

Predicting offensive plays by NFL teams.

Rafael Andrade Ruíz Capetillo

Facultad de Ingeniería
Universidad Panamericana Aguascalientes, México

03/26/2023
Email: 0225854@up.edu.mx

Abstract— The following writing will describe the process of guessing what kind of offensive play a football team will use next. This will be done by feeding a neural network with data from the "NFL Play by Play 2009-2018" dataset and other official statistical data.

Keywords—*football, plays, classification, neural network.*

I. INTRODUCTION

American football is a play-by-play sport which means that each play you have some time to prepare yourself for the next one, so making a correct call is essential to gain control over the game. Traditionally coaches use their own experience to make these play calls (Mielke, D. 2007), however taking into account all the factors that can influence this decision can be somewhat complicated. Taking into account the abilities of your team with which you train daily is something that is considered more easily, but you also have to analyze the opponent and the moment that is being experienced in the match.

This is where the implementation of some type of model that allows you to predict the type of play that the opposing team (and specifically the offensive team) will use and with that information be able to propose a better defensive formation and tactic. This is said based on multiple examples of how artificial intelligence has helped with the prediction of sporting events. Such is the case of the Google Cloud platform, which was able to predict a large part of the matches of the 2014 World Cup, including the final (Google Blog, 2014). We also have the research presented by Ronal Yurko et al. in 2020 where he exposes different models to predict sports events in real time.

Said model must be fed with historical and statistical data on the behavior of a team throughout a complete season (effectiveness of the offense on the ground and in the air, effectiveness of the defense itself in the same areas, team's shape, record of both teams, etc.) as well as data that describe the context of the game at the specific moment (yardage of the field, number of down and yards to go, last plays executed, formation, advanced yards for each type of play, etc.).

This type of information can be found in the dataset used for the "Big Data Bowl" competition organized by Kaggle in 2019. This contains detailed information on each play of NFL games from the 2009 season to the 2018 season (Max Horowitz, 2018). In addition, various data sets from the official NFL website will be used to obtain the moment of form that a team had in a given game (record, opposing offensive effectiveness, own defensive effectiveness, etc.).

The types of play are divided mainly into pass or run. Each one has its respective subdivisions. Predicting a specific play would be almost impossible nowadays due to the high number of routes that exist and their combinations. So, the initial focus will be to do a binary classification (to predict if the offense is going to choose to pass or run) using a neural network.

In similar projects we can find the one presented by Cameron Taylor, a graduate of the Stanford University Business School. Where the use of convolutional neural networks is

exposed to predict plays of the pass or ground type and the yards that would be obtained for that play. In addition to using the same dataset as in this project, the one presented by Cameron used images for the context of the play, however he pointed out that they did not provide valuable information for the final result (*Cameron Taylor, 2020*). For this reason the statistical information was selected to feed the model instead. It is also information that a coach can obtain more quickly during a game.

II. THEORETICAL FRAMEWORK

American football:

Soccer and American football are two sports with different rules and regulations. In soccer, there are two teams consisting of eleven players each who play on a rectangular field. American football, on the other hand, is played on a rectangular field that is 120 yards long and 53 1/3 yards wide. The field has a field goal on either end and a ten-yard long end zone. The field goals are located at the end of the end zone, not inside. There are single-yard markers on both sidelines, and two sets of hash marks inside the field. The lines are spaced five yards apart, and every second line has a number indicating the distance to the closest end zone.

