Theory and Applications of Difference Evaluation Functions

(Doctoral Consortium)

Mitchell Colby
Oregon State University
442 Rogers Hall
Corvallis, OR 97331
colbym@engr.orst.edu

ABSTRACT

The credit assignment problem (which agents get credit or blame for system performance) is a key research area. For a team of agents collaborating to achieve a goal, the effectiveness of each individual agent must be calculated in order to give adequate feedback to each agent. We use the *Difference Evaluation Function* to find agent-specific feedback. The Difference Evaluation Function has given excellent empirical results in many domains, including air traffic control and mobile robot control. Though there has been some theoretical work that shows why Difference Evaluation Functions improve system performance, there has been no work to show when and under what conditions such improvements are realized. We apply an evolutionary game-theoretic analysis to show the theoretical advantages of the Difference Evaluation Function. We then focus on how to apply these multiagent learning methods to optimize distributed sensor networks in advanced power generation systems.

Categories and Subject Descriptors

H.3.4 [**Systems and Software**]: Distributed Systems

Keywords

Multiagent learning; Coordination

1. MULTIAGENT LEARNING SYSTEMS

The following sections describe the Difference Evaluation Function and its theoretical advantages over traditional control methods.

1.1 Difference Evaluation Function

The *Difference Evaluation Function* is defined as in [2]:

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

where $C(z)$ is the global evaluation, and $C(z_{-i})$ is the global evaluation without the influence of agent i. Intuitively, $D_i(z)$ gives agent i's specific impact on $C(z)$. Note that the second term removes all effects on $C(z)$ not related to agent i, so an agent acting to increase $D_i(z)$ will also act to increase $C(z)$. This property is termed *factoredness* [1]. Further, because the Difference Evaluation only depends on the actions of agent i, noise from other agents is reduced in the feedback given by $D_i(z)$. This property is termed *learnability* [1]. Difference Evaluation Functions provide *structural* information and can be used in both multiagent reinforcement learning and cooperative coevolutionary algorithms [2].

An extension of the Difference Evaluation Function is the Expected Difference Evaluation Function, defined as in [2]:

$$ED_i(z) = C(z) - \sum_{a \in A} P_i(a) C_i(z_a) \tag{2}$$

where $P_i(a)$ is the probability that agent i takes action a, and $C_i(z_a)$ is the global evaluation when agent i takes action a. The Expected Difference Evaluation Function gives the expected impact of agent i on the global reward. If we assume a uniform probability of taking each action, we get the Average Difference Evaluation $AD_i(z)$.

1.2 Theoretical Advantages of D(z)

One avenue of current research involves showing the theoretical advantages of the Difference Evaluation Function. Using an evolutionary game-theoretic setting with two cooperating agents, we have shown that the probability of reaching the globally optimal Nash equilibrium with cooperative coevolution when using Difference Evaluation Functions is greater than when using the global evaluation function under benign conditions (work submitted to GECCO '13). Specifically, suppose that we have a payoff matrix C where c_{ij} is the payoff for the joint action (i, j), and c_{avg} is the average payoff for all joint actions. We prove the following: **Theorem 1:** if:

$$E[i^*] < c_{avg} \tag{3}$$

$$E[j^*] < c_{avg} \tag{4}$$

Then:

$$\lim_{t\to\infty} P_D(x_{i^*}^{(t)} = 1) > \lim_{t\to\infty} P_C(x_{i^*}^{(t)} = 1) \tag{5}$$

$$\lim_{t\to\infty} P_D(y_{j^*}^{(t)} = 1) > \lim_{t\to\infty} P_C(y_{j^*}^{(t)} = 1) \tag{6}$$

where $P_D(x_{i^*} = 1)$ and $P_C(x_i^* = 1)$ are the probabilities that the Difference Evaluation Function and the global evaluation function yield the optimal policy for the first agent, respectively. This essentially means that if the optimal actions are deceptive (i.e. the payoff is comparatively low

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

unless the collaborating agent selects the correct action), then use of the Difference Evaluation Function increases the probability of convergence to the optimal Nash equlibrium. Figure 1 shows relative performance of $D(z)$ vs $C(z)$ when the conditions in Equations 3 and 4 are satisfied. When the assumptions of Theorem 1 are satisfied, the Difference Evaluation Function leads to convergence to the optimal Nash equilibrium, while the global evaluation function leads to convergence to a suboptimal solution. When the assumptions of Theorem 1 are not met, both Difference Evaluation Functions and global evaluation functions can be good or bad depending on system dynamics. These empirical results support the theoretical conclusions in Theorem 1.

