Zhaoying Lei, Yu Qi Zhou, Yu Shi Wang

Prisoner's Dilemma

Introduction

In our project, our main objective is to evolve the best solution, in other words, the solution where we would yield the most points, for the iterated prisoner's dilemma using a genetic algorithm (GA). Therefore, to see if our strategy can "win" at the prisoner's dilemma by our definition. Another objective of our project is to verify if our solution evolved is similar to TIT FOR TAT, the best solution according to literature. This problem is essential to study because it affects many real-life situations where two self interested parties are facing against each other. This situation has been used to study behaviors in international politics, environmental studies, economics and addiction research. In literature, it is stated that although in a classic Prisoner's dilemma where the player knows the number of outcomes, always defecting will yield a higher result, in the case where the prisoners do not know the number of rounds, a better solution would be TIT FOR TAT. In our case, to simulate more realistic situations, we have chosen to make a model where there would be an indeterminate number of rounds, as in real life, it is impossible for any party to know exactly the number of times they will confront the opposing party. There are two parties facing-off against each other and cannot communicate. Thus, in our model, we tried to evolve an optimal strategy for this type of case and defined win as where in each case, the maximisation of the amount of points accumulated overtime. Our hypothesis is that we would evolve a strategy similar to the TIT FOR TAT model as this strategy has been proven to be the most effective when played against other strategies. It has been shown to be effective under most situations to maximise the outcomes of each player in a prisoner's dilemma face-off. It is a strategy that is both able to reach the optimal solution where both parties' outcomes are maximised and is also able to defend itself against a very "mean" player that will always defect. Our reasoning is thus based on predictions made in literature and the results of each strategy in literature.

Description of Model

In our version of the prisoner's dilemma, there are 2 parties involved. These two players have no way of communicating with each other and will play a high number of rounds. In each round, the player can use one of 2 actions: to defect or to cooperate. In the case both defects, each player will obtain only 1 point each. If both players cooperate, they will both obtain 3 points and in the case where one cooperates and the other defects, the one who cooperates will obtain 0 points while the one who defects will obtain 5 points. These outcomes can be put in a payoff matrix. (Figure 1)

Red / Blue	Cooperate	Defect
Cooperate	R / R	T / S
Defect	S / T	P / P

Figure 1. Payoff matrix for two players blue and red. R represents the number of points they will obtain if both cooperate, T is the number of point the player will cooperate if he/she will defect alone. P is the number of point the player will obtain if both defect and S is the number of points the player will obtain if they defect alone. In this payoff matrix, T>R>P>S. R > P symbolises that both cooperating will yield a higher result than if both defects and T > R, P > S implies that for a rational agent in a set number of games, defecting will be a more dominant strategy.

This model is different from the traditional prisoner's dilemma as, in that case, the prisoners know the number of rounds they will play. In our case, the number of rounds played is only a representative number. This number is only symbolic because we want to simulate a situation similar to real life where a player cannot know how many rounds they will play. Therefore, our model tries to evolve an optimal solution where the points of the players are maximised in the iterated prisoner's dilemma. In other words, in a prisoner's dilemma where the players do not know how many rounds they will play.

We will evaluate the overall score overtime instead of simply evaluating which member in the confrontation has obtained the highest score. Winning simply against the opponent by having a higher number of points at the end does not mean necessarily that it is a good strategy. For example, in the traditional case of defecting all the time, it is true that one can win a lot if one encounters an opponent that cooperates all the time and who can be exploited. However, if one encounters an opponent that always also defects, the number of points by both cannot be maximised. Thus, a truly good strategy should not exploit its opponent, but should rather try to enforce cooperation by being nice and by punishing the un-cooperation of the opponent. Thus, if a strategy has a score that is much higher than his opponent, it means that he is exploiting his opponent, which is undesirable. Therefore, in our case, a player cannot win individually, and always win with their opponent, as a team of two. As it is explained, it is more advantageous to both players cooperate with each other and they will thus maximise the advantages of both. (See Figure 1) We define winning as the following: both players will have a score that, when summed up will be superior to 400 and both players will have a difference of score that will be inferior to 5. The first criteria to have a score superior to 400 will indicate a certain maximisation of the score of both players and that cooperation indeed occurred. The minimal difference between the scores of both players will indicate a situation where the strategy was able to defend itself against exploitative strategies and that no player effectively completely exploited by the other.

