Approximating Difference Evaluations with Local Information

(Extended Abstract)

Mitchell Colby
Oregon State University
Corvallis, OR, USA
colbym@engr.orst.edu

William Curran
Oregon State University
Corvallis, OR, USA
curranw@engr.orst.edu

Kagan Tumer
Oregon State University
Corvallis, OR, USA
kagan.tumer@oregonstate.edu

ABSTRACT

Difference evaluations can effectively shape agent feedback in multiagent learning systems, and have provided excellent results in a variety of domains, including air traffic control and distributed sensor network control. In addition to empirical evidence, there is theoretical evidence demonstrating how difference evaluations help shape agent utilities/objectives in order to promote system-wide coordination. However, analytically calculating difference evaluation functions requires knowledge of the states of all agents in the system, as well as the mathematical form of the system evaluation function. In practice, neither of these elements are typically available. In this work, we demonstrate that each agent can locally approximate difference evaluations using only local state and action information, as well as a broadcast value (rather than the mathematical form) of the system evaluation function, allowing for difference evaluations to be implemented in multiagent systems where global state information is unavailable. This approximation technique is tested in a multiagent congestion problem, and the results demonstrate that approximate difference evaluations perform similarly to analytically computed difference evaluations, while using far less system information.

Categories and Subject Descriptors

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

General Terms

Algorithms

Keywords

Multiagent reinforcement learning, difference rewards

1. INTRODUCTION

Difference evaluation functions have been shown to significantly improve learning in multiagent systems, and have produced excellent results in many multiagent domains [1]. Difference evaluations are defined as [1]:

$$D_i(z) = G(z) - G(z_{-i} + c_i) \quad (1)$$

where $G(z)$ is the system evaluation function, z is the system state vector, z_{-i} is the system state vector without the contributions of agent i, and c_i is a *counterfactual* term which replaces agent i. Intuitively, the difference evaluation function determines the impact of agent i on the overall system evaluation function, by removing all elements of the system evaluation function not related to agent i. Difference evaluations have two key theoretical properties which lead to their effectiveness. First, they are *aligned* with the system objective function, meaning that any agent i which acts to increase the value of $D_i(z)$ also increases the value of $G(z)$. Second, as the last term in $D_i(z)$ removes all elements of $G(z)$ not related to agent i, difference evaluations are *sensitive* to the actions of a particular agent, resulting in a favorable signal to noise ratio allowing for agents to easily discern the effects of their actions on system performance.

Although difference evaluations provide excellent learned performance, they are often difficult to compute in practice. Computing the second term in Equation 1 requires the global state of the system as well as the mathematical form of $G(z)$. In practice, this information is typically unavailable in multiagent systems. In order to allow for the implementation of difference evaluations, they must be approximated when global knowledge is unavailable. Difference evaluations have been approximated in past work [**?**], but this approach relied on expert domain knowledge and global knowledge of the system state, and thus did not address the key motivating factors for approximating difference evaluations.

2. DOMAIN AND APPROACH

In order to approximate difference evaluations, we assume that each agent has access to its local state and action, as well as a broadcast value of $G(z)$. This information is typically available in a multiagent system, as some type of system performance metric is generally used to provide feedback to learning agents. At each time step, each agent records its local state s_i and action a_i (combined in vector z_i), as well as the broadcast value of the system evaluation function $G(z)$. Each agent maintains a local approximation $\hat{G}_i(z_i)$, which is updated with the $\{s_i, a_i, G(z)\}$ tuple according to:

$$\hat{G}_i(z_i) \leftarrow (1 - \alpha_A)\hat{G}_i(z_i) + \alpha_A G(z) \quad (2)$$

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

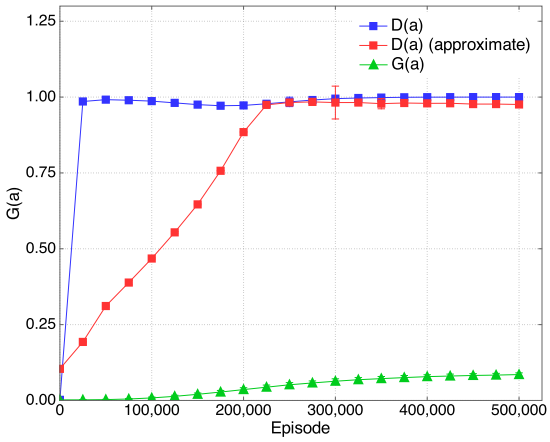

Figure 1: Bar problem results. When approximating $D_i(a)$, the learning rate is slower than when using $D_i(a)$, but the converged performance is nearly identical. $D_i(a)$ performs almost 10 times better than the overall system evaluation function $G(a)$.

