Project Proposal

Chris Fenton

DATA 698

My goal for this project is to fit a classification model to NFL plays to derive a metric called “Win Probability”, and an additive metric “Win Probability Added”. My hypothesis is that WPA will be a better indicator of player performance than traditional player metrics. I will test this on Quarterback play, since QB is generally considered to be the most important player on a football team.

One of the challenging things about doing data analysis of NFL Football is that since the sport is highly situational, it makes many of the basic game data hard to value without proper context. Certainly, a scoring touchdown is generally going to be a positive data point for a team, as it scores points for that team, and the object of the game is to score more than other team.

However, consider a player running for 4 yards (one of the more common plays of a football game). Is this considered a positive or negative play for that player’s team? The answer completely depends on the context. Is the player’s team facing a 3rd and 1, where four yards will bring a new set of downs? Than very likely it is a positive play. However, if the players team is facing a 3rd and 20, they will likely have to punt, thus giving the ball away, and it is a negative play. So the data that we have that a player ran for four yards is impossible to assign proper value to without additional, contextual data.

Despite this truth, the way the NFL community has typically analyzed game statistics ignores context, out of either a lack of availability, ignorance, or both. After all, it is much easier to calculate the total amount of yards a player has run for, than to go play by play and manually assign each play a value of “positive” or “negative”. And generally speaking, gaining yards is an indicator of success, since it states that the ball was moved toward the opponent’s goal, and scoring is the object of the game. But if we stop there, much of the true information about what happened is lost.

In recent years solid progress has been made addressing this via a variety of efforts that attempt to encapsulate critical “game state” information that can be used to contextual game data. One of the first was Defense-adjusted Value Over Average (DVOA) by Football Outsiders (<http://www.footballoutsiders.com/info/methods)> , which manually encoded each play according to a proprietary scoring scale, and summed the results by player, game, or season.

A more recent attempt has been to incorporate the idea of Win Probability (WP), and its close companion, Win Probability Added (WPA). The idea of Win Probability goes back to mid 20th century baseball.