Use of NFL Tracking Data to Identify Defensive Coverages and Corresponding Offensive Play Outcomes

Taarak Shah

Abstract

Analyzing on-field performance is essential for any NFL team. There are many strategies to limit opposing performance from both an offensive and defensive perspective. We use NFL tracking data on all passing plays from the 2018 regular season, provided by the 2021 NFL Big Data Bowl. This paper looks into an extension of an unsupervised classification model to identify defensive coverage to be either man coverage or zone coverage. There is further examination into the defensive schemes and analysis about how the offense performs against these coverages, quantified by receiver performance and play outcomes.

Contents

List of Figures

List of Tables

1 Introduction and Objectives

The main objective of the game of American football is to score as many points as possible and limit the opposing team from scoring. At its highest level in the NFL, there have been countless years of study, analysis, and strategizing on the best methods to do this as effectively as possible. With the advent of the NFL's Next Gen Stats division, there are opportunities to use detailed player tracking data to rigorously examine strategies of the game from a statistical perspective. Previously, quantitative analysis of NFL data has been subject to heuristic, situational data about the passer, receiver, and outcome of the play, or statistics such as completions, tackles, sacks, and other basic counting measures.

There was not a way to determine what happened through the course of the play, much less any usable method to determine avenues for improvement until this data became available for use. This data was released to the public domain with the *NFL Big Data Bowl* [4]. The Big Data Bowl is a Kaggle competition held each year for students and professionals alike who are tasked with analyzing trends and performance based on the vast amount of data collected by the NFL Football Operations Division. The datasets consist of details about each play and tracking information of each player. Now, with this available data, there are many opportunities to closely examine the tendencies of individual players and ways teams can exploit advantageous matchups.

In this essay, I aim to examine two key matchups that define a majority of play outcomes. The first focuses on analysis of coverage schemes for defensive players, specifically cornerbacks and safeties. This analysis was built on an extension of an essay by Dutta, Yurko, and Ventura [2]. In that paper, the authors examine unsupervised learning methods to evaluate classification of cornerbacks into man coverage or zone coverage based on traits such as position on the field, speed, direction, and distance from their opponent throughout the play. Their paper was constructed with data from the inaugural 2019 Big Data Bowl, which was limited and did not have information regarding player orientation. This forced them to only consider coverage assignments for cornerbacks, while forgoing analysis for

other important positions like free safeties, strong safeties, and linebackers. I construct my own implementation of their methods later in the paper, but include additional constructed features and consider analysis for safeties. I chose to forgo analysis for linebackers due to computational constraints and knowledge that linebackers rarely cover offensive players one-to-one and offer support in run defense, so any insights regarding man or zone coverage from this perspective may be largely redundant or ineffective.

The second aim of this analysis focuses on how offensive players, specifically wide receivers, perform against the different types of coverage. We are provided with ground-truth labels of the routes that receivers are running, and we compare these with their individual performance against the labels assigned via the clusters in our reproduced unsupervised learning approach. The intent of this analysis was to provide an exploratory overview of how routes and individual receiver performance can vary largely with the type of coverage they face. Another interesting discussion is to what degree the route influences the movement of the defender and dictates coverages, instead of the classic school of thought indicating that the coverages are predetermined assignments prior to the play.

We will first detail the in-depth feature creation and consolidation for the coverage analysis, comparing our methodology to Dutta, Yurko, and Ventura [2], detailing key differences and improvements. We then consider the extension of the model and how we use the model results for an exploratory analysis of corresponding offensive performance.

2 Methods

2.1 Data

This dataset was retrieved from the 2021 NFL Big Data Bowl [4]. There are multiple files in the dataset; as a whole, it contains all information about every passing play in the 2018-2019 NFL season. Additional information about accessing the data can be found in Section 6.2. There is information on 19,239 passing plays in the season. There are additional

files separated by each of the 17 weeks that contain individual frame-by-frame data for each of these 19,239 plays, totalling to 18,309,388 frames of plays. The tracking data contains information on player position, speed, direction, distance traveled since previous frame, and player orientation. This data was collected and verified by standardized procedures through NFL's Next Gen Stats [6]. Player / object tracking data is monitored via RFID tags in each player's shoulder pads and an RFID tag in each NFL football. Over 200 data points are created on every play of every game. Details on how these features were adapted for our analysis are found later in Section 2.

2.1.1 Predictors in Dataset

A full list of predictors can be found in Table 2 in Section 6.4.

2.2 Clustering for Coverage Assignments

The visualization is provided by *NFL Big Data Bowl - Plotting Player Position* [5]. The GMM framework is initialized by Banfield and Raftery [1].

