Multiagent Metareasoning Through Organizational Design (Extended Abstract)

Jason Sleight and Edmund H. Durfee
Computer Science and Engineering
University of Michigan
Ann Arbor, MI 48109, USA
{jsleight,durfee}@umich.edu

ABSTRACT

We describe an approach to multiagent metareasoning that uses organizational design to focus each agent's reasoning on the aspects of its local problem to which it can make the most worthwhile contributions to joint behavior. We summarize an organizational design problem that explicitly considers the quantitative impact that a design has on both the quality of the agents' behaviors and their reasoning costs. We overview an automated organizational design process that can approximately solve our design problem via incremental search, and outline techniques that efficiently estimate the incremental impact of a candidate organizational influence.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Coherence & Co-ordination, Multiagent Systems*

General Terms

algorithms, performance

Keywords

organizational design; multiagent metareasoning

1. INTRODUCTION

When agents operate in large, complex, and dynamic problem domains, the amount of computation time needed to make provably optimal decisions can exceed the time available before action must be taken. Research into metareasoning—reasoning about reasoning—studies mechanisms that agents can use to make principled decisions about whether the benefits of additional reasoning to make better decisions are expected to outweigh the costs of delaying enacting decisions. (See Cox and Raja [2] for a thorough discussion of work in this area.) Metareasoning becomes even more complicated in multiagent settings, since the benefits of additional reasoning might depend on the reasoning and behaviors of other agents. For example, if one agent assumes responsibility for (reasoning about) performing a task, then there could be no benefit for other agents to also reason about that task.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

In contrast to typical metareasoning approaches that try to dynamically predict the net benefit of additional reasoning, the fundamental idea of our approach is to have an organizational design process (ODP) utilize a global view of the problem domain to identify high-performing, long-term behavior patterns (both within and across problem instances), and then influence the agents to avoid *even thinking* about behaving counter to those patterns. For example, the ODP might identify that certain tasks should typically be the responsibility of one agent, and codify this pattern by preventing other agents from considering those tasks. Such influences trade computational speedup (since other agents consider smaller task spaces) for reduced performance quality (in cases where another agent should perform the now-removed tasks).

To quantitatively measure the expected performance of an organization, we consider an organizational influence as a modification to an agent's local problem description that either constrains the agent's local policy space, or re-prioritizes the agent's preferential ordering over its local policy space. We define an organization, Θ, as a set of such influences for each agent. The performance of Θ is determined by the expected quality of the agents' behaviors w.r.t. Θ, which we refer to as the operational reward, $OpR(\Theta)$, and the agents' expected computational costs to calculate those behaviors, which we refer to as the operational reasoning costs, $OpC(\Theta)$. An ODP's objective is to create an organization with maximal operational performance, $\Theta^* \equiv \text{argmax}_\Theta\, OpP(\Theta)$, where $OpP(\Theta)$ balances (the optimal balance is defined by the problem domain) $OpR(\Theta)$ and $OpC(\Theta)$. Unfortunately, the space of possible Θs (a distinct Θ for every permutation of every subspace of the joint policy space) is intractably large even for simple domains, making direct enumeration infeasible.

2. ORGANIZATIONAL DESIGN PROCESS

Since enumerative search is infeasible, we instead focus on incrementally searching the organizational design space to create an approximately optimal design. A key observation for efficient incremental search is that the operational performance of a candidate organization can be factored into the performance of an individual influence, Δ, and the performance of the current organization, Θ (from the previous search iteration). That is, by computing only the conditional, incremental impact of Δ w.r.t. Θ, an ODP can avoid redundantly computing Θ's contribution to the operational performance. In our problem formulation, an ODP must compute the incremental impact of Δ on both the operational reward, OpR, and operational reasoning costs, OpC.

Our methodology for computing Δ's incremental impact on OpC is to identify the marginal cost of adding a state and/or edge to an agent's planning problem, and then compute how Δ alters the number of states and edges. We developed a methodology for empirically estimating the marginal cost of a state and/or edge by having an agent solve problems with two related domain models: its original local model, and a modified version of that model which contains the minimal number of edges such that the original reachable state space and optimal local policy are each preserved. The relative difference between the computational costs using these two models estimates an edge's marginal cost. A state's marginal cost can be estimated by having the agent solve various problems (i.e., with different numbers of states) using the modified version of its local model since that model disentangles the costs associated with states and the edges connecting them. Each type of influence has a well-defined impact on the number of an agent's states and edges (e.g., a Δ preventing an agent from considering an action removes an edge for each possible successor state of that action, and removes any now-unreachable states).

