Machine Learning in Iterated Prisoner's Dilemma using Evolutionary Algorithms

Shashi Mittal

Department of Computer Science and Engineering

Indian Institute of Technology, Kanpur.

Kanpur Genetic Algorithms Laboratory

OUTLINE:

1. Description of the game
2. Strategies
3. Machine Learning with Single Objective EA
4. Machine Learning with Multi-Objective EA
5. Results obtained
6. Comparison of the two methods
7. Analysis of results, Implications
8. Conclusions

THE GAME

1. Two players in a game
2. Each player can either "Cooperate" (C) or "Defect" (D).
3. The score received by the players depends on the move made by both of them

Kanpur Genetic Algorithms Laboratory

PAYOFF MATRIX

		Player 2	
	Decision	Cooperate	Defect
Player 1	Cooperate	R=3 R=3	S=0 T=5
	Defect	T=5 S=0	P=1 P=1

R: REWARD S: SUCKER T: TEMPTATION P: PENALTY

$$T > R > P > S, \ (T+S)/2 < R$$

Kanpur Genetic Algorithms Laboratory

WHY DILEMMA ?

If the opponent plays **Defect** then

1. If I play **Cooperate** - I will get the Sucker's Payoff - Worst Situation
2. If I play **Defect** - Low score for mutual defection, but better than cooperation

If the opponent plays **Cooperate** then

1. If I play **Cooperate** - Reward for mutual cooperation
2. If I play **Defect** - Will get Temptation award - best scenario

Hence the best option is to play **Defect**

WHY DILEMMA ? (continued)

By the same reasoning, the opponent also plays **Defect**, and hence both get the same score of punishment.

- *But both of them could have done better by cooperating !*

Hence the dilemma !!

Kanpur Genetic Algorithms Laboratory

ITERATED PRISONER'S DILEMMA

The situation is more interesting when the players play the game iteratively for a certain number of moves.

- The number of moves should *not* be known to the two players.
- The winner is the player with the highest score in the end.
- Usually, there are many players in the fray, and there is a Round Robin Tournament among all the players - the player with the highest score wins.

PROPERTIES OF IPD

- *Non zero Sum Game*
  - Both the players can simultaneously win or lose !!
- *No Universal Best Strategy*
  - The optimal strategy depends upon the opponents in the fray

IPD IN LIFE AND NATURE

- Symbiotism, Prey-Predator etc. behavior all can be captured by IPD.
- Business strategies
- Politics
- War scenarios
- Even zero sum games sometimes can get degenerated to Prisoner's dilemma !! (Dawkins, *The Selfish Gene*).

Kanpur Genetic Algorithms Laboratory

STRATEGIES

- Always Cooperate, Always Defect
- Random
- Tit For Tat, Tit for Two Tats
- Naive Prober, Remorseful Prober
- and many more ...

AXELROD'S STUDY

To study the best strategy for IPD, Axelrod in 1985 held a competition in which strategies were submitted by Game theorists.

- In the first competition, 15 entries were submitted. The winner was *Tit for Tat*, submitted by Anatol Rapoport.
- Axelrod held a second round of competition. Again the winner was *Tit for Tat* !

Many complex strategies performed badly !!

MEASURING PERFORMANCE OF PLAYERS

Suppose there are 15 players, each player plays against each other as well as itself, and each game is of 200 moves.

- Maximum score : 15000
- Minimum score : 0

However, neither extreme is achieved in practice

- **Benchmark score** : The score a player would have received against an opponent if both the players always cooperated
- Divide the score of a player by the number of players, then express it as a percentage of Benchmark Score.
- For Example : If a player scores 7500, then he has scored (500/600), i.e. 83% of the Benchmark Score in this case.

ENCODING STRATEGIES

Axelrod's method :

- The next move depends upon the behavior of both the parties during *previous three moves.*
- Four possibilities :
  - CC or R for Reward
  - CD or S for Sucker
  - DC or T for Temptation
  - DD or P for Penalty
- Code the particular behavioral sequence as a 3-letter string.
  - e.g RRR represents the sequence where both parties cooperated over the first three moves
  - SSP : The first player was played for sucker twice and defected.

Kanpur Genetic Algorithms Laboratory

ENCODING STRATEGIES (contd . . .)

Use the 3-letter sequence to generated a number between 0 and 63 by interpreting it as an integer base 4.

- e.g. if $CC = R = 0$, $DC = T = 1$, $CD = S = 2$ and $DD = P = 3$ then RRR will decode as 0 and SSP will decode as 43.

Strategy string : 64-bit binary string of C's and D's where he ith bit corresponds to the ith behavioral sequence.

PREMISES

Behavior in the first three moves of the game undefined in the above scheme.

- Add six bits to coding to specify a strategy's *premises*, i.e. assumption about the pregame behavior.

Together, each of the 70-bit strings thus represents a particular strategy.

How to find the optimal strategy from the all the possible strings ??

WHY USE GA ?