2.3 Assessing Receiver Performance

3 Results

- Discussion of results of our GMM algorithm
- Some visuals of quantifiable differences
- What ways could we improve our results?

3.1 Coverage Types

3.2 Receiver Routes vs Coverage Types

4 Conclusion

- Why is this question important?

- What was the main key result of our project?

- Does this result provide value practically? Where can we use these results in a practical setting?

- Where could we expand on this project and provide more detailed insights? (Further work, additional things left undone, etc)

4.1 Discussion

4.2 Limitations / Future Work

5 References

[1] Jeffrey D. Banfield and Adrian E. Raftery. "Model-Based Gaussian and Non-Gaussian Clustering". In: *Biometrics* 49.3 (1993), pp. 803–821. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/2532201 (visited on 05/02/2022).

[2] Rishav Dutta, Ronald Yurko, and Samuel Ventura. *Unsupervised Methods for Identifying Pass Coverage Among Defensive Backs with NFL Player Tracking Data*. 2019. DOI: 10.48550/ARXIV.1906.11373. URL: https://arxiv.org/abs/1906.11373.

[3] *Glossary of Terms*. URL: https://operations.nfl.com/learn-the-game/nfl-basics/terms-glossary/.

[4] *NFL Big Data Bowl*. 2021. URL: https://www.kaggle.com/c/nfl-big-data-bowl-2021/.

[5] *NFL Big Data Bowl - Plotting Player Position*. Oct. 2020. URL: https://www.kaggle.com/code/robikscube/nfl-big-data-bowl-plotting-player-position/notebook.

[6] *NFL Next Gen Stats*. 2022. URL: https://operations.nfl.com/gameday/technology/nfl-next-gen-stats/.

6 Appendix

6.1 Reproduction

The data and code used for the analysis, as well as the resources for the paper can be found at: https://github.com/taarakshah/ms-thesis.

6.2 Downloading NFL Tracking Data

The data used in this analysis be downloaded from the Kaggle website here: https://www.kaggle.com/c/nfl-big-data-bowl-2021/data. The data will be downloaded in a zipped file of roughly 2.33 GB. There are 20 .csv files, 17 of which represent tracking data for the 17 weeks of the NFL season, and 3 of which provide supplementary information and keys to join on about individual games, players, and plays.

6.3 Glossary of NFL Terms

Some definitions were generously provided by the NFL Football Operations [3] division.

Term	Definition
Cornerback (CB)	A defensive player who focuses on guarding the opposing offense's wide receiver. Primarily serves to prevent pass completions.
Defensive Lineman (DL)	A defensive player who focuses on disrupting the quarterback's processing and attempts to tackle him for a loss of yardage.
Extra Point	A scoring play that results in 1 point for the offensive team. Can only occur after a successful touchdown.

Field Goal	A scoring play that results in 3 points for the offensive team. The ball must be kicked through the field goalposts for a successful try.
Line of scrimmage	The imaginary line marking the beginning yard line of a specific play. The ball is positioned here upon the start of each play.
Linebacker (LB)	A defensive player who is responsible for run defense and pass defense.
Offensive Lineman (OL)	An offensive player who focuses on protecting the quarterback from oncoming defenders.
Pass play	A type of play initiated when the quarterback receives the ball and throws forward to an eligible receiver, typically a tight end, running back, or wide receiver.
Play	A moment in the game where the ball is in action. Can be either a run play or pass play.
Quarterback (QB)	An offensive player responsible for initiating the play, throwing to a receiver or handing the ball off to a running back, and completing passes.
Run play	A type of play initiated when the quarterback receives the ball and hands it to a player in the backfield so that player can advance the ball, typically a running back.
Running back (RB)	An offensive player responsible for running the ball forward. Typically lines up behind or next to the quarterback.

Safety (S)	A defensive player who is responsible for covering zones to prevent passes being completed down the field. There are two versions of this position, known as "strong safety" (SS) or "free safety" (FS). Typically, strong safeties assist linebackers in run support, while free safeties are assisting cornerbacks in pass coverage.
Tight End (TE)	An offensive player that functions similar to a wide receiver, but is generally larger and assists in
Touchdown	A scoring play that results in 6 points for the offensive team. Can come in the form of a passing play, rushing play, or special teams play.
Wide Receiver (WR)	An offensive player responsible for running in predetermined patterns (routes) and catch the ball on passing plays from the quarterback.

Table 1: Glossary

6.4 Predictors in Tracking Data

Predictor	Mean, (25%, Median, 75%) / Unique Levels

Table 2: Predictors available in NFL tracking data