Predictive Tool for American Football Defensive Positioning Using Machine Learning to Aid Coaches in Design of Offensive Formations and Plays

CPSC 502 Project Proposal

Alejandro Escobar

Student, University of Calgary, alejandro.escobar1@ucalgary.ca

Jonathan Hudson

Assistant Professor, University of Calgary, jwhudson@ucalgary.ca

CCS CONCEPTS • Computing methodologies ~ Machine learning ~ Machine learning approaches ~ Neural networks

1 INTRODUCTION

American Football is the most popular and highest viewed yearly athletic sport in North America. A game of American[1] football consists of two teams which field eleven players at a time [1]. A team wins a football game if they can outscore the opposition. Most of the game is played in a series of downs, where the offensive team has possession of the ball and the defensive team aims to prevent the offensive team from progressing to their end, which would result in a touchdown[2]. These downs are subunits of an entire football game, where both teams line up in opposing formation depending on the play they are executing. The number of downs the offensive team gets is reset if the offensive team can progress 10 or more yards before they run out of the previously awarded number of downs.

For our project, we will be focusing on a single down and the formational plays used by a team when the ball is live. More specifically, we will be focusing on the response of the defensive team when the offensive team executes a chosen play. We aim to analyze the positional data associated with each player on either team at any given time from when the teams line up, to when the ball is live, and throughout the duration of the play until it is ruled complete. The overarching goal is to produce a tool for coaches to observe the most likely defensive player movement in response to different chosen offensive formations. The proposed tool is expected, at least in the non-private field of machine learning in sports, to be novel and represent a move beyond using machine learning for descriptive statistics towards predictive assistance in changing strategy. Although the tool may not be completed given the timeframe, we are willing to accept the progress towards it as just as valuable of research

Data analysts have become invaluable in modern sports and have become a huge contributing factor to an NFL team's success throughout a season [2]. Having the ability to predict defensive response and positioning to specific routes and plays helps offensive coaches design and analyze their plays against a specific opposition to have higher success rates. Designing tools that visualize and predict these responses would be an asset to any team with a sufficient and usable dataset,

[1] Other variants exist with different rules, such as Canadian.
[2] 6 points

such as professional NFL teams. Fortunately, the NFL has introduced tracking in equipment used during games [3]. RFID sensors embedded in the back of player protective equipment and game balls collect tracking data throughout each game [4]. The resulting data is complex enough for the creation of insightful prediction tools.

Each year, the NFL releases a portion of the large datasets collected in past seasons for a sports analytics competition called the Big Data Bowl [5]. In this competition, professionals, and aspiring data analysts (including students), are invited to contribute innovative solutions to yearly challenges using provided datasets. Previous Big Data Bowl submissions display the possibilities of machine learning techniques when applied to large datasets [6]. These datasets include positional data about the players from NFL games throughout a season. The datasets are a reliable and valid source for our project's required input. We will use this dataset to create and train a data model for our proposed tool.

The data model we will be using for this project is a Long Short-Term Memory (LSTM) Neural Network [7]. An LSTM is a type of recurrent neural network that is capable of learning order dependence in sequence prediction problems. In other words, we can use an LSTM to predict the future of a coordinate system path based on the past sequence of path location, speed, and orientation. LSTMs have been applied to solutions for Weather Forecasting, Stock Market Prediction, and many other predictive tools [7]. These are some examples that demonstrate the advantages of LSTM and what it can be used for. For this reason, we will implement an LSTM model for our defensive positioning predictive tool and find common trends in the large NFL dataset. We plan to create this model using TensorFlow for Python and other supporting libraries. We will train this model using relevant data extracted from the datasets provided by the Big Data Bowl.

We will measure the success of our project by evaluating the realization of a proof-of-concept interactive tool that predicts defensive positioning in response to offensive positions and movements. We plan to embed the model trained on the NFL tracking data in an interactive tool. This tool will receive the offensive play scheme and selected defensive formation as input. The tool will use the model to predict and simulate the defensive team's movement in response to the offensive play, which will aid coaches in offensive formation and play design. We will evaluate both the design of the model, and the interactive tool using it, at the completion of the project.

