A Replicator Dynamics Analysis of Difference Evaluation Functions

(Extended Abstract)

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

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

ABSTRACT

A key difficulty in Cooperative Coevolutionary Algorithms (CCEAs) is the credit assignment problem[1]. One solution to the credit assignment problem is the difference evaluation function, which produces excellent results in many multiagent domains. However, to date, there has been no prescriptive theoretical analysis deriving conditions under which difference evaluations improve the probability of selecting optimal actions. In this paper, we derive such conditions.

Categories and Subject Descriptors

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

Keywords

Multiagent learning; difference evaluation functions

1. DIFFERENCE EVALUATIONS

The difference evaluation function $D_i(z)$ is defined as [1]:

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

where $G(z)$ is the global evaluation function, $G(z_{-i} + c_i)$ is the global evaluation function without the effects of agent i, and c_i is the *counterfactual* term used to replace agent i. Difference evaluations have desirable theoretical properties and have provided excellent empirical results [1], but no prescriptive theory has been conducted to date.

2. EGT MODEL FOR CCEAS

The EGT-RD model for CCEAs is defined as in [2]:

$$u_i^{t,c} = \sum_{j=1}^{m} c_{ij} y_j^t$$

$$w_j^{t,c} = \sum_{i=1}^{n} c_{ij} x_i^t$$

$$x_i^{t+1,c} = \left(\frac{u_i^{t,c}}{\sum_{k=1}^{n} x_k^t u_k^{t,c}} \right) x_i^t$$

$$y_j^{t+1,c} = \left(\frac{w_j^{t,c}}{\sum_{k=1}^{m} y_k^{(t)} w_k^{t,c}} \right) y_j^t$$

Where:

- $u_i^{t,c}$: fitness of agent 1 taking action i at time t
- $w_j^{t,c}$: fitness of agent 2 taking action j at time t
- x_i^t: probability agent 1 takes action i at time t.
- y_j^t: probability agent 2 takes action j at time t.
- $x_i^{t+1,c}$: probability agent 1 takes action i at time $t+1$
- $y_j^{t+1,c}$: probability agent 2 takes action j at time $t+1$

For this analysis, we assume that: **(A1)** All elements of the payoff matrix are non-negative; **(A2)** Not all elements in the payoff matrix have the same value. A1 ensures that the system remains invariant in the simplex. A2 is needed in the proofs, but if this assumption does not hold we have a trivial payoff matrix where every element has equal value.

3. DIFFERENCE PAYOFF MATRICES

We define agent-specific *difference payoff matrices* D^1 and D^2 by applying the difference evaluation function to the global payoff matrix C:

$$d_{ij}^1 = c_{ij} - \frac{1}{n} \sum_{k=1}^{n} c_{kj} + c_{max}$$

$$d_{ij}^2 = c_{ij} - \frac{1}{m} \sum_{k=1}^{m} c_{ik} + c_{max}$$

The difference payoff matrices are implemented in the EGT model in the fitness assignment stage.

We now derive the fitness for an agent using the difference payoff matrices in terms of the global payoff matrix. The fitness for the first agent taking the action i while using the difference evaluation function at time t is:

$$\begin{aligned} u_i^{t,d} &= \sum_{j=1}^{m} d_{ij}^1 y_j^t \\ &= \sum_{j=1}^{m} \left(c_{ij} - \frac{1}{n} \sum_{k=1}^{n} c_{kj} + c_{max} \right) y_j^t \\ &= u_i^{t,c} - \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{m} c_{kj} y_j^t + c_{max} \end{aligned}$$

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.*

A similar derivation yields the normalized fitness for the second agent taking action j and using the difference evaluation at time t:

$$\bar{w}_j^{t,d} = \frac{w_j^{t,c} - \frac{1}{m}\sum_{k=1}^{m} w_k^{t,c} + c_{max}}{m \cdot c_{max}}$$

4. DIFFERENCE EVALUATIONS THEORY

We define the joint expected system payoff at time t as:

$$E_{tot}^t[C] = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} x_i^t y_j^t$$

We now prove that in cases where the optimal action (corresponding to the optimal Nash equilibrium) has a relatively low expected payoff, difference evaluations result in superior system performance compared to the global evaluation.

