Developing Targeted Communication through a Trust Factor in Multi-Agent Reinforcement Learning

Simone Di Rienzo[1], Francesco Frattolillo[1], Roberto Cipollone[1], Andrea Fanti[1], Nicolo' Brandizzi[1,2] and Luca Iocchi[1]

[1]*Sapienza University of Rome, Via Ariosto, 25, 00185 Roma RM, Italy*

[2]*Fraunhofer IAIS, Schloss Birlinghoven, 1, 53757 Sankt Augustin, Germany*

Abstract

The concept of trust has long been studied, initially in the context of human interactions and, more recently, in human-machine or human-agent interactions. Despite extensive studies, defining trust remains challenging due to its inherent complexities and the diverse factors that influence its dynamics in multi-agent environments. This paper focuses on a specific formalization of a trust factor: predictive reliability, defined as the ability of agents to accurately forecast the actions of their peers in a shared environment. By realizing this trust factor within the framework of multi-agent reinforcement learning (MARL), we integrate it as a criterion for agents to assess and select collaborators. This approach enhances the functionality of MARL systems, promoting improved cooperation and overall effectiveness.

Keywords

Multi-Agent Systems, Reinforcement Learning, Trust Factor, Computational Modeling,

1. Introduction and Background

With the advent of artificial intelligence, the number of applications that require co-existence and the interaction between intelligent agents and humans is increasing over time. Such applications include autonomous vehicles [1], industrial robotics [2], healthcare robotics [3], service robotics [4], agricultural robotics [5], and many more [6]. In this context, the concept of trust becomes essential, as it fosters cooperation and collaboration between humans and robots, enhancing efficiency and user satisfaction. It instills confidence in the reliability and predictability of robotic systems, which is crucial for their acceptance and adoption.

Trust is a concept that has been defined numerous times in the literature [7], yet there is still no single universally accepted definition. However, the numerous factors that influence trust are much easier to study when analyzed separately. Such factors may be associated with the trustor, which is the person that trusts, with the trustee, the ones being trusted, or could be dependent on the context [8]. In this article, to formalize one of such "trust factors", we take inspiration from the definition given in Gambetta [9]:

"trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects his own action"

According to this notion, the trust between two agents can be correlated to the trustor's expectations about the choices made by the trustee in a context of mutual interaction. We formalized this definition in a Multi-Agent Reinforcement Learning (MARL) setting. Here, autonomous agents can benefit from reasoning about other agents' intentions, and they can use this information to improve their performance and select which agent to communicate with.

Problem and Solution Formulation   we consider the common scenario in which agents do not have complete knowledge of the environment which is formalized by the Decentralized Partially Observable Markov Decision Process (DEC-POMDP) [10] framework, defined as a Tuple $\langle D, \mathcal{S}, \mathcal{A}, T, R, \Omega, O \rangle$, where D is the number of agents; $\mathcal{S}$ is the set of environment states shared by all agents; $\mathcal{A}$ is the set of joint actions; $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the transition function; $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function; Ω is the set of joint observations and $O : \mathcal{S} \times \Omega \rightarrow [0, 1]$ is a set of observation probabilities returning the probability of joint observation.

A MARL solution for a DEC-POMDP is a set of D functions, called *policies* $\pi^i : \Omega_i \rightarrow \mathcal{A}_i$, which map the local observations of each agent to its actions, in order to maximize the expected joint sum of discounted rewards: $\sum_{k=t}^{T} \gamma^k R(s_k, a_k)$, where $0 \leq \gamma < 1$.

Trust factors   We refer to Frattolillo et al. [11] for a general definition of trust factors in MARL systems. Specifically, any trust factor is computed with respect to a specific *Trustor* X, a *Trustee* Y, and a task Γ:

$$TrustFactor(X|Y, \Gamma) = f(o^X, a^Y, r, [b^{X \rightarrow Y}], [c^{X \rightarrow Y}], [c^{Y \rightarrow X}]) \tag{1}$$

In this template, o^X denotes the observation of the trustor, a^Y is the action of the trustee, r is the immediate reward, and $b^{X \rightarrow Y}$ is the current belief that the trustor currently maintains with respect to the trustee. Finally, $c^{X \rightarrow Y}$ and $c^{Y \rightarrow X}$ represent known facts that result from communication from trustor to trustee and vice versa. The brackets denote that these are optional components.

2. Method

In this work, we propose a specific instantiation of the function f in eq. (1) to capture the dependency identified above between the actions of the trustee and the trustor's expectations. Specifically, we model one of these trust factors as the ability of one agent, acting as a trustor, to predict the actions of another agent, acting as trustee. Therefore, among all agents, we select one to be considered as the **primary agent** who, in addition to learning its own policy, learns how to predict the actions computed by other agents. Specifically, for each trustee i, the primary agent estimates the *Trust Score* defined as the number of correctly predicted actions over the
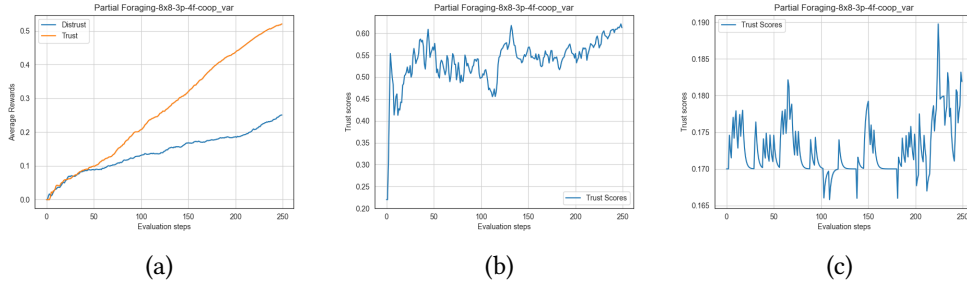
Figure 1: (a) comparison between the mean score of the primary agent when it uses the predictions from the trustable agent and the unreliable one. Trust score with respect to the (b) trustable agent and (c) unreliable agent

number of true actions. For predicting the others' actions, we adopt a simple neural network that we define **PredNet**, which is trained in a supervised fashion and that takes as input the observations of other agents and returns a prediction about their actions. The action predicted by the PredNet is concatenated to the state of the primary agent; this allows us to influence its decisions based on other agents' intentions. All agents used in the experiment are trained in a decentralized way through a MARL algorithm called Independent PPO [12].

