral cues while overlooking high-level semantic priors (e.g., language descriptions), resulting in inefficient intention identification. We introduce TALP, a Bayesian framework for intention-aware teammate adaptation. At deployment, it leverages language priors and interaction history to perform unbiased intention inference, enabling targeted cooperation within a single episode. Experiments demonstrate that TALP accurately infers teammate intentions and significantly boosts collaborative efficiency.

KEYWORDS

Multi-agent Reinforcement Learning, Coordination and Cooperation, Intention Alignment

ACM Reference Format:
Zhichao Wu, Ruiqi Xue, Yichen Li, Cong Guan, Jingwen Yang, Lei Yuan, and Yang Yu. 2026. Efficient Teammate Adaptation with Language-assisted Progressive Intention Alignment: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/HOBH9968

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/HOBH9968

1 INTRODUCTION

Cooperative multi-agent reinforcement learning has achieved remarkable success across complex domains, including autonomous driving [2] and embodied intelligence [3], driven by algorithms progress [6, 8, 11, 12]. However, enabling agents to collaborate effectively with diverse and previously unseen teammates remains a core challenge, particularly in open environments [9, 13]. While existing Zero-Shot Coordination (ZSC) research has made significant strides in adapting to diverse teammate behaviors [5, 7, 10, 15], these methods typically operate under a fixed shared reward assumption. Consequently, the critical challenge of collaborating with partners who pursue distinct and unobserved personal rewards (called **intentions** here) remains largely unexplored, limiting performance in tasks involving multi-modal coordination equilibria.

To address intention uncertainty, teammate modeling [1, 14] is commonly employed to infer partner goals. However, existing approaches rely primarily on low-level behavioral cues (e.g., trajectories), which suffer from limited information density and inherent ambiguity. Crucially, these methods overlook the role of high-level semantic priors, such as language descriptions of potential teammate intentions. Relying solely on behavioral observations without semantic guidance results in inefficient intention identification, as agents struggle to distinguish between complex goals quickly, leading to delayed or misaligned collaboration.

Addressing these challenges, we propose TALP, a novel Bayesian inference framework for intention-aware teammate adaptation. Unlike standard approaches, TALP leverages language descriptions as a semantic prior and interaction history as evidence to perform unbiased Bayesian intention inference within a single episode. This allows the agent to rapidly identify the teammate's specific intention and initiate targeted cooperation. Extensive experiments across Gridworld and Overcooked environments demonstrate that TALP not only accurately infers intentions but also consistently outperforms existing context-based meta-learning and teammate modeling benchmarks in collaborative efficiency.

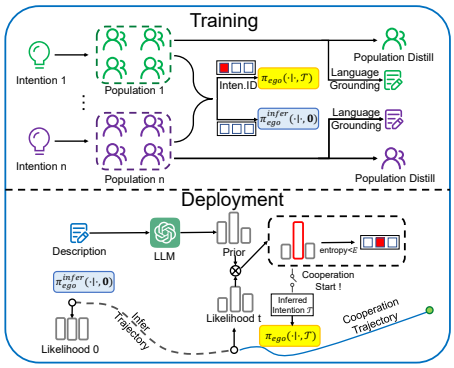

Figure 1: The overall workflow of TALP.

2 METHODS

We propose TALP, a Bayesian framework that synergizes high-level semantic priors with low-level behavioral cues for one-shot collaboration. As illustrated in Fig. 1, TALP operates by maintaining a belief distribution over M potential teammate intentions, $\mathcal{T} \in \{1, ..., M\}$, and dynamically switching policies based on inference confidence.

Bayesian Intention Inference. We formulate intention identification as a recursive Bayesian inference process. Given an interaction history $\tau_t = (s_0, a_0, ..., s_t, a_t)$ at step t, the posterior probability of the teammate having intention $\mathcal{T}$ is updated as:

$$P(\mathcal{T}|\tau_t) \propto P(\tau_t|\mathcal{T})P(\mathcal{T}), \tag{1}$$

where $P(\mathcal{T})$ is the semantic prior from language, and $P(\tau_t|\mathcal{T})$ is the behavioral likelihood.

