MDP-Based Analysis of Agent Interactions: From Collaborative to Adversarial Dynamics

Jurij Ružejnikov, Tatiana Valentine Guy

Czech University of Life Sciences Prague, Kamýcká 129 165 00 Praha - Suchdol, Czech Republic, Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic

{ruzejnikov, guy}@utia.cas.cz

Abstract

In multiagent systems (MAS), agents often share policy information to influence one another's decisions. Agent interactions can be categorized as either adversarial or cooperative, and these behaviors can be intentional or unintentional. In the intentional case, agents may share misleading policies to either hinder or support other agents' decision-making, whereas in the unintentional case, the interaction is merely incidental. From a single-agent perspective, the agent must be able to adapt to various interaction types. This work models MAS using the Multiagent Markov Decision Process (MMDP) and introduces a necessary condition for both intentional cooperative and adversarial interactions. We classify possible policy communications by their truthfulness and intent, and we lay the groundwork for a dynamic, trust-based framework that allows an agent to evaluate and incorporate shared policy information. The proposed approach enables robust and adaptive behaviour in both cooperative and adversarial environments.

Keywords

Markov decision process (MDP), multiagent systems, trust modelling, policy inference, agent interaction dynamics, information fusion

1 Introduction

Multiagent systems (MAS) consist of multiple autonomous entities, known as agents, that interact within a shared environment. Each agent follows a policy that determines its behaviour in the environment. These policies can be communicated to other agents to influence their decision-making process. Agents operating in a shared environment are able to interact in various ways, as seen in many real-world scenarios (examples are provided in Section 2.1). Such interactions can be categorised as cooperative, where agents work together towards common goals, or adversarial, where agents have conflicting objectives. Furthermore, interactions can be intentional or non-intentional. In the case of intentional interactions, agents may share a policy different from the one they actually follow in order to influence the behaviour of others. From the perspective of a single agent, the ability to distinguish and model these interaction dynamics is crucial for designing effective decision-making strategies in MAS applications, such as robotics [1], economics [2], or distributed computing [3]. In this work, we adopt the Multiagent Markov Decision Process (MMDP) within a setting where agents exchange policy information. We present a necessary condition for intentional interaction types and lay the groundwork for how a single agent can dynamically integrate the shared policy with its own estimate of another agent's policy, while accounting for the possibility that the shared policy may be misleading.

2 Main contribution

Popular approach to modelling decision-making in MAS is the Multiagent Markov Decision Process (MMDP). An MMDP provides a mathematical framework for modelling decision-making in situations where outcomes are partly random and partly under the control of the decision-makers. In this work, we formally define the MMDP as a discrete-time stochastic process represented by the following sequence:

$$(\mathcal{S}, \mathcal{A}, \mathcal{T}, \{\mathcal{R}_i\}_{i=1}^{n}, \{\pi_{i,t}\}_{i=1}^{n})_{t=0}^{\infty}\,, \tag{1}$$

where:

- $\mathcal{S}$ is a discrete set of states s, representing all possible configurations of the environment, shared by all of the agents;
- $\mathcal{A}$ is the discrete set of actions a available to all agents;
- $\mathcal{T} : \mathcal{S} \times \prod_{i=1}^{n} \mathcal{A}_i \rightarrow \mathcal{P}(\mathcal{S})$ is the system transition model defining the probability distribution over next states s' given the current state s and the joint actions of all agents $\mathbf{a} = \{a_1, a_2, \ldots a_n\}$.
- $r_i : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ is the reward function of agent i, which remains unchanged over time, and represents the reward that the decision maker receives in given state s when action a is implemented and the environment transitions to state s';
- $\pi_{i,t} : \mathcal{S} \rightarrow \mathcal{A}_i$ is the policy of agent i at time step t, mapping each state $s \in \mathcal{S}$ to an action $a_i \in \mathcal{A}_i$. The policy $\pi_{i,t}$ defines the agent's behaviour in the environment, determining the action to be taken in each state and may change over time.

For simplicity let us consider a two agent setup, the following diagram illustrates the interaction between agents and the environment at time step t.

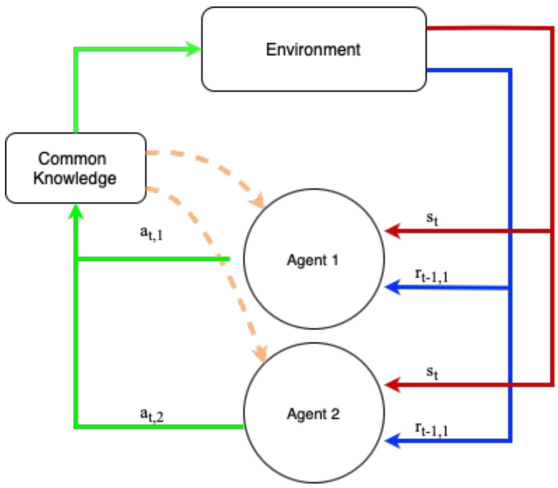

Figure 1: Diagram illustrating the interaction dynamics in a two-agent MMDP.

Additionally, no agent has knowledge of the other's reward function, nor can it observe the rewards received by other agents. Agents are only aware of the current state of the environment and can observe other agents' actions only after they have been executed.

2.1 Agent interaction categorisation

We distinguish all possible combinations of policy communication between agents based on the truthfulness of the communicated policy and its actual intent.

