A Co-evolution model of Scores and Strategies in IPD games: toward the understanding of the emergence of the social morals

Yoshiki Yamaguchi, Tsutomu Maruyama and Tsutomu Hoshino
Institute of Engineering Mechanics and Systems, University of Tsukuba
1-1-1 Ten-ou-dai Tsukuba Ibaraki, 305-8573 JAPAN
E-mail: {yoshiki, maruyama, hoshino}@darwin.esys.tsukuba.ac.jp

Abstract

It was shown that the evolution of a world consisting of agents that play Iterated Prisoner Dilemma (IPD) games each other is open-ended by Lindgren. The behavior of the world is very sensitive to the values of the payoff matrix used in IPD games, because the values have great influence on the population dynamics of the world. In general, the values are fixed throughout the simulations. In the real world, however, morals and the behaviors of individuals that follow the morals have been evolved influencing mutually.

In this paper, we propose a co-evolution model of agents and scores of IPD games toward the understanding of emergence of social morals. The co-evolution model consists of two layers. In the first layer, scores for IPD games are evolved using a genetic algorithm. Scores vary within the range of dilemma games, and scores that attract more agents in the second layer will gradually increase. In the second layer, agents play IPD games with all other agents following the scores that they believe and are evolved using Lindgren’s model.

Simulation results showed that the values of the scores evolve toward the score which gives more payoffs for cooperative strategies and less payoffs for defective strategies as the strategies of the IPD agents are evolved. The results also showed that small colonies of defective strategies repeatedly appear and disappear throughout the simulations.

Introduction

It was shown that the evolution of a world consisting of agents that play Iterated Prisoner Dilemma (IPD) games each other is open-ended by Lindgren (1). In the model, strategies of the agents are evolved by mutations, and the agents that obtained more payoffs (using a given payoff matrix) will gradually increase its ratio in the total population. The behavior of the evolution of the world is very sensitive to the values of the payoff matrix, because the values have great influence on the population dynamics of the world. For instance, when a defective strategy is created by mutations, the balance of the values of the payoff matrix decides whether the strategy can increase in co-operative worlds or not. In general, the values are fixed throughout the simulation. In the real world, however, it seems that morals (which have strong influences on the behaviors of individuals like the values of the payoff matrix in IPD games) are gradually refined as the behaviors of individuals in the world are evolved and vice versa.

In this paper, we propose a co-evolution model of agents and scores of IPD games toward the understanding of emergence of social morals. The co-evolution model consists of two layers. In the first layer, scores for IPD games are evolved using genetic algorithms. Scores vary within the range of dilemma games, and scores that attract more agents in the second layer and obtain more payoffs (total payoffs obtained by the agents that follow the scores) will gradually increase. In the second layer, agents play IPD games with all other agents following the scores that they believe and are evolved using Lindgren’s model. With this co-evolution model, we can observe how the scores are evolved as the strategies of the IPD agents are evolved, and how the scores lead the evolution of the behaviors (strategies) of the IPD agents.

Co-evolution Model

In this section, we describe the details of the co-evolution model. Our model has two layers; agents’ layer and scores’ layer. Figure 1 shows the overview of the model. In each layer, the size of the circle shows the ratio of each score and agent in the total population.

Figure 1: Co-evolution Model
The Agents’ Layer

In the agents’ layer, each agent plays IPD games with all other agents following Lindgren’s IPD model (1), and agents that obtained more payoffs will increase its ratio in the next generation. Each agent follows one of the payoff matrix (score) in the score’s layer, and their payoffs are given based on the payoff matrix (score). Therefore, agents with same strategies may get different payoffs even if they play games against same agents.

The Scores’ Layer

Table 1 shows the payoff matrix used in our model.

<table>
<thead>
<tr>
<th>Agent 1 (C)</th>
<th>Cooperate (C)</th>
<th>Defect (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reward</td>
<td>3 + \gamma_i</td>
<td>0 + \beta_i</td>
</tr>
<tr>
<td>Sucker</td>
<td>-3 + \gamma_i</td>
<td>-1 + \delta_i</td>
</tr>
</tbody>
</table>

Our matrix has four parameters; \( \alpha_i, \beta_i, \gamma_i \) and \( \delta_i \). These parameters are evolved using a simple genetic algorithm, and the scores (matrices) that attracted more agents and obtained more payoffs (total payoffs obtained by agents that follow the scores) will increase its ratio in the next generation.