Language Priors Modeling. Unlike standard methods that start with a uniform prior, TALP leverages Large Language Model (LLM) to ground intention language L (e.g., "go to the bottom-left") into the intention prior $P(\mathcal{T})$. We utilize high return trajectories to fetch language descriptions of each intention $Text(\mathcal{T})$. At deployment, an LLM compares the semantic similarity between teammate's language L and each intention description $Text(\mathcal{T})$, normalizing the similarity scores to form prior distribution $P(\mathcal{T})$.

Likelihood Modeling. Formally, computing the likelihood of observing a teammate's action a_t^{tm} under a specific intention $\mathcal{T}$ requires marginalizing over the entire space of potential teammate policies $\Pi^{\mathcal{T}}$:

$$P(a_t^{tm}|s_t, \mathcal{T}) = \mathbb{E}_{\pi_{tm} \sim P(\cdot|\mathcal{T})}[\pi_{tm}(a_t^{tm}|s_t)] \tag{2}$$

Directly computing this expectation is intractable as it involves enumerating an infinite or vast policy space. To address this, we first approximate the space by generating a diverse teammate population $\{\pi_i^{\mathcal{T}}\}_{i=1}^{G}$ for each intention. We then distill this population into a single distilled policy $\pi_{dis}^{G}(\cdot|s, \mathcal{T})$ by minimizing the total variation distance. This enables efficient real-time calculation of the trajectory likelihood $P(\tau_t|\mathcal{T}) \approx \prod_{i=0}^{t} \pi_{dis}^{G}(a_i^{tm}|s_i, \mathcal{T})$ without the expensive cost of querying individual population models.

Two-Phase Execution Strategy. TALP employs a progressive intention alignment strategy within a single episode:

- Inference Phase: Initially, the agent executes an intention-agnostic policy $\pi_{ego}^{infer}(\cdot|\cdot, \mathbf{0})$, trained to interact with teammates from any intention group. This policy focuses on gathering information without biasing the interaction sequence.
- Cooperation Phase: At each step, we calculate the entropy of the posterior $H(P(\mathcal{T}|\tau_t))$. Once H drops below a threshold E (indicating high confidence), TALP commits to the most likely intention $\hat{\mathcal{T}}$ and switches to the intention-aware policy $\pi_{ego}(\cdot|\cdot, \hat{\mathcal{T}})$, which is optimized to maximize cooperative returns for that specific task.

3 EXPERIMENTS

We evaluated TALP on our custom GridNav and Overcooked [4] against baselines from ZSC (FCP [10]), teammate modeling (Fastap [14]), and context-based intention inference (VariBAD [16]).

Main Results. As shown in Tab. 1, TALP consistently outperforms all baselines in collaborative efficiency. In complex Overcooked tasks involving multi-modal intentions, TALP achieves a 28% performance boost (49.2 vs. 38.4) over the strongest baseline when precise language is available.

Robustness to Vague Language. A critical advantage of TALP is its ability to leverage ambiguous semantic priors. When provided with vague language instructions (e.g. "go to the bottom" instead of "bottom-left"), baselines incorporating language embeddings often degrade in performance due to grounding failures—for instance, Fastap's reward in GridNav drops significantly from 0.428 to 0.305. In contrast, TALP utilizes LLMs to generate a probabilistic prior, effectively narrowing the hypothesis space and accelerating intention identification, thereby maintaining high rewards and even achieving performance gains (e.g., increasing from 43.9 to 45.4 in Overcooked) even with vague instructions.

Table 1: Average Episode Rewards. TALP maintains high performance even with vague language.