- Total number of all possible strategies very high - 2^{70}.
  - *Exhaustive search will take more time than lifetime of the Universe !!*
- Fitness function : Non continuous, classical methods will not work
- GAs emulate the natural process of evolution.

Kanpur Genetic Algorithms Laboratory

WHY USE MULTI OBJECTIVE GA ?

- *Non zero sum game* : My score and Opponent's score may not be directly related !!!
- Treat Self Score and Opponent Score as two different objectives : Maximize the first and minimize the second.

SEARCHING OPTIMAL STRATEGIES

Axelrod's approach :

- Maximize the self score using Single Objective GA

Our Approach :

- Maximize self score and minimize opponent score *separately* using Single Objective GA
- Optimize *both* the objectives *simultaneously* using NSGA-II and compare with the results found above.

USING SINGLE OBJECTIVE GA

Single Objective GA used to

- *Maximize* self score
- *Minimize* opponent score

Use the strategies obtained above in the Round Robin Tournament and observe their behavior.

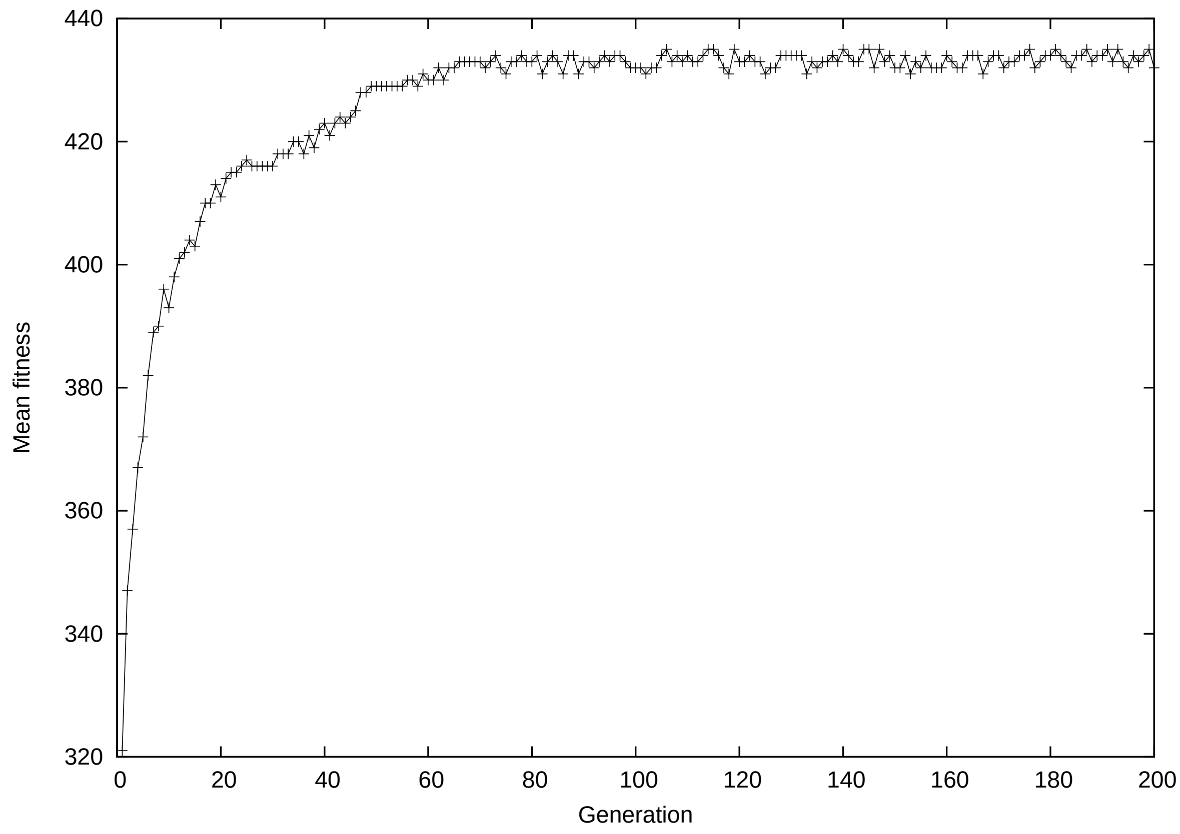

Figure 1: Plot of the mean fitness of population when self score is maximized.