Bit	History	Move
0	First Move	C
1	Opponent C	C
2	Opponent D	D
3	Opponent CC	C
4	Opponent CD	D
5	Opponent DC	C
6	Opponent DD	D
7	C C C C C C	C
8	C C C C C D	D
9	C C C C D C	C
10	C C C C D D	D
11	C C C D C C	C
12	C C C D C D	D
13	C C C D D C	C
14	C C C D D D	D
15	C C D C C C	C
16	C C D C C D	D
17	C C D C D C	C
18	C C D C D D	D
19	C C D D C C	C
20	C C D D C D	D
21	C C D D D C	C
22	C C D D D D	D
23	C D C C C C	C
24	C D C C C D	D
25	C D C C D C	C
26	C D C C D D	D
27	C D C D C C	C
28	C D C D C D	D
29	C D C D D C	C
30	C D C D D D	D
31	C D D C C C	C
32	C D D C C D	D
33	C D D C D C	C
34	C D D C D D	D
35	C D D D C C	C
36	C D D D C D	D
37	C D D D D C	C
38	C D D D D D	D
39	D C C C C C	C
40	D C C C C D	D
41	D C C C D C	C
42	D C C C D D	D
43	D C C D C C	C
44	D C C D C D	D
45	D C C D D C	C
46	D C C D D D	D
47	D C D C C C	C
48	D C D C C D	D
49	D C D C D C	C
50	D C D C D D	D
51	D C D D C C	C
52	D C D D C D	D
53	D C D D D C	C
54	D C D D D D	D
55	D D C C C C	C
56	D D C C C D	D
57	D D C C D C	C
58	D D C C D D	D
59	D D C D C C	C
60	D D C D C D	D
61	D D C D D C	C
62	D D C D D D	D
63	D D D C C C	C
64	D D D C C D	D
65	D D D C D C	C
66	D D D C D D	D
67	D D D D C C	C
68	D D D D C D	D
69	D D D D D C	C
70	D D D D D D	D

Table 2: Prisoner Chromosome

Figure 2. 71 prisoner chromosomes of TIT FOR TAT

This definition of allows the theoretically best solution supported by literature, TIT FOR TAT, to win. In fact, the winning rate of TIT FOR TAT (fitness of TIT FOR TAT), using this strategy is 80%.

TIT FOR TAT is a strategy that won in the prisoner's dilemma experiment done by Robert Axelrod. The main purpose of this strategy is to maximize the scores (benefits) of both players, instead of maximizing its own score (benefits). Basically, what it does is to cooperate on the first turn and then, on subsequent turns, does whatever the opponent did for its move on the previous turn. In other words, if the opponent chooses to cooperate, TIT FOR TAT will continue cooperating until the opponent defects. Once the opponent defects, TIT FOR TAT will punish that defection with a defection of its own on the next turn, but it will forgive the opponent by cooperating again once the opponent starts to cooperate.

As the random strategies used to "train" the GA best strategy are generated in a way such that their genes have 50% chance of being cooperation and 50% chance of being defecting, the random strategies most often tend to cooperate with the GA best strategy to a certain extent. However, when encountering a strategy with 0% cooperation, which is the All DEFECT strategy, the GA best strategy becomes inefficient. It is just like in Robby the Robot where Robby was trained with rooms containing 50% of garbage, but becomes inefficient when put in a room with 100% garbage. To solve the problem of specificity, we have to put some All DEFECT amidst the set of random strategies against which the GA best strategy is trained. It is like to help the GA best strategy develop a resistance against the ALL DEFECT strategy. Then, a new part should be added to the already existing definition of winning. In the case the genes of the opponent are over 65% defect and if, at the same time, difference between the scores of the two people is inferior or equal to 5, it is also considered a win. Notice that in this part of definition, there is no restriction on the sum of both players' scores. It is lenient, because we understand that if the opponent refuses to cooperate, then it is impossible for the sum to be high.However, our goal is for the GA best strategy to be able to defend himself against this kind of player, so the restriction of a small difference between the two players' scores is still present. Thus to win,if either the first definition and the second definition apply, it is considered a win. With both parts of the definition of winning, now the trained GA best strategy should behave similarly to TIT FOR TAT in all simulations, including against the ALL DEFECT.

