Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination

Stone, P., G. A . Kaminka, S. Kraus, and J. S. Rosenschein (2010). Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination, in the *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'2010)*, Atlanta, GA, July 11-15, pp. 1504-1509.

Introduction

Ad Hoc Team Setting

Multiple agents with *different* knowledge and capabilities find themselves in a situation such that their goals and utilities are perfectly aligned yet they have had *no* prior opportunity to coordinate.

Introduction 2

Examples

May arise among robots or software agents that have been programmed by different groups and/or at different times such that it was not known at development time that they would need to coordinate

- Rescue robots at disaster response
- E-commerce agents interacting with legacy agents

Introduction 3

A Good "Ad Hoc Team Player"

Must be adept at assessing (1) *the capabilities of other agents, especially in relation to its own capabilities*; and (2) *the other agents' knowledge states*

- If you are trained in first aid, what would you do? But if there is a doctor around, what would you do?

Must also be proficient at estimating *the effects of its actions on the other agents*

The Challenge

To create an autonomous agent that is able to efficiently and robustly collaborate with previously unknown teammates on tasks to which they are all individually capable of contributing as team members

Evaluation

An Empirical Challenge

Suppose there are two agent designs: which one is better at being an ad hoc team player?

- Plug each agent into a Domain and a set of existing agents, and measure the rewards that it achieves
- Whichever agent has a higher reward is the better design

Evaluation

Some quantitative performance measure, or “score”

Expected score

Minimum threshold expected performance (e.g., foraging vs. pushing heavy boxes)

Evaluate(a_0, a_1, A, D)

- Initialize performance (reward) counters r_0 and r_1 for agents a_0 and a_1 respectively to $r_0 = r_1 = 0$.
- Repeat:
  - Sample a task d from D.
  - Randomly draw a subset of agents B, $|B| \geq 2$, from A such that $E[s(B, d)] \geq s_{min}$.
  - Randomly select one agent $b \in B$ to remove from the team to create the team B^-.
  - increment r_0 by $s(\{a_0\} \cup B^-, d)$
  - increment r_1 by $s(\{a_1\} \cup B^-, d)$
- If $r_0 > r_1$ then we conclude that a_0 is a better ad-hoc team player than a_1 in domain D over the set of possible teammates A.

Evaluation

Breadth of the domain *D* and breadth of teammate capabilities in *A*

- Assumptions
  - Agents are aware of the domain *D*
  - Agents are aware of the set of potential teammates *A* (but *A* may have infinite cardinality)
- Agents may not be aware of teammates at all

Example Theoretical Approach

Framework of Game Theory

A good ad hoc team agent should be able to learn to interact with a previously unknown teammate in a fully cooperative (common payoff) iterative normal form game

- If the teammate plays a fixed (possibly stochastic) strategy, the ad hoc team agent should be simply learn what that strategy is and play the best response

Example Theoretical Approach

Collaborative Multi-Armed Bandits

The ad hoc team player interacts repeatedly in a stochastic environment with a teammate that is both less capable and less knowledgeable than itself

- Teammate can only execute a subset of actions, is unaware of relative utilities of actions, and is unaware of team

K-Armed Bandit
At each time step, a learning agent selects one of the k-arms to pull. The arm returns a payoff according to a fixed, but generally unknown, distribution. The agent's goal is to maximize the sum of the payoffs it receives over time. The setting is well-suited for studying exploration vs. exploitation: at any given time, the agent could greedily select the arm that has paid off the best so far, or it could select a different arm in order to gather more information about its distribution.

Example Theoretical Approach
Collaborative Multi-Armed Bandits 2

Assumptions

- Results of actions are fully observable to both agents *(why is this important?)*
- Number of rounds (actions per agent) remaining is finite and known to the teacher *(why is this important?)*
- Learner's behavior is fixed and known: acts greedily *(why is this important?)*

Extension to Ad Hoc

Two distinct agents, known as the teacher and the learner , who select arms alternately, starting with the teacher. Consider a bandit with just three arms such that the teacher is able to select from any of the three arms, while the learner is only able to select from among the two arms with the lower expected payoffs. Consider the fully cooperative case such that the teacher's goal is to maximize the expected sum of the payoffs received by the two agents over time (the teacher is risk neutral).

Example Theoretical Approach
Collaborative Multi-Armed Bandits 3

Decision Making

The teacher must decide whether to do what is best in the short term, namely pull the arm with the highest expected payoff; or whether to increase the information available to its teammate, the learner, by pulling a different arm.

Note that if the teacher were acting alone, trivially its optimal action would be to always pull the arm with highest expected payoff.

Example Empirical Approach

Human Soccer

Robot Soccer

Playing pick-up games, finding out roles, assessing teammates' abilities, adjusting strategies and tactics accordingly

Controlling the Scope

Teammate Characteristics

- Action capabilities, sensing capabilities, decision making and learning capabilities, whether they can communicate directly, and prior knowledge

Team Characteristics

- Whether teammates are homogeneous or heterogeneous, how many teammates are on the team, and whether they can observe each other's actions.

Task Characteristics

- The goal, the time horizon, whether it is turn-taking, and how closely coordinated the agents need to be in their actions.
- Can teammates divide the task at a high level and then act independently, or do they need to coordinate low-level actions?

Discussion & Related Work

At odds with most prior treatments of teamwork (Grosz & Kraus 1996, Tambe 1997, Decker & Lesser 1995, Stone & Veloso 1999)

Multi-armed bandit work similar to (Brafman & Tennenholtz 1996) but "teacher" not embedded in the environment as teammate to "learner"

Suggested techniques to tackle the challenge

- Game theory
- Intended plan recognition
- Opponent modeling
- Reinforcement learning

Human ad hoc teams (Kildare 2004)

References

Brafman, R. I., and Tennenholtz, M. 1996. On partially controlled multi-agent systems. JAIR 4:477–507.

Decker, K. S., and Lesser, V. R. 1995. Designing a family of coordination algorithms. In ICMAS-95, 73–80. Menlo Park, California: AAAI Press.

Grosz, B. J., and Kraus, S. 1996. Collaborative plans for complex group actions. AIJ 86:269–358.

Kildare, R. 2004. Ad-hoc online teams as complex systems: agents that cater for team interaction rules. In Proceedings of the 7th Asia-Pacific Conference on Complex Systems.

Stone, P., and Kraus, S. 2010. To teach or not to teach? Decision making under uncertainty in ad hoc teams. In AAMAS-10. International Foundation for Autonomous Agents and Multiagent Systems.

Stone, P., and Veloso, M. 1999. Task decomposition, dynamic role assignment, and low-bandwidth communication for real-time strategic teamwork. AIJ 110(2):241–273.

Tambe, M. 1997. Towards flexible teamwork. JAIR 7:81–124.