

EECS349 Machine Learning – Project Proposal

by: Weilin Ma

My project for this course will be focusing on learning the statistics from NBA players and games to determine which position suits one specific player the most, or if one specific team is at the best chance to win the title this season.

I chose this topic because I am a basketball super fan and I have been studying the NBA players, teams, organizations, games all the time, based on my experience watching the games. I am always obsessed with the statistical analysis by big professional sports agencies like ESPN, TNT, Bleacher Report, FS1, NBC sports, NBA TV, etc. But now that I have learned some machine learning techniques, I would love to study these stats on my own “mathematically”.

Because I know the NBA games and all related things (players, organizations) so well, I know which attributes will filter out the best result and exactly how important each one of them will be. Also because of my reach experience, I can evaluate the accuracy very easily on my own so that it will be much easier for me to optimize the training algorithms.

I will be acquiring data sets most likely from sports data base websites (kaggle.com, nba.com, basketball-reference.com). These data base websites are official and trustworthy. The agencies I mentioned above have been using their stats all the time.

The likely attributes that I would use are the usual stats categories. They include: player height, weight, wingspan, vertical, acceleration, lateral quickness, points per game, rebounds per game, assist per game, steals per game, blocks per game, shot percentage, playing minutes per game, games played per season. Like I said before, I know which attributes are crucial and which are redundant depending on the results I am trying to get. So I will choose only the most important attributes when I am training the data sets.

My initial approach would be to assign one specific player at two best positions on court. Because nowadays almost every player has a dominant position to play and can also sub in at another position sometimes. Two positions out of five is also much easier and faster to train. I will most likely use nearest neighbor for this project because it's a natural and more reliable method used by human sports analyzers. They compare similar players all the time and they will always assign a “model player” already actively playing in the league, to each rookie on their draft night. I will evaluate my success easily by comparing to the player's real life playing positions and behavior.