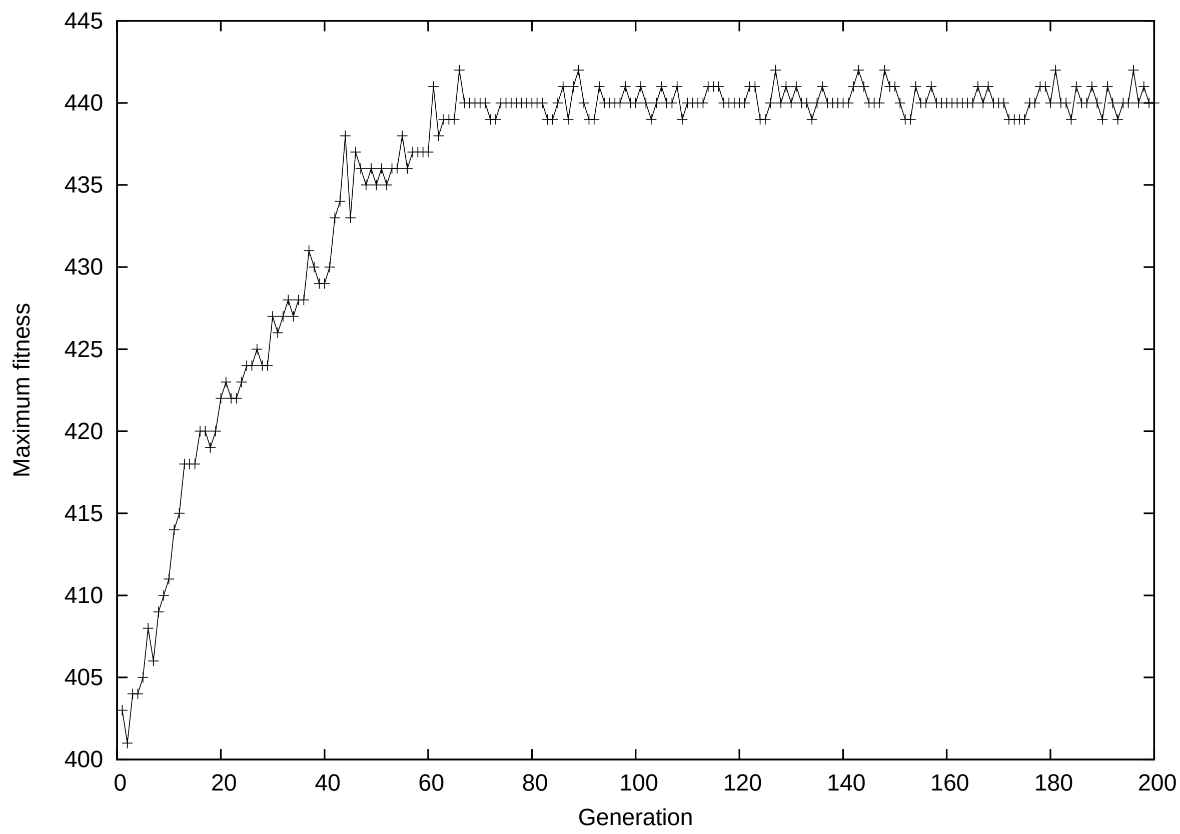

Figure 2: Plot of the maximum fitness of a sample in the population when self score is maximized.