THEOREM 1. *If the fitness values for the best response actions i^* and j^* are less than the joint expected system payoff, then difference evaluations result in higher probabilities of selecting the best response actions as compared to the global evaluation function.*

$$E_{tot}^t[C] > u_{i^*}^{t,c} \Rightarrow x_{i^*}^{t+1,d} > x_{i^*}^{t+1,c} \quad (1)$$

PROOF. Starting with Equation 1, we have that:

$$E_{tot}^t[C] > u_{i^*}^{t,c}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} x_i^t y_j^t > u_{i^*}^{t,c}$$

Noting that $u_i^{t,c} = \sum_{j=1}^{m} c_{ij} y_j^t$, we have that:

$$\sum_{i=1}^{n} u_i^{t,c} x_i^t > u_{i^*}^{t,c}$$

We now multiply both sides of the inequality by a positive constant A:

$$A\sum_{i=1}^{n} u_i^{t,c} x_i^t > A u_{i^*}^{t,c}$$

Noting that $\sum_{i=1}^{n} A x_i^t = A$, we have that:

$$A\sum_{i=1}^{n} u_i^{t,c} x_i^t > u_{i^*}^{t,c} \sum_{i=1}^{n} A x_i^t$$

We add $u_{i^*}^{t,c}\sum_{i=1}^{n} u_i^{t,c} x_i^t - u_{i^*}^{t,c}\sum_{i=1}^{n} u_i^{t,c} x_i^t = 0$ to the right hand side of the inequality in order to allow factoring of terms, yielding:

$$A\sum_{i=1}^{n} u_i^{t,c} x_i^t > u_{i^*}^{t,c}\left[\sum_{i=1}^{n} u_i^{t,c} x_i^t + \sum_{i=1}^{n} A x_i^t - \sum_{i=1}^{n} u_i^{t,c} x_i^t\right]$$

$$\Rightarrow A\sum_{i=1}^{n} u_i^{t,c} x_i^t > u_{i^*}^{t,c}\left[\sum_{i=1}^{n}\left(u_i^{t,c} + A\right) x_i^t - \sum_{i=1}^{n} u_i^{t,c} x_i^t\right]$$

$$\Rightarrow A\sum_{i=1}^{n} u_i^{t,c} x_i^t + u_{i^*}^{t,c}\sum_{i=1}^{n} u_i^{t,c} x_i^t > u_{i^*}^{t,c}\sum_{i=1}^{n}\left(u_i^{t,c} + A\right) x_i^t$$

$$\Rightarrow \left(u_{i^*}^{t,c} + A\right)\sum_{i=1}^{n} u_i^{t,c} x_i^t > u_{i^*}^{t,c}\sum_{i=1}^{n}\left(u_i^{t,c} + A\right) x_i^t \quad (2)$$

We now focus on the term A from Equation 2. Recall that:

$$\begin{aligned} u_i^{t,d} &= \sum_{j=1}^{m}\left(c_{ij} - \frac{1}{n}\sum_{k=1}^{n} c_{kj} + c_{max}\right) y_j^t \\ &= \sum_{j=1}^{m} c_{ij} y_j^t - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^t + \sum_{j=1}^{m} c_{max} y_j^t \\ &= u_i^{t,c} - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^t + c_{max} \end{aligned}$$

We thus define A as:

$$A = -\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^t + c_{max}$$

Note that A is strictly positive by assumptions A1 and A2 from Section 2. With this definition of A, we have that:

$$u_i^{t,d} = u_i^{t,d} + A \quad (3)$$

Combining Equations 2 and 3 yields:

$$\begin{aligned} & u_{i^*}^{t,d}\sum_{i=1}^{n} u_i^{t,c} x_i^t > u_{i^*}^{t,c}\sum_{i=1}^{n} u_i^{t,d} x_i^t \\ \Rightarrow & \frac{u_{i^*}^{t,d}}{\sum_{i=1}^{n} u_i^{t,d} x_i^t} > \frac{u_{i^*}^{t,c}}{\sum_{i=1}^{n} u_i^{t,c} x_i^t} \end{aligned} \quad (4)$$

Note that the terms in the inequality from Equation 4 are equivalent to the coefficients in the population update rules from the EGT model. We thus have:

$$x_{i^*}^{t+1,d} > x_{i^*}^{t+1,c}$$

A similar derivation for the second agent yields:

$$E_{tot}^t[C] > w_{j^*}^{t,c} \Rightarrow y_{j^*}^{t+1,d} > y_{j^*}^{t+1,c} \quad (5)$$

□

Thus, if the condition in Equation 1 holds, then difference evaluations improve the probability of selecting the optimal action. Equation 1 holding corresponds to the optimal Nash equilibrium being*deceptive*, meaning that if one agent deviates from this equilibrium then the system payoff is dramatically increased. In these cases, extremely tight coordination is required to reach the optimal Nash equilibrium, because unless both agents select the best response action simultaneously, the system payoff will be low. In these difficult games, we have shown that difference evaluations can improve system performance.

Acknowledgments

This work was partially supported by NASA NNX14AI10G and NETL DE-FE0012302.

REFERENCES

[1] A. 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] L. Panait, K. Tuyls, and S. Luke. Theoretical Advantages of Lenient Learners: An Evolutionary Game Theoretic Perspective. *J. Mach. Learn. Res.*, 9:423–457, June 2008.