Building a Winning Fantasy NBA Team

Project Report

as part of IE537: Discrete Optimization Algorithms

Vamsi Vipparla, Malfrine Das

May 3, 2018

Table of Contents

1. Introduction

Since its inception in the 1980s, daily fantasy sports have developed into a multi-billion dollar industry. The major companies, DraftKings and FanDuel, are valued at $0.75 and $1 billion respectively. Fantasy sports leagues can often be very competitive as they offer as much as $100,000 for league winners. In fact, several businesses like, Rotogrinders and DailyFantasyNerd, have been created to provide fantasy sports enthusiasts with sports analytics and lineup optimality assessments. It is estimated that fantasy sport contestants each spend approximately $50 in order to improved their drafting methodology.

The National Basketball Association (NBA) was established in 1946 and has since risen as one of the most profitable sports entertainment leagues in the world. Fans of the NBA seeking a more immersive entertainment experience have turned to fantasy sports. In the NBA fantasy leagues, the user takes on the role of the general manager and creates a fantasy roster consisting ten current NBA players. The roster must meet constraints regarding player positions and salaries.

The objective of this study is to build a fantasy NBA team that accumulates the maximum number of fantasy points in a season. The players are chosen based on their past performance, salary, and position. Performance is based on fantasy points accumulation and it is a function of the number of points, assists, blocks, rebounds, steals and turnovers that a player makes in a given season.

The project objective will be addressed in this paper in the following manner:

1. Background:
    a. An explanation of multiple fantasy NBA contest types is provided in Section 2
2. Model Development:
    a. The problem approach and assumptions are found in Section 3.1 and 3.2, respectively
    b. The construction of the objective function and constraint equations can be found in Section 3.3 and 3.4
3. Model Solution:
    a. The methodology behind the model solution is provided in Section 4

4. Results:
    a. The results of the performance prediction model and roster selection is in Section 5.1 and 5.2 respectively
5. Discussion:
    a. Further recommendation based on the results of the project is provided in Section 6

2. Background

There are three major variations of fantasy sports:

1. Daily fantasy rosters: On a daily basis, players pick 10 players to their roster. Over the course of the season, the total performance based on the players they have chosen is calculated and the contestant with the largest sum is considered the winner. The player salary in these tournaments is determined by the fantasy sports league and is subject to change on a daily basis.
2. Annual rosters with player replacement: At the beginning of the season, contestants select any 10 current players based on salary and positional restriction. In some contests, contestants "dress" only five players per day and only those players' performance contribute to the roster's overall performance for that day. In others, the performance of all players is accounted for every day.
3. Annual rosters without player replacement: This contest is similar to the Contest 2, above. The only exception being that the roster selection is held in a manner similar to real sports drafts. Contestants pick players in a given order. Once a player is "drafted" to a contestant's team, they can no longer be picked by the other contestants in the league. Since, the number of star players in the league are quite limited, these contests are often much smaller and are usually held by a group of friends and moderated by a third-party website.

For the purpose of this project, the contest that will be simulated as Contest 2 described above. It will be assumed that the performance of all 10 players in a roster will contribute to its overall performance.

3. Model Development

3.1. Problem Approach

In mathematical terms, the given problem can be viewed as the knapsack problem. Given a set of resources with a given weight and value, determine the resources that will be selected in order to meet the weight limitations of a knapsack while maximizing value. For the purpose of the fantasy basketball roster problem, the resources are the players and the knapsack is the roster. The player value is the predicted performance and their associated weight is their salary and position. To solve the problem, the problem will be modeled as an integer linear program and solved using a Gurobi solver in Python.

Annual player statistic data from 2010-2016 will be used as the training data for the polynomial regression model that will be employed to predict player performance in the 2016-2017 season. The 2016-2017 data will be used as test data to determine the accuracy of the performance prediction model.

3.2. Assumptions

In order to simplify the model development, assumptions regarding player performance, salary, position, injuries, and selection were made as follows:

1. Player Performance: In reality, player performance is hard to predict and affected by many factors. For the purpose of this model, players perform as specified by a cubic polynomial regression. The focus of this project is centered on the solution of the optimization algorithm. Therefore, an in-depth analysis on player performance did not occur.
2. Player Salaries: Player salaries and the roster salary cap in fantasy leagues are not the same as the actual salaries that players receive in the NBA.
3. Player Positions: It is common in the NBA for players, especially bench players, to play multiple positions. For the purpose of this project, only the position the player is assigned to on basketball-reference.com will be used.
4. Player Injuries: It is difficult to determine if a player is injury prone. However, it is assumed that the performance prediction model will predict players who are more injury-prone to

have less points because they will have played fewer games and scored fewer points than their healthier counterparts.