In our model, to evolve our solution, we initialize a population of 100 individuals each containing 71 chromosomes and evolve this population over 500 generations. The 71 chromosomes are In each generation, each member of the population will play 500 rounds against a random strategy and their capacity to win will be evaluated. There are 71 chromosomes. (Figure 2) The reason for this number is because bit 0 stores the first move in game 1 (1 chromosome). Bit 1 & 2 store the data of game 2 by looking at the opponent's first move. (2 more chromosome because 2 possible routes to take). Bit 3-6 store the data of game 3 by looking at the opponent's first two move. (2^2=4 chromosome or possible actions to take). Bit 7-70 store the data of game 4 and after by looking at the opponent's three past move and the player's three past move (2^6=64 chromosome). Thus, 1+2+4+64=71 chromosomes or possible histories. Then, a strategy is evolved from 500 generations by the use of the 2 points cross over method. Finally, we test this strategy against 1000 random strategies to obtain the fitness, in other words, the degree to which the strategy will win.

Our model also contains an evaluation of the TIT FOR TAT strategy fitness and also allows the TIT FOR TAT strategy to battle against different strategies to act as a control group for the strategy that we have evolved, as it is the best strategy mentioned in literature. The moves of our TIT FOR TAT control strategy come from the solution that Andrew Errity has evolved.

Description of Computational Method

In the boolean section at the beginning, there is a choice to keep elites or not. It is also possible to choose to add the strategies "TIT FOR TAT", "all cooperate", and "all defect" in the initial strategies (in the initial population).

When running the main method, two choices (excluding the exit choice) appear. It is possible to either start the genetic algorithm or to start with the TIT FOR TAT zone. It is recommended to start with the TIT FOR TAT zone, because TIT FOR TAT is the control strategy as it is the one of the theoretically known best strategies. Beginning by playing with the TIT FOR TAT strategy can give you an idea about what a good strategy looks like, and about whether the strategy developed by the GA is a good one or not.

After entering the TIT FOR TAT zone, four choices appear. It is possible to evaluate winning rate (fitness) of TIT FOR TAT strategy by making it play against 1000 random strategies, to simulate round of TIT FOR TAT against All COOPERATE, to simulate TIT FOR TAT against All DEFECT, or to simulate TIT FOR TAT against RANDOM STRATEGY. It is recommended to try all of them. Then, it is possible to exit the TIT FOR TAT zone by pressing -1.

Then, we go back to the genetic algorithm to evolve the best strategy. To evolve the best genetic strategy, we are using the two points cross-over strategy that is used in Robby the Robot.

Then, we re-evaluate the fitness of the best solution as many times as wished, with newly generated 1000 random strategies at each time.

Afterward, we continue with the simulation. Now, you still have the choice to simulate TIT FOR TAT against various opponents, but in addition, you will be able to simulate the games of the GA best solution against opponents.

In the simulation method, there are history arrays to store the information from the game. The 6 integers in the array "history" represent, in the following order, player's first move, opponent's first move, player's second move, opponent's second move, player's third move, opponent's third move. The third move is the move in the last game with respect to the game that is about to be played. For example, if game 45 is about to be played, the opponent's third move is his move in game 44, whereas the opponent's first move is his move in game 42.

In the situation method, the number of situations (of different histories) was supposed to be 2^6 = 64, because there are 6 integers in the history, which are all either 0 or 1 (either defect or cooperate). However, for the three first games of a round, there is no complete history of the three past games. Therefore, 7 new situations were added to account for this problem.
The first situation, the first game of the round has not been played yet, therefore there is no history. In the second and third situation, the first game was played, and the second game is about to be played. In situation 2, the opponent's last move was cooperate. In situation 3, the opponent's last move was defect. In situation, 4 to 7, the first and second game were played, third game is about to be played. Situation 4 refers to when opponent's last two moves were cooperate and cooperate. Situation 5 refers to when the opponent's last two moves were respectively to cooperate and to defect. Situation 6 is used to refer to the case where the opponent's last two moves were respectively to defect and to cooperate. Situation 7: opponent's last two moves were defect and defect. In the code, the integer -1 is used to represent no history (neither defect nor cooperate).

