**NBA Players and Teams Analysis**

Zibin Chen, Weifeng Lyu, Zeyu Ni, Jingya Qin, Jian Zhan, Yang Zhang

Georgia Institute of Technology

**Introduction – Motivation**

We plan to analyze various types of NBA players, teams and games statistics. We will perform Machine Learning analysis to evaluate the players, players’ teamwork and teams, and eventually predict the champion of 2018–19 NBA season. We will perform the analysis results with user interface such that the users can view the ratings of their favorite players or teams. The reason we choose to analyze NBA is that NBA is one of the most popular sport associations attracting international audience. Our analysis system can provide meaningful domain knowledge and interesting analytic facts to NBA fans, and maybe generating real profits.

**Problem definition**

1. What are you trying to do?

This project implementation is divided into three steps:

The first step targets to evaluate NBA players’ player efficiency rating (PER) [1], which is the overall rating of a player's per-minute statistical production. We aim to compute the PER using players’ regular statistics in official algorithm, and implementing machine learning techniques such as decision tree and non-linear regressions [2] [3] [4] to make the rating more convincing.

The second step involves the evaluation of the teamwork among players with a team. As a team sport, the cooperation among players and the team chemistry are the important factor for winning the championship. Win-share [5] [6] is a classical measurement to determine how much a player contribute to the team’s wining, comparing to the fact that PER is computed without considering game result. We will also adjust the Win-share with the players’ position to justify their contributions with cluster analysis [7].

The third step is to evaluate the teams with the results integrated from previous steps. We will combine the PER and Win-share measurements from previous steps [8] [9], together with Elo ratings [10] as an adjustment to evaluate the overall team ratings. The Elo ratings is for calculating team’s relative competitive levels in one on one situation and many researches are applied into NBA analysis scenario [11]. It can be improved by the SVM and logistic regressions [11]. Once the ratings of each team are determined, the probability of wining the final champion can be calculated based on regression analysis [12], Monte Carlo Simulations [13] and Machine Learning technics [14].

Beyond the evaluation steps, we also need to scrape NBA players and team’s statistics from ESPN, NBA-stats and basketball-reference with Python [15] [16]. These three websites store consolidated and updated data of NBA players and teams. We will also utilize D3 [17] and JavaScript [18] to present our analysis results to user. Users can then check the ratings of their favorite players or teams. We will also show the probability of each teams that could win the final champion.

**Survey**

2. How is it done today; what are the limits of current practice?

Either evaluate a player or a team, or predict NBA champion are not trivial, as many factors will affect their performance. Scientific researches and predictions nowadays are well-conducted and can guide NBA team to make better decisions for trades, contracts, rookie selections [19] [20] [21]. However, most of the researches are conducted by the experts or scouts, which is private for themselves to generate profit for NBA teams. Those complex analysis and valuable researches are inaccessible and not understood from public.

3. What's new in your approach? Why will it be successful?

Please refer Innovation.

4. Who cares?

Scouts, coaches and fans have their attitudes to players and teams, with positive and negative opinions and even judgements. NBA team even hires data analysts to evaluate players and help them make trading decisions. Our data presentation can convey meaningful implication to the NBA fans through cool user interaction.

5. If you're successful, what difference and impact will it make, and how do you measure them?

The impact of this project is a direct delivery to end-users. With sufficient data analytics, we aim to disseminate objective judgement about NBA to users with great experience. Every time users finish browsing our application, a simple well-designed survey will populate to ask for their experiences and suggestions.

6. What are the risks and payoffs?

Data preparation and result explanation is among the risks in this project. The risk will be managed by limiting the scope of data preparation to trusted data source and relying on team member’s NBA domain knowledge. We expect to learn NBA domain knowledge, analytic skills and data analysis techniques along with this project experience as the payoff.

7. What are the midterm and final "exams" to check for success? How will progress be measured?

Our targets is to establish the players rating system by midterm and eventually complete the champion prediction system. The success will be measured on how meaningful data is presented from the graph, and how effective our conclusion based on the back-test of our predictions with historical data.

