To appear in *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS), Paris, France, May 2014.

The RoboCup 2013 Drop-In Player Challenges: A Testbed for Ad Hoc Teamwork

(Extended Abstract)

Patrick MacAlpine, Katie Genter, Samuel Barrett, Peter Stone
Department of Computer Science, The University of Texas at Austin, Austin, TX 78712, USA
{patmac,katie,sbarrett,pstone}@cs.utexas.edu

ABSTRACT

As the prevalence of autonomous agents grows, so does the number of interactions between these agents. Therefore, it is desirable for these agents to be capable of collaborating without pre-coordination. While past research on ad hoc teamwork has focused mainly on relatively simple domains, the long-term vision has been to enable robots and other autonomous agents to exhibit the sort of flexibility and adaptability on complex tasks that people do. This research introduces a series of pick-up robot soccer experiments that were carried out in three different leagues at the international RoboCup competition in 2013. In all cases, agents from different labs were put on teams with no pre-coordination. This abstract summarizes the structure of these experiments and analyzes the results. The work describes a new large-scale ad hoc teamwork testbed that can serve as a starting point for future experimental ad hoc teamwork research.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

Keywords

Ad Hoc Teams; Multiagent Systems; Teamwork; Robotics

1. INTRODUCTION

The increasing capabilities and decreasing costs of robots makes it increasingly possible to study the interactions among teams of heterogeneous robots. To date, most such research on multi-robot teamwork assumes that robots share a common coordination protocol. However, as the number of different companies and research labs producing robots grows, and especially as long-term autonomous capabilities become more common, it becomes increasingly likely that robots will have the occasion to collaborate with previously unknown teammates in pursuit of a common goal. When engaging in such *ad hoc teamwork* [2], robots must recognize and reason about their teammates' capabilities.

Although much of the initial research on ad hoc teamwork has taken a theoretical perspective, it has been argued that ad hoc teamwork is "ultimately an empirical challenge" [2]. In order to facilitate such empirical ad hoc teamwork research, this research summarizes a series of "drop-in player challenges" that the authors helped to organize at RoboCup 2013, a well established multi-robot soccer competition. In each game of the challenges,[1] robots were drawn from participating teams and combined to form a new team. Robots were not informed of each other's identities, and thus had to adapt quickly, and without prior coordination, to their unknown teammates during the course of a game. Previous work by Bowling and McCracken [1] investigates ad hoc team agents in RoboCup, where the agent's playbook differs from that of its teammates.

This research introduces drop-in player challenges as a novel testbed for ad hoc teamwork and facilitates future research in this area. The abstract's main purpose is to serve as a basis for future large-scale experimental ad hoc teamwork research, both in RoboCup, and other multi-robot domains.

2. DOMAIN DESCRIPTION

In the RoboCup soccer domain, teams of autonomous robots compete with each other in a complex, real-time, noisy and dynamic environment, in a setting that is both collaborative and adversarial. RoboCup consists of several leagues, each emphasizing different research challenges. Our research takes place in three different RoboCup leagues: the Standard Platform League (SPL), the 2D Simulation League, and the 3D Simulation League. In the SPL, teams compete with identical Aldebaran Nao humanoid robots. In the 2D Simulation League, autonomous agents play soccer on a simulated 2D soccer field. 2D soccer abstracts away many of the low-level behaviors required for humanoid robot soccer, including walking and computer vision, instead focusing on higher-level aspects of playing soccer such as multiagent coordination and strategy. In the 3D Simulation League, soccer takes place in a 3D simulated environment with realistic physics. Simulated Aldebaran Nao robots receive abstract perceptual information and send torque commands for their motors to a central game server.

