Evolving Cooperation for the Iterated Prisoner's Dilemma

Jessica Finocchiaro
University of Colorado at Boulder
Department of Computer Science
Boulder, Colorado 80309
jessica.finocchiaro@colorado.edu

H. David Mathias
University of Wisconsin - La Crosse
Department of Computer Science
La Crosse, Wisconsin 54601
dmathias@uwlax.edu

ABSTRACT

The Iterated Prisoner's Dilemma (IPD) is an intriguing problem for which the Nash Equilibrium is not globally optimal. Typically treated as a single-objective problem, a player's goal is to maximize their own score. In some work, minimizing the opponent's score has been added as an additional objective. We explore the role of mutual cooperation in IPD player performance. We implement a genetic algorithm in which the population is divided into four multi-objective sub-populations: selfish, communal, cooperative, and selfless, the last three of which use a measure of mutual cooperation as an objective. Game play occurs among all members, without regard to sub-population, while crossover and selection occur only within a sub-population. Testing is against a population of well-known strategies and is single objective, using only self score. We find that players evolved to cooperate perform very well, in some cases dominating the competition. Thus, learning to play nicely with others is a successful strategy for maximizing personal reward.

CCS CONCEPTS

•Applied computing → Multi-criterion decision-making;

KEYWORDS

Iterated Prisoner's Dilemma Nash Equilibrium Genetic Algorithms Multi-Objective Optimization

ACM Reference format:
Jessica Finocchiaro and H. David Mathias. 2019. Evolving Cooperation for the Iterated Prisoner's Dilemma. In *Proceedings of Genetic and Evolutionary Computation Conference Companion, Prague, Czech Republic, July 13–17, 2019 (GECCO '19 Companion),* 3 pages.
DOI: 10.1145/3319619.3322021

1 INTRODUCTION

The Prisoner's Dilemma is a two-player game in which each player chooses whether to implicate (defect) or not implicate (cooperate) their criminal accomplice. A penalty (symmetrically, reward) is given depending on the choices of both players. Thus, the behavior of each affects the other. In the Iterated Prisoner's Dilemma (IPD), the participants play the game repeatedly, utilizing a strategy to optimize their aggregate scores. Though populations achieve the best aggregate result by acting communally, most research has focused on players that act only in their own self-interest.

Traditionally, IPD has been treated as a single objective problem, with players attempting only to maximize their own score. Mittal and Deb [2] implemented a multi-objective version in which each player attempts to maximize their own reward and minimize that of their opponent. They showed that players trained in this way were superior to those trained only to maximize their own reward.

We deviate from the typical game-theoretic assumption that success requires acting selfishly. We explore the hypothesis that cooperation is not only mutually beneficial but a better strategy for self-success. To this end, we create a population in which each player has one of four pairs of non-contradicting objectives. The population is initialized with an equal number of players with each objective pair and trained against each other or a benchmark population. We then evaluate all players in a tournament against the benchmark population. Independent of the objectives used during training, players are evaluated solely on self-score.

The central question is this: Does learning to cooperate benefit individuals more than learning to be selfish? We find that players with cooperative objectives outperform their selfish counterparts, even when evaluated by self-score. That is, players trained to play to the benefit of others win tournaments in which the only metric is personal reward.

2 OUR MODEL

We implement a multi-objective genetic algorithm inspired by that of Mittal and Deb [2]. An important difference in our algorithm is that each individual is assigned one of four distinct objective pairs.

`Selfish` players have the same objectives as those in the top-scoring algorithm implemented by Mittal and Deb [2]. `Communal` players attempt to maximize their personal score and maximize their opponent's score. The remaining two objective pairs explicitly reward *mutual* cooperation: games in which both players cooperate. This is achieved via an objective that calculates the fraction of games in which *both* players choose cooperation. These two pairs are called `Cooperative` and `Selfless`.

The genome for each player consists of a 70-bit string: a 64-bit *decision string*, followed by a 6-bit *history string*. Additionally, each player contains an integer in [0..3] indicating its objective pair, though this is not part of the genome. The history string stores the results of the last three games played and serves as an index into the decision string, determining the player's next decision.

In each generation, play consists of a round-robin tournament in which every pair of players play 150 consecutive games. After each game, players learn the outcome and update their history strings so that they store the outcomes of the three previous games. Thus, in any game, if the decimal representation of a player's history string

GECCO '19 Companion, Prague, Czech Republic
© 2019 ACM. 978-1-4503-6748-6/19/07…$15.00
DOI: 10.1145/3319619.3322021

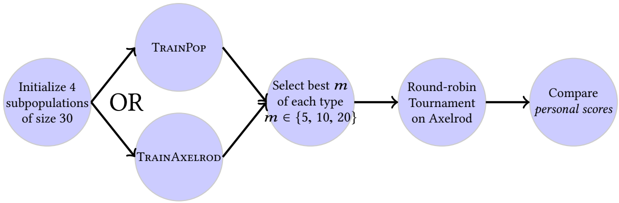
Initialize 4 subpopulations of size 30
OR
TrainPop
TrainAxelrod
Select best m of each type $m \in \{5, 10, 20\}$
Round-robin Tournament on Axelrod
Compare *personal scores*

Figure 1: Structure of our experiments

is i, its decision in that game is given by the i^{th} bit of its decision string, where 0 indicates cooperation and 1 indicates defection.