Results

Number of evaluations	Fitness of TIT FOR TAT	Fitness of GA strategy
1	81.50%	78.90%
2	80%	81.20%
3	82%	79.80%
4	78.90%	79.10%
5	81.70%	80.30%
6	79.90%	82.40%
7	80.50%	80.10%
8	80.30%	79.70%
9	78.10%	80.60%
10	77.40%	76.40%
11	80.90%	80.10%
12	80.60%	79.10%
Average	80.15%	79.81%

Table 1. Fitness of TIT for TAT and of GA Strategy evaluated 12 times

When we evaluate the effectiveness of the GA, the GA would also plateau at around 85% which is absolutely normal, since 80% is already considered as very high. Then when we re-evaluate the fitness of the GA best strategy against newly generated random solutions, it would lower itself to around 80%, which is again perfectly normal because the best strategy TIT FOR TAT has 80% of fitness. As it is indicated in Table 1, the fitness of our evolved strategy and of TIT for TAT have very similar averages, indicating their similarity.

Only to verify the similarity of TIT FOR TAT and our evolved GA strategy, we simulate our GA strategy against all cooperate, TIT FOR TAT, all defect and random strategies. When we simulate GA best strategy against TIT FOR TAT, they almost tie and both have a really high score. The GA strategy however seems to have a slight edge of 5 points over TIT FOR TAT, as in TIT FOR TAT, the first move is to cooperate while in the strategy evolved by the GA, the first move is to cooperate and for the GA strategy, the first move is to defect. The high score would however indicate that GA best strategy tends to be cooperative and evolves to become cooperative.

```
Scores for TIT_FOR_TAT vs ALL_COOPERATE:
TIT_FOR_TAT: 300 points
ALL_COOPERATE: 300 points
Sum of the two players's scores: 600/600 (Ideal case: 600/600. Would mean that both players cooperated at each of the 100 games)
```

```
Scores for BEST_STRATEGY vs ALL_COOPERATE:
BEST_STRATEGY: 302 points
ALL_COOPERATE: 297 points
Sum of the two players's scores: 599/600 (Ideal case: 600/600. Would mean that both players cooperated at each
of the 100 games)
```

Figure 3. Best Strategy vs All Cooperate and TIT FOR TAT vs All Cooperate

When the GA best strategy is simulated against All COOPERATIVE, there is a slight edge on the side of the GA strategy, as its first move is to defect instead of cooperating however, their score are very close and both should have a high score, as it is demonstrated in Figure 4. Again, this would indicate that GA best strategy tends to be cooperative. This score is extremely similar to when we simulate TIT FOR TAT against ALL COOPERATE, in which case, we obtain 300 points for both parties. This is also a logical outcome as TIT FOR TAT always mimics the cooperation of the other party.

```
Scores for BEST_STRATEGY vs ALL_DEFECT:
BEST_STRATEGY: 99 points
ALL_DEFECT: 104 points
Sum of the two players's scores: 203/600 (Ideal case: 600/600. Would mean that both players cooperated at each
of the 100 games)
```

```
Scores for TIT_FOR_TAT vs ALL_DEFECT:
TIT_FOR_TAT: 99 points
ALL_DEFECT: 104 points
Sum of the two players's scores: 203/600 (Ideal case: 600/600. Would mean that both players cooperated at each of the 100 games)
```

Figure 4. Best Strategy vs ALL Defect and TIT FOR TAT vs ALL Defect

However, when we simulate GA best strategy against All DEFECT, contrarily to TIT FOR TAT who is able to avoid being exploited by All DEFECT, GA best strategy may get heavily exploited by all defect. This is caused by the problem of specificity that has been explained in the description of our model. By using the solution proposed in the description of our model, we can see that our strategy can also very well defend itself against ALL DEFECT as its difference of scores is smaller or equal to 5. Thus, even though the result is not superior to the one obtained in ALL DEFECT, the GA strategy is not completely exploited. Furthermore, when we simulate TIT FOR TAT against all defect, we can see that the scores are exactly the same, thus making our strategy similar to the TIT FOR TAT strategy.

```
 Scores for BEST_STRATEGY vs RANDOM_STRATEGY:
BEST_STRATEGY: 281 points
RANDOM_STRATEGY: 281 points
Sum of the two players's scores: 562/600 (Ideal case: 600/600. Would mean that both players cooperated at each
of the 100 games)
```