In all drop-in player challenges, teams were composed of randomly selected drop-in players from teams competing in the main RoboCup competition. The SPL challenge[2] was scored using two metrics: average goal difference and average score from three judges where each judge was asked to award each drop-in player a teamwork score ranging from 0

[1] Videos of the challenges are at `http://www.cs.utexas.edu/~AustinVilla/sim/3dsimulation/AustinVilla3DSimulationFiles/2013/html/dropin.html`

[2] Full SPL challenge rules at `http://www.tzi.de/spl/pub/Website/Downloads/Challenges2013.pdf`

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

to 10. The two scoring metrics were normalized and combined to determine the total score. In both of the 2D[3] and 3D[4] drop-in player challenges, agents were scored by the average goal difference received across all games it played.

3. RESULTS AND ANALYSIS

In this section, we summarize the results of the drop-in player challenges. As winners of the challenges should display the best adaptive teamwork abilities, not necessarily the best pre-coordinated teamwork and low level skills for playing soccer, we also compare teams' performances in the challenges to their performance in the main competition. Additionally, as the number of games played at the RoboCup competition do not provide statistically significant results, we also provide data from playing many games in the simulation leagues with binaries released after the competition.

SPL Results

The SPL drop-in player challenge consisted of four games. Table 1 shows challenge scores and rankings, as well as relative rankings in the main RoboCup competition.

	Drop-In					Main
Team	**AGD**	**NGD**	**AJS**	**Score**	**Rank (G,J)**	**Rank**
B-Human	1.17	10.00	6.67	16.67	1 (1,1)	1
Nao Devils	0.57	4.90	6.24	11.14	2 (3,2)	4
rUNSWift	0.67	5.71	5.22	10.94	3 (2,4)	3
UTAustinVilla	-0.29	-2.45	6.00	3.55	4 (4,3)	2
UPennalizers	-0.57	-4.90	4.48	-0.42	5 (5,5)	6
Berlin United	-1.29	-11.02	3.38	-7.64	6 (6,6)	5

Table 1: Final scores (average goal difference (AGD), normalized goal difference (NGD), average judge score (AJS)) and rankings (goal (G) and judge (J)) for the SPL drop-in challenge and also relative rankings in the main RoboCup competition.

Overall, the challenge's results were well correlated with the main competition's results. UPennalizers and Berlin United finished near the bottom in the drop-in challenge, and they also were in the lower ranks for the main competition. Notably, B-Human was first for the drop-in and main ranks as well as the human judges' scores, indicating their teamwork and adaptability performed well in both settings.

2D Simulation Results

Seven games were played for the 2D drop-in player challenge. Following the competition, we also replayed the challenge with the released binaries across 4,200 games which included all combinations of the teams contributing two agents each to a game. Additionally we ran 1,000 games of our team's main competition binary (UTAustinVilla) against each of the other teams' released main competition binaries. Results for the both the drop-in player challenge and main RoboCup competition, as well as the results for each of these computed over many games run after the competition, are in Table 2.

The difference between drop-in player challenge results, and results across many games run after the RoboCup competition, show that only playing seven games does not reveal the true rankings of teams (e.g. the last place team during the challenge at RoboCup, AUTMasterminds, finished third overall when playing thousands of games). Although there is not a direct correlation between ranking in the drop-in player challenge compared to standard team soccer, there

	Drop-In				Main		
	RC		Many Games		RC	Vs UTAustinVilla	
Team	**R**	**AGD**	**R**	**AGD**	**R**	**R**	**AGD**
FCPerspolis	1	2.40	1	3.025 (0.142)	5	4	3.127 (0.059)
Yushan	2	2.25	2	2.583 (0.141)	2	3	4.034 (0.065)
ITAndroids	3	2.00	5	1.379 (0.152)	7	7	0.505 (0.063)
Axiom	4	1.20	6	1.315 (0.148)	3	5	1.803 (0.074)
UTAustinVilla	5	0.25	4	1.659 (0.153)	8	8	0.000 (self)
HfutEngine	6	-0.20	7	-2.076 (0.153)	9	9	-6.027 (0.184)
WrightEagle	7	-1.60	9	-6.218 (0.129)	1	1	6.176 (0.287)
FCPortugal	8	-2.20	8	-3.379 (0.150)	6	6*	*
AUTMasterminds	9	-2.80	3	1.711 (0.152)	4	2	5.111 (0.117)

