On the Utility of External Agent Intention Predictor for Human-AI Coordination

Extended Abstract

Chenxu Wang
Tsinghua University
Beijing, China
wcx21@mails.tsinghua.edu.cn

Zilong Chen
Tsinghua University
Beijing, China
czl22@mails.tsinghua.edu.cn

Huaping Liu*
Tsinghua University
Beijing, China
hpliu@tsinghua.edu.cn

ABSTRACT

Reaching a consensus on the team plans is vital to human-AI coordination. We suggest incorporating external models to assist humans in understanding the intentions of AI agents when the AI has no explainable plan to communicate. In this paper, we propose a two-stage paradigm that first trains a Theory of Mind (ToM) model from collected offline trajectories of the target agent and utilizes the model in the process of human-AI collaboration by real-timely displaying the future action predictions of the target agent. We further implement a transformer-based predictor as the ToM model and develop an extended online human-AI collaboration platform for experiments. Experimental results validate that our ToM model can significantly improve team performance, demonstrating the potential of our paradigm in human-AI collaboration.

KEYWORDS

Human-AI Cooperation; Deep Reinforcement Learning; Human-Agent Interaction; Intention Prediction

ACM Reference Format:
Chenxu Wang, Zilong Chen, and Huaping Liu*. 2024. On the Utility of External Agent Intention Predictor for Human-AI Coordination: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Human-AI Collaboration has been receiving increasing research interest [2, 5, 19] and the techniques are applied in various domains, such as robots [7, 9], data science [22], and decision making [10, 14].

Unlike competitive or solo tasks, AI agents and humans share the same goal in collaboration tasks, naturally encouraging the team to reach a consensus on the joint plan to maximize the common payoff. An intuitive way to resolve this dilemma is communication, which is natural in human-human collaboration and has been widely applied in Human-Agent Collaboration through graphical or textual ways [1, 13, 17, 18, 21, 24]. However, in Deep Reinforcement Learning (DRL), a class of algorithms that have been widely used in many challenging robotic tasks [4, 9, 15], the model may be

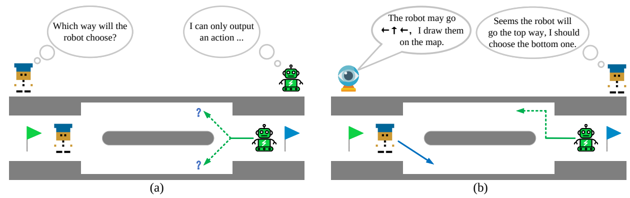

Figure 1: A representative case where the human and the robot want to switch their positions. The potential difficulty in the collaboration is reaching an agreement on the selection of routes without collision.

opaque and only output an atomic action at each time step with no human-explainable plans to show. Though the explainable AI has been proven important for human-AI collaboration [16], there may exist a trade-off between the explainability and performance [6, 8]. Therefore, how to conduct communication when the AI agent is not explainable is still an open problem.

To mitigate this predicament, we suggest building an external Theory of Mind (ToM) model to predict the future actions of the AI agent and assist humans in recognizing its intentions. As the example shown in Figure 1, the ToM model that has watched records of the robot can predict its future actions and helps the human better coordinate. Carrying on this idea, we propose a two-stage paradigm to assist humans in the human-AI collaboration. For the agent we want to predict, which is denoted as the target agent hereinafter, we first collect the trajectories by pairing the target agent with other agents to form a dataset and then train a stand-alone ToM model from its historical data to model the behavior of the target agent. In the human-AI collaboration stage, we utilize the ToM model to predict the future action sequence of the target agent from the states of the ongoing task and visually display the predictions.

Our proposed paradigm is distinguished from existing techniques such as plan communication [13, 17] by the following characteristics: (1) It does not require any prior knowledge of the environment and predicts at the action level, ensuring its availability in general DRL scenarios. (2) The ToM model is trained from offline data and can be regarded as a complete post-process, providing compatibility for all DRL algorithms. (3) We treat the agent as a gray or black box, making the ToM model eligible to be a third-party assistant.

To evaluate our paradigm, we implement a transformer-based ToM model and develop an extended online experimental platform that is capable of prediction display and user study based on the Overcooked environment [2], which is a two-player collaborative game that has been widely used for studying human-AI collaboration [3, 20, 23, 25]. Our paradigm and model are comprehensively

*Corresponding author.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand.

tested with two types of DRL agents across multiple layouts with an attached user study. The results demonstrate that our method can improve collaboration performance in various situations.

2 METHODOLOGY

We gather trajectory data from both self-play and cross-play. In self-play, the target agent plays with a copy of itself, where the trajectories demonstrate the ideal plan in the agent's opinion. In cross-play, the target agent is paired with a group of partners to obtain the behavior pattern when cooperating with others, where collaboration may not be ideal. Specifically, we train an independent group of agents with various DRL algorithms and pick 3 checkpoints with different skill levels from each agent to form the partner population. For each selected pair of agents, we collect trajectories under different settings, including the roles of agents and whether the partner agent is deterministic.

