An Analysis Framework for Ad Hoc Teamwork Tasks

Samuel Barrett Peter Stone

The University of Texas at Austin
{sbarrett,pstone}@cs.utexas.edu

AAMAS 2012
June 8, 2012

UT Austin Learning Agents Research Group

Ad Hoc Teamwork

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

Examples in humans:

- Pick up soccer
- Accident response

Motivation

- Agents are becoming more common and lasting longer
- Pre-coordination may not be possible
- Previous work focuses on specific subsets of the ad hoc teamwork problem
- Unify research in ad hoc teamwork

Overview

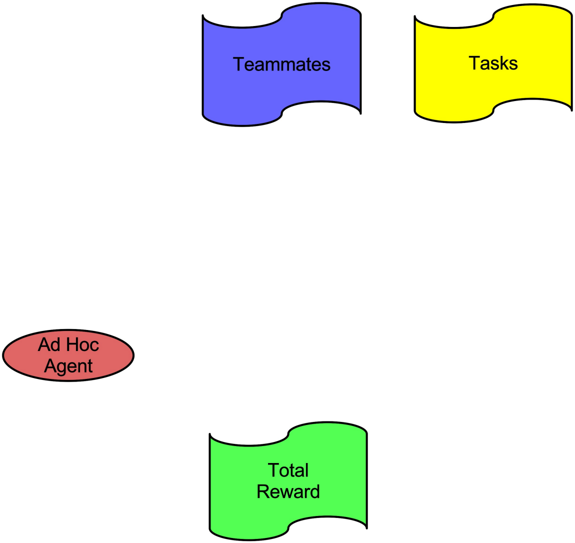
- Analyze ad hoc team problems in terms of 3 dimensions
- Analysis helps for reusing prior algorithms on new domains
- Better identify areas for future research

Ad Hoc Agent Evaluation

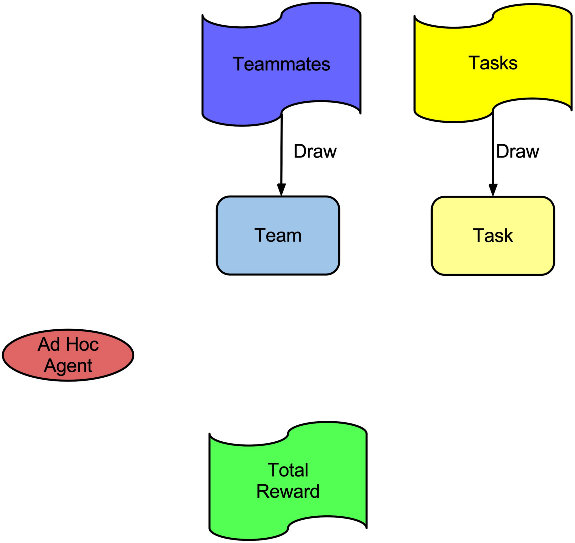
- Not whether they win, but how well they cooperate
- Compare against other ad hoc agents
- Depends on possible tasks
- Depends on possible teammates

