Too Many Cooks: Coordinating Multi-agent Collaboration Through Inverse Planning*

Extended Abstract

Rose E. Wang*
MIT
rewang@mit.edu

Sarah A. Wu*
MIT
sarahawu@mit.edu

James A. Evans
U. Chicago
jevans@uchicago.edu

Joshua B. Tenenbaum
MIT
jbt@mit.edu

David C. Parkes
Harvard
parkes@eecs.harvard.edu

Max Kleiman-Weiner
Harvard, MIT, & Diffeo
maxkleimanweiner@fas.harvard.edu

ABSTRACT

Humans collaborate in dynamic and flexible ways. Collaboration requires agents to coordinate their behavior on the fly, sometimes jointly solving a single task together and other times dividing it up into sub-tasks to work on in parallel. We develop *Bayesian Delegation*, a learning mechanism for decentralized multi-agent coordination that enables agents to rapidly infer the sub-tasks that other agents are working on by inverse planning. These inferences enable agents to determine, in the absence of communication, whether to plan jointly with others or work on complementary sub-tasks. We test this model in a suite of decentralized multi-agent environments inspired by cooking problems. To succeed, agents must coordinate both their high-level plans (sub-task) and their low-level actions (avoiding collisions). Including joint sub-tasks in the prior of Bayesian delegation enables agents to carry out sub-tasks that neither agent can finish independently. The full system outperforms lesioned systems that are missing one or more of these capabilities.

KEYWORDS

multi-agent coordination; inverse planning; Bayesian inference

ACM Reference Format:
Rose E. Wang*, Sarah A. Wu*, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner. 2020. Too Many Cooks: Coordinating Multi-agent Collaboration Through Inverse Planning. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION AND ENVIRONMENT

Real world collaboration is challenging as it requires people to coordinate their behaviors. In this work, we build collaborative agents that coordinate to solve hierarchical tasks inspired by the video-game Overcooked [2]. These problems are challenging because of the variation that is present between problems: the specifics of the environment and task objectives can drastically change the coordination strategy, so successful collaboration requires flexible and abstract mechanisms. Specifically, our agents unify solutions to three challenges of coordination: (A) Divide and conquer: agents work in parallel when sub-tasks can be carried out individually, (B) Cooperation: agents work together on the same sub-task when required or most efficient, (C) Spatio-temporal movement: agents avoid getting in each others way while working separately or together.

We study decentralized Markov decision processes (Dec-MDPs) with a partial order of sub-tasks over object-object interactions [3, 4]. We use this formalism to develop a test suite based on simple kitchen cooking tasks where sub-tasks are different parts of a non-linear recipe. The left column of Figure 1 shows the different kitchen layouts and the top row shows the different recipes which can be composed together. Agents move simultaneously in any cardinal direction or can stay still. The kitchens have counters that contain movable foods and plates and immovable knives. Agents pick-up objects by moving into them and can put them down on counters by moving into it. The goal is to deliver the completed recipe to the star. Agents optimize reward by completing delivery in as few time steps and movements as possible.

2 MODEL

We introduce a novel a hierarchical planning algorithm for multi-agent coordination based on probabilistic inference over sub-tasks. At a *high-level*, each agent must decide which sub-task they should do next. We develop a new algorithm called *Bayesian Delegation* which enables agents to probabilistically take into account the unobserved intentions of other agents in order to dynamically decide whether to divide and conquer on different sub-tasks or cooperate on the same one. At a *low-level*, agents use model-based reinforcement learning to find approximately optimal policies for a specific sub-task. Planning is decentralized at both levels i.e., each agent plans and learns for itself without sharing any information with others. This contrasts with existing work on multi-agent task allocation, which rely on communication or pre-coordination to centralize beliefs or share the global state [1, 7].

