Cricket Team Selection Using Evolutionary Multi-Objective Optimization

Faez Ahmed[1], Abhilash Jindal[1], Kalyanmoy Deb[1]
[1] Kanpur Genetic Algorithms Laboratory (KanGAL),
Indian Institute of Technology, Kanpur,
Kanpur, India
{faez,ajindal,deb}@iitk.ac.in
http://www.iitk.ac.in/kangal

Abstract. Selection of players for a high performance cricket team within a finite budget is a complex task which can be viewed as a constrained multi-objective optimization problem. In cricket team formation, batting strength and bowling strength of a team are the major factors affecting its performance and an optimum trade-off needs to be reached in formation of a good team. We propose a multi-objective approach using NSGA-II algorithm to optimize overall batting and bowling strength of a team and find team members in it. Using the information from trade-off front, a decision making approach is also proposed for final selection of team. Case study using a set of players auctioned in Indian Premier League, 4th edition has been taken and player's current T-20 statistical data is used as performance parameter. This technique can be used by franchise owners and league managers to form a good team within budget constraints given by the organizers. The methodology is generic and can be easily extended to other sports like soccer, baseball etc.

1 Introduction

Formation of a good team for any sports is vital to its success. Team selection in most sports is a subjective issue using commonly accepted notions to form a good team. In this work we have taken game of cricket as an example to demonstrate applicability of multi-objective optimization methodology to subjective issue of team formation from a set of players using available statistics. Cricket is a game played between two teams of 11 players where one team bats, trying to score as many runs as possible while the other team bowls and fields, trying to limit the runs scored by the batting team [1, 4]. Batting and bowling strength of a team are the major criteria affecting its success along with many other factors like fielding performance, captaincy, home advantage etcetera. We have explored the problem of building a 'good' team out of a set of players given the past performance statistics and suggested a new methodology from the perspective of multi-objective genetic optimization. Optimization studies have been done in many sports [6, 9, 11], and also has been done in various fields in cricket [10,12]. Just as in most league competitions a pool of players is provided as an input along with their performance statistics. Each player is paid a certain amount of money by the team owners for playing for their team, which we refer to as player's cost. League organizers impose an upper limit on budget for each franchise/club to avoid giving undue advantage to rich franchises. Player cost is either fixed by organizers as salary, decided through auction or determined by some form of contract agreement. We have considered Indian Premier League (IPL) as a test case for our analysis. IPL is a professional league for T-20 cricket competition in India. As of now, not much literature is available for any team selection methodology in cricket. In IPL, the franchise managers have the task of building a good team within budget cap. Individual players are bought by the franchises during a public auction of the players. Since the total number of players in the market pool is large, the challenge of finding the optimal teams becomes increasingly complicated and common sense methods, mostly employed, may fail to give a good team. Data used for this work (uploaded on [2]) has a pool of 129 players from IPL 4th edition. We have used performance statistics of each player in international T-20. The need of an effective optimization technique can be justified by rough calculation of the size of the decision space. From the given data of only 129 players from IPL-4 auction, containing 10 captains and 15 wicket-keepers, the total number of possible teams under the constraints of at least one wicketkeeper and one captain is as follows

$$Total\ Teams = ({}^{10}_{1}C)\ ({}^{15}_{1}C)({}^{127}_{9}C)$$

Considering the large number of different possible team combinations (order of 10^{15}), finding optimal teams under the added constraint of budget is not trivial. Currently most of the team selections are done using different heuristics or greedy algorithms. Usually, two or three high performance batsmen or bowlers are picked and the remaining team slots are filled according to budget constraints. But, this approach may not always give an optimal solution since matches are won by team effort. In such scenarios, overall quality of team may be poor. For example, a team with the best bowlers in the world may not win due to their inability to chase even a small target due to their poor batting performance. Hence, our aim is to investigate formation of an overall optimal team.