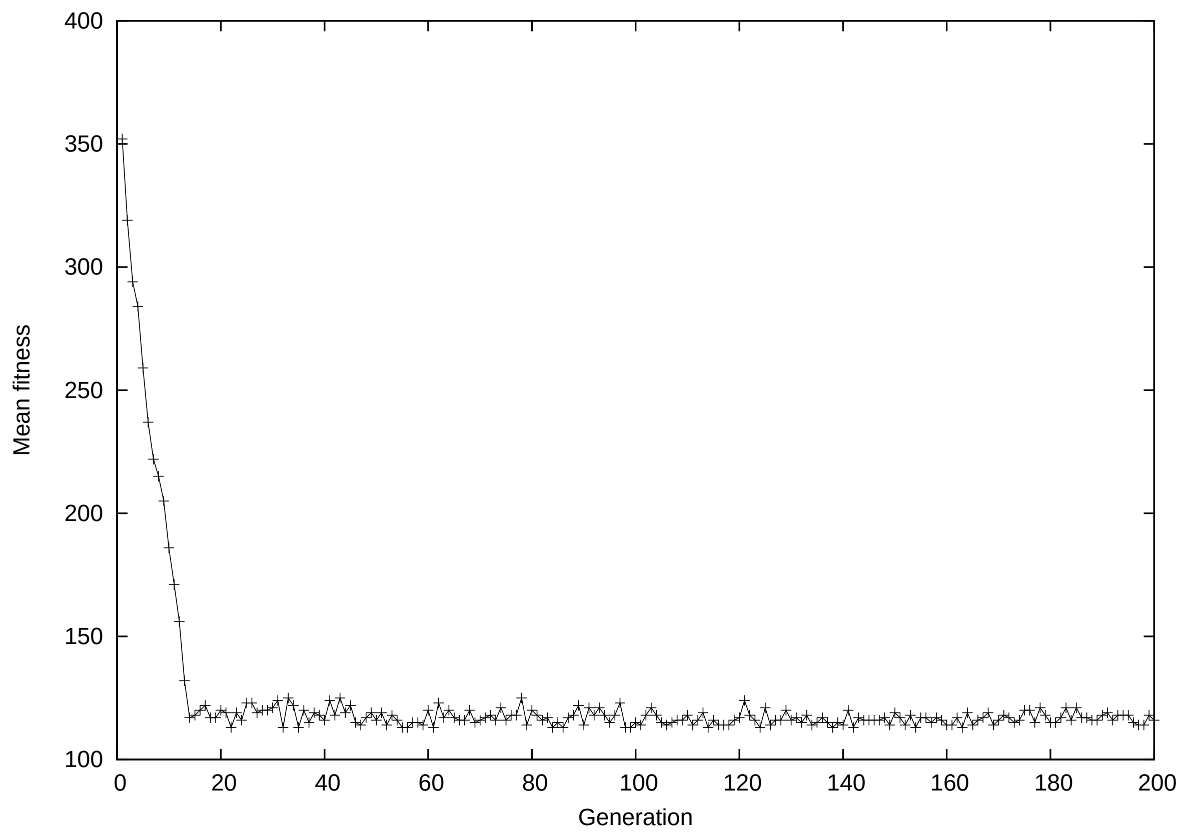

Figure 3: Plot of the mean fitness of population when opponent's score is minimized.

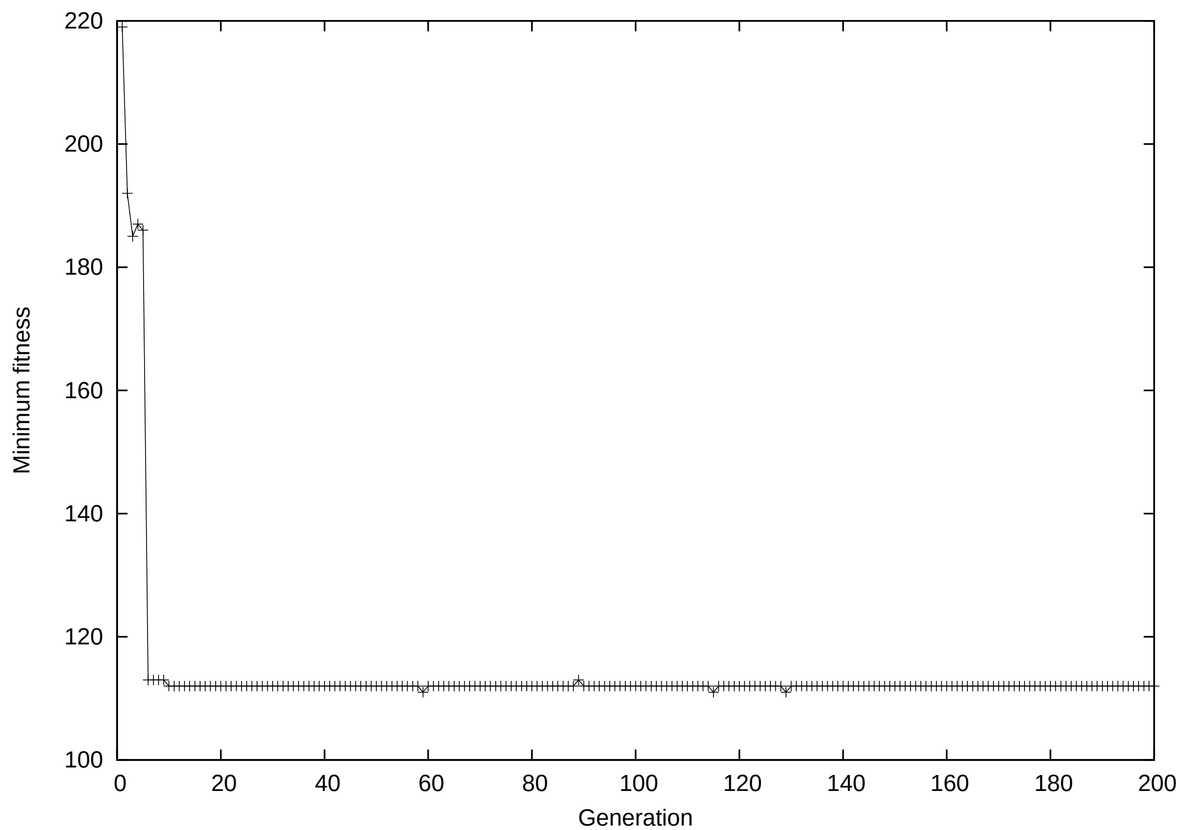

Figure 4: Plot of the minimum fitness of a sample in the population when self score is minimized.

PERFORMANCE OF SINGLE OBJECTIVE GA

When the strategies obtained from maximizing self score is used in the round robin tournament

- These strategies always emerge as the winner - defeats the *Tit For Tat* strategy
- Score in the tournament is nearly 98% of the benchmark score !

USING NSGA-II

Two objectives :

- *Maximize* self score
- *Minimize* opponent's score

NSGA-II algorithm with niching used.