UT Austin Learning Agents Research Group

Ad Hoc Agent Evaluation

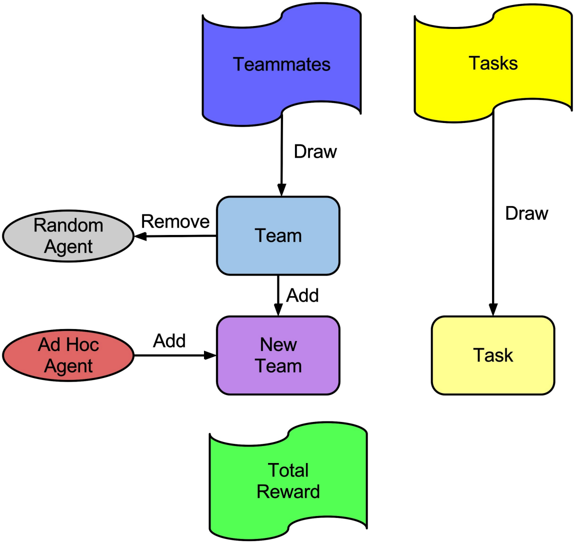


Ad Hoc Agent Evaluation



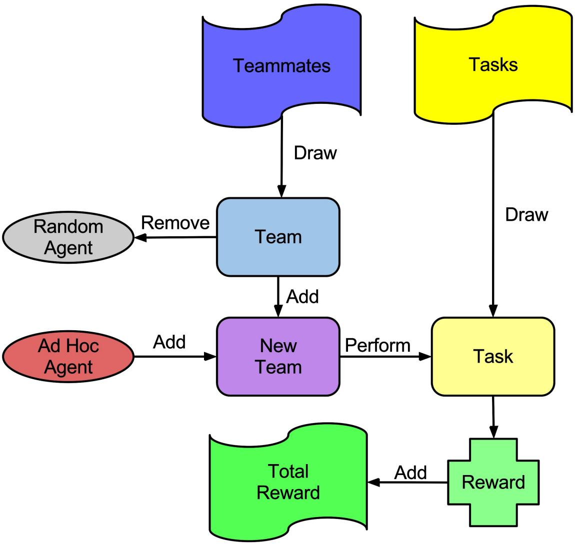
UT Austin Learning Agents Research Group

Ad Hoc Agent Evaluation



UT Austin Learning Agents Research Group

Ad Hoc Agent Evaluation



Dimensions

- Analyze ad hoc team problems
- Identify informative dimensions
- Give explicit measures to compare problems

Team Knowledge

Does the ad hoc agent know what its teammates' actions will be for a given state, before interacting with them?

Team Knowledge

Does the ad hoc agent know what its teammates' actions will be for a given state, before interacting with them?

- What the ad hoc agent knows ahead of time, not what it can learn
- Compare expected distribution of teammates' actions to true distribution
- Averaged over all states and teammates
- Higher values $\rightarrow$ more knowledge $\rightarrow$ easier planning

Team Knowledge

$$K(T,P) = \begin{cases} 1 & \text{if } \mathrm{JS}(T,P) = 0 \\ 1 - \dfrac{\mathrm{JS}(T,P)}{\mathrm{JS}(T,U)} & \text{if } \mathrm{JS}(T,P) < \mathrm{JS}(T,U) \\ -\dfrac{\mathrm{JS}(P,U)}{\mathrm{JS}(U,\mathrm{Point})} & \text{otherwise} \end{cases}$$

$$\text{Team Knowledge} = \frac{\sum_{s=1}^{n} \sum_{t=1}^{k} K(T_t(s), P_t(s))}{nk}$$

T - True distribution

P - Predicted distribution

JS - Jensen-Shannon divergence, a symmetric variant of KL

Point - distribution with all weight on one point

Environment Knowledge

Does the ad hoc agent know the transition and reward distribution given a joint action and state before interacting with the environment?

Environment Knowledge

Does the ad hoc agent know the transition and reward distribution given a joint action and state before interacting with the environment?

- 2 parts - transition and reward are separate
- Compare expected next state/reward distribution to true distribution
- Given full information of joint action
- Averaged over all states
- Higher values $\rightarrow$ more knowledge $\rightarrow$ easier planning

Reactivity of Teammates

How much does the ad hoc agent's actions affect those of its teammates?

Reactivity of Teammates

How much does the ad hoc agent's actions affect those of its teammates?

- Compare resulting joint action distribution for different ad hoc agent's actions
- One step effects
- Similar to empowerment
- Higher values $\rightarrow$ more reactive $\rightarrow$ harder planning, but more control

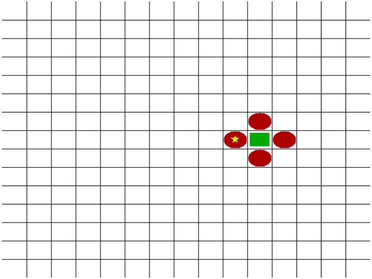
Overview

- Analyze variations on an example domain
- Identify how to reuse prior work
- Analyze more domains
- Identify areas for future research

Pursuit Domain

- Grid world - Torus
- N Predators and 1 Prey
- Predators' goal is to capture the prey as quickly as possible
- Act simultaneously
- Collisions randomly decided - loser stays still

4-Predator Known Behaviors



- 4 predators
- Teammates use 1 of 4 behaviors
- Teammate behavior is known

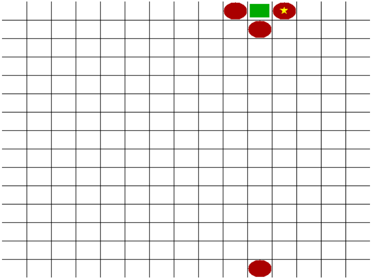
UT Austin Learning Agents Research Group

Problem Analysis

Domain	Team Knowledge	Environment (Trans, Reward)	Teammate Reactivity
4 Known	1	(1,1)	0.00105–0.501