5. Player Selection: Rookie and sophomore players will not be selected for the roster because it is assumed that there is not enough data points to predict their performance accurately. Additionally, players who have retired the previous season will not selected.

3.3. Objective Function Development

In NBA fantasy sports, player performance is a direct function common player statistics. Table 1 displays each statistic and the weightage that most fantasy leagues associate with it.

Table 1: Weightage of player statistics in fantasy leagues

Statistic	Fantasy League Weightage
3-point field goal	3 points
2-point field goal	2 points
Free throw	1 point
Rebound	1 point
Assist	1 point
Block	2 point
Steal	2 point
Turnover	-1point

For this project, performance will be a weighted linear function of the above function and can be seen in the equation below.

$$Perf = 3 \times three\ pointer + 2 \times field\ goal + free\ throw + \ rebound + assists + block + steal + turnover$$

3.4. Constraint Equation Development

The constraint on the model optimization take the following factors into account.

1. Position Constraint:
   The 5 positions employed by a basketball team is the shooting guard, center, small forward, power forward, and point guard. In most fantasy leagues, each team must have two players that play each position.
2. Salary Constraint:
   Fantasy leagues provide as player salary cap at the start of the season. Using that salary cap, user select players for their roster such that the sum of all the players' salaries on their roster are less than the salary cap.

4. Model Solution

There are many ways of solving the problem. It can be solved using recursion, integer linear programming (ILP) and greedy methods.

Initially, a greedy approach was chosen to solve the problem. Since there are multiple constraints, the players are segregated into different sets based on their playing positions. The salary constraint is divided so that each positional set has an equal amount of salary that can be allotted to two players. The knapsack algorithm was run for each positionally separated set of players and the top two players were selected for the optimal roster. This approach significantly reduced the computational complexity of the algorithm. However, this approach does not provide an optimal roster because salaries are spread out evenly between each position. In order to produce an optimal roster the problem was formulated as an ILP. The ILP has a significantly higher run time. It was determined that the increase in run time was negligible because the list of players eligible to be selected for the roster was small. A detailed overview of the ILP formulation can be found in Appendix B. The ILP was solved using a Gurobi solver in Python.

5. Results and Discussion

Player statistics from the 2010 - 2017 was collected from basketball-reference.com on March 20, 2018 and can be found in Appendix A. Python notebooks used for data cleaning, calculations and model development can be found in Appendix A.

5.1. Performance Prediction

The results of the polynomial regression can found in "fitter.ipynb" of Appendix A. The average discrepancy between predicted and actual player performance was calculated to be 15%. This discrepancy is quite significant and will hinder the optimality of the predicted roster. The roster optimization was carried out in order to determine how significant this discrepancy would affect player selection and the overall performance between the true optimal and theoretically optimal roster.

5.2. Roster Selection

Table 2 lists the players, performance and total salary for the theoretically optimal and truly optimal rosters. Table 3 shows the difference between actual and predicted performance for the theoretically optimal and true optimal rosters. The theoretically optimal roster is the roster that was predicted to be optimal in 2017 based on their performance in 2010-2016. The true optimal roster is the roster that was actually performance based on their performance in 2017.

Table 2: List of players, overall perforamnce and total salary for theoretically optimal and truly optimal rosters

	Theoretical Optimal Roster	True Optimal Roster
Predicted Overall Performance	22066	-
Actual Overall Performance	19661	35135

Salary	$97,699,494.00	$95,832,163.00
Player 0	David Lee	Bojan Bogdanovic
Player 1	David West	Buddy Hield
Player 2	DeMarcus Cousins	DeMarcus Cousins
Player 3	Karl-Anthony Towns	Ersan Ilyasova
Player 4	LeBron James	Giannis Antetokounmpo
Player 5	Marcus Thornton	Lou Williams
Player 6	Michael Carter-Williams	Mason Plumlee
Player 7	Monta Ellis	Russell Westbrook
Player 8	Rudy Gay	Serge Ibaka
Player 9	Stephen Curry	Stephen Curry

Table 3: Difference between actual and predicted performance for the theoretically optimal and true optimal rosters