By changing the values of these parameters, categories of games become one of the followings.

- \( T > P > R > S \): Deadlock
- \( T > R > P > S \): IPD
- \( T > R > S > P \): Chicken
- \( R > T > P > S \): Stag Hunt

Figure 2 shows the relationship between a score (matrix) and categories of the games.

![Figure 2: Scores and the Category of Games (\( \beta_i = 0 \))](image)

In order to simplify the figure (a figure by four parameters is four dimensional), we fix the value of \( \beta_i \) to zero. In Figure 2, the projected figure on the floor \((\gamma_i \text{ and } \delta_i \text{ plain})\) shows the relationship between the two parameter when \( \alpha_i \) and \( \beta_i \) are fixed to -1.5 and 0. The black parts show areas out of dilemma games (CC > DC > CD > DD).

We limited the range of parameters as shown in the equation below. By this limitation, scores move only in the range of dilemma games.

\[
-1.5 \leq \alpha_i, \beta_i, \gamma_i, \delta_i \leq 1.5
\]

Evolutionary Operations for Scores

Mutation At the initial state, there is only one score (matrix) in the scores’ layer, and its parameters are all zero. Therefore, all agents in the agents’ layer follow the score.

Figure 3 shows how a new score is created by mutations. When a mutation happens to a score, its ratio \( y_i \) in the total population is divided to \( (1 - \tau) y_i \) and \( \tau \times y_i \) and \( \tau \times y_i \) is given to the new score created by the mutation. The ratio of the agents that follows the score is also divided to \( (1 - \xi) x_a \) and \( \xi \times x_a \), and \( \xi \times x_a \) follows the new score.

![Figure 3: Mutation of Scores](image)

In the mutation of the score, the parameters are changed using a gaussian function \( f(s, t, u, v) \) below.

\[
f(s, t, u, v) = \frac{1}{4 \pi \sigma_s^2 \sigma_t^2 \sigma_u^2 \sigma_v^2} \exp \left( -\frac{(s-a)^2}{2 \sigma_s^2} - \frac{(t-b)^2}{2 \sigma_t^2} - \frac{(u-c)^2}{2 \sigma_u^2} - \frac{(v-d)^2}{2 \sigma_v^2} \right)
\]

We fixed the value of \( \sigma_s, \sigma_t, \sigma_u \) and \( \sigma_v \) to \( \frac{1}{3} \).

Crossover Figure 4 shows how new scores are created by the crossover operation.

![Figure 4: Crossover of Scores](image)

First, two scores are selected at random, and then parameters in the payoff matrices are crossed over. The values of parameters are changed using the function \( g(s, t, u, v) \) below.

\[
g(s, t, u, v) = \frac{1}{4 \pi \sigma_s^2 \sigma_t^2 \sigma_u^2 \sigma_v^2} \exp \left( -\frac{(s-a)^2}{2 \sigma_s^2} - \frac{(t-b)^2}{2 \sigma_t^2} - \frac{(u-c)^2}{2 \sigma_u^2} - \frac{(v-d)^2}{2 \sigma_v^2} \right)
\]

We fixed the value of \( \sigma_s, \sigma_t, \sigma_u, \sigma_v \) as shown below.

\[
\sigma_s = \frac{1}{2} \left| \frac{a - a_i}{2} \right|, \sigma_t = \frac{1}{2} \left| \frac{b - b_i}{2} \right|, \sigma_u = \frac{1}{2} \left| \frac{c - c_i}{2} \right|, \sigma_v = \frac{1}{2} \left| \frac{d - d_i}{2} \right|
\]
Agents that follow the scores which are crossovered are divided into two groups \((1 - \lambda)x_e\) and \(\lambda \times x_e\), respectively, and \(\lambda \times x_e\) and \(\lambda \times x_e\) follow new scores created by the crossover operation as shown in the figure 4.