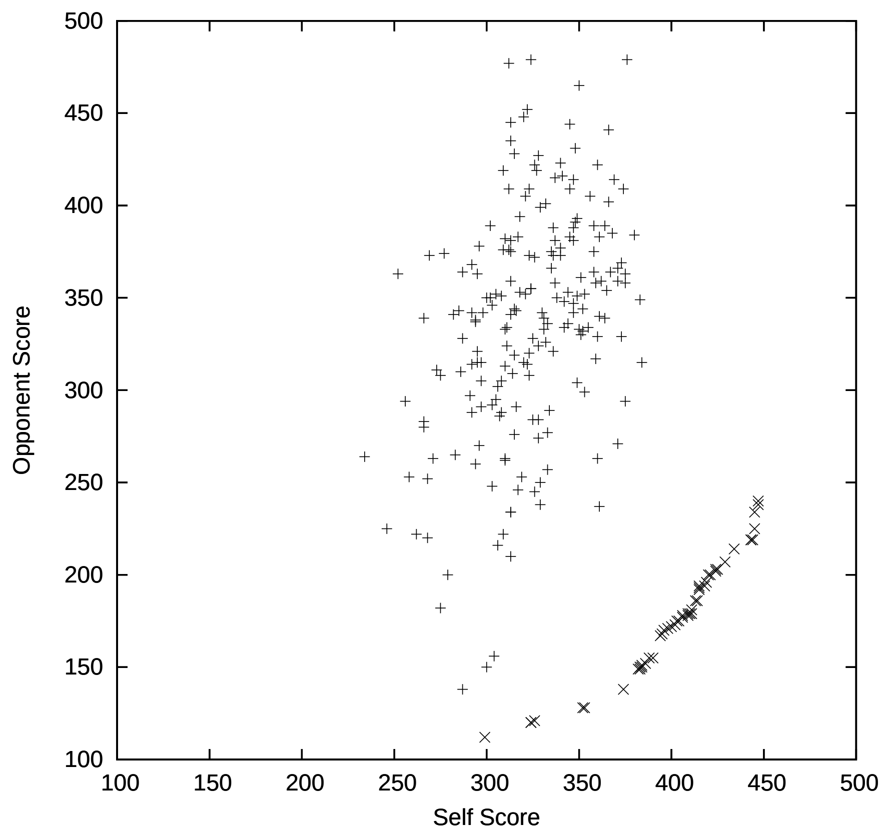

Figure 5: The initial random solution (shown with '+') and the Pareto optimal solutions(shown in 'x').

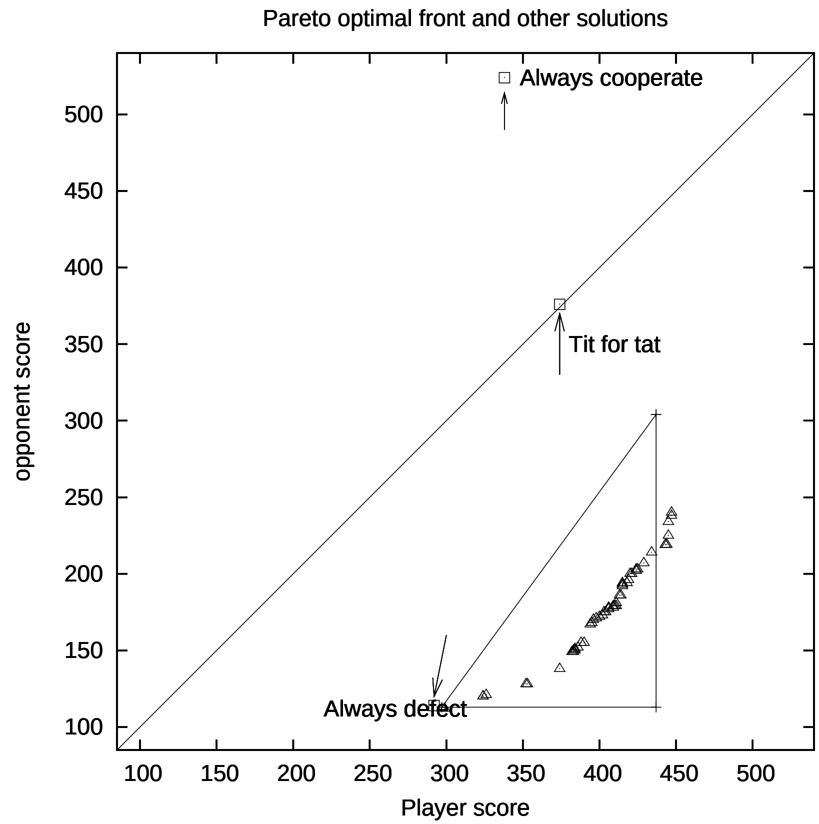

Figure 6: Pareto Optimal front, together with the single objective EA results (the upper and the left vertexes of the triangle) and a few other strategies.