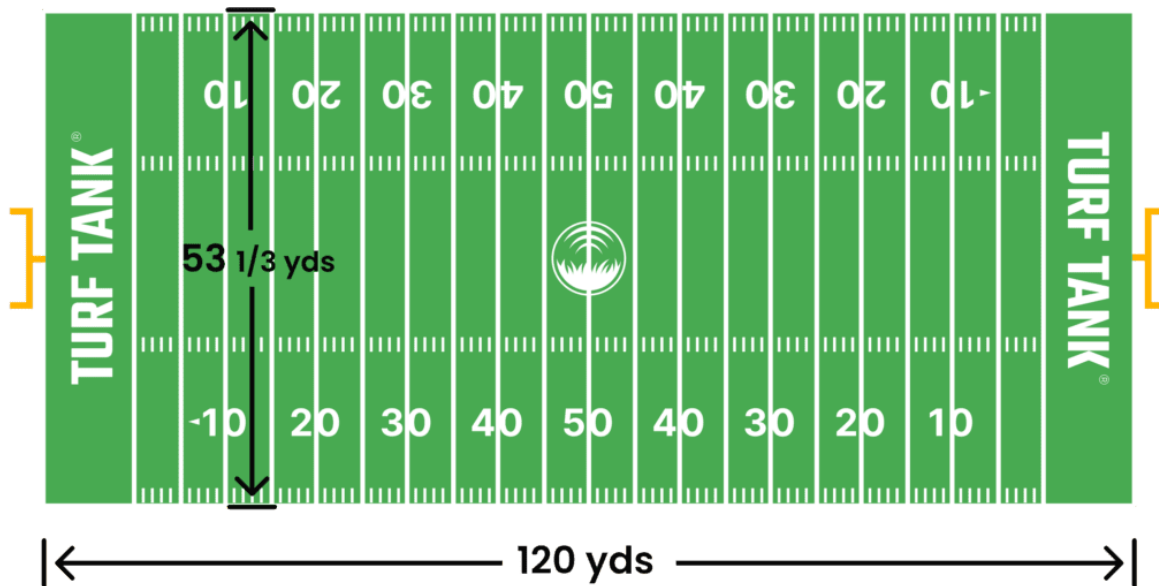


Figure 1: American Football Field (Turfank, 2023)

American football is played play-by-play, which means that the offense of one team executes a run or pass play. The play ends when a defender tackles the ball-carrying player successfully, or the ball carrier leaves the field of play. A play can also end if the quarterback throws an incomplete pass or a defender intercepts the ball, in which case ball possession changes, and the next play is carried out. Each play starts at the line of scrimmage (LOS), which is at the position of the ball when the previous play terminated. The offensive line, consisting of five players, protects the quarterback, and the center holds the ball on the LOS with one hand before the play. On a signal from the QB, the center

hands or throws the ball through his legs to the QB. This is called the snap or snapping the football, and it starts a play. After the snap, both teams can cross the LOS, but not before. The goal of the team in possession of the ball is to advance the ball down the field into the end zone of the opposing defense. The offense has four chances (also called downs) to advance the ball ten yards. If they manage to gain those ten yards, they get a new set of downs. If not, ball possession changes. Dickmanns, L. (2021).

Neural Network:

Neural networks are a subset of machine learning that are fundamental to deep learning algorithms. They are modeled after the human brain and its biological neurons, with each node or artificial neuron connected to another through associated weights and thresholds. These networks are made up of layers, which include an input layer, one or more hidden layers, and an output layer. When a node's output exceeds a specified threshold, it is activated and sends data to the next layer of the network. Neural networks rely on training data to improve their accuracy over time and can be used in various applications such as speech and image recognition. Once fine-tuned, they can classify and cluster data quickly, and are powerful tools in computer science and artificial intelligence. Google's search algorithm is an example of a well-known neural network.

To train a neural network, you need a dataset with known inputs and outputs. The network is adjusted by changing the weights and biases until it produces the correct output for each input in the dataset. This process is called supervised learning. The goal is to minimize the cost function, which measures how far off the network's output is from the correct output.

The most common type of neural network is a feedforward network, which means the data flows in one direction through the network. However, it's also possible to train a network using backpropagation, which involves adjusting the weights and biases from the output layer back to the input layer. This helps to improve the accuracy of the network.

Overall, neural networks are a powerful tool for solving complex problems, such as image recognition or natural language processing, and can be trained using a variety of techniques to produce accurate results. (IBM, 2023).