UT Austin Learning Agents Research Group

4-Predator Unknown Behaviors



- 4 predators
- Teammates are drawn from set of 4 behaviors
- Teammate behavior is **not** known

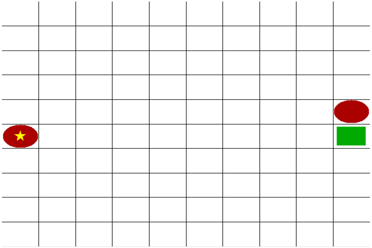
Problem Analysis

Domain	Team Knowledge	Environment (Trans, Reward)	Teammate Reactivity
4 Known	1	(1,1)	0.00105–0.501
4 Unknown	0.155–0.807	(1,1)	0.00105–0.501

UT Austin Learning Agents Research Group

2-Predator Simultaneous

- 2 predators
- Choose high level behavior to play for an episode
- Both know expected capture time for behaviors
- Teammate plays best response
- Capture by both predators neighboring the prey

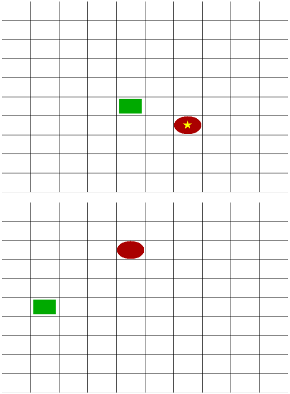


Problem Analysis

Domain	Team Knowledge	Environment (Trans, Reward)	Teammate Reactivity
4 Known	1	(1,1)	0.00105–0.501
4 Unknown	0.155–0.807	(1,1)	0.00105–0.501
2 Simul	1	(1,1)	0.198

UT Austin Learning Agents Research Group

2-Predator Teaching



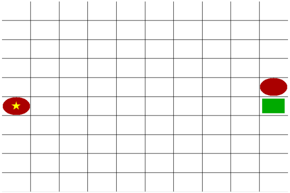
- 2 predators taking turns
- Choose high level behavior to play for an episode
- Ad hoc agent knows expected capture times for behaviors
- Teammate chooses greedily based on observed capture times

Problem Analysis

Domain	Team Knowledge	Environment (Trans, Reward)	Teammate Reactivity
4 Known	1	(1,1)	0.00105–0.501
4 Unknown	0.155–0.807	(1,1)	0.00105–0.501
2 Simul	1	(1,1)	0.198
2 Teach	1	(1,1)	0.0342–0.118

2-Predator Simultaneous Revisited

- Repeated normal-form game
- Shared payoffs
- Choosing behaviors corresponds to choosing row/column
- Try to find lowest cost path to optimal cell



=

	TA	PD	GR
TA	-4.583	-5.123	-5.152
PD	-5.123	-4.946	-4.615
GR	-5.152	-4.615	-4.379

2-Predator Simultaneous Revisited

- Efficient algorithm when memory size of 1
- Exponential algorithm when memory size of >1
- Can handle non-deterministic teammates

Related Work

UT Austin Learning Agents Research Group

Problem Descriptions

- Flocking control - controlling a flock with a “shill” agent
  - Han, Li, and Guo 2006
- Unknown teammates (UTM) - cooperative box pushing, meeting in a 3x3 grid, and multi-channel broadcast
  - UTM-1 - follow a fixed set of actions
  - UTM-2 - attempt to play optimally, but have limited observations
  - Wu, Zilberstein, and Chen 2011
- Simulated pickup soccer - ad hoc agent given a different playbook
  - Bowling and McCracken 2005

Problem Analysis

Domain	Team Knowledge	Environment (Trans, Reward)	Teammate Reactivity
Flocking control	1	(1,1)	0.0732–0.880
Cooperating with UTM-1 teammates	0	(1,1)	0
Cooperating with UTM-2 teammates	0	(1,1)	>0
Simulated pickup soccer	>0	(>0,1)	>0

Future Research

- Most prior work focused on planning to interact with teammates
- Low team knowledge - must explore teammates' behaviors
- Low environment knowledge - must explore environment
- Trade-off between exploiting current knowledge, exploring teammates, and exploring the environment
- More complex domains

Conclusions

- Can analyze ad hoc team problems in terms of:
  - Team Knowledge
  - Environmental Knowledge
  - Teammate Reactivity
- Analysis helps for reusing prior algorithms on new domains
- Better identify areas for future research

Thank You!

- Analyzing domains can help ad hoc team agents cooperate with a variety of teammates