

CS91r Final Submission

Matt Goldberg

December 15, 2016

1 Introduction

Much of the work being done in the world of NBA analytics attempts to answer the problem of how an individual player should be valued. In particular, a common insight is that basketball is inherently a team sport, and the players with whom and against whom one plays can significantly alter one's traditional box score statistics. Therefore, a major problem in player evaluation is teasing apart an individual player's contribution to his team without falsely attributing contributions to poor players who merely play with great players (or against even worse players), and without overlooking great players on mediocre teams. The major breakthrough in player evaluation that addressed these issues was Adjusted Plus/Minus[3], which controls for a player's teammates and opponents to determine what effect a player has on point differential, independent of who else is on the court. Since this framework was introduced, many others have expanded on it and adopted the mindset of breaking down plus/minus [1][5].

However, NBA general managers and coaches do not make decisions about a player based on his value in a vacuum; rather, they make decisions based on a player's value in the context of the team they already have. Notably, chemistry is often critical to success in the NBA, so there is a strong incentive to ensure that players fit well together[?]. Therefore, my goal is to build a principled model that, at its core, predicts the point differential between a given two lineups, based on the talent and play style of the players involved. Using this model, a

coach would be able to determine how player's specific skillset would make an impact given their current roster. Moreover, a coach could determine which players on a roster would complement each other well and should play a significant portion of their minutes together.

2 Background

- Explain Adjusted Plus/Minus and RAPM
- Clustering is very common - "new positions"
- Explain and compare/contrast [2]
- Explain idea of regression with RAPM and play styles

3 Clustering Approach

3.1 Feature Selection

* Explain that we want style as much as possible, without taking into account minutes, actual contribution, etc. * Go over different areas of play style and stats to represent each

3.2 Clustering Algorithms

* Go over different clustering algorithms briefly * Use of silhouette score for evaluation [4]

KMeans

GMM

PCA + KMeans

PCA + GMM

4 Results

* See notebook * Visualizations (?)

5 Future Work

* Regression

- Given such a system, possible insights and analysis: Section 5

What is each team's best lineup? Especially relevant when starters are resting

Are certain types of players more valuable, or perhaps more replaceable, than others?

Which coaches are the best at playing players that complement each other?

Given a roster and a possible acquisition, how valuable is the player in question?

How well do they fit?

Are there any notable mutually-beneficial trades that could happen?

How have these effects changed over time?

References

- [1] Jeremias Englemann. Youtube, Nov 2015.
- [2] Allan Maymin, Philip Maymin, and Eugene Shen. Nba chemistry: Positive and negative synergies in basketball. *Sloan Sports Analytics Conference*, Mar 2012.
- [3] Dan T. Rosenbaum. Picking the difference makers for the all-nba teams, Apr 2004.
- [4] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:5365, 1987.

- [5] Joseph Sill. Improved nba adjusted +/- using regularization and out-of-sample testing.
Sloan Sports Analytics Conference, Mar 2010.