3 EXPERIMENTS AND RESULTS

Each match in the tournament consists of 150 prisoner's dilemma rounds. The population size is 120, divided evenly across the four sub-populations. We use single-point crossover with probability 0.9. Each child is then mutated by flipping bits with probability 1/70, so that the expected number of bits flipped per player is 1. Both the decision string and the history string are evolved. Selection is via NSGA-II [1]. The GA is run for 2500 generations.

After each training trial, the top 20 members of each objective pair are logged. These members are combined across 32 trials and sorted with objective pair as the primary key and self score as the secondary key. The top 160 members for each objective pair are used to create a candidate pool for use during testing.

Testing is via a round-robin tournament. m members of each of the 4 evolved types are chosen at random, from the candidate pool described above, for $m \in \{5, 10, 20\}$. The population also includes m members for each benchmark strategy. Thus, the population size is $21m$. Each member plays 150 rounds of prisoner's dilemma against $20m$ players, as they do not play against members of their own type. The sole criterion for evaluating the players is self score. Our experiments are depicted in Figure 1.

Our primary goal in this project is to investigate if players trained to value cooperation can be competitive when playing IPD against a population of standard players, even when cooperation is not used as a metric of success in the competition.

A test consists of 100 round-robin tournaments. Figure 2 shows representative results. In this stacked bar graph, each player type is depicted using a color/pattern pair. The magnitude of a player's color/pattern in the i^{th} bar indicates the proportion of the 100 tournaments in which that player type ranked in i^{th} place based on average score of all m players of that type.

In testing of players trained within the evolving population, we find that our Cooperative players (maximizing personal score and cooperation score) win every tournament. Other top finishers are our Communal players, Tit-for-Tat, and our Selfless players. The success of Selfless players is notable. Their optimization objectives during training are to maximize opponent score and cooperation, yet they perform well in a tournament in which the only metric is personal score. Selfish players occasionally break into the top 5 finishers, but are typically the lowest-scoring of our players.

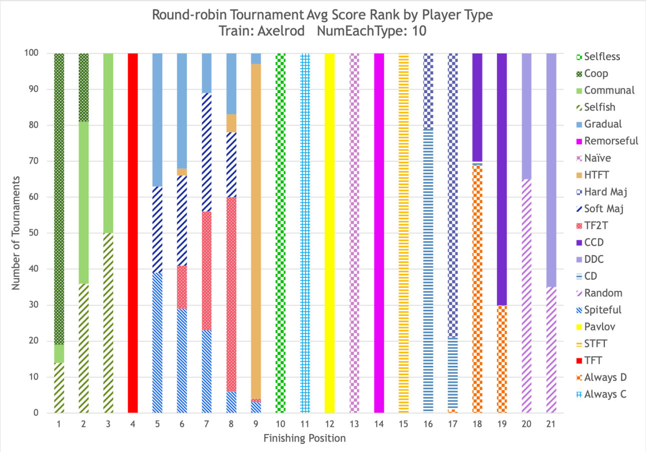

Figure 2: Results for 100 tournaments. Within a torunament, scores for players of each strategy are averaged and used to determine finishing position for that strategy.

For players trained against the Axelrod population and Gradual strategy (as in the graph), we find a less decisive ranking. As before, Cooperative and Communal players are consistently among the top finishers. A key difference from tests of population-trained players is that our Axelrod-trained Selfish player finishes much higher. In fact, these players win a majority of the tournaments though our cooperative players dominate based on average score.

4 CONCLUSION

We find that cooperation does indeed pay off: when players are trained to cooperate, they tend to outperform selfish players in a round-robin Iterated Prisoner's Dilemma tournament in which the measure of success is personal score. We compare the performance of 21 different strategies for IPD and observe the success of cooperative players in such settings. Even when trained with inclusion of an always defecting sub-population, we still observe the success of cooperative players, suggesting robustness of our players' strategies. Our results support the traditional strength of selfish players and introduce cooperative strategies that outperform these selfish strategies. Perhaps these results suggest the viability of cooperative behaviors in other domains, such as International Environmental Agreements and Nuclear Standoffs. In future work, we hope to extend our model to such games.

ACKNOWLEDGEMENTS

This work utilized the Extreme Science and Engineering Discovery Environment (XSEDE) [3], supported by National Science Foundation grant ACI-1548562. We used XSEDE resource Comet at the San Diego Supercomputer Center through allocation TG-CIE170029.

REFERENCES

[1] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation* 6, 2 (2002), 182–197.

[2] Shashi Mittal and Kalyanmoy Deb. 2009. Optimal strategies of the iterated prisoner's dilemma problem for multiple conflicting objectives. *IEEE Transactions on Evolutionary Computation* 13, 3 (2009), 554–565.

[3] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkins-Diehr. 2014. XSEDE: Accelerating Scientific Discovery. *Computing in Science and Engineering* 16, 5 (2014), 62–74.