Communicating with Unknown Teammates

Samuel Barrett[1] Noa Agmon[2] Noam Hazon[3]
Sarit Kraus[2,4] Peter Stone[1]

[1]University of Texas at Austin
{sbarrett,pstone}@cs.utexas.edu

[2]Bar-Ilan University
{agmon,sarit}@macs.biu.ac.il

[3]Ariel University
noamh@ariel.ac.il

[4]University of Maryland

ECAI
Aug 21, 2014

UT Austin Learning Agents Research Group

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

UT Austin Learning Agents Research Group

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

UT Austin Learning Agents Research Group

Example

UT Austin Learning Agents Research Group

Example

Ad Hoc Agent

Teammates

Learning Agents Research Group

Example

How long does the first road take?

Ad Hoc Agent

Teammates

Learning Agents Research Group

Outline

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

Outline

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

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

ε-Greedy

UCB(c)

Teammate Behaviors

ε-Greedy

- Track arm means
- Usually choose greedily
- ε - fraction of time to explore

UCB(c)

Teammate Behaviors

ε-Greedy

- Track arm means
- Usually choose greedily
- ε - 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
- ε - fraction of time to explore

UCB(c)

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

- Have probability of following suggestion sent by ad hoc agent

Outline

Research Question

Can an ad hoc agent approximately plan to communicate optimally with these teammates in polynomial time?

UT Austin Learning Agents Research Group

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)

UT Austin Learning Agents Research Group

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

Full version

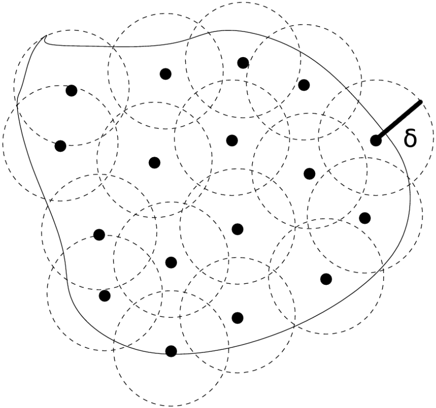
- Do **not** fully know teammates' behaviors
- Know teammates are either ε-greedy or UCB(c)
- Do not know ε or c
- Problem is a POMDP

UT Austin Learning Agents Research Group

Background

- POMDPs can be approximately solved in polynomial time in terms of the number of δ-neighborhoods that can cover the belief space (aka the covering number)
  - H. Kurniawati, D. Hsu, and W. S. Lee. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *In Proc. Robotics: Science and Systems*, 2008

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 ε 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 ε 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 ε can be represented as beta distribution
  - Belief space of c can be represented by the upper and lower possible values
  - Can track probability of ε-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 ε can be represented as beta distribution
  - Belief space of c can be represented by the upper and lower possible values
  - Can track probability of ε-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 ε can be represented as beta distribution
  - Belief space of c can be represented by the upper and lower possible values
  - Can track probability of ε-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