2 Strategy and Optimization Methodology

In cricket, player statistics has multiple parameters like number of matches played, total runs made, strike rate, number of wickets taken, number of overs bowled etc. It is important to identify those statistical parameters which reliably indicate player’s performance. The overall aim of a franchise is to build a team of 11 players with optimum bowling, batting as well as fielding performance within budget and rule constraints. Rule based constraints like presence of at least one player capable of wicket-keeping or maximum 4 overseas players in playing 11 also have to be taken in account. Considering the large amount of statistical data denoting various cricketing attributes that is available for each of the players, we first tend to reduce the dimension of data. One approach can be to use standard batting and bowling rating of cricketers obtained after exhaustive statistical analysis. Such ratings like the ICC world cricket rating [5], takes into account multiple factors of performance. But, such a rating system is currently available only for one day and test matches, so, we cannot apply it for T-20 format. For simplicity, we have used batting average and bowling average of a player in international T-20 cricket as a measure of their performance in batting and bowling.

Now, we redefine the team selection as bi-objective optimization problem as follows:

$$\max_{t=\{c,w,p_1 \dots p_9\}} \sum_{i=c,w,p_1 \dots, p_9} Batting\ Performance(i)$$

$$\max_{t=\{c,w,p_1 \dots p_9\}} \sum_{i=c,w,p_1 \dots, p_9} Bowling\ Performance(i)$$

The team is subject to the constraints

$$\sum_{i=c,w,p_1 \dots p_9} Cost(i) < Total\ Budget$$

where, t represents a team comprising of c, the captain of the team, w, the wicket keeper of the team, $p_1 \dots p_9$, the other 9 players of the team.

After problem formulation, we run multi-objective genetic optimization over the team using the elitist non-dominated sorting genetic algorithm (NSGAII) [8]. The players are sorted according to player cost and each player is assigned a unique integer number (tag). A team is represented as a chromosome of 11 real variables with each variable denoting the tag of the players present in the team. Fitness values of each population member (i.e. team) is evaluated as the total bowling strength and total batting strength as explained in Sec. 3. Also, the maximum total budget constraint is mentioned as a constraint for the total cost of the team. Additional constraints include that no two players of the team can be same i.e. duplicates not allowed. IPL specific rules are also taken as constraints i.e. a maximum of four foreign players can be taken into the squad.

3 Performance Measures

3.1 Batting Performance

A player's batting average is the total number of runs scored divided by the number of times he has been out [1]. Since the number of runs a player scores and how often he gets out are primarily measures of his own playing ability, and largely independent of his team mates. Thus, batting average is a good metric for an individual player's skill as a batsman. The objective function in our analysis has been taken as the summation of batting averages of all players. The problem with this approach is that some new players, even bowlers, may have batting average comparable to few of the best established batsmen due to good performance in few matches played. Hence, to avoid this scenario, the concept of primary responsibility of a player is used. A player is identified as a batsman only if he has scored at least 300 runs in international T-20 format. In calculation of team batting performance, the batting average of players identified as batsmen are only added to find net batting average. This condition is used in order to exclude batsmen who have not played enough games for their skill to be reliably assessed. So the overall batting average of team is maximized.

3.2 Bowling Performance

A bowler's bowling average is defined as the total number of runs conceded by the bowlers divided by the number of wickets taken by the bowler. So, the lower average is better. Again to avoid including misleading high or low averages resulting from a career spanning only a few matches, we qualify a player as bowler only if he has taken at least 20 wickets in T-20 format. Total bowling average of a team is taken as a measure of bowling performance and is minimized. Using such a strategy in optimization, results in exclusion of all bowlers from a team so that net bowling average of team goes to zero. Hence an artificial penalty bowling average for all non-bowlers needs to be added to the objective function. For our simulations, we have taken the bowling average of all non-bowlers as 100. This value is chosen to be worse than the bowler with highest bowling average.

