

NBA Hackathon Paper

Team Name: 4Data

Team Members: Yizhe Ge, Tianye Song, Kevin Sun

Introduction

The prompt that our team tackled for the NBA Hackathon was prompt 1: Develop a new method or tool for evaluation of defensive performance in the NBA. In the NBA, every player has a position associated with their role on the court (point guards, shooting guards, small forwards, power forwards, and centers). Traditionally, point guards protect the ball and pass the ball to players in an optimal position to score; shooting guards shoot the ball at a high percentage and frequency; small forwards are athletic and tend to be versatile; power forwards play with their back to the basket; and centers grab rebounds. However, players are not labeled based on their defensive roles.

This research aims to use multivariate clustering analysis to label players based on their defensive performances. The research then tries to discern what type of defensive players are more prominent in winning teams.

Methods Used

The method we used was multivariate clustering analysis, specifically the EM algorithm for Gaussian mixture models. The variables we considered for each players were **BLK_Rate**, **STEAL_Rate**, **MP**, **DRB**, **Percent_Contesting**, and **DREB_CHANCE_DEFER** for the 2015-2016 NBA season. The table below summarizes the variables:

BLK_RATE	"Block_Rate": Average blocks per game/ Average Minutes per game
STEAL_Rate	"Steal_Rate": Average steals per game/ Average Minutes per game
MP	"Minutes Per Game"
DRB	"DRB": Average Defense Rebounds per game
Percent_Contesting	"Percent of Contested Shots": Number of times the players contested within 2 feet of the defender / Number of times the player was the closest person to the defender
DREB_CHANCE_DEFER	"Defensive Rebound Chance Deferred": Number of times the player deferred the rebound to a teammate.

Findings:

The table below summarizes the results the research found. Four clusters were identified. Cluster 1 was identified as “premier rebounders” (such as Tim Duncan). Cluster 2 was identified as “athletic defenders” (such as Russell Westbrook). Cluster 3 was identified as “weak defenders” or “old athletes” (Andre Miller). Cluster 4 was identified as “Shot contesters” such (such as JaVale McGee).

```
# A tibble: 4 x 7
  CLUSTERS BLK_RATE STEAL_RATE      MP      DRB PERCENT_CONTESTING DREB_CHANCE_DEFER
    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>          <dbl>          <dbl>
1       1  0.7581163 -0.22290098  0.8253188  1.54293338    0.5469509    1.41070805
2       2 -0.5710643  0.29199603  0.8759408 -0.05308492   -0.5195146   -0.07724017
3       3 -0.6063848  0.09655061 -0.7699747 -0.85418713   -0.4636611   -0.84860681
4       4  0.9116171 -0.33599443 -0.9454875 -0.16803613    0.8402493   -0.02759127
```

We then looked at which type of clusters were more frequent on winning teams. We identified winning teams as the top 5 teams that limited scoring of opponent points to the lowest. A summary of what we found are below:

	Cluster 1	2	3	4
Total # of players	69	111	102	80
Avg % on winning teams	26.3%	25.8%	33.1%	14.8%

We found that cluster 3 was the most prominent.