Disappearance of Scores  The total number of scores (number of scores with difference values of the parameters because scores with same values are grouped) in the scores' layer is limited, and when the number of scores exceeds the limit, scores that have less followers are deleted. When a score is deleted, the agents that follow the score select one of the scores in the scores' layer at random.

Figure 5: Disappearance of Scores

Population Dynamics  
The ratio of score \(k\) (matrix \(k\)) in the total population is expressed as \(y_k\). The population dynamics of scores are decided using the total payoffs of agents that follow the scores. Therefore, the point for score \(k\) in one generation is

\[ p_k = \sum s, \text{ (added only if agent; follows the score)} \]

The ratio of the score in generation \(t+1\) is decided by the following equation, where \(y_i(t)\) is the ratio of the agent in generation \(t\) and \(d_{\text{score}}\) is a constant.

\[ y_i(t+1) = y_i(t) + d_{\text{score}} \cdot y_i(t) \left( \frac{x_{\text{max}}}{x_{\text{ave}}} \right) \]

Simulation Results  
We simulated the behavior of our co-evolution model changing the range of parameters in the payoff matrix (table 1) as follows.

1. changing the parameters without any limitation within the range below.
   \[-1.5 \leq \alpha, \beta, \gamma, \delta \leq 1.5\]
2. adding the limitations as follows.
   \[\alpha + \delta = 0, \quad \beta + \gamma = 0\]
3. narrowing the range as shown below and with the limitation above.
   \[-1.0 \leq \alpha, \beta, \gamma, \delta \leq 1.0\]

By changing the range and the limitations, we can observe several kinds of behaviors of the evolution, though most of them proceed toward cooperative worlds.

Other Parameters  
Table 2 shows the parameters used for the IPD agents, and table 3 shows the parameters for the scores. In the tables, number of agents/scores means the number of agents/scores with different strategies/values. In the simulations, agents/scores with same strategies/values are grouped and treated as one strategy/score.

Table 2: Parameters for Agents

<table>
<thead>
<tr>
<th>(N_{\text{agent}})</th>
<th>iteration times of games</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>90000</td>
</tr>
</tbody>
</table>

Table 3: Parameters for Scores

<table>
<thead>
<tr>
<th>(p_{\text{score}})</th>
<th>score mutation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.1\times 10^{-4})</td>
<td></td>
</tr>
</tbody>
</table>

Simulation Results with no Limitation  
In this case, parameters in the matrix table can move within the range below without any limitation.

\[-1.5 \leq \alpha, \beta, \gamma, \delta \leq 1.5\]

Figure 6 shows the average point of agents. In this simulation, agents following scores with larger payoffs (namely larger values of parameters) can get more payoffs in any situations of the IPD games. Therefore, all parameters of scores go to \(+1.5\) in despite of strategies of agents that follow the scores. A table in figure 6 (right-side) shows this payoff matrix. With this table, a defective strategy created in a cooperative world by mutation can obtain more payoffs compared with the original matrix (all parameters are zero). However, cooperative strategies can also get more payoffs even if they are defected, and the whole world always proceeds to cooperative world.

Figure 6: Average Points of Agents (No Limitation)

Simulation Results with the Limitation  
By adding the following limitations, all agents are faced with more strict dilemma situations. With the first equation, scores that give more payoffs to cooperative strategies also give smaller payoffs when defected, and with the second equation, scores that give more payoffs...
to defective strategies also give smaller payoffs when defected.

\[ \alpha_1 + \delta_1 = 0, \quad \beta_1 + \gamma_1 = 0 \]

Figure 7 shows the relationship between scores and categories of games (left-side) and a payoff matrix with this limitation (right-side). In the figure 7, a point in the center shows the original matrix used in Lindgren’s model. The two vertical axes show the payoff obtained when the players’ moves are CC and CD, while the two horizontal axes show the payoff obtained when the players’ moves are DC and DD.

Figure 7: Scores and the Categories of Games

With the limitation above, simulations show that the world will fall into one of the followings and become stable.