	Discrepancy Compared with Predicted Optimal (2010 – 2016)
Actual Performance of True Optimal (2016-2017)	+60%
Actual Performance of Theoretical Optimal (2016-2017)	-11%

6. Discussion

Based on the results of the project, it was found that an average performance prediction discrepancy of 15% resulted in a 60% difference between the actual performance of the theoretical optimal and true optimal roster. In a real life contest setting, it is unlikely that this roster would place well. The player performance prediction was the factor that limited the optimality of the roser. In order to improve the roster selection the following recommendations should be considered for future work:

1. Improve the accuracy of player performance prediction. This can be achieved with any of the following recommendations:
    a. Collect game-by-game data of player performance. Though this data exists, it is often expensive and does not guarantee that a polynomial regression will be more accurate. However, with the advent of the data collection API recently released by the NBA statistics committee. This data is much more cheaper and accessible than before.
    b. Explore implementing more nuanced and complex machine learning algorithms. With the advent of Big Data, machine learning algorithms have become useful for tasks like predicting player performance.
    c. It may be effective to use the well-established predictive technologies of third-party companies that exist in the NBA fantasy sports market. Though this is expensive, it may still be a financially viable idea based on the payouts from the fantasy sport contests.
    d. There has been extensive work done by NBA analysts in the department of analytics in order to determine the underlying trends in player performance. Perhaps integrating analytics metrics into the player performance prediction model would improve its accuracy
2. Player performance will always have a given amount of uncertainty. Players may be injured, adjusting to the systems of new teams, or simply be having an off year. Similar to

trading in the stock market, it might be effective to develop a portfolio of rosters that account for fluctuations in player performance.

3. Once a satisfactory strategy has been developed based on the above recommendations, it would be useful to implement similar strategies for the other NBA fantasy contests listed in the background section of this report.

1.

7. Conclusions

In this report, a fantasy NBA roster was created by predicting player performance using a polynomial regression model and selecting players for the roster using an integer linear program. Due the inaccuracy of the performance prediction model, the roster that was selected was 60% less optimal than the true optimal roster. Based on the results of this study, the following conclusions were drawn:

1. Improve the player performance prediction model by integrating trending analytics, integrating machine learning, integrating third-party company predictive technologies or improving data quality.
2. Develop a portfolio of rosters that account for the stochasticity in player performance.
3. Modify the model to develop rosters for other NBA fantasy sport contests.

8. References

1. https://en.wikipedia.org/wiki/National_Basketball_Association
2. https://en.wikipedia.org/wiki/Fantasy_sport
3. Kellerer, H., U. Pferschy and D. Pisinger (2005) Knapsack Problems , Springer, Berlin, Germany.
4. 2010-17 Hollinger NBA Player Statistics. (n.d.). Retrieved March 23, 2018, from http://insider.espn.com/nba/hollinger/statistics
5. . 2016-17 NBA Player Contracts. (n.d.). Retrieved March 23, 2017, from http://www.basketballreference.com/contracts/players.html#player-contracts::none

9. Appendices

Appendix A

Units

Salary	Fantasy League Dollars ($)
Player Performance	Fantasy League Performance

Sets

PLY = {ply | ply is a player available in the 2016-2017 roster}

1	Steph Curry,
2	Lebron James,
…	…
N	James Harden

POS = {**pos** | pos is one of the five basketball positions}

SF	Small Forward,
PF	Power Forward,
PG	Point Guard,
SG	Shooting Guard,
C	Center,

Parameter

Sal_{ply}	Salary of player, *ply* ∈ PLY
$Pos_{ply,pos}$	1 if player *ply* ∈ PLY plays position *pos* ∈ POS 0 otherwise
$Perf_{ply}^{pred}$	Predicted performances of player, *ply* ∈ PLY

Decision Variable

x_{ply} 1 if player is selected for fantasy roster, *ply* ∈ PLY
0 otherwise

Objective Function

$$\max \sum_{PLY} x_{ply} \times Perf_{ply}^{pred}$$

Constraint Equations

Position Constraint

(C1) $$2 = \sum_{PLY} x_{ply} \times Pos_{ply,pos} \qquad \forall\, pos \in POS$$

Salary Constraint

(C2) $$S_{max} \geq \sum_{PLY} x_{ply} \times Sal_{ply}$$

Appendix B

https://github.com/malfrine/NBAFantasyKnapsack