We then train a transformer-based ToM model to predict the next l actions of the target agent from recent game history, including a short sequence of actions and states. The model aims to perform action-level predictions, employing an MLP as the output layer, and is optimized by minimizing the Cross-Entropy loss: $\mathcal{L} = -\sum_{i=1}^{l}\sum_{j=1}^{|\mathcal{A}^{AI}|} y_{ij} log(\hat{y}_{ij})$. Where $\mathcal{A}^{AI}$ is the action space of the AI agent, $\hat{y}_{ij}$ denotes the predicted probability of j-th action, y_{ij} is 1 for the correct action and 0 otherwise.

3 EXPERIMENTS

3.1 Experimental setup

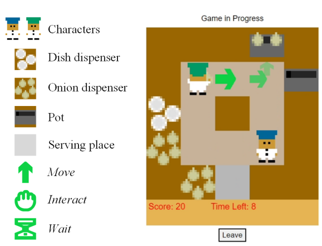

Figure 2: Extended platform based on Overcooked [2].

To inspect the effectiveness of our human-AI collaboration framework, we conduct experiments in the Overcooked environment [2], which has been widely used for studying zero-shot coordination and human-AI collaboration [2, 3, 11, 12, 20, 23, 25]. To provide action predictions in the human-AI collaboration process, we extend the front-end user interface to show predictions by depicting successive arrows and icons with gradually varied sizes and transparency at estimated locations, as illustrated in Figure 2.

We incorporate two types of agents in the experiment: Self-play (SP) and Fictitious Co-Play (FCP) [20] and run experiments on 5 layouts: *Coordination Ring*, *Double Rings*, *Double Counters*, *Matrix*, *Clear Division*, testing the capabilities of agents across various challenges. Human subjects are divided into 3 groups regarding the predictor settings: ToM model, random, and no predictor. They then play with the two agents in all layouts. After each episode, participants are asked to give a subjective assessment by filling out a questionnaire with questions corresponding to four indicators: *satisfaction with the partner*, *satisfaction with the predictor*, *situational awareness*, and *cooperation efficiency*.

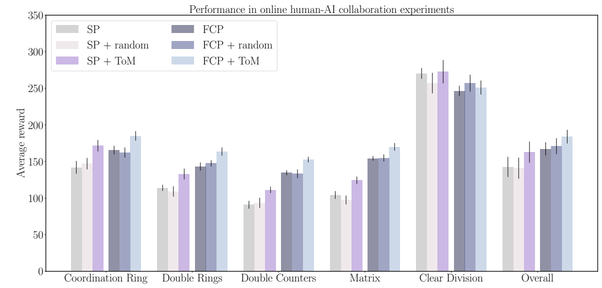

Figure 3: Average rewards with standard error bar.

3.2 Results

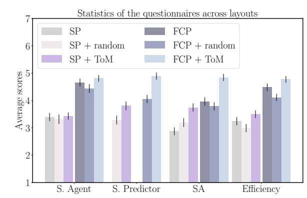

Figure 4: Average scores in user assessment across layouts.

We present the average reward on each layout and across layouts in Figure 3. We first perform an ANOVA to compare the rewards in different ToM settings and the results show statistically significant differences for both SP ($H(2, n = 430) = 10.83, p = .004$) and FCP ($H(2, n = 430) = 15.23, p < .001$). Subsequent ANOVA conducted on the results in each layout show that significant differences exist on the first 4 layouts (with a significance level of $p < .05$, same hereinafter), and pairwise comparisons with the T-test show that the ToM model improves the performance than the no-predictor baseline, even after the Benjamini-Hochberg correction for false discovery control.

To investigate the factors of performance improvements in positive situations, we scrutinize user assessments and present the average scores on the first 4 layouts across settings in the Figure 4. Humans are significantly more satisfied with the ToM model than the random baseline (SP: $t(224) = 2.33, p = .021$; FCP: $t(226) = 4.16, p < .001$), underscoring that humans care about the predictions. The ToM model also significantly improves situational awareness compared to both the random baseline and the no-predictor setting, indicating the efficacy of our paradigm in assisting humans in predicting the intentions of the AI agent.

4 CONCLUSION

In this paper, we introduce a novel paradigm for human-AI collaboration that integrates a ToM model, trained from the offline trajectories of the target AI agent, into the collaboration process. Experimental results validate our paradigm and the efficacy of the ToM models implemented in both quantitative rewards and subjective measurement. Our method and platform lay the groundwork for further research on assisting humans to understand AI intentions in human-AI collaboration.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Fund for Distinguished Young Scholars under Grant 62025304.

REFERENCES

[1] Arash Ajoudani, Andrea Maria Zanchettin, Serena Ivaldi, Alin Albu-Schäffer, Kazuhiro Kosuge, and Oussama Khatib. 2018. Progress and prospects of the human–robot collaboration. *Autonomous Robots* 42, 5 (2018), 957–975.

[2] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. In *Advances in neural information processing systems*, Vol. 32.

[3] Rujikorn Charakorn, Poramate Manoonpong, and Nat Dilokthanakul. 2020. Investigating partner diversification methods in cooperative multi-agent deep reinforcement learning. In *International Conference on Neural Information Processing*. Springer, 395–402.