3. Experiments

The environment used is a customized version of the Level-based Foraging (LBF) [13]. This is a grid-world multi-agent environment in which agents should navigate and cooperate to collect food, which can be collected only if the sum of the levels of agents is equal to or higher than the level of the food. In our experiments, the primary agent selects which agent to communicate with among the agents in its field of view based on their trust score learned during the training. We did some experiments in a *three agents* environment, where the other two agents are defined respectively as *trustable* and *unreliable*. The trustable agent executes its actions according to its learned policy, and its goal is to cooperate with the primary agent. On the other side, the unreliable agent performs actions according to a *bad policy* that, with a certain probability, leads to incorrect action performed. The results of the experiment are shown in Figure 1. Here, the trust score with respect to the trustable agent (b) is much higher than the one referred to the unreliable agent (c), and additionally, the average return of the primary agents is drastically better when relying on the former (a). In conclusion, we showed that using the trust score as a mechanism to select which agent to communicate with improves the performance in the case where an agent is not reliable.

4. Acknowledgments

References

[1] D. Parekh, N. Poddar, A. Rajpurkar, M. Chahal, N. Kumar, G. P. Joshi, W. Cho, A review on autonomous vehicles: Progress, methods and challenges, Electronics 11 (2022). URL: https://www.mdpi.com/2079-9292/11/14/2162. doi:10.3390/electronics11142162.

[2] V. Villani, F. Pini, F. Leali, C. Secchi, Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications, Mechatronics 55 (2018) 248–266. URL: https://www.sciencedirect.com/science/article/pii/S0957415818300321. doi:https://doi.org/10.1016/j.mechatronics.2018.02.009.

[3] M. Kyrarini, F. Lygerakis, A. Rajavenkatanarayanan, C. Sevastopoulos, H. R. Nambiappan, K. K. Chaitanya, A. R. Babu, J. Mathew, F. Makedon, A survey of robots in healthcare, Technologies 9 (2021). URL: https://www.mdpi.com/2227-7080/9/1/8. doi:10.3390/technologies9010008.

[4] G. A. Zachiotis, G. Andrikopoulos, R. Gornez, K. Nakamura, G. Nikolakopoulos, A survey on the application trends of home service robotics, in: 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2018, pp. 1999–2006. doi:10.1109/ROBIO.2018.8665127.

[5] J. P. Vasconez, G. A. Kantor, F. A. Auat Cheein, Human–robot interaction in agriculture: A survey and current challenges, Biosystems Engineering 179 (2019) 35–48. URL: https://www.sciencedirect.com/science/article/pii/S1537511017309625. doi:https://doi.org/10.1016/j.biosystemseng.2018.12.005.

[6] A. Dahiya, A. M. Aroyo, K. Dautenhahn, S. L. Smith, A survey of multi-agent human–robot interaction systems, Robotics and Autonomous Systems 161 (2023) 104335. URL: https://www.sciencedirect.com/science/article/pii/S092188902200224X. doi:https://doi.org/10.1016/j.robot.2022.104335.

[7] S. Shahrdar, L. Menezes, M. Nojoumian, A survey on trust in autonomous systems, Advances in Intelligent Systems and Computing 857 (2019) 368–386. URL: https://link.springer.com/chapter/10.1007/978-3-030-01177-2_27. doi:10.1007/978-3-030-01177-2_27/TABLES/4.

[8] P. Hancock, T. T. Kessler, A. D. Kaplan, K. Stowers, J. C. Brill, D. R. Billings, K. E. Schaefer, J. L. Szalma, How and why humans trust: A meta-analysis and elaborated model, Frontiers in psychology 14 (2023).

[9] D. Gambetta, Can We Trust Trust?, Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford (2000) 213–237.

[10] D. S. Bernstein, R. Givan, N. Immerman, S. Zilberstein, The Complexity of Decentralized Control of Markov Decision Processes, Mathematics of Operations Research 27 (2002) 819–840. URL: https://pubsonline.informs.org/doi/10.1287/moor.27.4.819.297. doi:10.1287/moor.27.4.819.297.

[11] F. Frattolillo, N. Brandizzi, R. Cipollone, I. Luca, Towards computational models for reinforcement learning in human-ai teams, 2nd International Workshop on Multidisciplinary Perspectives on Human-AI Team Trust (2023). URL: https://ceur-ws.org/Vol-3634/paper9.pdf.

[12] C. S. de Witt, T. Gupta, D. Makoviichuk, V. Makoviychuk, P. H. S. Torr, M. Sun, S. Whiteson, Is independent learning all you need in the starcraft multi-agent challenge?, CoRR

abs/2011.09533 (2020). URL: https://arxiv.org/abs/2011.09533. arXiv:2011.09533.

[13] F. Christianos, L. Schäfer, S. V. Albrecht, Shared experience actor-critic for multi-agent reinforcement learning, in: Advances in Neural Information Processing Systems (NeurIPS), 2020.