3.3 Other performance measures

Final team selection from the trade-off front may require various other measures as explained in Sec. 5. The captaincy performance of a player is calculated as fraction of matches won out of the total number of matches played in the role of captain. It is also an important criterion in decision making process. Similarly, a player's fielding performance is measured by

$$Player's\ Fielding\ Performance = \frac{Total\ Catches\ Taken}{Total\ Number\ of\ Innings}$$

Team's net fielding performance is summation of all individual players fielding performance. Number of stumping by a wicketkeeper can be taken as his wicket keeping performance measure.

4 Bi-objective Optimization Results

Here we present the simulation results of the above mentioned algorithms applied on our player database. The budget constraint is taken as 6 million. At least one wicketkeeper, one captain and maximum four foreign players can be included in the squad. Fig. 1 shows the Pareto-optimal front obtained. Each point on the Pareto-optimal front represents a team of 11 players. Few solution teams corresponding to points marked on the Pareto-optimal front are shown in table. The right extreme of Pareto-optimal front shows teams with very good overall batting averages while left extreme shows bowling dominant teams with low net bowling average.

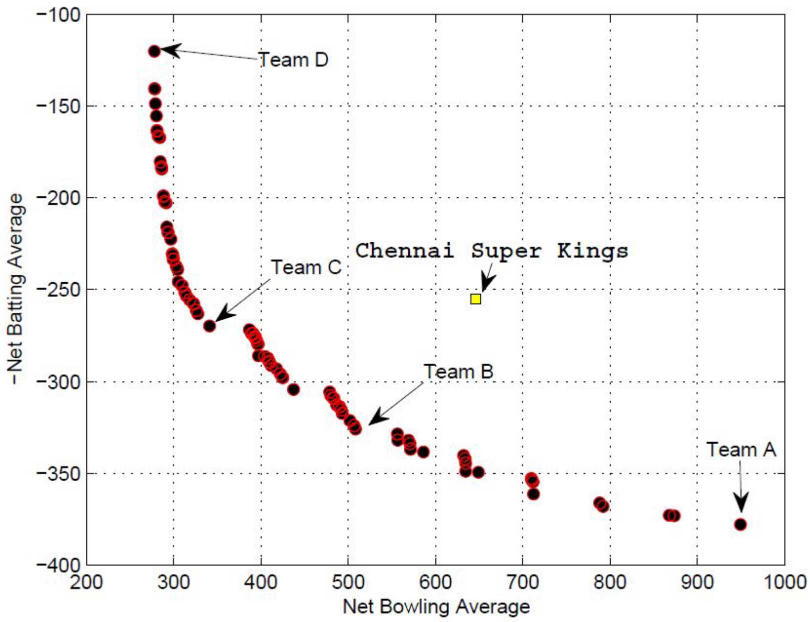

Figure 1 - Bi-objective trade-off front. CSK team is outperformed by Teams B and C on batting and bowling performances

To compare our results with common sense team selection methods we took the playing 11 cricketers of Chennai Super Kings (CSK), the franchise which won the finals in IPL-4. The bowling and batting fitness of the team was calculated using rules defined above along with total budget. The point representing CSK team is shown in Fig. 1. The total cost of CSK playing 11 members is estimated to be around 7.5 million. It can be seen that the team is non-optimal as well as costlier. Similar results were found for other franchise also.

4.1 Budget Sensitivity Analysis

To analyze the effect of budget constraint on team's performance we have done a sensitivity analysis where the optimization process is run for a range of budget constraints and Pareto-optimal front is plotted each time. It is seen from the Fig. 2 that budget constraint affects batting dominant teams more than bowling. This is because the price difference among batsmen with high batting average and those with low average is significant. The same effect is not observed among bowlers. It also ceases to significantly affect Pareto-optimal front above a limit. Such an analysis can guide the team owners when actual prices of players are variable and the optimization is done using maximum estimated price of each player.

