

Christopher Cheung
UROP Faculty Supervisor: Ben Shields
Term: Spring 2020
2/10/2020

Using ML to Predict Successful Sports Teams Based on Player Personality and Team Dynamics

Project Overview

Provide an explanation/background of your UROP project that includes with whom and where you are conducting research.

There are plenty of instances of sports analytics that have existed for the past decade or so. Prevalent examples include the MLB “shift”, the NBA “Hack-a-Shaq”, or the NFL’s recent rise of 4th down conversions. What I want to discover is the analytics behind sports team dynamics, and how team members’ relationships with each other may drive success. Specifically, I hope to define team lessons that leaders of the sports industries, such as head coaches or GMs, may leverage for their own teams to be successful.

This may involve using statistics to help coaches better understand and lead individuals with varying types of personalities, empirically predicting whether inserting a player into the starting lineup may emotionally rally a team, or even the statistical importance of an “alpha dog” leader, and how vital/harmful to a team they may be.

I will be conducting machine learning research to predict success in team cohesion in sports. I will be researching with Ben Shields, Senior Lecturer at MIT Sloan, and one of the leaders in the field of sports analytics and management at MIT. Research will mainly occur at MIT Sloan, on MIT’s campus, or in my living residence (Zeta Beta Tau).

Personal Role & Responsibilities

Describe what you are contributing to the project. Be specific about what your personal duties are and what you will be responsible for accomplishing throughout the term.

It would be very useful to use a sports player’s personality data to analyze team dynamics and potential success. However, in sports, this type of data is incredibly hard to encode and quantify in order to use objectively. For one, professional sports leagues like the NBA or NFL don’t record personality statistics of their players - no Myers Briggs, no time spent talking to their teammate, and so on. But even if we did try to track some sort of statistic like this, it would to a large extent be subjective, rather than objective.

So, my role will be to apply machine learning and/or deep learning to mitigate the subjectivity of personality data that has been numerically encoded, in other words to assume that there is or that there will be extensive “soft” statistics available that would allow us to accurately define

personality traits, behaviors, especially interaction between players. My hope is to find some non-random pattern from the intricate mappings of the “soft” datasets or correlations that the ML algorithms would find, using the constant big data number crunching and black-boxing technologies. I can’t be sure that ML will have success in tackling this question, because this work has largely not been done before. But it’s worth it to find out.

I’d like to structure my attempts at this problem in the following way:

Step 1: Create my own metric that measures NBA team cohesion, using player history data including: where a player has played in the past, what position he has played in, who he has played with, how long for, etc. I will then create a machine learning model that is trained on these features involving player history (for now). This information is in the public domain, and should be readily available for work.

This model will be able to track the importance of each and every feature that I feed in, so that however granular I choose to go with my metrics, the machine can still handle it. In addition, using a machine learning model can do interesting things like make predictions about current NBA teams today, make analyzations about player trades, among other things. I will be responsible for the data scraping for NBA data online, the algorithm churning to develop these features, and the research into what my metric might empirically mean.

Step 2: In addition to features based off of player history data, I’ll input more “soft” personality data into my model. I will start by creating a survey for the world wide web, for people to access with their phones or computers. This is a realistic start for trying to gauge a numerical statistic for a player’s personality, and will provide the large amount of data necessary to train a machine learning model. There are many ways and many personality traits that I can survey people to ask about a player’s personality. One interesting soft characteristic is:

Many teams have an “alpha dog” player - the player who calls a lot of the shots, who literally may get all of the shots, and who tends to dominate on the floor and over the rest of the teammates. Is there a way we can quantify this “alpha dog-ness”? A simple survey could show images/videos of various NBA players, and have a random group but a large enough cohort to rate on a scale the “alpha dog”-ness of the player they’re seeing play. The rating could be based on their play, or how they interact with their teammates (potentially more interesting). Of course, this is not the only way to measure a player’s level of “alpha dog-ness”. But it is a question that I will be starting off with.

Then, the ratings for a certain player (or team) would serve as another feature encoding to be used in my machine learning model. I will also use this to answer a variety of questions regarding team “cohesion” with an alpha dog. Is a team more successful if the “alpha dog” is the team captain? What about the teammate who is the most “team player” of them all? Does giving an alpha dog player all the shots harm team chemistry and performance?

Goals

Explain what your personal goals for the UROP are, as well as what the overall aim is of the project.

My goals for this UROP include first being able to identify the major factors in player history data that play a role in determining team chemistry and cohesion. I hope to answer the currently unanswered questions regarding how to know if a player is truly a good “fit” for a team. To apply this to today’s current NBA landscape, and to understand how the individual personalities of NBA players can develop a championship-winning team, is my goal.

I also hope to discover how machine learning can be used in topics I am particularly passionate about. I have been exposed to and understood material taught in my computer science and machine learning classes, but I have not been able to apply it to a realistic situation such as this UROP. From a career standpoint, my vocational goal is to hopefully pursue a profession that delves deep into sports analytics, and this UROP would give me an introduction into what that would be like.

Of course, following Step 2 of my plan, I would expand greatly on this project. An overall aim is to increase knowledge and awareness of the relevance of these “soft” statistics in sports and how they can play a major role in determining a sports team’s success. The higher the prevalence of “soft” statistics that become available, or that our model identifies as being relevant, the better for my model and for the sports analytics world in general. My hope is that it would open the door to entice leagues like the NBA, for example, to consider officially collecting these “soft” statistics for more analyses like this one.

Personal Statement

Briefly state why you are interested in this UROP and explain what you hope to learn from it.

My personal interests for this UROP are to delve deep into a topic that has always been a passion of mine, but has never really been explored. Sports analytics is itself a growing field, but the majority of sports analytics revolves more around “hard” statistics such as when to shoot a 3-pointer vs. a 2-pointer. I would like to begin research on the “soft” statistics which currently has little to no data or findings for the professional sports realm.

I am also very excited to combine my knowledge and experience in computer science with my love for sports. As an athlete and a past member of the MIT Varsity Volleyball team, I have always been interested in sports data, but was never fully able to use what I had learned in Course 6 and apply it to such a burgeoning field. There are so many realizations in sports that computer science and machine learning can unlock, and I believe this UROP will help me explore how powerful the intersection of these two fields can be.