KEY OBSERVATIONS

- Pareto - optimal front *is* obtained
  - *There is a trade off between self score and opponent's score*
- Convergence towards Pareto optimal front
  - *NSGA-II is able to search a complex front, starting from a random population*

Kanpur Genetic Algorithms Laboratory

PERFORMANCE OF NSGA-II

- When the strategy with the highest self score is fielded in the round-robin tournament, it emerges as the winner, winning by a significant margin.
  - Scores about 99% of the benchmark score
  - Also defeats the strategy obtained by Single Objective GA
- The later strategies, with lower self score, do not perform as well, and rank lower in the round robin tournament.
- The strategy with the lowest self score behaves like *Always Defect* strategy.

RELATIONSHIP AMONG STRATEGIES

As such, there does not appear to be any relationship among the different Pareto-optimal strategies.

- DDCDDDCCCDDCDDDDDDCDDDCCCDCDDCDDDDCDCDCCDCDC...
- DDCCDDDCDCCDDDCCDDCDCDDDDDDDDCDDDDCCCDCCDCDD...
- DDCDDDDCCCCCDDDDDDCDDDDDCDDCDCDDDDCCCDDDDDDC...

Kanpur Genetic Algorithms Laboratory

FINDING RELATIONSHIP AMONG STRATEGIES

- Play the strategy in the round robin tournament
- Keep track of how many times a particular bit position is used in the tournament
- Plot the frequency versus bit position,and then analyze

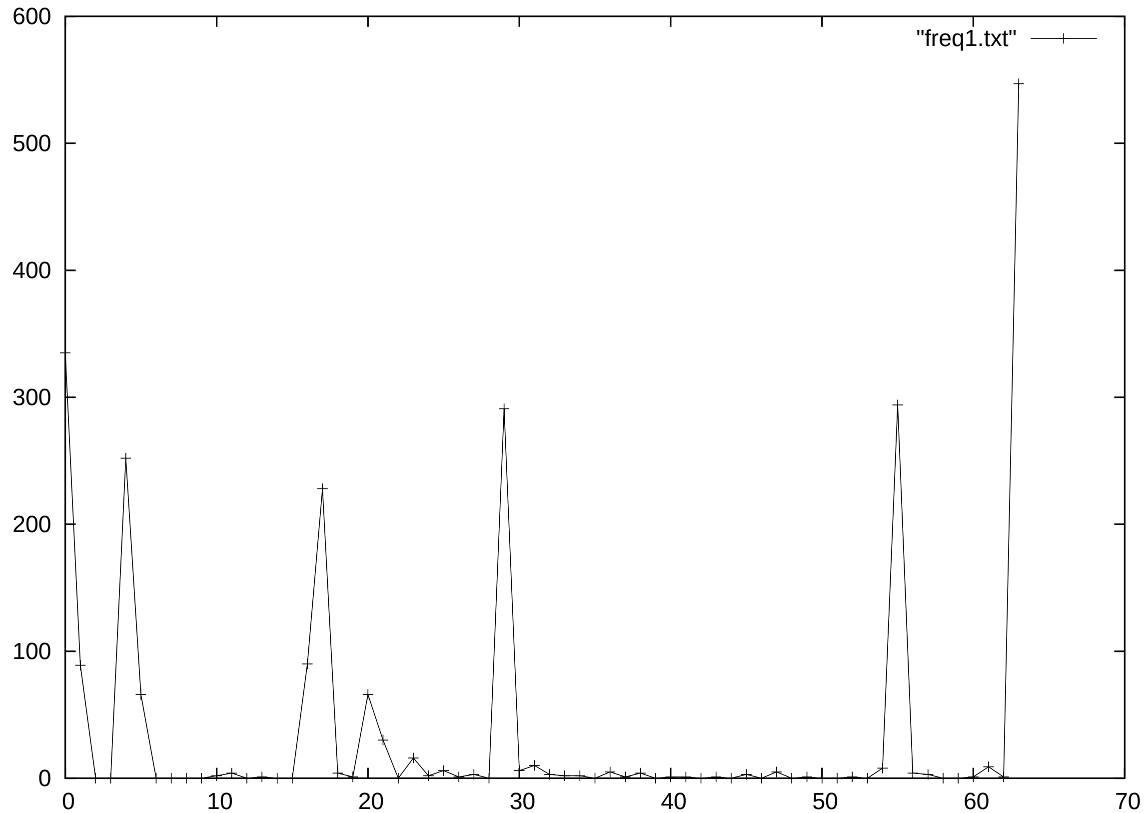

Figure 7: Plot for the move distribution with a particular strategy. Self score = 441, Opponent Score = 220