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

UT Austin Learning Agents Research Group

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

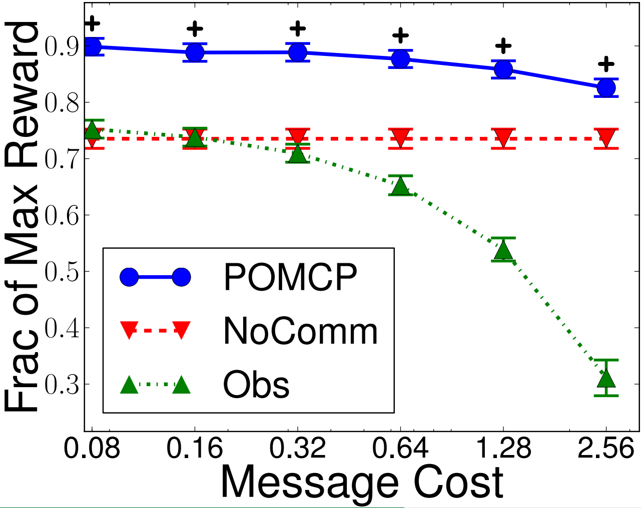
Ad Hoc Agent Behaviors

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

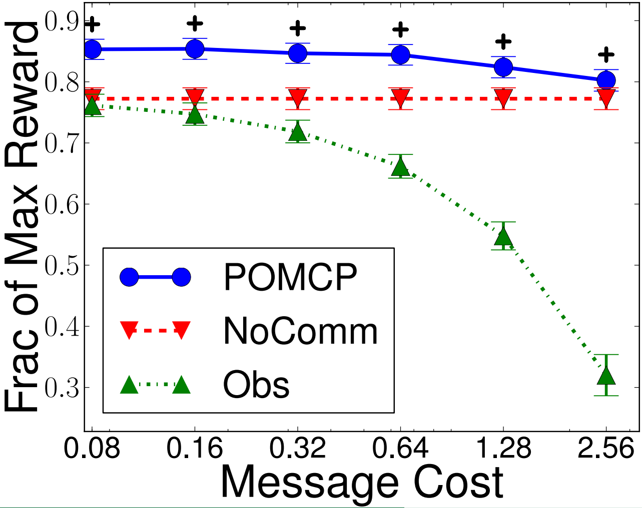
Problem Description

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

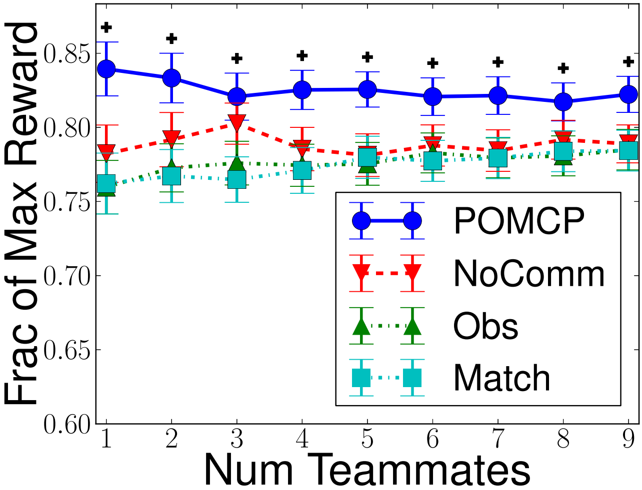
ε-Greedy Teammates



UCB(c) Teammates



Unknown arms - ε-greedy or UCB(c)



Externally-created Teammates

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

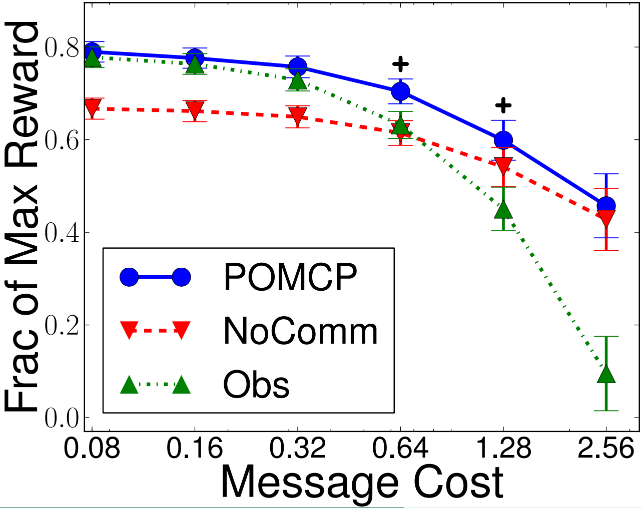
Externally-created Teammates

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

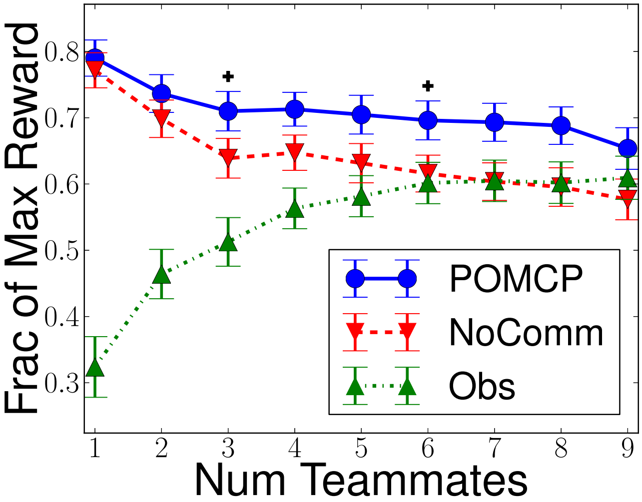
Externally-created Teammates

- True ad hoc teamwork scenario
- Models are incorrect
- Theoretical guarantees do **not** hold

Externally-created Teammates – Cost



Externally-created Teammates – Num Teammates



Learning Agents Research Group

Outline

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

UT Austin Learning Agents Research Group

Related Work

- S. Liemhetcharat and M. Veloso. Modeling mutual capabilities in heterogeneous teams for role assignment. In *IROS '11*, pages 3638 –3644, 2011
- F. Wu, S. Zilberstein, and X. Chen. Online planning for ad hoc autonomous agent teams. In *IJCAI*, 2011
- M. Bowling and P. McCracken. Coordination and adaptation in impromptu teams. In *AAAI*, pages 53–58, 2005
- J. Han, M. Li, and L. Guo. Soft control on collective behavior of a group of autonomous agents by a shill agent. *Journal of Systems Science and Complexity*, 19:54–62, 2006
- M. Knudson and K. Tumer. Robot coordination with ad-hoc team formation. In *AAMAS '10*, pages 1441–1442, 2010
- E. Jones, B. Browning, M. B. Dias, B. Argall, M. M. Veloso, and A. T. Stentz. Dynamically formed heterogeneous robot teams performing tightly-coordinated tasks. In *ICRA*, pages 570 – 575, May 2006

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

Thank You!

In some cases, ad hoc agents can optimally plan about how to communicate with their teammates.

UT Austin Learning Agents Research Group