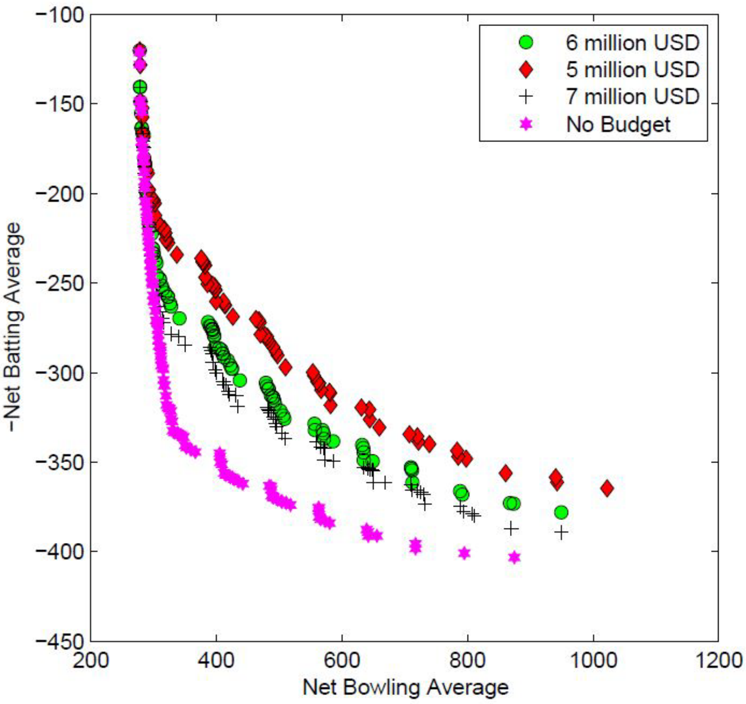

Figure 2 - Budget Sensitivity Analysis

5 Team Selections

The objective of the entire process is to obtain a single team of 11 members rather than a set of teams. Hence multi-criteria decision making methods need to be employed which helps in selecting any single team from the given set of mutually non-dominated teams on the trade-off front. After the initial optimization analysis, we get a Pareto-optimal front as shown in Fig. 1. Now the following method is proposed for the final team selection.

5.1 Knee Point Approach

Obtained trade-off front comprises of a set of points in the objective space representing various teams. For the given problem, we prefer selecting the team represented by a knee point present in the knee region of the Pareto-optimal front. Such a region is preferred because deviating from the knee region by a small change in the value of one of the objectives will compromise a large change in the value of the other objectives. Knee point can be identified by various methods [7]. Team C and

Team B shown in Fig. 1 are taken as the knee points and corresponding teams are shown in Table 1.

Table 1. Four teams chosen from the trade-off front (Fig. 1) obtained by NSGA-II.

Team A	Team B	Team C	Team D
Sachin Tendulkar	Yuvraj Singh	Yuvraj Singh	Yuvraj Singh
Michael Hussey	Nathan McCullum	JP Duminy	R Ashwin
Manoj Tiwary	Manoj Tiwary	Sudeep Tyagi	Sudeep Tyagi
Rahul Dravid	Ravindra Jadeja	Ravindra Jadeja	Nathan Rimmington
Suresh Raina	Suresh Raina	Suresh Raina	Paul Collingwood
Shaun Marsh	James Franklin	James Franklin	Steven Smith
Wriddhiman Saha	Wriddhiman Saha	Wriddhiman Saha	Wriddhiman Saha
Aaron Finch	Brad Hodge	Brad Hodge	Pragyan Ojha
Andrew McDonald	Andrew McDonald	Andrew McDonald	Shakib Al Hasan
Shikhar Dhawan	Shikhar Dhawan	Jaidev Unadkat	Jaidev Unadkat
Naman Ojha	Amit Mishra	Amit Mishra	Amit Mishra