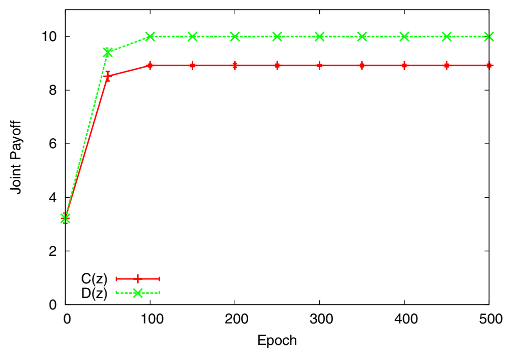

Figure 1: Equations 3 and 4 are satisfied. As predicted by Theorem 1, the difference payoff matrices outperform the standard payoff matrix

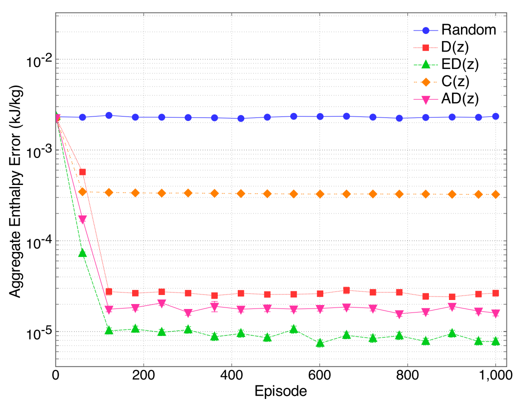

Figure 2: DSN measurement error with 1000 sensors. Variants of the Difference Evaluation Function give almost two orders of magnitude improvement.

2. POWER PLANT APPLICATIONS

A key focus of our research is the design of optimal sensor control strategies in advanced power systems. A Distributed Sensor Network (DSN) consists of multiple autonomous sensing agents which monitor environmental conditions such as temperature or pressure. DSNs have many advantages over traditional sensing implementations, such as:

- Multiple sensors means no single point of failure.
- Sensors can preprocess information, making information from the sensor network more useful than simple aggregate information.
- Sensors in DSNs can collaborate to provide measurements which are more accurate than any single sensor.

DSNs are a very attractive sensing implementation for power plant applications, because the robust nature of the DSN is a critical property in high-risk applications such as power plants. In order for DSNs to be used in power plants, the sensors must be trained in order to provide accurate measurements. We use the Difference Evaluation Function in order to train sensors using multiagent reinforcement learning. When using the Difference Evaluation Function to train a DSN with 1000 sensors, the distributed sensor network was two orders of magnitude more accurate than when training with the global evaluation function (Figure 2).

3. CONCLUSION

One of the key difficulties in multiagent learning systems is that of credit assignment. Given a team of agents collaborating to complete an overall system objective, determining the inidividual contribution of each agent towards reaching that objective is crucial for providing adequate feedback to each learning agent. The Difference Evaluation Function aims to solve this credit assignment problem, and we have used evolutionary game theory to show the theoretical advantages of using the Difference Evaluation Function.

Our first avenue of future work involves further investigation the theoretical properties of Difference Evaluation Functions. Specifically, we will analyze the conditions given by Equations 3 and 4 to determine their impact on system performance, and how the degree to which a condition is met (or not met) can affect agent behavior. Next, we will attempt to provide a similar proof in a multiagent reinforcement learning setting, in order to show that the Difference Evaluation Function has advantages in both cooperative coevolution and multiagent reinforcement learning.

We also aim to apply cooperative coevolutionary algorithms using the Difference Evaluation Function on advanced power system control. System models for these plants are unavailable due to their complexity; further, these systems are comprised of many subsystems. The model-free and distributed nature of these systems creates an ideal testbed for multiagent learning techniques. The contributions of this work will be two-fold. First, we will show that a multiagent learning approach can provide solutions for sensing and control in advanced power systems. Second, we will show the effectiveness of the Difference Evaluation Function in complex real-world problems.

Acknowledgements

This work was partially supported by NETL grant DE-FE0000857.

4. 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] K. Tumer. Designing agent utilities for coordinated, scalable and robust multiagent systems. In P. Scerri, R. Mailler, and R. Vincent, editors, *Challenges in the Coordination of Large Scale Multiagent Systems*, pages 173–188. Springer, 2005.