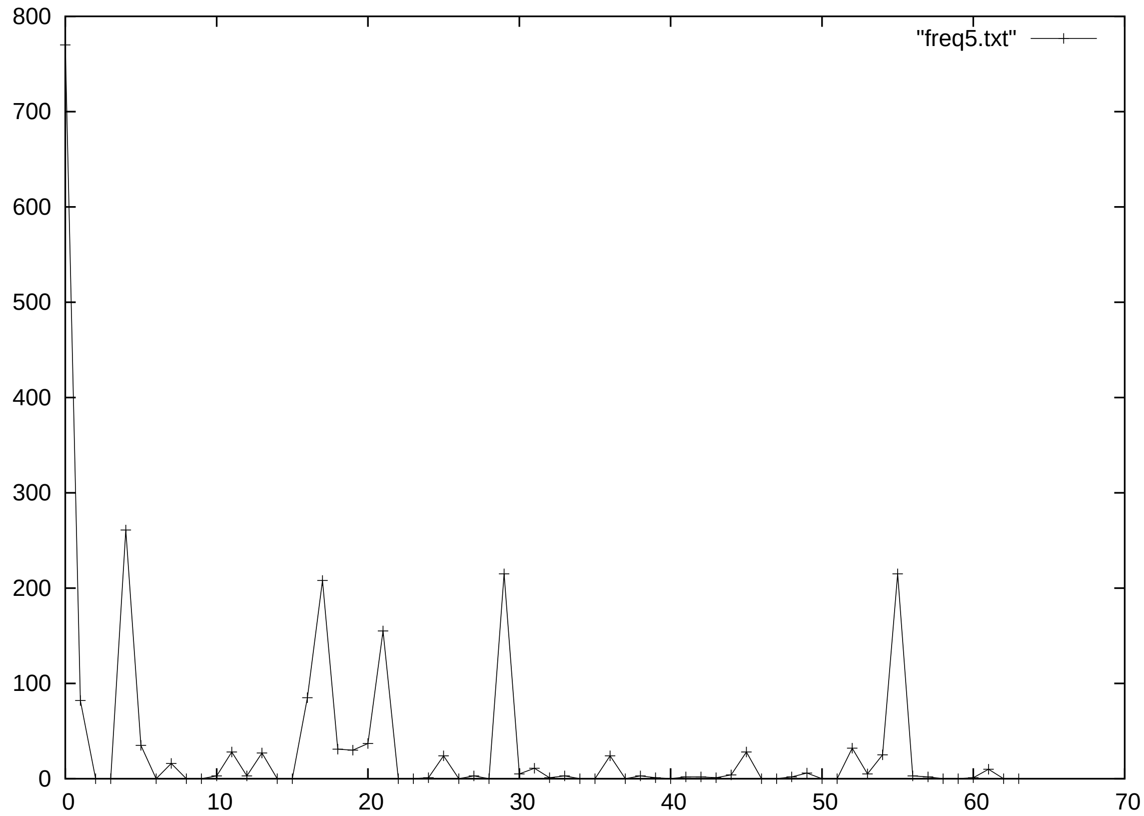

Figure 8: Plot for the move distribution with a particular strategy. Self score = 383, Opponent Score = 194

