Ad Hoc Teamwork

- Only in control of a single agent or subset of agents
- Unknown teammates
- No pre-coordination
- Shared goals

Examples in humans:
- Pick up soccer
- Accident response
Motivation

- Agents are becoming more common and lasting longer
  - Both robots and software agents
- Pre-coordination may not be possible
- Agents should be robust to various teammates
- Past work focused on cases with no communication
Motivation

- Agents are becoming more common and lasting longer
  - Both robots and software agents
- Pre-coordination may not be possible
- Agents should be robust to various teammates
- Past work focused on cases with no communication

Research Question:
How can an agent act and communicate optimally with teammates of uncertain types?
Example
Example
Example

Ad Hoc Agent

Teammates

Communicating with Unknown Teammates
Example

How long does the first road take?

Ad Hoc Agent

Teammates
Problem Description

- Multi-armed bandit
  - Two Bernoulli arms
  - Ad hoc agent observes all payoffs
Problem Description

- Multi-armed bandit
  - Two Bernoulli arms
  - Ad hoc agent observes all payoffs
- Multi-agent
  - Simultaneous actions
Problem Description

- Multi-armed bandit
  - Two Bernoulli arms
  - Ad hoc agent observes all payoffs
- Multi-agent
  - Simultaneous actions
- Limited communication
  - Fixed set of messages
  - Has explicit cost
Problem Description

- Multi-armed bandit
  - Two Bernoulli arms
  - Ad hoc agent observes all payoffs
- Multi-agent
  - Simultaneous actions
- Limited communication
  - Fixed set of messages
  - Has explicit cost
- Goal: Maximize payoffs and minimize communication costs
Communication

- Last observation
- Arm mean
- Suggestion
Communication

- **Last observation** - The last arm chosen and the resulting payoff
- **Arm mean** - The mean and number of pulls of a selected arm
- **Suggestion** - Suggest that your teammates should pull the selected arm
Teammates

- Limited number of types
- Continuous parameters
- Tightly coordinated
Introduction

Problem Description

Theoretical Results

Empirical Results

Conclusions

Overview

Communication

Teammates

Teammates

- Limited number of types
- Continuous parameters
- Tightly coordinated

  - Team shares knowledge through communication

  - Do not need to track each agent’s pulls
Teammate Behaviors

$\varepsilon$-Greedy

$\text{UCB}(c)$
Teammate Behaviors

**ε-Greedy**
- Track arm means
- Usually choose greedily
- $\varepsilon$ - fraction of time to explore

**UCB($c$)**
Teammate Behaviors

$\varepsilon$-Greedy

- Track arm means
- Usually choose greedily
- $\varepsilon$ - fraction of time to explore

UCB($c$)

- Track arm means and pulls
- Choose greedily with respect to bounds
- $c$ - weight given to bounds
Teammate Behaviors

**ε-Greedy**
- Track arm means
- Usually choose greedily
- $\epsilon$ - fraction of time to explore
- Have probability of following suggestion sent by ad hoc agent

**UCB($c$)**
- Track arm means and pulls
- Choose greedily with respect to bounds
- $c$ - weight given to bounds
Can an ad hoc agent approximately plan to communicate optimally with these teammates in polynomial time?
Model

- Model as a POMDP (teammates’ behaviors)
- State:
  - Pulls and successes:
    - Teammates’
    - Ad hoc agent’s
    - Communicated
Model

- Model as a POMDP (teammates’ behaviors)

- State:
  - Pulls and successes:
    - Teammates’
    - Ad hoc agent’s
    - Communicated
  - Types and parameters of teammates (partially observed)
Model

- Model as a POMDP (teammates’ behaviors)

- State:
  - Pulls and successes:
    - Teammates’
    - Ad hoc agent’s
    - Communicated
  - Types and parameters of teammates (partially observed)

- Actions are arms to choose and messages to send

- Transition function is based on arms’ distributions and teammates’ behaviors
Simple Version