2 RELATED WORK

Most plays in an American football game feature an offensive unit of 11 players from one team facing against the defensive unit of the opposing team [1]. A minority of plays will feature special team units of 11 players from each team. When the offensive team has possession, they are allowed four downs to progress the ball down the field to the opposition's end. Downs are attempts to snap the ball into play and progress the ball down the field or attempt a point scoring play. Points in American football are scored through touchdown (6 points), field goal (3), safety (2) and a one or two extra point conversion (try) after touchdown. The number of downs the offensive team gets is reset if the offensive team can progress 10 or more yards before they run out of the previously awarded number of downs.

Before the play begins (the ball is snapped), players from each team lineup in a particular formation. In each lineup there are different positions/roles that a player can occupy which determine what purpose they serve and affects their positioning and movement. Player roles are heterogeneous and may have one or multiple players filling them. This indicates that player ability and purpose in a play can differ for individual players and will be a factor in our model.

Next Gen Stats' RFID sensors in game balls and player protective equipment have provided a large and accurate dataset for Big Data Bowl competitions and for predictive model usage [3, 4]. In 2020, during the Big Data Bowl competition, the first-place team created a convolutional neural network using the collected data for predicting the expected rushing yards of a ball-carrier from the moment of handoff [6, 8]. This solution utilized the positional data in the Big Data Bowl dataset and a variant of it is now deployed by the NFL statistics group to produce statistics presented through official league media and during live broadcasts [8]. Although the predictive challenge was different, this solution is evidence of successful neural network machine learning application. There have been several other applications in previous years of the Big Data Bowl. For just one other example, the collected dataset has also been used to solve problems such as 'defensive player coverage' evaluation through data analysis and defining metrics for evaluating a defender's performance in different scenarios [9].

In our project, we plan to utilize an LSTM neural network for predicting defensive player positioning. LSTMs are a type of recurrent neural network that are well known for their ability to store previous results and use them for future predictions [7]. LSTMs are very powerful for problems that require recognizing trends in sequential data and making predictions based off the previous results fed to the model, such as path prediction. A good example of an application of LSTM is the prediction of pedestrian trajectory by taking destination and other factors into consideration [10]. In this project, LSTM was utilized to classify and predict pedestrian routes/trajectories which proved to be highly accurate in trajectory prediction when compared to other methods. This is one example of the utilization of LSTM to identify trends in sequential data and create accurate predictions while considering many different factors that may affect the result.

Although our proposed work differs from the mentioned projects we discovered, it is important to note that it is impossible to know if professional NFL teams have engineered a private project like our proposed solution. To our knowledge, an interactive coaching tool that we aim to create by utilizing an LSTM for prediction is not publicly advertised as existing or otherwise publicly available.

3 PROPOSED WORK

The first step of this project is the processing of Big Data Bowl data into a format that is consumable by our LSTM model. We will do this by using supporting libraries in Python such as NumPy or Pandas. Initially, we will extract the positional data of single players at a time to reduce the complexity, rather than attempting to process an entire team's data. Once we have this extracted data, we will be able to build an LSTM in TensorFlow for Python and feed it a portion of the extracted data in multiple iterations (epochs) to train the model. Once the model has processed the training dataset for training, we will test the accuracy of the LSTM for a single player and adjust accordingly, until the prediction accuracy is satisfactory for the next steps.

Once we have accurate results for a single player, we will improve our model to predict for two players and repeat the process of training until reaching a threshold accuracy, before adding additional players. We will repeat this process until our LSTM model can create a prediction for a full defensive team. In each iteration we would be making use of the extracted data and have separated testing and training datasets which will be subsets of the data we processed using NumPy/Pandas. We will proceed in this manner to slowly increase the complexity of our problem and maintain a high prediction accuracy throughout the improvements of our model. This will be especially important particularly in our problem, since the model will have to factor players on each team that fill specific roles for the deployed formation/play. This will allow us to have simpler iterations of our target tool in case we encounter any other setbacks throughout the process of development.

When our model can predict results for a full defensive team, we will develop the coaching tool that will implement the final, trained LSTM model. This will be a visual interface that will receive offensive play schemes and a selected defensive formation as input and will utilize the LSTM model to predict the corresponding defensive movement in response to the play. In our tool, once the user has provided the required input, this will play out as a simulation that will provide the user with a clear visual of the expected response from the defensive team. This interactive coaching tool will be the last step of our project which will aim to provide coaches with accurate predictions to aid in designing offensive formations and plays.