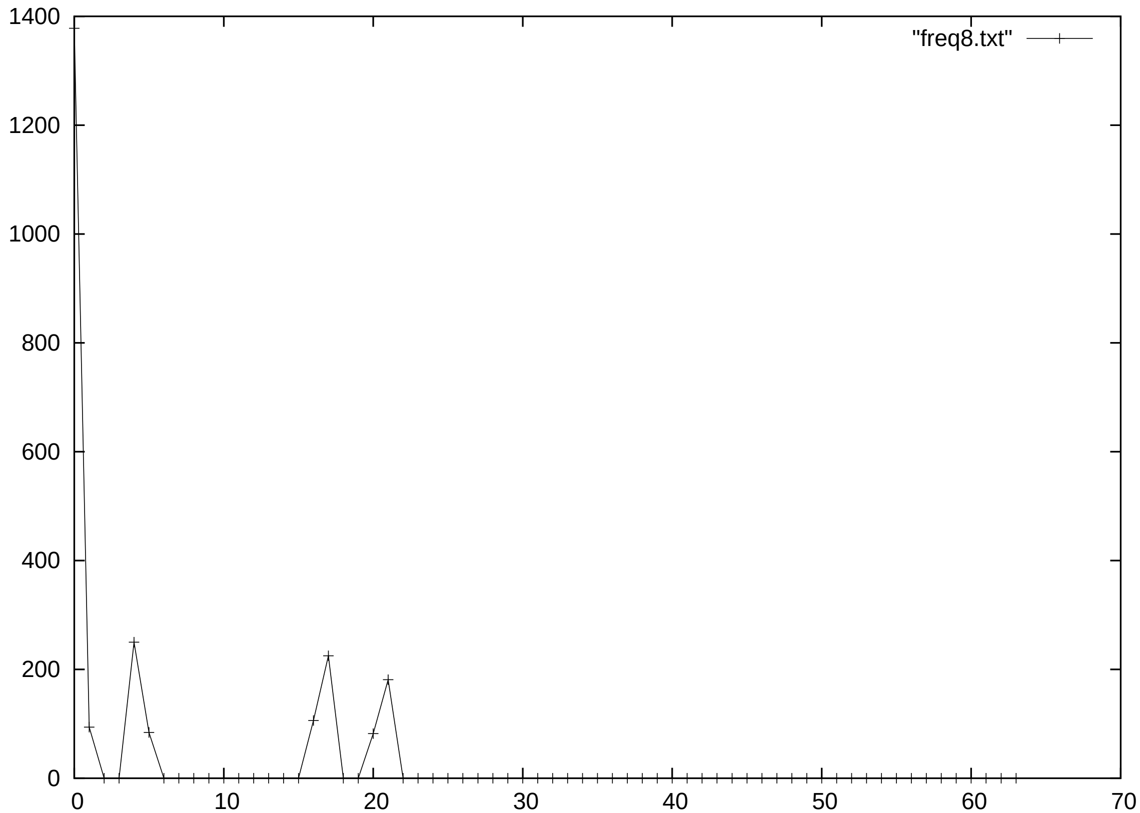

Figure 9: Plot for the move distribution with a particular strategy. Self score = 290, Opponent Score = 114

Kanpur Genetic Algorithms Laboratory

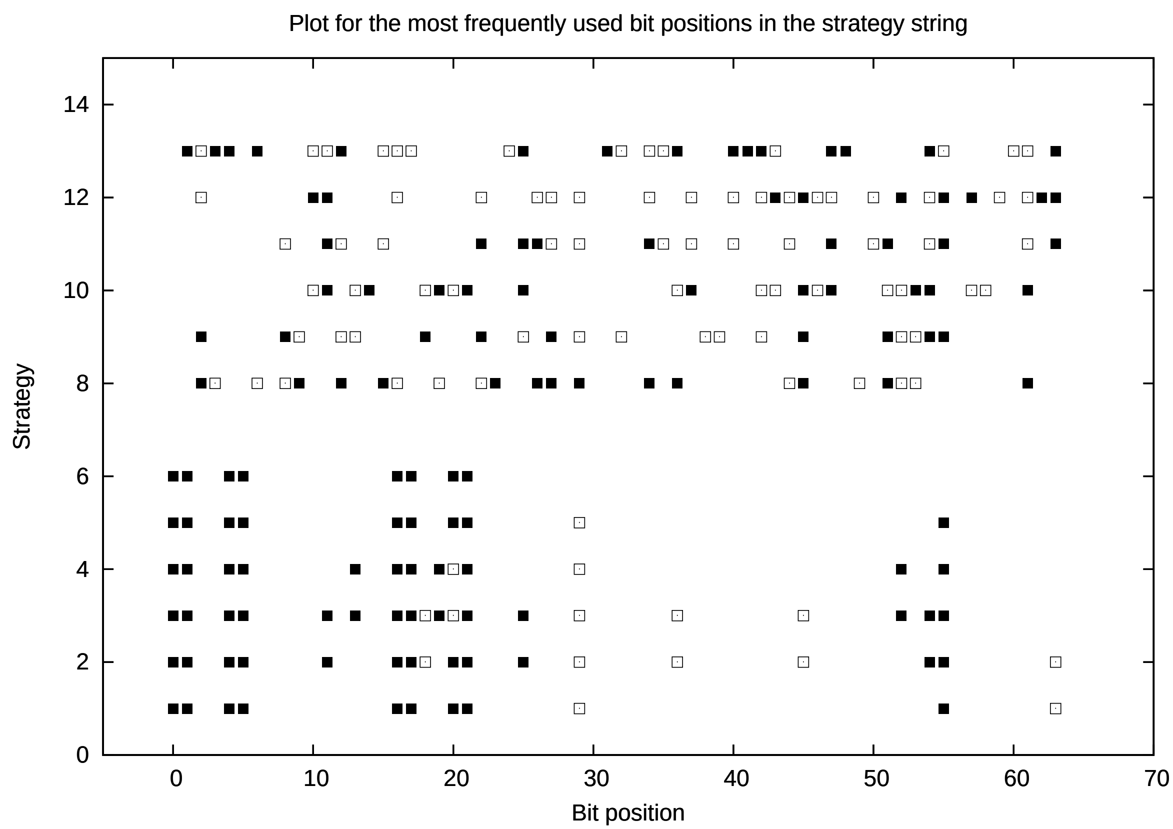

Figure 10: Plot of the frequently used bit positions in strategy strings. Lower one is for optimized strategies, upper is for random strategies.

KEY OBSERVATIONS

- Only *some* of the bit positions are used by the optimal strategies, and rest are not used.
- Random strategies *do not* show such a trend, the distribution is practically uniform.
- Thus, the Pareto-optimal strategies share some common properties.

CONCLUSIONS

- Self score and Opponent score are not directly related, both need to be optimized
- Multi-objective algorithm is better suited to find optimal strategies in this game
- Game theoretic result : Desired behavior of strategies can be studied by observing which bit positions of the strategy string are frequently used.

THANK YOU !

Questions are welcome.

Kanpur Genetic Algorithms Laboratory