# NBA Lottery Prediction

黃允暘 杜方辰 林湛悅 林晁璿 鄭幃謙 顏訓哲
computer science, National TsingHua University
Hsinchuy, Taiwan

## I.    ABSTRACT

The result of a basketball game is influenced by many factors. Most significantly, the performance of players of the team. But is the best player always win? The answer is no and the "Chemistry" between players in a team is also very important.

## II.    INTRODUCTION

Since we are all basketball lover and are fascinated by NBA. Final result is the important part that we desire to know. Thus, we decided to write a model that can predict the win team in an incoming NBA match.

We have tried to analyze the player's performance to get the final result through linear regression, SVN, random forest, XGBoost etc.

But the results from those methods are always awful. Therefore, rather than considering a player's performance, we try to compute the special relation between two players and find out how it can influence a game. The relation between two players is the "Chemistry" we have mentioned above.

We will explain our method's implementation in detail in 3rd paragraph. In 4th paragraph there will be comparison between the result from out methods and

the result from other methods. Analyzing our methods will be in 5th paragraph.

## III.    METHOD

The experiment data is come from the website "basketball reference", which contain the personal information of every NBA players(Table I , Table II).

Table I. The personal information of LeBron James



Table II. Player's data attribute.



After collecting the data we need, we pre-process them and get the information we want including: Date of the game, 2 teams name(Team A ,Team B), Winner of the game, Player for each match(A1~A20, B1~B20). This is just a game data. And Then, we try to build the training data by the following rules:

- One game is one data
- According to the player of the game, we select the 20 players of two teams
- Furthermore, we select previous 3 games data of a player, respectively so, the input shape will be X.shape = (num_of_game, 20*3*attribute_num) attribute_num is the number of attribute of a player's data including : age, GS, MP, FG. And the ground truth is result of every game.

We fit the training data to the model in the following method:

1. Method1: We first used some of the models that introduced in this course, like svm(support vector machine), Gradient Booting, (extreme)XGBoost and RandomForest. The result is not as good as we expected, so we try another method.

2. Method2: We found another related paper, NBA game Predictions based on Player Chemistry, which researched about the relationship between the result of basketball games and player chemistry. The main idea is that each team has some intrinsic symmetry amongst its players (some players play better with other players on the same team) and an anti-symmetric component that represents some adversity between the two players. The symmetry matrix will be represented by S and the adversity matrix with A.



Fig 1. Interpretation of the elements of the S (left) and A (right) matrices.

We only need to preprocess the data to denote which player were in the same game. And then we use $h(x) = g(x^T dx)$ to predict the result, where x is [t1, t2] (team1's player and team2's player) and d is compose of S and A matrix.

## IV.    RESULT

1. Method I.
   - SVC: 0.594
   - Gradient boosting: 0.596
   - Random forest: 0.603
   - XGBoost: 0.573

2. Method II.

We got a better performance with 64% accuracy. Because the input data is very simple(only two teams' players), we consider this method is better than our method.

## V.    DISCUSSION

We have tried two different data selecting way:
1.    Using the previous game data of the player to predict the result, which means each training data include each player's previous three games data
2.    Using the combination of each teammate and opponents to predict the result, which means each training data just include every player's name.

In this project we do these two methods respectively, and the second method has the higher performance. We are wondering that if we combine these methods together, we can get higher performance. Taking the advantage of the two methods, we can think over player's condition of the previous games and also bring the chemical reaction between each player into consideration

## VI.    CONCLUSION

In this paper, we show many different ways to predict the NBA game result, we find out that random forest and chemical reaction between each player get the better performance in method 1 and method 2, respectively. The future outlook is the combination of the two methods in order to get optimal prediction.

## VII.    AUTHOR CONTRIBUTION STATEMENTS

杜方辰(17%): Method I implementation, data preprocess, report.
黃允暘(17%): Method I implementation, data preprocess, report.
林晃璿 (17%):Method II implementation, data preprocess, report.
林湛悅 (17%): data collection, data analysis, report.
顏訓哲(17%): data collection, data analysis, report.
鄭幃謙(17%): data collection, data analysis, report.

### REFERENCES

[1]    NBA Game Predictions based on Player Chemistry Prastuti Singh and Bai Yang Wang.

[2]    https://www.basketball-reference.com.