[4] Yu Fan Chen, Miao Liu, Michael Everett, and Jonathan P How. 2017. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 285–292.

[5] Dominik Dellermann, Adrian Calma, Nikolaus Lipusch, Thorsten Weber, Sascha Weigel, and Philipp Ebel. 2019. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

[6] Mark Edmonds, Feng Gao, Hangxin Liu, Xu Xie, Siyuan Qi, Brandon Rothrock, Yixin Zhu, Ying Nian Wu, Hongjing Lu, and Song-Chun Zhu. 2019. A tale of two explanations: Enhancing human trust by explaining robot behavior. *Science Robotics* 4, 37 (2019), eaay4663.

[7] Ali Ghadirzadeh, Xi Chen, Wenjie Yin, Zhengrong Yi, Mårten Björkman, and Danica Kragic. 2020. Human-centered collaborative robots with deep reinforcement learning. *IEEE Robotics and Automation Letters* 6, 2 (2020), 566–571.

[8] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science robotics* 4, 37 (2019), eaay7120.

[9] Julian Ibarz, Jie Tan, Chelsea Finn, Mrinal Kalakrishnan, Peter Pastor, and Sergey Levine. 2021. How to train your robot with deep reinforcement learning: lessons we have learned. *The International Journal of Robotics Research* 40, 4-5 (2021), 698–721.

[10] Lujain Ibrahim, Mohammad M Ghassemi, and Tuka Alhanai. 2023. Do Explanations Improve the Quality of AI-assisted Human Decisions? An Algorithm-in-the-Loop Analysis of Factual & Counterfactual Explanations. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 326–334.

[11] Paul Knott, Micah Carroll, Sam Devlin, Kamil Ciosek, Katja Hofmann, Anca Dragan, and Rohin Shah. 2021. Evaluating the Robustness of Collaborative Agents. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. 1560–1562.

[12] Xingzhou Lou, Jiaxian Guo, Junge Zhang, Jun Wang, Kaiqi Huang, and Yali Du. 2023. PECAN: Leveraging Policy Ensemble for Context-Aware Zero-Shot Human-AI Coordination. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 679–688.

[13] Simone Macciò, Alessandro Carfì, and Fulvio Mastrogiovanni. 2022. Mixed Reality as Communication Medium for Human-Robot Collaboration. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2796–2802.

[14] Avleen Malhi, Samanta Knapic, and Kary Främling. 2020. Explainable agents for less bias in human-agent decision making. In *Explainable, Transparent Autonomous Agents and Multi-Agent Systems: Second International Workshop, EXTRAAMAS 2020, Auckland, New Zealand, May 9–13, 2020, Revised Selected Papers 2*. Springer, 129–146.

[15] Hai Nguyen and Hung La. 2019. Review of deep reinforcement learning for robot manipulation. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE, 590–595.

[16] Rohan Paleja, Muyleng Ghuy, Nadun Ranawaka Arachchige, Reed Jensen, and Matthew Gombolay. 2021. The utility of explainable ai in ad hoc human-machine teaming. In *Advances in Neural Information Processing Systems*, Vol. 34. 610–623.

[17] Julie Porteous, Alan Lindsay, and Fred Charles. 2023. Communicating agent intentions for human-agent decision making under uncertainty. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 290–298.

[18] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. 2020. Communicating robot arm motion intent through mixed reality head-mounted displays. In *Robotics Research: The 18th International Symposium ISRR*. Springer, 301–316.

[19] Konrad Sowa, Aleksandra Przegalinska, and Leon Ciechanowski. 2021. Cobots in knowledge work: Human–AI collaboration in managerial professions. *Journal of Business Research* 125 (2021), 135–142.

[20] DJ Strouse, Kevin R McKee, Matthew Botvinick, Edward Hughes, and Richard Everett. 2021. Collaborating with Humans without Human Data. In *Advances in Neural Information Processing Systems*, Vol. 34. 14502–14515.

[21] Aaquib Tabrez, Matthew B Luebbers, and Bradley Hayes. 2022. Descriptive and prescriptive visual guidance to improve shared situational awareness in human-robot teaming. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 1256–1264.

[22] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.

[23] Mesut Yang, Micah Carroll, and Anca Dragan. 2022. Optimal Behavior Prior: Data-Efficient Human Models for Improved Human-AI Collaboration. *arXiv preprint arXiv:2211.01602* (2022).

[24] Luyao Yuan, Xiaofeng Gao, Zilong Zheng, Mark Edmonds, Ying Nian Wu, Federico Rossano, Hongjing Lu, Yixin Zhu, and Song-Chun Zhu. 2022. In situ bidirectional human-robot value alignment. *Science Robotics* 7, 68 (2022), eabm4183.

[25] Rui Zhao, Jinming Song, Hu Haifeng, Yang Gao, Yi Wu, Zhongqian Sun, and Yang Wei. 2021. Maximum Entropy Population Based Training for Zero-Shot Human-AI Coordination. *arXiv preprint arXiv:2112.11701* (2021).