Robot Coordination with Ad-hoc Team Formation
(Extended Abstract)

Matt Knudson
Oregon State University
knudsonm@engr.orst.edu

Kagan Tumer
Oregon State University
kagan.tumer@oregonstate.edu

ABSTRACT

Coordinating multiagent systems to maximize global information collection both presents scientific challenges (what should each agent aim to achieve?) and provides application opportunities (planetary exploration, search and rescue). In particular, in many domains where communication is expensive (for example, because of limited power or computation), the coordination must be achieved in a passive manner, without agents explicitly informing other agents of their states and/or intended actions. In this work, we extend results on such multiagent coordination algorithms to domains where the agents cannot achieve the required tasks without forming teams.

Categories and Subject Descriptors

I.2.6 [**AI**]: Learning

General Terms

Algorithms, Experimentation

Keywords

Agent Cooperation; Teamwork; Learning (Multiagent)

1. INTRODUCTION

Coordinating multiple robots to achieve a system-wide objective in an unknown and dynamic environment is critical to many of today's relevant applications, including the autonomous exploration of planetary surfaces and search and rescue in disaster response [3, 4]. In such cases, the environment may be dangerous, uninhabitable to humans all together, or sufficiently distant from central control that response times require autonomous, coordinated behavior.

In this work, we focus on problems where robots need to coordinate their actions to achieve high levels of performance. We investigate the use of difference objective functions to promote team formation [2, 1]. The key contribution of this work is to extend those results to problems requiring coordination through the coupling of the robots' objective functions (e.g., no explicit coordination directives).

Cite as: Robot Coordination with Ad-hoc Team Formation (Extended Abstract), Matt Knudson and Kagan Tumer, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1441-1442

The application domain we selected is a distributed information gathering problem. First we explore the case where unless a particular point of interest is observed by n robots, the point of interest is not considered as observed. Second we explore the case where there is an optimal number of robots (n) that need to observe a point of interest, but where the system receives some value for observations by teams with other than n members.

2. ROBOT COORDINATION

The multi-robot information gathering problem we investigate in this work consists of a set of robots that must observe a set of points of interest (POIs) within a given time window [2]. The POIs have different importance to the system, and each observation of a POI yields a value inversely related to the distance the robot is from the POI.

Each robot uses a two layer sigmoid activated artificial neural network to perform this mapping. The weights of the neural network are adjusted through an evolutionary algorithm for ranking and subsequently locating successful networks within a population. The algorithm maintains a population of ten networks, utilizes mutation to modify individuals, and ranks them based on a performance metric specific to the domain. The inputs to this function approximator are four POI sensors and four robot sensors, providing the POI and robot "richness" of each quadrant. Two outputs from the function approximator indicate the velocity of the robot (in the two axes parallel and perpendicular to the current robot heading) [2, 1].

3. REQUIRING TEAM FORMATION

In the first problem we examine, the robots need to form teams to perform a task and contribute to the system objective. Here, if more than two robots visit a POI, only the observations of the closest two are considered and their visit distances are averaged in the computation of the system objective (G), which is given by:

$$G(z) = \sum_i \sum_j \sum_k \frac{V_i \; N^1_{i,j} \; N^2_{i,k}}{\frac{1}{2}(\delta_{i,j} + \delta_{i,k})} \tag{1}$$

where V_i is the value of the ith POI, $\delta_{i,j}$ is the closest distance between jth robot and the ith POI, and $N^1_{i,j}$ and $N^2_{i,k}$ determine whether a robot was within the observation distance δ_o and the closest or second closest robot, respectively, to the ith POI. The single robot objective used by each robot only focuses on the value a robot receives for observing a particular POI:

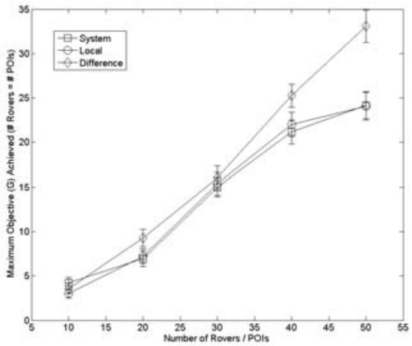

Figure 1: **Maximum objective achieved for equal numbers of robots and POIs. Performance of system, local, and difference objectives *requiring* teams of two robots.**