where α_A is the update rate of the approximator. The approximate difference evaluation function is defined as:

$$\hat{D}_i(z) = G(z) - \hat{G}_i(c_i) \tag{3}$$

where c_i is the counterfactual state and action used to replace agent i. This approximation requires only local state and action information, as well as a broadcast value of $G(z)$. This approximation approach is tested in the multi-night modification of the El Farol bar problem [2]. A set of agents must each choose a night of the week to go to a bar, where there is an optimal capacity for each night. Agents are trained using multiagent reinforcement learning, and agent rewards are assigned with either $G(z)$, $D_i(z)$, or $\hat{D}_i(z)$.

3. RESULTS

The bar problem was initialized as follows. There are 1000 agents and 10 nights, where each night has a capacity of 10. The update rate α_A for the approximation of the system evaluation function is set to 0.1. The learning rate α for the Q-table is set to 0.1. Each experimental run lasted for $500,000$ timesteps, and there were 100 statistical runs conducted. The experimental results are shown in Figure 1, and the reported error bars are error in the mean σ/N^2.

As seen in Figure 1, approximating the difference evaluation function results in almost 10 times better performance than using the system evaluation function $G(a)$. When approximating $D_i(a)$, the solution takes much longer to converge than when analytically computing $D_i(a)$. However, converged performance of actual and estimated difference evaluation functions is nearly identical.

It is of note that although the analytical calculation of $D_i(a)$ results in faster learning, it is often impossible to analytically compute the system evaluation function in multiagent learning systems. Often, the mathematical form of $G(a)$ is unknown, meaning $D_i(a)$ cannot be directly computed. However, agents can still use local knowledge to approximate $G(a)$ and thus estimate difference evaluation functions, which drastically improve system performance. So, even though the analytical computation of $D_i(a)$ provides faster learning than when approximating $D_i(a)$, it is not always possible to perform this analytical computation.

The key result is that agents with only local knowledge converge to the same performance (although in more computational time) as agents with global knowledge about the system. In cases where global knowledge is available, it is often beneficial to use this knowledge while shaping agent feedback signals. However, in many cases, agents' knowledge is often limited to what they can observe, and constructing meaningful agent feedback based on this limited information is critical for ensuring high system performance. These results demonstrate that in some cases, approximate difference evaluations result in no significant loss in converged system performance when only local knowledge is available.

4. DISCUSSION

Difference evaluations have been empirically shown to improve coordination in multiagent systems in a many domains, including air traffic control, rover control, sensor network control, and congestion games. Further, difference evaluations are aligned with the system evaluation function, and are typically low in noise. However, a key limitation of difference evaluations is the requirement for global knowledge about the state of the system as well as the system evaluation function. Thus, directly implementing difference rewards in generic multiagent domains is often a difficult task.

In this work, we demonstrate that difference evaluations may be approximated by each agent using only local state and action information. The only assumption is that the value of the system evaluation function $G(s, a)$ can be broadcast to each agent, which in most cases is a reasonable assumption. We present an approach for approximating difference evaluation functions in order to provide agent-specific feedback to improve coordination, and demonstrate in two domains the effectiveness and scalability of our approach.

The key contribution of this work is presenting a novel method to implement difference evaluation functions in any generic multiagent system, without requiring global knowledge about the state of the system or the mathematical form of the system evaluation function. Further, this approximation approach significantly outperforms methods which use the overall system evaluation function. As each agent maintains a local approximation of the system evaluation function, increases in computational cost are insignificant, because the computation is parallelized across each agent in the system.

Acknowledgments

This work was partially supported by NETL DE-FE0012302 and NSF CMMI-1363411.

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] W. B. Arthur. Inductive Reasoning and Bounded Rationality (The El Farol Problem). In *Amer. Econ. Review 1994*, volume 84, pages 406–411, 1994.