- What if we know the teammates’ behaviors?
Simple Version

- What if we know the teammates’ behaviors?
- Problem simplifies to an MDP
- What is the size of the state space?
Simple Version

- What if we know the teammates’ behaviors?
- Problem simplifies to an MDP
- What is the size of the state space?
  - Team is tightly coordinated \( \Rightarrow \) only track pulls and successes of team
  - Track team’s, ad hoc agent’s, and communicated pulls
Simple Version

- What if we know the teammates’ behaviors?
- Problem simplifies to an MDP
- What is the size of the state space?
  - Team is tightly coordinated $\Rightarrow$ only track pulls and successes of team
  - Track team’s, ad hoc agent’s, and communicated pulls
  - Polynomial in terms of number of teammates and rounds
- Solvable in polynomial time
Do not fully know teammates’ behaviors

Know teammates are either $\varepsilon$-greedy or UCB($c$)

Do not know $\varepsilon$ or $c$

Problem is a POMDP
Background

- POMDPs can be approximately solved in polynomial time in terms of the number of $\delta$-neighborhoods that can cover the belief space (aka the covering number)
δ-neighborhood
Proof Sketch

- Observable part of the state adds a polynomial factor
- Only need to worry about the partially observed teammates
Proof Sketch

- Observable part of the state adds a polynomial factor
- Only need to worry about the partially observed teammates
  - Belief space of $\varepsilon$ can be represented as beta distribution
Proof Sketch

- Observable part of the state adds a polynomial factor
- Only need to worry about the partially observed teammates
  - Belief space of $\varepsilon$ can be represented as beta distribution
  - Belief space of $c$ can be represented by the upper and lower possible values
Proof Sketch

- Observable part of the state adds a polynomial factor
- Only need to worry about the partially observed teammates
  - Belief space of $\varepsilon$ can be represented as beta distribution
  - Belief space of $c$ can be represented by the upper and lower possible values
- Can track probability of $\varepsilon$-greedy vs UCB using Bayes updates
Proof Sketch

- Observable part of the state adds a polynomial factor
- Only need to worry about the partially observed teammates
  - Belief space of $\varepsilon$ can be represented as beta distribution
  - Belief space of $c$ can be represented by the upper and lower possible values
  - Can track probability of $\varepsilon$-greedy vs UCB using Bayes updates
- Covering number of belief space is polynomial $\Rightarrow$ POMDP can be solved in polynomial time
Proof Sketch

- Observable part of the state adds a polynomial factor
- Only need to worry about the partially observed teammates
  - Belief space of $\varepsilon$ can be represented as beta distribution
  - Belief space of $c$ can be represented by the upper and lower possible values
  - Can track probability of $\varepsilon$-greedy vs UCB using Bayes updates
- Covering number of belief space is polynomial $\Rightarrow$ POMDP can be solved in polynomial time
- Results carry over into case of unknown arm means
Outline

1. Introduction
2. Problem Description
3. Theoretical Results
4. Empirical Results
5. Conclusions

S. Barrett, N. Agmon, N. Hazon, S. Kraus, P. Stone
Communicating with Unknown Teammates
Approach

- POMDP problem is tractable $\Rightarrow$ we can use existing POMDP solvers
- POMCP
  - Particle filtering to track beliefs
  - Monte Carlo tree search to plan

- D. Silver and J. Veness. Monte-Carlo planning in large POMDPs. In *NIPS ’10*, 2010
Approach

- POMDP problem is tractable $\Rightarrow$ we can use existing POMDP solvers
- POMCP
  - Particle filtering to track beliefs
  - Monte Carlo tree search to plan
  - Fast
  - Handles large state-action spaces
  - Approximate

- D. Silver and J. Veness. Monte-Carlo planning in large POMDPs. In *NIPS '10*, 2010
Empirical Setup

- Vary message costs
- Vary number of rounds
- Vary number of arms
- Vary number of teammates
Ad Hoc Agent Behaviors