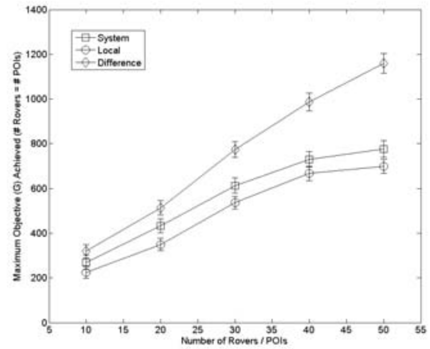

Figure 2: **Maximum objective achieved for equal numbers of robots and POIs. Performance of system, local, and difference objectives *encouraging* teams of two robots.**

$$P_j(z) = \sum_i \frac{V_i}{\delta_{i,j}} \text{ if } \delta_{i,j} < \delta_o \tag{2}$$

where notation is the same as above. This objective promotes selfish behavior only, providing a clear, easy-to-learn signal, but one not aligned with the system objective as a whole. Finally, the difference objective for a robot provides system-wide beneficial behavior, while remaining sensitive to the actions of a robot [2]:

$$D_j(z) = \begin{cases} \sum_i \left(\frac{V_i}{\frac{1}{2}\left(\delta_{i,j}+\delta_{i,k}\right)} - \frac{V_i}{\frac{1}{2}\left(\delta_{i,j}+\delta_{i,l}\right)} \right) & \text{if } \delta_{i,j}, \delta_{i,k} < \delta_{i,l} < \delta_o \\ \sum_i \frac{V_i}{\frac{1}{2}\left(\delta_{i,j}+\delta_{i,k}\right)} & \text{if } \delta_{i,j}, \delta_{i,k} < \delta_o \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where l is the third closest robot to POI i (robots j and k are the closest two).

The environment was highly dynamic, where 10% of the POIs (selected randomly) changed location and value at each episode. This was done to encourage specific coordination behavior based on sensor inputs rather than specific x-y coordinates. The results are based on 2000 episodes of 30 timesteps each, and are averaged for significance. Figure 1 shows that performance was similar for small problems, but that the difference objective provided a better signal to promote team formation for larger problems.

4. ENCOURAGING TEAM FORMATION

In the second problem we examine, multiple robots are encouraged (rather than required) to form teams to perform a task. In this problem, a POIs value is optimized for n robots observing it, but the system receives lesser value for other numbers of robots observing the POI. For these objectives, δ_o remains the same, and as before, three objectives are defined, beginning with the system objective given by:

$$G(z) = \sum_i \alpha V_i x e^{\frac{-x}{\beta}} \tag{4}$$

where i indexes POIs, x is the number of robots within δ_o, β is the observation capacity, and α is a constant chosen to be 1.37 such that the maximum of the exponential curve approximates the POI value V_i. For this new system objective, the selfish robot objective is defined as:

$$P_j(z) = \sum_{i_j} \alpha V_{i,j} x e^{\frac{-x}{\beta}} \tag{5}$$

where indexing and constant selection is the same as above. This objective includes no information regarding contribution to the system as a whole, rather indicating only what robot j can directly observe (the component of the system objective for which robot j was within δ_o). Finally, the difference objective for this system results in:

$$D_j(z) = \sum_{i_j} \alpha V_{i,j} \left[x e^{\frac{-x}{\beta}} - (x-1) e^{\frac{-(x-1)}{\beta}} \right] \tag{6}$$

where indexing and constant selection is the same as above. This objective aims to provide the contribution of robot j to the system.

Here again Figure 2 shows that as the system increases in complexity, the difference objective, through providing a better learning signal, provides consistent behavior through the increased complexity of the system. The system and local learning objective performance tapers off, where using the difference objective maintains its' performance slope, clearly indicating that when the number of robots within the system becomes large, the difference objective is able to maintain successful dynamic team formation.

Acknowledgments

This work was partially supported by AFOSR grant FA9550-08-1-0187 and NSF grant IIS-0910358.

5. REFERENCES

[1] A. K. Agogino and K. Tumer. Analyzing and visualizing multiagent rewards in dynamic and stochastic environments. *Journal of Autonomous Agents and Multi Agent Systems*, 17(2):320–338, 2008.

[2] A. K. Agogino and K. Tumer. Efficient evaluation functions for evolving coordination. *Evolutionary Computation*, 16(2):257–288, 2008.

[3] C. Guestrin, M. Lagoudakis, and R. Parr. Coordinated reinforcement learning. In *Proceedings of the 19th International Conference on Machine Learning*, 2002.

[4] M. Tambe. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7:83–124, 1997.