4 TIMELINE

This section provides a tentative timeline for project milestones and a general guideline we hope to follow, not including bi-weekly status update meetings. This timeline will only include submission dates for CPSC 502.01A[3].

1. **June 15th – August 28th** Project Supervisor communications via email, initial meeting, and planning. Project Proposal drafts, proofreading and review from the Project Supervisor. Project Proposal submission by August 28th.
2. **August 28th – August 31st** Skim through the available selection of research opportunities and discover interesting opportunities that are or will be available. Complete research opportunity summaries for 3 selected opportunities.
3. **August 31st – October 10th** Collect academic papers for previous related work. This includes previous findings of Big Data Bowl projects, as well as academic articles of findings using LSTMs for predictive tools.
4. **October 10th – October 21st** Complete literature review for submission before the October 21st deadline.
5. **October 21st – November 15th** Collect provided data from previous NFL games through the Big Data Bowl Competition. Process and convert this data in a Python notebook using supporting libraries into a readable format for our LSTM. Start building LSTM Neural Network Data Model and setting up model and dataset for training. Create a first draft for the Interim Report and complete portions of the report that can be completed to date.
6. **November 15th – November 31st** Train the Neural Network with the transformed data and begin testing accuracy of results with a separate testing dataset. Continue working on the Interim Report and log the results to date.
7. **November 31st – December 7th** Complete and proofread Interim Report for submission.
8. **December 7th – March 15th** Continued work on improving the LSTM for more accurate predictions by training and adjusting data entries. Design a user interface that takes an offensive scheme and a selected defensive starting formation as input and simulates the predicted defensive response with visual display.
9. **March 15th – End of Winter 2023** Related additional work to project goals after revision and testing. Final Report and Presentation completion.

REFERENCES

[1] 2022. 2022 Official Playing Rules of the National Football League. (2022). Retrieved September 17, 2022 from https://operations.nfl.com/media/5kvgzyss/2022-nfl-rulebook-final.pdf

[2] Liam Fox. 2021. How the NFL uses analytics, according to the lead analyst of a Super Bowl Champion. (August 2021). Retrieved September 17, 2022 from https://www.forbes.com/sites/liamfox/2021/08/12/how-the-nfl-uses-analytics-according-to-the-lead-analyst-of-a-super-bowl-champion/?sh=2b38c1a0424e

[3] 2022. NFL Next Gen Stats. (2022). Retrieved September 17, 2022 from https://operations.nfl.com/gameday/technology/nfl-next-gen-stats/

[4] Vince Guerrieri. 2022. How RFID Chips Are Bringing Next-Gen NFL Stats to Your Couch. (August 2022). Retrieved September 19, 2022 from https://www.popularmechanics.com/technology/a40848969/rfid-chips-next-gen-football-stats/

[5] 2022. Big Data Bowl: The annual analytics contest explores statistical innovations in football — how the game is played and coached. (2022). Retrieved September 17, 2022 from https://operations.nfl.com/gameday/analytics/big-data-bowl/

[6] 2022. Past Big Data Bowl Recaps: Recap of the NFL's first and second annual Big Data Bowl competitions. (2022). Retrieved September 20, 2022 from https://operations.nfl.com/gameday/analytics/big-data-bowl/past-big-data-bowl-recaps/

[7] Aleia Knight. 2021. LSTM neural network: The basic concept. (April 2021). Retrieved September 17, 2022 from https://towardsdatascience.com/lstm-neural-network-the-basic-concept-a9ba225616f7

[8] The Next Gen Stats Analytics Team. 2020. Next Gen Stats: Intro to Expected Rushing Yards. (July 2022). Retrieved September 20, 2022 from https://www.nfl.com/news/next-gen-stats-intro-to-expected-rushing-yards

[9] 2022. 2021 Big Data Bowl Recap: Recap of the NFL's third annual Big Data Bowl. (2022). Retrieved September 20, 2022 from https://operations.nfl.com/gameday/analytics/big-data-bowl/2021-big-data-bowl-recap/

[10] Hao Xue, Du Q. Huynh, and Mark Reynolds. 2020. PoPPL: Pedestrian Trajectory Prediction by LSTM With Automatic Route Class Clustering. (March 2020). Retrieved September 20, 2022 from https://ieeexplore-ieee-org.ezproxy.lib.ucalgary.ca/document/9031707

[3] Retrieved from CPSC 502.01A - Research Project in Computer Science