Figure 5. Best Strategy vs Random Strategy (one instance)

When we simulate GA best strategy against RANDOM STRATEGIES, most of the time sum of the two players' scores should be larger than 400 and the difference between the two scores should be equal or less than five for the reason that most strategies evolved are 50% cooperation and 50% defection. This indicates that GA best strategy tends to enforce cooperation and does not exploit the opponent or let the opponent exploit him.

	TFT vs Random Strategies					GA vs Random Strategies			
Round #	Score of TFT	Score of Random Strategies	Sum	Difference	Round #	Score of TFT	Score of Random Strategies	Sum	Difference
1	183	188	371	5	1	174	179	353	5
2	108	113	221	5	2	212	212	424	0
3	239	244	483	5	3	295	295	590	0
4	247	252	499	5	4	274	274	548	0
5	249	254	503	5	5	230	230	460	0
6	248	253	501	5	6	197	197	394	0
7	232	232	464	0	7	280	280	560	0
8	261	261	522	0	8	298	298	596	0
9	199	199	398	0	9	235	235	470	0
10	299	299	598	0	10	251	251	502	0
11	204	204	408	0	11	265	265	530	0
12	264	269	533	5	12	251	251	502	0
13	160	165	325	5	13	275	275	550	0
14	198	203	401	5	14	244	249	493	5
15	241	241	482	0	15	200	200	400	0
Average	227.75	230.67	447.27	3.00	Average	245.40	246.07	491.47	0.67
Standard dev	46.03	43.92	91.35	2.45	Standard dev	36.08	35.44	71.51	1.70

Table 2. TIT FOR TAT vs Random Strategies and GA strategy vs Random Strategies

For our strategy against random strategy, even though it is difficult to control the outcomes for random strategies, we can observe that the average sum of TIT FOR TAT against random strategies (447.27) and the score of our GA evolved strategy against random strategy (491.47) are quite similar and well within the standard deviation range. Therefore, we can conclude that these two strategies generate almost the same outcome when confronted to random strategies. Such a large error is normal as the strategies generated are quite random and the difference of score that a strategy will yield if the population is predominantly cooperative or defecting is quite large.

Discussion

Our evolved strategy first has an average fitness of more than 80%, therefore, making our strategy a very viable strategy for most possible situations. Furthermore, when evolving our strategy, we have found a similar effectiveness as TIT FOR TAT and also have found very similar scores produced when they are confronted to random strategies, the ALL COOPERATE strategy and the ALL DEFECT strategy. Therefore, we can conclude that our strategy is quite similar to the TIT FOR TAT strategy.

People might think that ALL DEFECT is the best strategy in prisoner’s dilemma game since it never “losses”. However, many studies and experiments show that TIT FOR TAT is the best and the most successful strategy in direct competitions so far. In this section, we are trying to find out the reasons why TIT FOR TAT is better.

Firstly, we don't care about who wins between TIT FOR TAT and ALL DEFECT when those two play against each other, because our goal is to develop an idea strategy among humans, so necessarily we will play TIT FOR TAT against humans and play ALL DEFECT against human, and then determine based on those two rounds, which is the best between TIT FOR TAT and ALL DEFECT. We will also compare the results from TIT FOR TAT vs Random Strategy and from ALL DEFECT vs Random Strategy, in case people might wonder how the game ends with if the opponent is computer.

1. Compare TIT FOR TAT vs Humans with ALL DEFECT vs Humans

   A normal human usually, on his first turn, will choose to defect because the worst possible situation for him would be to only obtain 1 point. In the case where he obtains 1 point, it will still be better than to cooperate and obtain no points at all. ALL DEFECT will defect on the first turn as well. The normal human will then notice that by defecting he only gets 1 points at each game, so maybe he will try to cooperate, and see if his opponent would cooperate as well, so that he could get a higher score (3 points per game). If the human tries to cooperate, he will very soon realize that ALL DEFECT always defects no matter what, therefore, the human will come back to defecting. At the end of the game, ALL DEFECT will always get 5~15 more points than humans (maximum 104~112 points), depending on how quickly humans realize his opponent always defects. (I believe that a normal human would realize this fact in no more than 3 turns, because we are SMART!!!)