High-level planning (sub-task): *Bayesian Delegation* uses theory-of-mind action understanding to plan over sub-tasks. Under Bayesian Delegation, ta is the set of all possible allocations of agents to sub-tasks where all agents are assigned to a sub-task. For example if there are two possible tasks ($[\mathcal{T}_1, \mathcal{T}_2]$) and two agents ($[A_1, A_2]$), then $ta = [(A_1 : \mathcal{T}_1, A_2 : \mathcal{T}_2), (A_1 : \mathcal{T}_2, A_2 : \mathcal{T}_1), (A_1 : \mathcal{T}_1, A_2 : \mathcal{T}_1), (A_1 : \mathcal{T}_2, A_2 : \mathcal{T}_2)]$ where $A_1 : \mathcal{T}_1$ means that agent A_1 is assigned to sub-task $\mathcal{T}_1$. Thus ta includes the both the possibility that agents

*indicates equal contribution

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

will divide and conquer (work on separate sub-tasks) as well as the cooperate (work on the same sub-task). Each element of ta is a high-level plan. If all agents pick the same element, they will easily coordinate. However, in our environments, this was not possible. Agents maintain uncertainty about which element of ta the group should execute, $P(ta)$.

Each time step, the agent selects the most likely sub-task allocation ta^* and plans actions according to that sub-task allocation (low-level planning): $ta^* = \arg\max_{ta} P(ta|H_{0:T})$, with $P(ta|H_{0:T})$ the posterior over ta after having observed a history of actions, $H_{0:T} = [(s_0, \mathbf{a_0}), \ldots (s_T, \mathbf{a}_T)]$ where T is the number of time steps in the history. This posterior $P(ta|H_{0:T})$ is computed at time step T according to Bayesian inference:

$$P(ta|H_{0:T}) \propto P(ta)P(H_{0:T}|ta) = P(ta)\prod_{t=0}^{T} P(\mathbf{a_t}|s_t, ta)$$

where $P(\mathbf{a_t}|s_t, ta)$ is the likelihood of action for all agents. Note that these belief updates do not explicitly consider what each agent knows about their own sub-tasks at time $T-1$. Rather, the model only considers the information that is known by all, i.e., a third-party observer would know. The likelihood is computed by inverse planning using a soft-max to account for non-optimal and variable behavior: $P(\mathbf{a_t}|s_t, ta) \propto exp(\beta * Q^*_{ta}(s, \mathbf{a}))$, where $Q^*_{ta}(s, \mathbf{a})$, the expected future reward of actions towards the completion of sub-tasks ta, is computed by low-level planning and β controls the degree to which an agent believes other agents are optimizing.

Low-level planning (action): Low-level planning takes a goal induced by the sub-task selected next by the high-level planner, $\mathcal{T}_i$, and outputs a sequence of low-level actions for the agent to execute. Agents use bounded real-time dynamic programming (BRTDP) to approximates an optimal policy, $\pi_{\mathcal{T}_i}(s)$, under the Dec-MDP induced by the chosen sub-task [8]. To avoid collisions, agents best-respond to non-strategic (level-0) models of their partners. To cooperate on a sub-task, agents simulate a fictitious joint plan and then act according to their role in that shared intention [6, 9].

3 RESULTS AND DISCUSSION

Agents play with each other in nine levels (3 kitchens x 3 recipes). Experiments were replicated with 50 random seeds and each episode was 100 time steps. The softmax in the Bayesian inference is $\beta = 0.3$. With lesioned versions of our model, we investigate the importance of Bayesian Delegation and joint planning for successful multi-agent coordination. The first lesion (NJP) has Bayesian Delegation but agent's cannot simulate fictitious joint plans, i.e., no cooperating on the same of sub-task. The second lesion (NJP+NBD) has neither. These agents make no inferences about others and optimize sub-tasks without concern for what others are doing.

Figure 1 shows empirical results for the time it takes to complete the level for all three models. On simpler planning problems (row 1, column 1), the models are comparable to each other, but when faced with more complex recipes NJP and NJP+NBD take significantly longer to finish. For column 2, NJP maxes out because the agents are unable to jointly plan, yet they can still perform inference over each other. As a result, they often simultaneously yield to each other (assuming the other will pass through) and generate a deadlock. We hypothesize that this is also why NJP+NBD performs comparable