The resultant team obtained shows a reasonable compromise between batting and bowling performance as compared to Team A and Team D. Since knee point is not clearly defined in the Pareto-optimal front obtained in this analysis hence we select points in the approximate knee region and apply further selection analysis to them. The knee point approach does not take into account many other factors which define a good team like fielding performance, wicketkeeper performance, expenditure, brand value of players, captain's success rate etcetera. To take into account such factors we take the solution set obtained from knee region analysis and calculate their fitness values on all such measures. New fitness value on all such factors is assigned to each team. The factors are sorted in order of importance. For example, fielding performance may be most important criteria among the other aspects mentioned above in our team selection strategy. So, we sort the solution set with respect to fielding performance and pick the solution having the best fielding side. If some other factor is also to be taken into account then we can apply in a lexicographic manner. A domination approach can also be followed. After picking a few teams from original non-dominated set with high fielding performance we shall then sort them according to other important factors, say, captaincy performance or brand value of players. Hence the team with good captain can be selected or a team with branded players can be chosen. If all preferences are exhausted and still more than one team is present in the final pool, we can get back to the expenditure criteria where the cheapest team will be preferred. Taking fielding, wicket-keeper and captain as the criteria for further analysis, the solution team obtained from knee region is mentioned below:
Suresh Raina, Wriddhiman Saha, Yuvraj Singh, Manoj Tiwary, Roelof van der Merwe, Amit Mishra, Brad Hodge, Shikhar Dhawan, Nathan McCullum, Andrew McDonald, Ravindra Jadeja

Using different criteria different teams can be obtained as per the requirement of attributes. The above is just an example of a good team obtained by our systematic analysis.

6 Conclusions

We have proposed a new methodology for objective evaluation of cricket team selection using a bi-objective genetic algorithm. An analysis of the obtained trade-off solution has been shown to result in a preferred team that has been found to have better batting and bowling averages than the winning team of the last IPL tournament. Such a methodology can be extended to include many other criteria and constraints and a better pragmatic team can be selected by the systematic procedure suggested in the paper. A standard methodology for team selection can be developed by integrating this approach with statistical analysis and using a dynamic optimization technique to be applied in auction. Abstract factors like team coordination etc. can also be used for decision making process. Nevertheless, the procedure suggested in the paper clearly demonstrates the advantage of using a bi-objective computing methodology for the cricket team selection in major league tournaments. The proposed methodology now needs some fine tuning based on other criteria for it to be used in practice.

References

1. Cricket, http://en.wikipedia.org/wiki/Cricket.
2. Cricket database, http://home.iitk.ac.in/~ajindal/cricket.
3. Espn cricinfo, http://www.espncricinfo.com/.
4. An explanation of cricket, http://www.cs.purdue.edu/homes/hosking/cricket/explanation.htm.
5. International cricket council, http://icc-cricket.yahoo.net.
6. S. Butenko, J. Gil-Lafuente, and P.M. Pardalos, Economics, management and optimization in sports, Springer Verlag, 2004.
7. K. Deb and S. Gupta, Understanding knee points in bicriteria problems and their implications as preferred solution principles, Engineering Optimization 99999 (2011), no. 1, 1–30.
8. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: Nsga-ii, Evolutionary Computation, IEEE Transactions on 6 (2002), no. 2, 182–197.
9. A. Duarte, C. Ribeiro, S. Urrutia, and E. Haeusler, Referee assignment in sports leagues, Practice and Theory of Automated Timetabling VI (2007), 158–173.
10. I. Preston and J. Thomas, Batting strategy in limited overs cricket, Journal of the Royal Statistical Society: Series D (The Statistician) 49 (2000), no. 1, 95–106.
11. J.C. R´egin, Minimization of the number of breaks in sports scheduling problems using constraint programming, Constraint programming and large scale discrete optimization: DIMACS Workshop Constraint Programming and Large Scale Discrete Optimization, September 14-17, 1998, DIMACS Center, vol. 57, Amer Mathematical Society, 2001, p. 115.
12. T.B. Swartz, P.S. Gill, D. Beaudoin, and B.M. desilva, Optimal batting orders in one-day cricket, Computers & operations research 33 (2006), no. 7, 1939–1950.