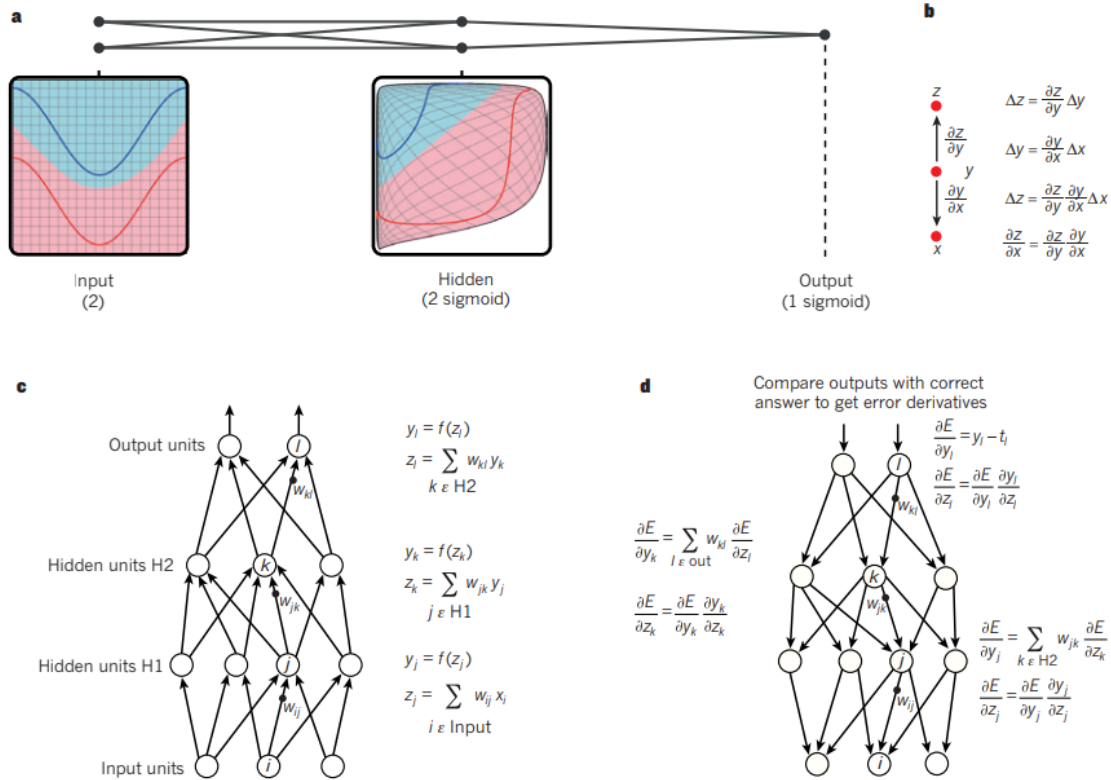


Figure 2:Diagram of a Neural network (LeCun et al., 2015)

III. METHODOLOGY

IV. RESULTS

V. CONCLUSION

REFERENCES

- Taylor, C. N. (2020). Using Machine Learning to Predict Property Maintenance Code Violations in Syracuse, New York (Final report). Retrieved from https://cameronntaylor.github.io/pdfs/final_report.pdf
- Yurko, R., Matano, F., Richardson, L. F., Granered, N., Pospisil, T., Pelechris, K., & Ventura, S. L. (2020). Going deep: Models for continuous-time within-play valuation of game outcomes in American football with tracking data. Journal of Quantitative Analysis in Sports, 0(0), January.
- Dickmanns, L. (2021). Pose Estimation and Analysis for American Football Videos.

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- IBM. (n.d.). Neural Networks. Retrieved March 17, 2023, from <https://www.ibm.com/topics/neural-networks>
- Turftank. (n.d.). How big is a football field? Turftank Academy. Retrieved from <https://turftank.com/us/academy/how-big-is-a-football-field/>
- Mielke, D. (2007). Coaching experience, playing experience and coaching tenure. *International Journal of Sports Science & Coaching*, 2(2), 105-108.
- Google Cloud Platform. (2014, July 14). Google Cloud Platform predicts the world cup winner. Google Cloud Blog. <https://googleblog.blogspot.com/2014/07/google-cloud-platform-predicts-world.html>
- Max Horowitz. (2018). NFL Play by Play 2009-2016. Kaggle. Retrieved February 17, 2023, from <https://www.kaggle.com/datasets/maxhorowitz/nflplaybyplay2009to2016>
- NFL. (n.d.). Team Stats. Retrieved February 17, 2023, from <https://www.nfl.com/stats/team-stats/>
-