**Proposed method**

1. Intuition

2. Description of your approaches

a. Algorithms

(1). Player Evaluation

Blah, blah, blah

(2). Team Evaluation

Blah, blah, blah

(3). Champion Evaluation

Blah, blah, blah

b. User Interfaces

(1). Player Evaluation

Blah, blah, blah

(2). Team Evaluation

Blah, blah, blah

(3). Champion Evaluation

Blah, blah, blah

It is rare to integrate machine learning techniques with many existing complicated models to analyze statistics and implications in NBA research. Besides, the NBA researches are usually conducted on the player level by scouts, while the quantitative evaluation of team is rare. This project will utilize our ratings of players and players’ contributions to appraise the teams, with the official algorithmic computations, such as Elo ratings, NBA Win Shares and PER. The machine learning methods, such as K-means, decision tree and Bayesian Linear Regression [22], are used to process the analysis effectively. The project also aims to combine data analysis with data visualization. A well-designed user interface will demonstrate our analysis results with awesome user experience.

**Experiments/ Evaluation**

1. Description of your testbed

a. Player Evaluation

list of questions your experiments are designed to answer

b. Team Evaluation

list of questions your experiments are designed to answer

c. Champion Evaluation

list of questions your experiments are designed to answer

2. Details of the experiments

a. Player Evaluation

observations (as many as you can!)

b. Team Evaluation

observations (as many as you can!)

c. Champion Evaluation

observations (as many as you can!)

The expected results will be the graphical interface showing the ratings of the players and teams. When the user chooses a team, the team’s winning probability and related data will be shown on the screen. When the user chooses a player, the related historical data and a bubble chart showing player’s interaction with others will be presented on the screen.

**Conclusions and discussion**

Blah, blah, blah

**Work Plan**

|  |  |
| --- | --- |
| Task | Team Member |
| **Phase I: Preparation** |  |
| Task 1: Recruit Project Team Member | J.Zhan, Y.Zhang, Z.Chen |
| Task 2: Research topics and brainstorm methods | All members |
| Task 3: Manage and Organize Weekly Meeting | J.Zhan |
| **Phase II: Proposal** |  |
| Task 1: Integrate thoughts into project planning | W.Lyu, J.Zhan |
| Task 2: Complete Proposal and Task Distribution | W.Lyu, J.Zhan |
| Task 3: Consolidate Video Recording and Presentation | J.Qin, Z.Chen |
| **Phase III: Implementation** |  |
| Task 1: Data Collection and Integration (Python Scrape) | All members |
| Task 2: Players Evaluation (Python, sci-kit learn) | W.Lyu, J.Zhan, |
| Task 3: Teams Evaluation (Python, sci-kit learn) | J.Qin, Z.Chen |
| Task 4: Champion Prediction  (Python, sci-kit learn, Simulation) | Y.Zhang, Z.Ni |
| Task 5: Result Visualization  (D3, Tableau) | All members |
| **Phase IV: Summary** |  |
| Task 1: Progress Report | W.Lyu, J.Zhan |
| Task 2: Poster Design | J.Qin, Z.Chen |
| Task 3: Peer Grading | Y.Zhang, Z.Ni |
| Task 4: Final Report | All members |

**Distribution of team member effort**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Zibin Chen | Weifeng Lyu | Zeyu Ni | Jingya Qin | Jian Zhan | Yang Zhang |
| 16% | 18% | 16% | 16% | 18% | 16% |

