

STAT428-final project-proposal

Group6

2019/4/3

Basic information for our project

- Title: Simulating NBA match results and predicting NBA playoff teams
- Group Number: 6
- Group members:
 1. Zepeng Xiao, zepengx2
 2. Shuogong, shuog2
 3. Xiaopin Hua, xh3
 4. Dongfan Li, dongfan2
 5. Sixin Ma, sixinma2

Basic description

Application Area

Our application area can be described as **Sports - Computational Statistics**. On the one hand, the topic of our project lies in the area of sports, on the other hand, the investigating method of our project is related to computational statistics.

Problems to be investigated

The National Basketball Association (NBA) is one of the most worldwide popular sporting event that is held every year in the world. A huge amount of statisticians work for it, using historical data to analyze the conditions for NBA players and NBA teams, and give predictions to the result of a single match, a series and even the whole season. Hence the attraction of NBA drives our group to do some statistical work related to it.

The main purpose of our project is simulating match results and predicting NBA playoff teams for season 2017-18 based on this season's real data, and comparing our results with this season's real data to calculate our method's accuracy.

Methods to be used

- Bayesian Method: We consider the match result as a binary distribution with probability θ . We use past statistics to determine the prior binary distribution.
- Random number generation: We use random number generation to determine the match result based on prior binary distribution.
- Linear Regression: Use stepwise method to choose the most significant variables
- Permutation test: Check if there is collinearity between variables to improve the regression model (for example, use ridge regression instead of linear regression) and our simulation results.

Data to be used or simulated

The datasets we decided to use include the followings, but not limited to:

- NBA Elo : which includes historical data for game-by-game Elo ratings and forecasts back to 1946 in NBA, the following variables may be used:
 1. pts: Points scored by team
 2. elo_i: Team elo entering the game
 3. elo_n: Team elo following the game
 4. win_equiv: Equivalent number of wins in a 82-game season for a team of elo_n quality
 5. opp_elo_i: Opponent elo entering the game
 6. opp_elo_n: Opponent elo following the game
 7. game_location: Home (H), away (A), or neutral (N)
 8. game_result: Win or loss for team
 9. forecast: Elo-based chances of winning for the team, based on elo ratings and game location
- nba-enhanced-stats : which includes box score and standing statistics from the NBA, the following variables may be used:
 1. teamAbbr: Abbreviation of team
 2. rank: Rank by winning percentage behind leader in conference
 3. rankOrd: Ordinal rank by winning percentage behind leader in conference
 4. gameWon: Number of games won during season
 5. gameLost: Number of games lost during season
 6. homeWin: Number of games won playing as home team
 7. homeLoss: Number of games lost playing as home team
 8. awayWin: Number of games won playing as away team
 9. awayLoss: Number of games lost playing as away team
 10. confWin: Number of games won playing against teams in same conference
 11. confLoss: Number of games lost playing against teams in same conference
 12. lastFive: Record of team in last five games
 13. lastTen: Record of team in last ten games

Project plan

The meetings will be held weekly, generally on 4:30 p.m to 6:30 p.m Friday, the purpose of meeting include checking each member's work last week and making plans for next week's work. Besides, different members that are involved in different tasks will have their own meetings if necessary.

Our detailed project plan is showed below:

Tasks	Involved group members	Internal due
Data searching	Dongfan Li, Zepeng Xiao	4/10
Data preprocessing	Shuo gong, Xiaoping Hua, Sixin Ma	4/14
Regression for the probability of match result and make one prediction for playoff teams	Xiaoping Hua, Zepeng Xiao	4/19
Repeat regression and prediction 1000 times	Shuo gong, Dongfan Li, Sixin Ma	4/22
Improve our result if possible	The entire group	4/29
Final Report and Final Presentation	The entire group	5/3
Individual Report and Peer Evaluation	The entire group	5/10