- **POMCP** - Plan using POMCP
- **NoComm** - Act greedily and do not communicate
- **Obs** - Act greedily and communicate the last observation
Problem Description

- Problem tackled in the theory
- Teammates are either $\varepsilon$-greedy or UCB($c$)
- Need to figure out:
  - Type
  - Parameter ($\varepsilon$ or $c$)
  - Chance of following suggestion
\( \varepsilon \)-Greedy Teammates

---

**Setup**

- \( \varepsilon \)-Greedy Teammates
- UCB(\( c \)) Teammates
- Unknown arms
- Externally-created Teammates

---

**\( \varepsilon \)-Greedy Teammates\**

<table>
<thead>
<tr>
<th>Message Cost</th>
<th>Frac of Max Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>0.9</td>
</tr>
<tr>
<td>0.16</td>
<td>0.8</td>
</tr>
<tr>
<td>0.32</td>
<td>0.7</td>
</tr>
<tr>
<td>0.64</td>
<td>0.6</td>
</tr>
<tr>
<td>1.28</td>
<td>0.5</td>
</tr>
<tr>
<td>2.56</td>
<td>0.4</td>
</tr>
</tbody>
</table>

---

**UCB(\( c \)) Teammates**

- \( \varepsilon \)-Greedy Teammates
- Unknown arms
- Externally-created Teammates

---

**Unknown arms**

- \( \varepsilon \)-Greedy Teammates
- UCB(\( c \)) Teammates
- Externally-created Teammates

---

**Externally-created Teammates**

- \( \varepsilon \)-Greedy Teammates
- UCB(\( c \)) Teammates
- Unknown arms

---

**Communicating with Unknown Teammates**

S. Barrett, N. Agmon, N. Hazon, S. Kraus, P. Stone
Communicating with Unknown Teammates

Setup

- \( \varepsilon \)-Greedy Teammates
- UCB(\( c \)) Teammates

Unknown arms

Externally-created Teammates

**UCB(\( c \)) Teammates**

<table>
<thead>
<tr>
<th>Message Cost</th>
<th>Frac of Max Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>0.9</td>
</tr>
<tr>
<td>0.16</td>
<td>0.8</td>
</tr>
<tr>
<td>0.32</td>
<td>0.7</td>
</tr>
<tr>
<td>0.64</td>
<td>0.6</td>
</tr>
<tr>
<td>1.28</td>
<td>0.5</td>
</tr>
<tr>
<td>2.56</td>
<td>0.4</td>
</tr>
</tbody>
</table>

POMCP

NoComm

Obs

S. Barrett, N. Agmon, N. Hazon, S. Kraus, P. Stone
Unknown arms - \( \varepsilon \)-greedy or UCB(\( c \))

![Graph showing Frac of Max Reward vs Num Teammates for different algorithms: POMCP, NoComm, Obs, Match.](image-url)
Externally-created Teammates

- Teammates we did not create
- Created by students for project
Externally-created Teammates

- Teammates we did not create
- Created by students for project
- Not necessarily tightly coordinated
- Not considering ad hoc teamwork
Externally-created Teammates

- True ad hoc teamwork scenario
- Models are incorrect
- Theoretical guarantees do **not** hold
Externally-created Teammates – Cost

![Graph showing the relationship between message cost and frac of max reward for different communication methods, including POMCP, NoComm, and Obs. The graph indicates a decrease in frac of max reward as message cost increases.](image-url)
Externally-created Teammates – Num Teammates

Frac of Max Reward

Num Teammates

- POMCP
- NoComm
- Obs

S. Barrett, N. Agmon, N. Hazon, S. Kraus, P. Stone

Communicating with Unknown Teammates


Conclusions

- Can optimally plan best way to communicate with unknown teammates
- Can handle an infinite set of possible teammates
- Can cooperate with a variety of teammates not covered in theory
Future Work

- More complex domains
- Unknown environments
- Teammates that learn about us
In some cases, ad hoc agents can optimally plan about how to communicate with their teammates.