Football Synergy: Evaluating the Chemistry Between Football Players

Aryan Mittal

May 2023

Background

American football ("football") is by far the most popular sport in the United States [14], with viewership of the Super Bowl, the championship game of the National Football League (NFL), often reaching over 100 million viewers [2]. Football is a sport that relies heavily on a team's players working together as a cohesive unit to either move the ball down the field to score points (offense) or prevent the opponent's offense from scoring (defense). As such, an important consideration when evaluating any football team lineup is understanding how well the abilities of the players in that lineup complement each other. Obtaining such an understanding can provide a substantial competitive advantage to coaches in deciding which players to put on the field, as well as general managers in evaluating potential trades or player acquisitions. However, quantitative analysis of player performance in football rarely goes beyond individual box score statistics. For an offensive skill player (i.e. quarterback, runningback, wide receiver, or tight end), these can include yards gained and touchdowns scored via passing, rushing, and/or receiving, as well as completion percentage (for quarterbacks) or receptions (for receivers). For a defensive player, these may include tackles, sacks, interceptions, and others. Linear combinations of these figures (e.g. passer rating, [16]), overall team statistics, and individual player grades [1] fail to account for the effect that a player's teammates have on his performance or how his presence on the field can affect the performance of his teammates. For example, the presence of an elite runningback on the field can attract extra attention from defenders, leading to fewer scrimmage yards for the runningback but making it easier for his wide receiver teammates to accumulate receptions.

Recent research in football analytics has either focused on predicting game outcomes ([10], [3]) or evaluating players individually ([18], [13], [15]). While evaluating the synergy between players has been addressed for cooperative sports such as soccer ([5], [4]), basketball ([12], [11]), hockey ([9]), and baseball ([17], [6]), it still remains to be investigated for football.

Goal and Objectives

The goal of our project is to use publicly available data from games played at both the collegiate and NFL level to quantify the on-field chemistry between football players. We aim to study the value of the interaction effects between players, namely the amount by which the on-field effectiveness of a set of players differs from the mere sum of their individual talents. We accomplish this goal by achieving the following objectives:

1. Model individual players using a modified version of the models developed in [11]
2. Determine the proportion of a player's statistical output that is independent of his teammates
3. Compute the value players bring to particular lineups and identify mutually beneficial transfer/trade opportunities

Methods

We use NFL and college football play-by-play data available through the `R` packages `nflfastR` [7] and `cfbfastR` [8], respectively. These datasets record detailed information about each play (e.g.

offensive and defensive players involved, play type, yards gained before and after catch, etc.) from every game played since 2002. Of particular interest to our project is the expected points added (EPA) and win probability added (WPA) data included for each play. These figures, calculated using the methodology put forth in [18], encapsulate in a single number the impact of a play on the outcome of the game, making them highly useful metrics by which to classify plays.

For each season we wish to study, we first extract the top players at each position of interest by playing time. We classify the remaining players as "replacement level" and treat the average of their statistics as the baseline for comparison. Then, for each starting player, we classify every play he was involved in based on its play type and associated WPA to create "events". Using this event data, we model the probability that a player commits certain actions conditioned on both the probabilities of his teammates committing other actions and the probabilities of opposing players defending against that action, in accordance with a similar model for basketball players developed in [11].

After performing this analysis for each player of interest, we aim to use the data generated to quantify the amount by which a given player increased or decreased the overall production of his teammates. We will explore the applications of this framework to team management decisions, such as simulating trades and transfers, optimizing weekly starting lineups, and identifying top landing teams for collegiate players aiming to be drafted into the NFL.

References

[1] PFF player grades. `https://www.pff.com/grades`.

[2] Super Bowl ratings history (1967-present). `https://www.sportsmediawatch.com/super-bowl-ratings-historical-viewership-chart-cbs-nbc-fox-abc/`, Jan 2023.

[3] E. Cabral Balreira, Brian K. Miceli, and Thomas Tegtmeyer. An oracle method to predict NFL games. *Journal of Quantitative Analysis in Sports*, 10:183 – 196, 2014.

[4] Ryan Beal, Narayan Changder, Timothy Norman, and Sarvapali Ramchurn. Learning the value of teamwork to form efficient teams. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, page 7063–7070, 2020.

[5] Lotte Bransen and Jan Van Haaren. Player chemistry: Striving for a perfectly balanced soccer team. In *Proceedings of the 2020 MIT Sloan Sports Analytics Conference*, Boston, Massachusetts, United States, March 2020.

[6] Scott A. Brave, R. Andrew Butters, and Kevin A. Roberts. Uncovering the sources of team synergy: Player complementarities in the production of wins. *Journal of Sports Analytics*, 5(4):247–279, 2019.

[7] Sebastian Carl and Ben Baldwin. *nflfastR: Functions to Efficiently Access NFL Play by Play Data*, 2023. https://www.nflfastr.com/, https://github.com/nflverse/nflfastR.

[8] Saiem Gilani, Akshay Easwaran, Jared Lee, and Eric Hess. cfbfastR: The SportsDataverse's R package for college football data., 2022. R package version 1.9.0.

[9] TL Idson and LH Kahane. Team effects on compensation: An application to salary determination in the National Hockey League. *Economic Inquiry*, 38(2):345–357, 2000.

[10] Jason Kolbush and Joel Sokol. A logistic regression/markov chain model for American college football. *International Journal of Computer Science in Sport*, 16(3):185–196, 2017.

[11] Joseph Kuehn. Accounting for complementary skill sets: Evaluating individual marginal value to a team in the National Basketball Association. *Economic Inquiry*, 55(3):1556–1578, 2017.

[12] Allan Z. Maymin, Philip Maymin, and Eugene Shen. NBA chemistry: Positive and negative synergies in basketball. *International Journal of Computer Science in Sport*, 12(2), 2013.

[13] Jason Mulholland and Shane T. Jensen. Predicting the draft and career success of tight ends in the National Football League. *Journal of Quantitative Analysis in Sports*, 10(4), 2014.

[14] Jim Norman. Football still Americans' favorite sport to watch. `https://news.gallup.com/poll/224864/football-americans-favorite-sport-watch.aspx`, Nov 2021.

[15] R. Drew Pasteur and Kyle Cunningham-Rhoads. An expectation-based metric for nfl field goal kickers. *Journal of Quantitative Analysis in Sports*, 10(1), 2014.

[16] Don Smith, Seymour Siwoff, and Don Weiss. NFL's passer rating. `https://www.profootballhof.com/news/2005/01/news-nfl-s-passer-rating/`, 1973.

[17] Joel Sokol. A robust heuristic for batting order optimization under uncertainty. *Journal of Heuristics*, 9(4):353–370, 2003.

[18] Ronald Yurko, Samuel Ventura, and Maksim Horowitz. nflWAR: A reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, 15(3):163–183, 2019.