Task Selection for Human-Robot Teams in Dynamic Environments

Andrew J. Smith[1], and Geoffrey A. Hollinger[1]

This paper introduces the preliminary work developing Informed-Monte Carlo Tree Search (I-MCTS), a novel decentralized task assignment algorithm for combined multi-robot and human teams. This work builds upon ideas introduced in recent decentralized multiagent planning algorithms, such as [1] [2], by incorporating inferences about human team members to identify a joint policy that accounts for uncertainty in human team members. The joint policy consists of each robot's policy and the inferred policies of what human agents are likely to do. Humans act independently of the inferred policy while their actions are used to update each models of the human user. Robots then use a combination of the human model and broadcasts by other robots to determine the likely actions of other agents (human and robot) and subsequently, determine their own policies.

To infer human behaviors we represent the human agent's planning expertise as a Partially Observable Markov Decision Process (POMDP). This is done by maintaining a belief space of the human's ability to select tasks. Then, using that belief, each robot infers the probability of the human deviating from the optimal joint policy and alters its own policy to account for the inferred actions of the human user. While the POMDP formulation approach has been used to assist humans teleoperating AUVs [3] it is a novel addition to heterogeneous human-robot team planning.

For generality, the problem domain considers heterogeneous multi-robot teams interacting with multiple human users with varying task types. The objective is to minimize the cumulative task cost, where each time step a task is both available and incomplete a cost is incurred. Types of robots have different times to complete each type of task. Robots use MCTS to plan their policy using difference rewards to evaluate their contributions. After a predetermined time policies are broadcast to other agents. Each policy represents the probability of completing tasks by a certain time. In parallel, robots use MCTS to identify a policy for the human user. Then, the human's policy and model are used to infer the probability of the human user completing each task. Each robot then updates the expected benefit of each action using the probability of another agent, human or robot, completing tasks before them. This process is then iterated for a fixed amount of time or until convergence.

A preliminary result demonstrates the feasibility of the proposed method, Figure 1. For this example, a human user was coupled with four robots of three types and three types of tasks. Each robot had a different movement speed and task completion times. Each task had a different reward for completion. The simulation occurred in an environment with random and periodically generated tasks. Neither the robots or human agents are aware of the task generation schedule in advance. Users selected tasks for an avatar to complete for a fixed number of time steps while the robot assisted and the total cost was recorded. As baselines for comparison against the proposed I-MCTS with POMDP inference to infer human actions; MCTS optimal human task selection, random human task selection, and greedy human task selection models are tested. For each test iteration the team of robots is given the tested model of human action selection and the cumulative cost of incomplete tasks is counted for 30 seconds. Planners were evaluated by two users who attempted to act as an expert, POMDP gives mean $P(Expert) = 0.882$ for all trials, and a beginner, $P(Expert) = 0.712$, for three iterations with each algorithm. Notice that the I-MCTS planner was able to outperform the baselines for the beginner user and similar for the optimal model for the best performing for the expert user.

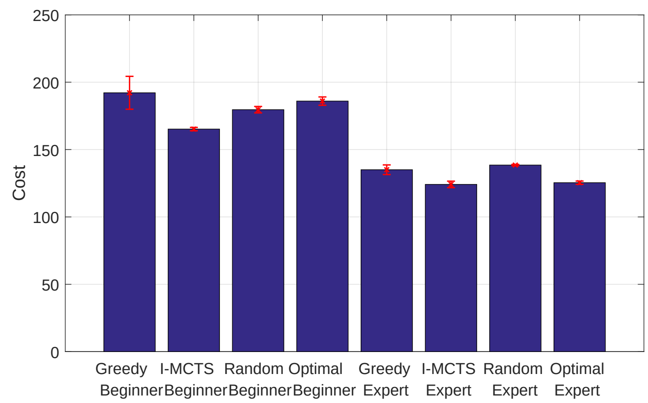

Fig. 1. Results for user trials demonstrate that including inferences about human teammate's ability to plan can improve the combined output of human-robot teams. The I-MCTS agent using a POMDP to infer human actions is able to either match or outperform the baseline models with beginner ($P(Expert) = 0.712$) and expert ($P(Expert) = 0.882$) users.

A. J. Smith and G. A. Hollinger are with the School of MIME at Oregon State University. {smithan7, geoffrey.hollinger}@oregonstate.edu

References

[1] G. Best, O. M. Cliff, T. Patten, R. R. Mettu, and R. Fitch, "Decentralised monte carlo tree search for active perception," in *Proc. of WAFR*, 2016.

[2] C. Amato, G. Konidaris, A. Anders, G. Cruz, J. P. How, and L. P. Kaelbling, "Policy search for multi-robot coordination under uncertainty," *The International Journal of Robotics Research*, vol. 35, no. 14, pp. 1760–1778, 2016.

[3] L. Milliken and G. A. Hollinger, "Modeling user expertise for choosing levels of shared autonomy," in *Proc. Planning for Human-Robot Interaction Shared Autonomy and Collaborative Robotics Workshop, Robotics: Science and Systems Conference*, 2016.