Form& CONTENT MEDIA LTD

DOI: sample

Aligning Agent Objectives for Learning and Coordination in Multiagent Systems

Kagan Tumer and Matt Knudson

Providing agents with difference objectives (a new class of aligned agent objectives) has led to coordinated behavior in many complex multiagent domains, including multi-robot coordination and air traffic control.

In large, distributed systems composed of adaptive and interactive components (agents), ensuring the coordination among the agents so that the system achieves certain performance objectives is a challenging proposition. The key difficulty to overcome in such systems is one of credit assignment: How to apportion credit (or blame) to a particular agent based on the performance of the entire system. This problem is prevalent in many domains including air or ground traffic, multi-robot coordination, sensor networks, and smart power grids [1,2]. In this article we provide a general approach to coordinating learning agents and present examples from the multi-robot coordination domain [3–5].

Many complex exploration domains (planetary exploration, search and rescue) require the use of autonomous robots. In addition, the use of multi-robot teams offers distinct advantages in efficiency and robustness over the use of a single robot. However, the potential gains also come at a cost: How to ensure that the robots do not work at cross-purposes and that the robots' efforts support a common, system level objective.

Directly extending single robot approaches to multi-robot systems presents difficulties in that the learning problem is no longer the same: the robots not only have to learn "good" actions, but actions that are complementary to one another in a constantly changing environment. Approaches that are particularly well suited to multi-robot systems include using Markov Decision Processes for online mechanism design [6], developing new reinforcement learning based algorithms [7–10], and domain based evolution [11]. In addition, forming coalitions for purposes of reducing search costs [12], employing multilevel learning architectures for the formation of coalitions [13], and market based approaches [14] have been examined. Finally, in problems with limited or no communication, devising agent-specific objective functions that implicitly include coordination components has proven very successful [3,4].

In this article, we summarize recent advances in developing such agent-specific objective functions. Given some system level objective function (e.g., number of areas explored), we aim to derive an objective function for the agents in such a way that when they achieve their own objectives, the system objective is also achieved. For some system-level objective $G(z)$, given as a function of the full system state z, consider the agent specific objective function for agent i:

$$D_i(z) = G(z) - G(z_{-i}) \tag{1}$$

where z_{-i} is the *counterfactual* state that does not depend on agent i's state. (In some systems it may not be practical to entirely remove an agent, in which case the counterfactual state is set to an "expected" action.)

This agent objective provides two benefits: First, each agent can ascertain the impact it's actions had on the system as a whole because the "difference" between the actual world and the counterfactual world removes many of the terms that do not depend on agent i. Because of this, this set of agent objectives have been called "difference objectives" [2,4]. Second, because the counterfactual term does not depend on the states of agent i, D_i and G have the same derivative with respect to changes in agent i's state. Intuitively, this means that an action that is beneficial for agent i is also beneficial for the system, though the agent does not explicitly need to know this.

This approach has been successfully applied to the multi-robot coordination domain. In this formulation, multiple robots are required to explore an environment where different points of interest for exploration have different values. The system objective is to maximize aggregate information collection and

Continued on next page

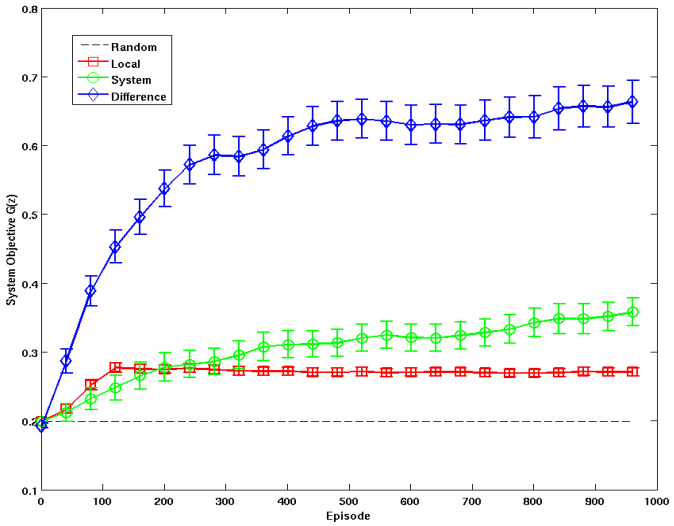

Figure 1. *The system objective is plotted over the number of training episodes for an environment with 10 robots and 40 points of interest. The robots are trained with the system-level objective, a local or selfish objective, and the difference objective.*

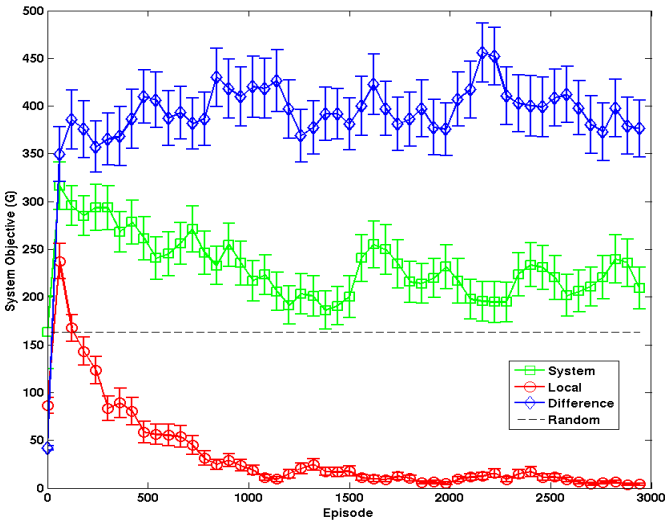

Figure 2. *The system objective is plotted over the number of training episodes for an environment with 40 robots and 50 points of interest. Two robots, one each of multiple types, must partner to observe a point of interest.*

the robots can observe the points of interest and each other, but they do not communicate. Instead, coordination is promoted through the use of the *difference objective* shown by Equation 1.