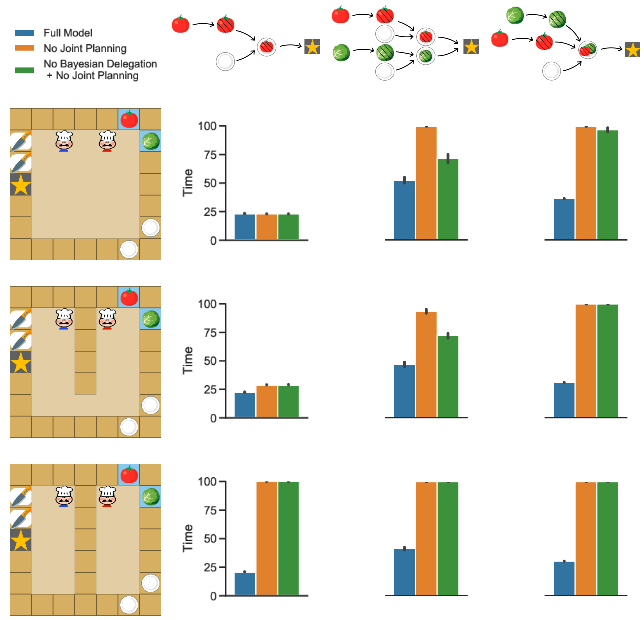

Figure 1: Performance of the Bayesian Delegation model by number of time steps until recipe completion (lower is better). The row shows the kitchen, the column shows the recipe. The full model outperforms two lesioned models. The error bars show the standard error of the mean.

or better than NJP: in certain situations, acting without regard for the other is more effective than considering the sub-tasks of other agents because it breaks symmetries.

Without the ability to reason about and for other agents, the lesioned models produce frequent spatial miscoordination in rows 1 and 2. For instance, they may inadvertently block each other from delivering separate dishes, or attempt to cross the kitchen in the lower bottleneck at the same time. Both lesions also take significantly longer than the full model because completion time multiplies with the number of sub-tasks, and without Bayesian Delegation each sub-task must be done individually.

We developed a new set of spatial and object-oriented cooking challenges that require nuanced coordination to successfully complete. In particular, subtle changes in the task or in spatial layout call for different multi-agent strategies. We developed a unified approach inspired by cognitive science, Bayesian Delegation, which rises to these challenges without communication between agents. It gives rise to many natural aspects of human cooperation, such as the emergence of and convergence of cooperative behavior when joint planning is deemed better than planning alone, as well as the natural decision to pursue one sub-task over another equally feasible sub-task. Coordination is achieved through two key mechanisms: (1) inverse planning that enables agents to rapidly infer the sub-tasks of other agents; and (2) joint planning that enables agents to mesh their intentions and complete sub-tasks in ways that neither agent could achieve on their own [5].

REFERENCES

[1] Luc Brunet, Han-Lim Choi, and Jonathan How. 2008. Consensus-based auction approaches for decentralized task assignment. In *AIAA guidance, navigation and control conference and exhibit*. 6839.

[2] Micah Carroll, Rohin Shah, Mark Ho, Thomas 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*.

[3] Daniel Claes, Philipp Robbel, Frans A Oliehoek, Karl Tuyls, Daniel Hennes, and Wiebe Van der Hoek. 2015. Effective approximations for multi-robot coordination in spatially distributed tasks. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 881–890.

[4] Carlos Diuk, Andre Cohen, and Michael L Littman. 2008. An object-oriented representation for efficient reinforcement learning. In *Proceedings of the 25th international conference on Machine learning*. ACM, 240–247.

[5] Barbara J Grosz and Sarit Kraus. 1996. Collaborative plans for complex group action. *Artificial Intelligence* 86, 2 (1996), 269–357.

[6] Max Kleiman-Weiner, Mark K Ho, Joseph L Austerweil, Michael L Littman, and Joshua B Tenenbaum. 2016. Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

[7] Mitchell McIntire, Ernesto Nunes, and Maria Gini. 2016. Iterated Multi-Robot Auctions for Precedence-Constrained Task Scheduling. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1078âĂŞ1086.

[8] H Brendan McMahan, Maxim Likhachev, and Geoffrey J Gordon. 2005. Bounded real-time dynamic programming: RTDP with monotone upper bounds and performance guarantees. In *Proceedings of the 22nd international conference on Machine learning*. ACM, 569–576.

[9] Michael Shum, Max Kleiman-Weiner, Michael L Littman, and Joshua B Tenenbaum. 2019. Theory of Minds: Understanding Behavior in Groups Through Inverse Planning. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*.