NRI: FND: Connected and Continuous Multi-Policy Decision Making

PI: Edwin Olson, University of Michigan

We're making robot teams better at coordinating by using real-time simulation to pick between a handful of simple strategies.

Key Problem - Our goal is to develop robot that exhibit robust behaviors under a wide range of conditions where rule-based approaches to specifying behavior are too brittle.

In this project, we're currently focusing on multi-robot coordination, including how robots pick their own behavior in order to maximize the *team* performance, and how/when they communicate with each other. The space of potential plans for a team of robots can be very large, and evaluating the quality of any given plan is very computationally costly due to the need to marginalize over uncertainty.

Solution - Our approach is based on Multi-Policy Decision Making (MPDM), in which the performance of a set of candidate policies is estimated online using a simulator; the policy with the best expected performance is "elected" to control the robot. This election cycle is repeated, allowing the robots to produce effective behavior by dynamically switching between simple polices that often perform poorly on their own.

A major advantage of this approach is that it is often possible to write policies that are simple and effective in at least *some* situations. Those simple strategies can encode long-horizon planning— far longer than can be achieved with a forward search. And critically, because the MPDM framework itself will pick the most effective policy for the given situation, it is not necessary for policies to be general-purpose. As a consequence, policies can be very simple— they can be designed to handle a particular situation well, and allowed to be lousy in other situations.

Problem Domain - A representative domain is "tag"— where a team of robots is attempting to capture an adversarial human. The position of the human can only be detected at short range, and robots can only learn about the location or policies of their teammates through unreliable communications. An example of a simple policy is for all the robots to converge upon a particular location (which can be effective in cornering the human). Another simple policy might be for the robots to explore the environment— which can result in either observing the location of the human or, at least, ruling out some locations for the human. Both of these policies reflect relatively well-coordinated, long-horizon plans. Neither is likely to be effective in capturing an adversarial human alone, but it is easy to imagine that some interleaving of these strategies (e.g. explore until you have a reasonable belief about where the human is, then have all robots converge upon that location) could be effective. The problem is determining *how and when* to switch between these policies— the best choices would likely depend on both the robots' belief states *and* the structure of the environment, making a rule-based system difficult to develop. With MPDM, the robots use online simulations to sample possible outcomes of each policy, and elect the policy with the largest expected reward.

New Contributions - We extend this MPDM framework to multi-robot teams, coordinating over unreliable radio links. We propose a "implicit consensus" approach, in which each robot's simulator assumes that all robot team-members use the same policy. When communications are working well, robots' belief states are also synchronized, so they all compute the same plan— which results in an implicit consensus. However, when communication drop outs result in divergence of belief states, robots may elect different policies according to their beliefs. A key advantage of this approach is that the robot team's performance degrades gracefully as communication degrades, and the team never halts due to a communication drop out.

Scientific Impacts- The techniques developed in this project are broadly applicable to planning problems where uncertainty or a large search space are obstacles to producing good behavior.

Broader Impacts - This work has broad applications to autonomous robots and vehicles, from cars, to in-home robots, to the factory floor. In addition to directly supporting one PhD student, this project indirectly supports a second PhD student and two undergraduate students.

Publications supported by this project:

Marcotte, Ryan J. and Wang, Xipeng and Mehta, Dhanvin and Olson, Edwin. (2019). Optimizing multi-robot communication under bandwidth constraints. *Autonomous Robots*.

Prescribing what a robot should do in every scenario isn't scalable. We "try out" a number of simple strategies in a simulator; the strategy with the highest expected reward is "elected", and controls the robot.

- You don't even have to know *why* the strategy works.
- The dynamically chosen strategy often works better than any single strategy.

We're focusing on teams of robots playing tag with a human. Long-horizon plans are needed to be successful, but large search space and uncertainty make it computationally difficult.

- Very low (and uncertain) communications budget.
- Good strategies are very sensitive to the shape of the environment and the behavior of the human