Figure 1 shows that when the robots use the system-level objective directly, learning is extremely slow. This is because all agents receive the same information, making it difficult for them to determine which of their actions is beneficial. Using only local information on the other hand, leads to agents competing for the points of interest, rather than cooperating. The difference objective on the other hand provides a signal that is both aligned with the system objective and sensitive to the actions of the agent. It therefore leads to the agents quickly learning the correct actions and coordinating successfully.

Figure 2 shows a domain where tighter coordination is required. In this case, points of interest provide higher value when observed by exactly two different types of robot (observations with one or more than two robots yields lower values), and the points of interest appear and disappear during the exploration stage. This domain severely tests the coordination of the robots as "incidental" coordination is not sufficient to achieve good behavior. The results show that the benefits of the difference objective are significantly more pronounced in this case, and that the agents form stable partnerships even though they do not communicate. Using the system-level objective, the agents struggle to do better than if they made random decisions. Quite expectedly, using a selfish objective produces entirely inappropriate behavior, resulting in almost no system benefit at all (a behavior similar to the tragedy of the commons [15]).

Whether the application is multi-robot exploration, distributed sensor networks, traffic management, or a host of other applications, many agents are tasked with coordinating to achieve a system-level goal. As the system grows more complex, becoming dynamic and containing hundreds or thousands of agents, the structural problem of assigning credit to individuals such that they can learn what benefits the system as a whole can quickly become intractable. Through the use of difference objectives, a system designer can develop a specific objective to give to each agent operating in the system that balances providing important global information with low signal to noise ratio. In addition to the benefits to robot coordination discussed in this article, the difference objective has proven successful in air traffic management, and continues development in the areas of distributed sensor networks, multi-objective optimization, and complex system decomposition.

This material is based upon work supported by AFOSR grant number FA9550-08-1-0187 and National Science Foundation grants 0910358 and 0931591.

Author Information

Kagan Tumer and Matt Knudson
Oregon State University
Corvallis, Oregon

Continued on next page

Dr. Kagan Tumer is a professor at Oregon State University. He received his Ph.D. from The University of Texas, Austin in 1996, and was a senior research scientist at NASA from 1997 to 2006. Dr. Tumer is the program co-chair of AAMAS 2011 and his research interests include learning and control in complex systems and multiagent coordination.

Dr. Matt Knudson received his Ph.D. in Mechanical Engineering in 2009, from Oregon State University. His research interests are dynamics and control, with a particular focus on model-free control techniques for autonomous robotics and coordination in teams of robots with limited resources.

References

1. K. Tumer and A. Agogino, *Distributed Agent-Based Air Traffic Flow Management*, **Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems**, pp. 330–337, Honolulu, HI, May 2007.
2. K. Tumer and D. Wolpert (eds.), **Collectives and the Design of Complex Systems**, Springer, New York, 2004.
3. A. K. Agogino and K. Tumer, *Efficient Evaluation Functions for Evolving Coordination*, **Evolutionary Computation 16** (2), pp. 257–288, 2008.
4. 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), pp. 320–338, 2008.
5. M. Knudson and K. Tumer, *Coevolution of Heterogeneous Multi-Robot Teams*, **Proceedings of the Genetic and Evolutionary Computation Conference**, pp. 127–134, Portland, OR, July 2010.
6. D. Parkes and S. Singh, *An MDP-Based Approach to Online Mechanism Design*, **NIPS 16**, pp. 791–798, 2004.
7. M. Ahmadi and P. Stone, *A Multi-Robot System for Continuous Area Sweeping Tasks*, **Proceedings of the IEEE Conference on Robotics and Automation**, pp. 1724–1729, May 2006.
8. C. Claus and C. Boutilier, *The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems*, **Proceedings of the Artificial Intelligence Conference**, pp. 746–752, Madison, WI, July 1998.
9. C. Guestrin, M. Lagoudakis, and R. Parr, *Coordinated Reinforcement Learning*, **Proceedings of the 19th International Conference on Machine Learning**, pp. 41–48, 2002.
10. J. Hu and M. P. Wellman, *Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm*, **Proceedings of the Fifteenth International Conference on Machine Learning**, pp. 242–250, 1998.
11. M. Alden, A.-J. van Kesteren, and R. Miikkulainen, *Eugenic Evolution Utilizing A Domain Model*, **Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2002)**, San Francisco, CA, 2002.
12. E. Manisterski, D. Sarne, and S. Kraus, *Enhancing MAS Cooperative Search Through Coalition Partitioning*, **Proc. Int'l Joint Conference on Artificial Intelligence**, pp. 1415–1421, 2007.
13. L. Soh and X. Li, *An Integrated Multilevel Learning Approach to Multiagent Coalition Formation*, **Proc. Int'l Joint Conference on Artificial Intelligence**, pp. 619–625, 2003.
14. Y. Ye and Y. Tu, *Dynamics of Coalition Formation in Combinatorial Trading*, **Proc. Int'l Joint Conference on Artificial Intelligence**, pp. 625–632, 2003.
15. G. Hardin, *The Tragedy of the Commons*, **Science 162**, pp. 1243–1248, 1968.