Communicating Intentions for Coordination with Unknown Teammates

(Extended Abstract)

Trevor Santarra and Arnav Jhala
University of California Santa Cruz
{tsarratt,jhala}@soe.ucsc.edu

ABSTRACT

Ad hoc multiagent teamwork introduces the challenge of coordinating with a variety of potential teammates, including teammates with unknown behavior. We examine the communication of policy information for enhanced coordination between such agents. The proposed decision-theoretic approach examines the uncertainty within a model of an unfamiliar teammate, identifying and acquiring policy information valuable to the collaborative effort.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Cooperation and coordination

General Terms

Experimentation

Keywords

Ad Hoc Teams; Teamwork; Coordination

1. INTRODUCTION

Coordinating a team of autonomous agents necessitates that agents must act in such a way that progresses toward the achievement of a goal while avoiding conflict with their teammates. In traditional multiagent systems literature, these teams of agents share an identical design for reasoning, planning, and executing actions, allowing perfect modeling of teammates. Ad hoc teamwork [4] further complicates this problem by introducing a variety of teammates with which an agent must coordinate. In these scenarios, one or more agents within a team can be unfamiliar, having unknown planning capabilities guiding their behavior.

Much of the existing ad hoc teamwork research focuses on reinforcement learning and decision theoretic planning. Agents employ models of known behavior to predict teammates' actions, using decision theory to maximize expected utility in instances where the predicted actions are uncertain [1]. Online learning refines these models with observations of behaviors during execution, increasing the accuracy of the models' predictions, permitting the team to coordinate more effectively [2, 3].

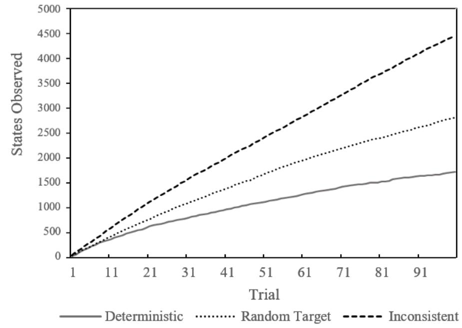
Our approach addresses the problem of planning under teammate behavior uncertainty by incorporating intentional multiagent communication within ad hoc teams. In partially observable multiagent domains, agents must share information regarding aspects of the environment such that uncertainty is reduced across the team, permitting better coordination. Similarly, we consider how communication may be utilized within ad hoc teams to resolve behavioral uncertainty. Querying a teammate for its intended actions in various states allows an agent to adjust its predictions of the teammate's course of action. With more accurate predictions of team behavior, an agent can better select its actions to support team cohesion. In contrast to traditional multiagent communication applications where communicative acts are few in number, we allow the agent to consider the entire set of states encountered when planning. In short, an ad hoc agent coordinating with an unknown teammate can identify uncertainties within its own predictive model of teammate behavior then request the appropriate policy information, allowing the agent to adapt its individual policy.

2. APPROACH

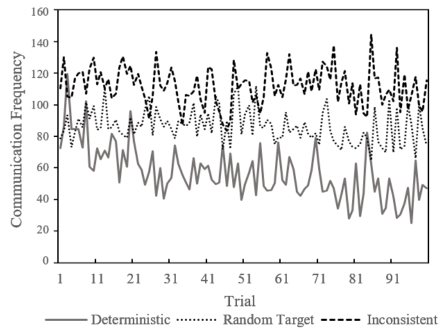
Having complete knowledge of a teammate's intended behavior would allow optimal planning on the part of the coordinating agent. However, online learning of a complete model of a teammate's behavior within relatively complex domains with thousands or millions of states is infeasible during limited instances of coordination. As Barrett et al. [1] observed, ad hoc agents can often collaborate effectively when observing the behavior of an unknown team in a comparatively small, recurring section of the domain's state space. This is particularly true in cases where the team attempts repeated trials with static initial conditions. It is, however, beneficial for agents to reason over what information about teammates they already possess and evaluate what subset of the team's behavior would be beneficial to know when coordinating in uncommon states.