Δ's incremental impact on OpR is determined by how the quality of the agents' joint behavior changes by adding Δ into Θ. In principle, an ODP could calculate this by determining the agents' joint policy w.r.t. $\Theta + \Delta$, and then measuring the expected quality of those behaviors; however, such an approach is computationally daunting given the complexity of computing joint policies and the possibly high number of search iterations. Instead, the insight we exploit is that an ODP can use its global view to compute or estimate (e.g., via sampling problem instances and/or using an abstract domain model) the optimal joint policy, and then only consider candidate designs that preserve this policy while steering agents away from taking, and even considering, behaviors outside of this policy. While this methodology (unavoidably) requires the ODP to determine what good behaviors are by calculating an optimal joint policy, the ODP only need do this costly calculation once, and then amortize those costs over all of the search iterations, which results in substantial computational savings.

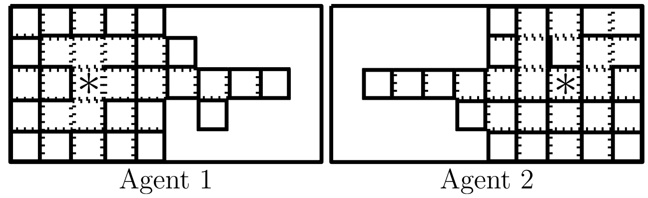

Figure 1: Example movement action influences our ODP creates. An agent can move into a cell in a direction where it first passes a dotted line, but not a solid line. $*$ designates the agent's initial position.

3. DISCUSSION

To illustrate how our ODP's designs impart a desired metareasoning regime upon the agents, we utilized a simplified firefighting grid world [4], where agents move through a grid to put out fires, and limited the ODP to constraining the agents' actions. We found that our ODP was able to encode surprisingly nuanced organizational designs despite our limiting it to only action constraints. For example, the ODP frequently imposes unidirectional movements (see Figure 1), where an agent is allowed to consider moving into a cell, but the action to move back and in effect "undo" the previous action is blocked from consideration. This type of influence imparts a good metareasoning regime by forcing the agent to reason about complete, irreversible behavior trajectories rather than needlessly reasoning about reversing prior actions. These unidirectional movements also improve coordinated behavior by discouraging an agent from rushing to the other side of the grid (where the other agent is located) to fight a high-intensity fire since it would then be unable to come back and fight a fire in its initial vicinity.

Our work in this paper resides in the intersection of three fields of study: multiagent metareasoning, organizational modeling, and multiagent sequential decision making. Prior research has largely treated multiagent metareasoning as a decentralized coordination problem, where agents model each others' reasoning processes and pass pertinent information among themselves so as to decide how best to coordinate the use of their reasoning resources (see [2] for a recent overview). In contrast, the work we present here centralizes the problem in the ODP, amortizing the costs associated with centralization by constructing long-term metareasoning regimes about which parts of the joint problem are worthwhile for each agent to reason about given its ongoing organizational role.

Organizational modeling research (see [3] for a recent overview) has typically focused on how to define the roles, norms, interaction protocols, etc. that agents should follow, and thus might simplify agent reasoning by focusing agents on considering particular tasks and interactions. Our work, while so far lacking in the richness of modeling constructs considered in much organizational modeling research, provides a basis for raising this otherwise overlooked impact of organizations on agent reasoning to explicit consideration.

Given the general intractability of optimally solving decentralized decision problems, multiagent sequential decision making research has investigated a variety of algorithmic techniques for approximating, simplifying, and decoupling agents' reasoning (see [1] for a recent overview). Rather than directly contributing to this body of techniques, our work instead emphasizes a strategy for analyzing patterns of joint behavior to selectively modify the problems agents solve. This idea has been used before to bias agents to separately find solutions that have joint benefit [1, 4], but that prior work did not explicitly factor quantitative impacts on agent reasoning when designing modifications to agents' local models.

4. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their suggestions. This work was supported by NSF grant IIS-0964512.

5. REFERENCES

[1] C. Amato, G. Chowdhary, A. Geramifard, N. K. Ure, and M. J. Kochenderfer. Decentralized control of partially observable Markov decision processes. In *The 52nd IEEE Conference on Decision and Control*, 2013.

[2] M. T. Cox and A. Raja. *Metareasoning: Thinking about thinking*. MIT Press, 2011.

[3] V. Dignum and J. Padget. Multiagent organizations. *Multiagent Systems, G. Weiss, ed., MIT Press*, 2012.

[4] J. Sleight and E. H. Durfee. Organizational design principles and techniques for decision-theoretic agents. In *AAMAS*, pages 463–470, 2013.