**Cooperative world**

The left graph in figure 8 shows the average point obtained by agents in this case. At the earlier generations, the world is not stable yet, and many strategies appear and disappear. As the strategies are evolved, the scores are also evolved and gather to the score shown in a table of the figure 9 (right-side). Figure 9 (left-side) shows the distribution of scores at generation 90,000. All of the scores are in Staghunt area, because it gives more point (4.5) for cooperative strategies, and smaller point (3.5) for a defective strategy even if it defects cooperative strategies.

**Defective world**

With much smaller probability, defective strategies dominate the world. The right graph in figure 8 shows the average point obtained by agents in this case. Even in the defective world, the scores also gather to Staghunt area shown in the figure 9. When defective strategies dominate the world after the parameters in scores are fully evolved toward the values shown in a table in figure 9, the defective strategies can get much payoff (2.5) by defecting each other, and cooperative strategies can not invade into the defective world unless many cooperative strategies are created at the same time by mutations, which are almost impossible with the parameters shown in the table 3.

Figure 8: Average point of Agents (from 0 to 90000)

Figure 9: Distribution of Scores (Generation 90000)

**Simulation Results with Narrower Range**

When moving the parameters in the scores as shown in the previous subsection, all scores gather to Staghunt area and they do not move any more. Therefore, we simulated the co-evolution model with narrower range of the parameters as shown below.

\[ \alpha_1 + \delta_1 = 0, \quad \beta_1 + \gamma_1 = 0, \quad -1.0 \leq \alpha_1, \beta_1, \gamma_1, \delta_1 \leq 1.0 \]

Figure 10 shows the new relationship between scores and categories of games, which is a part of the figure 7.

Figure 10: Scores and the Categories of Games

Figure 11 shows the average point and the number of scores, and the detail of the change from generation 25000 to 35000. In these figures, the left vertical axis shows the average point and the right vertical axis shows the number of scores. As shown in these figures, the average of the average point throughout the simulation is almost three, and this means that the world is
almost dominated by cooperative strategies. As shown in the figures, the average point becomes worse when the number of scores increase.

Figure 11: Average Points and the number of Scores

Figure 12 shows the ratio of the scores with in the range below which is close to the border with the Stag-hunt area. As shown in the figure, scores in the range always occupy the most of the population of scores.

$$-1.0 \leq \alpha, \beta, \gamma, \delta \leq +1.0$$

Figure 13 shows the distribution of scores. As shown in these figures, the scores repeat the cycles that the scores once gather to the area shown above, and then disperse. When the scores gather to the area, most of the strategies are cooperative, and the value of the parameter $\alpha$ in the score has no influence on their payoffs. Therefore, the value of $\alpha$ can be mutated freely (neutrality of mutation). However, this mutation makes it easier for defective strategies (created by mutation of strategies) to invade the cooperative world. Then, the number of the defective strategies increases and the average point becomes worse at this phase. By the invasion of the defective strategies, the scores that allows the invasion die out, and the world is dominated by the scores which are severe against defective strategies again.

Figure 13: Distribution of Scores

When the range of the values of the scores are limited to more strict dilemma game areas, we can observe cycles as follows. First, all values of all scores move to the cooperative scores which give less payoffs for defective strategies as described above. Under the cooperative circumstances, values of scores which have no influence on cooperative moves (namely values which decide the payoffs for defective strategies) are mutated freely (neutrality of mutation). Then, this mutation allows invasion by defective strategies created by mutations. Then the cooperative scores which allow the invasion die out, and the world is dominated by the cooperative scores which are severe against defective strategies again.

We believe that cycles we observed in our co-evolution model are also observed in the real world, though there are many future works left. The behavior of the model is very sensitive to all parameters used in the simulations. We need to analyze the behaviors of the model varying the parameters fixed throughout the simulations in this paper. Furthermore, we need to investigate how the scores are evolved when the agents are placed in a world with a concept of distance, and costs for maintaining the scores obtained by the co-evolution.

**Conclusions**

Simulations of the co-evolution model showed that scores of IPD games are gradually evolved toward cooperative scores, which give more payoffs for cooperative strategies and less payoffs for defective strategies as the strategies of the agents are evolved, under all three situations we have tested. With these cooperative scores, cooperative strategies dominate the world with very high probability.

**References**


Copyrighted Material