Capstone Project One

Classifying NBA players positions and performance

Proposal

by Yiding Weng
June17, 2018

Problem

Every off-season the management of each NBA team have to find a way to improve their roster, usually through recruiting new players on the free market, or signing new contracts with their current players. But how to decide which new player to sign or which player on the old roster to keep? Will that player be able to provide the desirable contribution for the team? As each team try to optimize the diversity and the integrity of the skill sets of the entire roster, there are problems they have to consider.

NBA teams are increasingly trotting out lineups with five players who can play and guard nearly any position. traditional positions don't accurately explain what a players skillset truly is, they incorrectly oversimplify the skill sets of NBA players. Simply plugging players into one of five positions does not accurately define a player's specific skill set. Moreover, the misclassification of a player's position may lead teams to waste resources on developing draft picks that do not fit their systems.

In light of these changes, we need an effective way to designate positions in the NBA not based on basic physical traits such as height and weight, but in terms of function, such as shooting and defense. A framework for modern NBA positions is important towards our understanding for how players have evolved, and effective roster construction.

Data

Proposed Data Collection:

aggregate NBA player's Statistics for 67 NBA seasons.

Data Location: https://www.kaggle.com/drgilermo/nba-players-stats/data

Historic NBA team records

Data Location: https://www.kaggle.com/druswick/nba-team-records-historic

Approach

The approach to the problem will be considered to have two parts:

1. Find an optimal way to classify players functions based on historical statistics
2. Test the level of completion of each team roster based on the total integrity of skill sets and functions of each individual players.

The first data collection includes the performance statistics of each individual player in each season. Using clustering method to find out the optimal ways of defining categories that specifies function or characteristics of players in the same category, compare that with the traditional way of defining player roles as PG, SG, SF, PF and C.

For the second part understand whether each team has a balanced number of players in each new defining roles, and test if the completion level is correlated to team regular season performance using the historic team records.