- **Intentional**
  - **Cooperative**
    - *Truthful communication of a cooperative policy (cooperative intent).*
      Example: Two autonomous vehicles approach an intersection. One truthfully communicates its intention to yield, allowing the other to proceed safely.
    - *Deceptive communication of a cooperative policy (adversarial intent).*
      Example: In a multi-robot exploration task, Robot A falsely claims it will cover a high-value region of the environment. Robot B, trusting this information, reallocates its path to a lower-value region to avoid redundancy. As a result, the high-value region remains unexplored, reducing Robot B's ability to maximise its reward.
  - **Adversarial**
    - *Truthful communication of an adversarial policy (adversarial intent).*
      Example: In a competitive trading scenario, a bot announces its intention to undercut prices, which it then follows through on to reduce competitors' profits.

  - *Deceptive communication of an adversarial policy (cooperative intent).*
    Example: In a multi-faction strategy game, Agent A publicly declares an aggressive policy toward Faction B, prompting B to reinforce its defenses. However, A is secretly collaborating with B to mislead a third faction, C, into believing that A and B are at odds. By diverting C's attention and resources away from a real threat, this deception ultimately supports A and B's shared objective of undermining C, illustrating a fundamentally cooperative (though deceptive) intent.

- **Non-intentional**: Agents' actions may result in non-intentional cooperative or adversarial interactions. In such cases, no deliberate misdirection occurs.
  Example: Two warehouse robots independently optimise for efficiency and unintentionally block each other's paths, causing delays. The resulting conflict is not deliberate.

2.2 Necessary condition for intentional interaction

We define a *goal-based state space* $\mathcal{S}_g \subset \mathcal{S}$ as the subset of states in which the reward function $r(s, a, s')$ is non-constant with respect to its arguments. Formally, for some state $s \in \mathcal{S}_g$, there exists at least one combination $(a, s', b, x') \in \mathcal{A} \times \mathcal{S} \times \mathcal{A} \times \mathcal{S}$ such that $r(s, a, s') \neq r(s, b, x')$. A necessary condition for intentional cooperative or adversarial interaction is that the goal-based state spaces of the agents must overlap. If these sets are disjoint, i.e., $S_{g,1} \cap S_{g,2} = \emptyset$, then no intentional cooperative or adversarial interaction can occur. In plain terms, if two agents have completely separate goals, neither has an incentive to cooperate nor to sabotage the other. Consequently, if the goal-based state spaces are disjoint, any information about an agent's policy $\pi_{i,t}$ cannot improve the other agent's decision-making task. This condition is useful for detecting situations when no adversarial or cooperation interaction occurs, particularly if and agent is equipped with reward structure estimation such as inverse reinforcement learning [4, 5].

2.3 Dynamic Trust-Based Policy Incorporation

In practical settings, agents may switch from non-intentional to intentional interaction (and vice versa), and thus must adapt to changing interaction types. From a single-agent perspective in described MMDP setting, the agent (for purposes of this section referred to as *the supported agent*) observes other agents actions after execution. Consequently, the supported agent can dynamically model the other agent's policy as the sequential decision-making process unfolds, enabling adaptive responses to changing interaction types. In this setting, agents also share their policies $\pi_{i,t}$. Incorporating a received policy into the supported agent's own decision-making can enhance its ability to achieve its goals. However, as noted, the other agent may share a policy different from the one it actually implements. To address this, we propose that the supported agent merge its internally estimated model of the other agent's policy with the communicated policy using dynamic probability-fusion methods [6, 7]. Moreover, to assess the truthfulness of the shared policy, we propose extending these fusion methods to include a similarity–dissimilarity measure, comparing the supported agent's estimate of the other agent's policy with the policy that is actually provided. A solution for dynamic merging probabilities while accounting for the changing reliability of the information source can be found in [8].

3 Conclusions

In this work, we categorized agent interactions in dynamic multiagent systems (MAS) using the MMDP framework, examining their truthfulness and intent. We then derived a necessary condition for intentional adversarial or cooperative interaction. Next, we proposed a method for incorporating shared policy information into the supported agent's estimate of another agent's policy, along with a trust-based extension that enables the supported agent to adapt to different interaction types. In the future, we plan to formalize these concepts in a unified framework and extend our approach to partially observable settings.

References

[1] Z. H. Ismail, N. Sariff, and E. G. Hurtado, "A survey and analysis of cooperative multi-agent robot systems: challenges and directions," *Applications of Mobile Robots*, vol. 5, pp. 8–14, 2018.

[2] P. Čech, P. Tučník, V. Bureš, and M. Husáková, "Modelling complexity of economic system with multi-agent systems," in *International Conference on Knowledge Management and Information Sharing*, vol. 2. SCITEPRESS, 2013, pp. 464–469.

[3] M. Štula, D. Stipaničev, and L. Šerić, "Multi-agent systems in distributed computation," in *Agent and Multi-Agent Systems. Technologies and Applications*, G. Jezic, M. Kusek, N.-T. Nguyen, R. J. Howlett, and L. C. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 629–637.

[4] S. Bergerson, "Multi-agent inverse reinforcement learning: Suboptimal demonstrations and alternative solution concepts," 2021. [Online]. Available: https://arxiv.org/abs/2109.01178

[5] M. Babes, V. Marivate, K. Subramanian, and M. L. Littman, "Apprenticeship learning about multiple intentions," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 897–904.

[6] A. Quinn, M. Kárný, and T. V. Guy, "Optimal design of priors constrained by external predictors," *International Journal of Approximate Reasoning*, vol. 84, pp. 150–158, 2017.

[7] P. K. Varshney, "Distributed detection and data fusion," *Signal Processing Series*, 1997.

[8] M. Kárný and F. Hůla, "Fusion of probabilistic unreliable indirect information into estimation serving to decision making," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 12, pp. 3367–3378, 2021.