**REFERENCES:**

|  |  |
| --- | --- |
| [1] | T. Zhang, J. Chen and X. Zhao, "Modeling and Analysis of Player Efficiency Rating for Different Positions: Case Study with NBA," in *Advances in Computer Science and Education Applications*, Beijing, Beijing University of Posts and Telecommunications, 2011, pp. 234-242. |
| [2] | A. Cheng, "Using Machine Learning to Find the 8 Types of Players in the NBA," Fastbreak Data, 2 March 2017. [Online]. Available: https://fastbreakdata.com/classifying-the-modern-nba-player-with-machine-learning-539da03bb824. [Accessed 8 October 2018]. |
| [3] | D. Rust, "An Analysis of New Performance Metrics in the NBA and Their Effects on Win Production and Salary," University of Mississippi, 2014. |
| [4] | I. Kohli, "Finding Common Characteristics Among NBA Playoff and Championship Teams: A Machine Learning Approach," York University, 2017. |
| [5] | Website, "NBA Win Shares," Basketball Reference, [Online]. Available: https://www.basketball-reference.com/about/ws.html. [Accessed 8 October 2018]. |
| [6] | A. Olson-Swanson, "Predicting NBA Winning Percentage," 9 August 2018. [Online]. Available: https://towardsdatascience.com/predicting-nba-winning-percentage-in-upcoming-season-using-linear-regression-f8687d9c0418. [Accessed 10 October 2018]. |
| [7] | D. Lutz, "A cluster analysis of NBA player, Sloan Sports Conference," MIT Sloan Sports Analytics, 2012. |
| [8] | S. Righini, "Actual vs. Perceived Value of Players of the NBA," Bryant University, Smithfield, 2013. |
| [9] | S. K. Deshpande and S. T. Jensen, "Estimating an NBA player’s impact on his team’s chances of winning," The Wharton School University of Pennsylvania, 2016. |
| [10] | S. P. Sorensen, "An overview of some methods for ranking sports teams," University of Tennessee, 2000. |
| [11] | V. Madhavan, "Predicting NBA Game Outcomes with Hidden Markov Models," University of California, Berkeley, 2015. |
| [12] | Y. Yang, "Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics," University of California at Berkeley, Berkeley, 2015. |
| [13] | A. Groll, C. Ley and G. Schauberger, "Prediction of the FIFA World Cup 2018 – A random forest approach with an emphasis on estimated team ability parameters," Technische University Dortmund, Dortmund, 2018. |
| [14] | R. Torres, "Prediction of NBA games based on Machine Learning Methods," University of Wisconsin Madison, 2013. |
| [15] | R. Mitchell, Web Scraping with Python: Collecting More Data from the Modern Web 2nd Edition, Newton: O'Reilly Media, 2018. |
| [16] | Q. E. McCallum, Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work 1st Edition, Newton: O'Reilly Media, 2012. |
| [17] | S. Murray, Interactive Data Visualization for the Web: An Introduction to Designing with D3, Newton: O'Reilly Media, 2017. |
| [18] | K. Dale, Data Visualization with Python and JavaScript: Scrape, Clean, Explore & Transform Your Data 1st Edition, Newton: O'Reilly Media, 2016. |
| [19] | S. Shea, Basketball Analytics: Objective and Efficient Strategies for Understanding How Teams Win, CreateSpace Independent Pub. Platform, 2013. |
| [20] | S. Shea, Basketball Analytics: Spatial Tracking, CreateSpace Independent Publishing Platform, 2014. |
| [21] | H. S. Bhat, L.-H. Huang and S. Rodriguez, "Learning Stochastic Models for Basketball Substitutions from Play-by-Play Data," University of California, Merced, 2015. |
| [22] | J. Perricone and I. Shaw, "Predicting Results for Professional Basketball Using NBA API Data," Stanford University, 2015. |
| [23] | S. Zhang, C. Tan and Q. Lv, " “This is why we play”: Characterizing Online Fan Communities of the NBA Teams," University of Colorado Boulder, 2018. |
| [24] | P. Moreno and S. Lozano, "A Network DEA Assessment of Team Efficiency in the NBA," *Annals of Operations Research,* vol. 214, no. 1, pp. 99-124, 2014. |
| [25] | S. Radovanović, M. Radojičić, V. Jeremić and G. Savić, "A Novel Approach in Evaluating Efficiency of Basketball Players," *Management,* vol. 67, pp. 37-45., 2013. |