Table 2: Rankings (R) and average goal difference (AGD) with standard error shown in parentheses for both the 2D drop-in player challenges and the main RoboCup competition with results given for both RoboCup (RC) and games played after the competition. *We were unable to run the released FCPortugal binary and thus used their relative ranking from the main competition.

is a trend for agents that perform better at standard team soccer to also perform better at the drop-in player challenge. Excluding the outlier team WrightEagle, which often showed odd behavior on our system during games, the top half of the teams for the drop-in player challenge over many games had an average rank of 4.25 when playing against UTAustinVilla, while the bottom half had an average rank of 6.75.

3D Simulation Results

The 3D drop-in player challenge was played across four games. We also replayed the challenge with released binaries across 630 games using all possible combinations of teams. Additionally we ran at least 100 games of our team's main competition binary (UTAustinVilla) against each of the other teams' released main competition binaries. Table 3 gives results for the both the drop-in player challenge and main RoboCup competition, as well as the results for each of these across many games run after the competition. There is not a strong correlation between rankings in the drop-in player challenge and ranking in the main competition. However, there is a trend that teams performing better at drop-in player soccer also do better at standard team soccer. The top half of teams for the drop-in player challenge over many games had an average rank of 3.4 against UTAustinVilla, while the bottom half's average rank was 7.6.

	Drop-In				Main		
	RC		Many Games		RC	Vs UTAustinVilla	
Team	**R**	**AGD**	**R**	**AGD**	**R**	**R**	**AGD**
BoldHearts	1	1.50	4	0.178 (0.068)	T5	6	-1.607 (0.029)
FCPortugal	T2	0.75	1	1.159 (0.060)	3	2	-0.465 (0.023)
Bahia3D	T2	0.75	7	-0.378 (0.068)	10	10	-9.800 (0.110)
Apollo3D	T2	0.75	5	0.159 (0.068)	1	3	-0.698 (0.027)
magmaOffenburg	5	0.25	3	0.254 (0.068)	T5	5	-1.447 (0.026)
RoboCanes	6	-0.50	6	-0.286 (0.068)	T5	7	-1.828 (0.031)
UTAustinVilla	T7	-0.75	2	0.784 (0.065)	2	1	0.000 (self)
SEUJolly	T7	-0.75	9	-0.613 (0.066)	4	4	-1.133 (0.027)
Photon	T7	-0.75	8	-0.425 (0.068)	8	8	-4.590 (0.081)
L3MSIM	10	-1.25	10	-0.832 (0.065)	9	9	-6.050 (0.098)

Table 3: Rankings (R) and average goal difference (AGD) with standard error shown in parentheses for both the 3D drop-in player challenges and the main RoboCup competition with results given for both RoboCup (RC) and games played after the competition.

4. CONCLUSIONS

This abstract summarizes the first drop-in challenges that occurred at RoboCup 2013. These challenges serve as a novel testbed for ad hoc teamwork, in which agents must adapt to a variety of new teammates without pre-coordination. We believe that these original drop-in challenges will provide a basis for many future drop-in challenges, and will also serve as a reference for designing new ad hoc teamwork testbeds.

5. REFERENCES

[1] M. Bowling and P. McCracken. Coordination and adaptation in impromptu teams. In *AAAI*, 2005.

[2] P. Stone, G. A. Kaminka, S. Kraus, and J. S. Rosenschein. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *AAAI '10*, July 2010.

[3] Full 2D challenge rules at `http://www.cs.utexas.edu/~AustinVilla/sim/2dsimulation/2013_dropin_challenge/2D_DropInPlayerChallenge.pdf`

[4] Full 3D challenge rules at `http://www.cs.utexas.edu/~AustinVilla/sim/3dsimulation/2013_dropin_challenge/3D_DropInPlayerChallenge.pdf`