Decision theory provides a mechanism for determining the expected change in expected utility for acquiring new information, as shown by Equation 1. During evaluation of communicative acts, for each potential response for an intention query, the agent reevaluates its expected utility (denoted by V') for both its original policy, π, as well as an updated policy, π', constructed under the new information. The difference in these two utilities is the increase in expected utility the agent would receive given the information in advance,

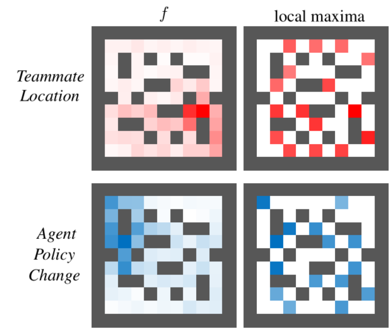
Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*



(a) Unique states encountered across successive trials.



(b) Average query frequency across successive trials.



(c) A heatmap of query frequencies, depicting the teammate's location and the corresponding location of the change in the agent's policy.

allowing the agent to preemptively adapt its policy.

$$U_{Comm}(s_t) = \sum_{a \in A_{s_t}} Pr(a|s_t) \left(V'_{\pi'|a}(s_0) - V'_{\pi|a}(s_0) \right) \quad (1)$$

3. SELECTED EXPERIMENTS

We test the communicating ad hoc agent in a variant of the multiagent pursuit domain, where a team of agents coordinate to trap one of several fleeing prey within a maze. The ad hoc agent in the team possesses a basic observation-based learning capacity, initially uniformly predicting the actions of its teammate. This model is updated by observing the teammate and predicting future actions with probability proportional to the frequency count with Laplace smoothing. In order to ensure a degree of uncertainty in the paired teammate's behavior, we test the coordinating ad hoc agent with a range of teammate types, varying in degrees of consistency of behavior, as follows:

1. Deterministic - This teammate consistently selects its target across trials and pursues it in an identical, deterministic manner each round.
2. Random Target - Here, the teammate begins each trial by uniformly sampling which target it will pursue.
3. Inconsistent - This agent periodically chooses an action at random as well as occasionally switches targets.

In contrast to existing work where unknown teammates may have largely deterministic behaviors from static initial conditions, we observe that more varied teammate behavior has a significant impact on the quantity of unique states encountered by the team. Uncommon states are consequently unlikely to have sufficient observations for the learned model to accurately predict teammate behavior. Analysis of the communication frequency to resolve intention uncertainty found that over time coordination attempts with the deterministic and random target teammate types resulted in decreased communication over successive trials. However, when the ad hoc agent coordinated with the inconsistent teammate, communication frequency was, on average, greater than for other teammate types and did not exhibit a statistically significant decline over time.

Furthermore, the relative frequencies of states queried for the teammate's intended actions demonstrate a bias toward states nearer to the initial state, where intention resolution may dictate which of three prey should be pursued. In addition, the local maxima of the state illustrate the agent correctly identifies branches within the maze as more commonly valuable to communicate over non-branch positions.

Teammate	Spearman's ρ	p
Deterministic	-0.627	<0.001
Random Target	-0.295	0.003
Inconsistent	-0.130	0.196

Table 1: Monotonicity of communicative frequency.

4. CONCLUSION

When cooperating with unknown teammates, agents with the capability to query policy information can act in a proactive manner, acquiring valuable team behavior information. Restricting an agent to learning purely through observation requires either a sufficient number of observations of teammate behavior or the ability to generalize predictions from prior experience [2]. The communication of intentional information complements learning agents, as such communicative behavior both requires a reflective analysis of the uncertainty within an existing teammate model as well as proactively expands the information an agent possesses about its teammates.

REFERENCES

[1] S. Barrett, P. Stone, and S. Kraus. Empirical evaluation of ad hoc teamwork in the pursuit domain. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 567–574. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[2] S. Barrett, P. Stone, S. Kraus, and A. Rosenfeld. Teamwork with limited knowledge of teammates. In *AAAI*, 2013.

[3] T. Sarratt and A. Jhala. Tuning belief revision for coordination with inconsistent teammates. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.

[4] P. Stone, G. A. Kaminka, S. Kraus, J. S. Rosenschein, et al. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *AAAI*, 2010.