   TIT FOR TAT will cooperate on the first turn, and a normal human will defect on the first turn for the same reason as mentioned above. In this case, the human gets +5 and TIT FOR TAT gets +0, so the human thinks he gets an advantage and will continue defecting. But after defecting for one turn or two, the human will realize that TIT FOR TAT doesn’t cooperate anymore, and that he can thus only obtain 1 point at each turn. He will want to try to see if he can get TIT FOR TAT back to cooperation by himself changing to cooperation, so that at least he gets 3 pts at each turn. So the human will try to cooperate once, and TFT will defect because TFT follows what the human did on the previous turn. In the following turns, human will come back to defecting as he finds that TIT FOR TAT is always defecting, but now TIT FOR TAT starts to cooperate. Then the human will notice the pattern that TIT FOR TAT follows what he did on the previous turns. Once the human notices the pattern, they will truly begin cooperating, such that at the end of the game, they will both get a score that is much higher than 104~112, which is what all defect usually gets when it plays against humans.

   This explains how in real life, while playing against real humans, TIT FOR TAT is the best strategy whereas ALL DEFECT is horrible.

2. Compare TIT FOR TAT vs Random Strategy with ALL DEFECT vs Random Strategy
   (to your interest)

Simulation	ALL DEFECT (pts)	Random Strategy (pts)	SUM of their scores (pts)
1	108	98	206
2	304	49	353
3	240	65	305
4	396	26	422
5	400	25	425
6	360	35	395
7	236	66	302
8	100	100	200
9	296	51	347
10	112	97	209
11	304	49	353
12	296	51	347
13	104	99	203
14	296	51	347
15	232	67	299
Average	252.2666667	61.93333333	314.2

Simulation	TIT FOR TAT (pts)	Random Strategy (pts)	SUM of their scores (pts)
1	251	251	502
2	300	300	600
3	227	227	454
4	204	204	408
5	273	273	546
6	99	104	203
7	248	253	501
8	260	265	525
9	262	267	529
10	250	250	500
11	300	300	600
12	247	247	494
13	275	275	550
14	225	225	450
15	248	253	501
	244.6	246.2666667	490.8666667

Table 2. The results of simulations of TIT FOR TAT vs RANDOM STRATEGY and ALL DEFECT vs RANDOM STRATEGY

(Note: The maximum SUM of their score is 600/600, in the case when both players cooperate at each of the 100 turns)

As mentioned above, the purpose of TIT FOR TAT is not to maximize its score; instead, it is to maximize the scores of both players. This is also shown in Table 2. We have simulated TIT FOR TAT vs RANDOM STRATEGY and ALL DEFECT vs RANDOM STRATEGY for 15 times. From the table, we can see that ALL DEFECT always gets a higher score than RANDOM STRATEGY, while TIT FOR TAT sometimes gets 5 points less than RANDOM STRATEGY. However, the sum of the scores of ALL DEFECT and RANDOM STRATEGY seldom exceeds 400 points, while the sum of the scores of TIT FOR TAT and RANDOM STRATEGY always exceeds 400 points, and sometimes reaches the maximum sum (600 points). Since the goal of studying Prisoner's Dilemma is to better help ourselves in real-life situations, where both parties want to find a way to maximize their benefits, the higher average sum of the scores of TIT FOR TAT and RANDOM STRATEGY explains why TIT FOR TAT is the best strategy, as it corresponds to our goal of studying Prisoner's Dilemma.

Reference

Golbeck, Jennifer. "Evolving Strategies for the Prisoner's Dilemma", *Advances in Intelligent Systems, Fuzzy Systems, and Evolutionary Computation*, pages 299–306. February 2002, Interlaken, Switzerland.

Mitchell, Melaine. "Chapter 14: Prospects of Computer Modeling." *Complexity: A Guide Tour*. New York: Oxford University Press, 2009.209-222. Print.

Errity, Andrew. "Evolving Strategies for the Prisoner's Dilemma", www.academia.edu/1878726/Evolving_Strategies_for_Prisoners_dilemma.