Method	GridNav			Overcooked		
	No Lang.	w/ Vague	w/ Lang.	No Lang.	w/ Vague	w/ Lang.
FCP	0.459	-	0.160	36.8	-	38.4
Fastap	0.428	0.305	0.376	38.4	39.3	37.7
VariBAD	0.374	0.352	0.354	39.1	36.5	38.2
TALP	**0.535**	**0.552**	**0.573**	**43.9**	**45.4**	**49.2**

4 FINAL REMARKS

In this work, we introduced TALP, a framework that synergizes high-level linguistic priors with low-level behavioral cues for efficient teammate adaptation. By grounding natural language into a Bayesian inference process, TALP demonstrating superior robustness and efficiency even with vague instructions. Currently, TALP relies on the passive observation of teammate actions to update likelihoods, which may result in latency during highly dynamic interactions. Future work could mitigate this by incorporating inverse dynamics models to predict teammate behaviors proactively, or by extending the framework to more complex, real-world embodied coordination tasks.

REFERENCES

[1] Stefano V Albrecht and Peter Stone. 2018. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence* 258 (2018), 66–95.

[2] Eduardo Candela, Leandro Parada, Luis Marques, Tiberiu-Andrei Georgescu, Yiannis Demiris, and Panagiotis Angeloudis. 2022. Transferring multi-agent reinforcement learning policies for autonomous driving using sim-to-real. In *IROS*. 8814–8820.

[3] Lorenzo Canese, Gian Carlo Cardarilli, Mohammad Mahdi Dehghan Pir, Luca Di Nunzio, and Sergio Spanò. 2024. Design and Development of Multi-Agent Reinforcement Learning Intelligence on the Robotarium Platform for Embedded System Applications. *Electronics* 13, 10 (2024), 1819.

[4] Ghost Town Games. 2016. Overcooked. *Team17: Wakefield, UK* (2016).

[5] Lihe Li, Lei Yuan, Pengsen Liu, Tao Jiang, and Yang Yu. 2025. LLM-Assisted Semantically Diverse Teammate Generation for Efficient Multi-agent Coordination. In *ICML*.

[6] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 6382–6393.

[7] Andrei Lupu, Brandon Cui, Hengyuan Hu, and Jakob Foerster. 2021. Trajectory diversity for zero-shot coordination. In *ICML*. 7204–7213.

[8] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2020. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research* 21, 178 (2020), 1–51.

[9] Peter Stone, Gal Kaminka, Sarit Kraus, and Jeffrey Rosenschein. 2010. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *AAAI*. 1504–1509.

[10] DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. 2021. Collaborating with humans without human data. In *NeurIPS*. 14502–14515.

[11] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2018. Value-Decomposition Networks For Cooperative Multi-Agent Learning Based On Team Reward. In *AAMAS*. 2085–2087.

[12] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. In *NeurIPS*. 24611–24624.

[13] Lei Yuan, Ziqian Zhang, Lihe Li, Cong Guan, and Yang Yu. 2023. A survey of progress on cooperative multi-agent reinforcement learning in open environment. *arXiv preprint arXiv:2312.01058* (2023).

[14] Ziqian Zhang, Lei Yuan, Lihe Li, Ke Xue, Chengxing Jia, Cong Guan, Chao Qian, and Yang Yu. 2023. Fast teammate adaptation in the presence of sudden policy change. In *UAI*. 2465–2476.

[15] Rui Zhao, Jinming Song, Yufeng Yuan, Haifeng Hu, Yang Gao, Yi Wu, Zhongqian Sun, and Wei Yang. 2023. Maximum entropy population-based training for zero-shot human-ai coordination. In *AAAI*. 6145–6153.

[16] Luisa Zintgraf, Sebastian Schulze, Cong Lu, Leo Feng, Maximilian Igl, Kyriacos Shiarlis, Yarin Gal, Katja Hofmann, and Shimon Whiteson. 2021. Varibad: Variational bayes-adaptive deep rl via meta-learning. *Journal of Machine Learning Research* 22, 289 (2021), 1–39.