A Disturbance in the Force? Modeling QB Pressure with Force-based Metrics

Annie Liang, Harrison Lian, Jingchun Quan, Joshua Choi
Project Advisor: Ron Yurko

Background & Introduction

Quarterbacks often get all of the attention, but a key to their success often lie in the linemen in front of them. The offensive line protects the QB, giving time for the team to carry out the play. At the same time, the opposing team’s defensive line attempts to break through and hit, hurry, or sack the QB. We are interested in seeing how the position and force of offensive and defensive linemen lead to a negative outcome of a hit, hurry, or sack. We analyzed player, play, game, scouting, and tracking data from NFL and Pro Football Focus in order to create features that would be predictive of a negative outcome and reveal new insights into how football coaches can integrate this information into plays.

Feature Engineering

Hypothesis: If defense exerts greater force, higher chance of negative outcome for QB

1. Net X & Y Force
2. Distance Weighted
3. Partitioned

Net Force exerted on field

Pass Block
Pass Rush

T. Brady pass incomplete (deep right to C. Godwin)

1. Calculated force exerted by player
2. Determined x and y forces exerted by direction for pass rushers and pass blockers
3. Force exerted was summed together to get net force

a. Net force > 0 = offense exerted more force
b. Net force < 0 = defense exerted more force

Analysis & Modeling

- Force features and distance features are orthogonal, revealing different sources of variability in the data
- Dist_to_qb and frame_id have strong negative correlation, since as plays go on, players move closer to the QB, giving more opportunities to hit/hurry/sack

Top 3 features in terms of information gain: frameld, net_x_force_middle and dist_to_qb_rush

1. Fit XGBoost and GLM as baseline models on the entire dataset using frameld + all engineered features as covariates
2. Fit separate models (logistic regression and random forest) for each frameld
   a. Football intuition: Frameld is most predictive variable but not something that a coach/player can control
   i. We can take away the influence of frame ID by having separate models trained for each frame to better observe other features
   b. Statistical Intuition: Response is at the play level but our observations are at the frameld level
   i. If we fit model on all frame IDs, we will be adding unnecessary error terms

We successfully modeled the likelihoods of the QB getting hit, hurried, or sacked using the tracking data of linemen, specifically focusing on the distance features as well as feature engineered net force calculations. We also made connections between player attributes, which coaches can control, to distance and time attributes, which tend to evolve over time. Throughout our modeling and analysis, we also observed:

1. Our models are currently heavily skewed by distance to QB which usually decreases as time passes, making the probability of a negative outcome higher
2. When distance features are dropped, similar performance is observed, which gives us confidence in the predictive ability of the net force features
3. Marginal increases in weight decrease a player’s force due to the uneven tradeoff between lower acceleration and higher weight

Conclusion & Takeaways

Figure 9 illustrates the relative importance of the current features used in our 2 modeling approaches. We note that many of our features are heavily correlated with each other (distance features and force features) as confirmed by the PCA. Besides exploring additional features to see trends on the spatial temporal data of linemen, other possible directions to expand our research include:

1. Model occurrence of bad outcome in next 10 frames instead of at end of entire play
2. Explore models that capture autocorrelation between frames
3. Used survival models and hazard functions to model how factors that players can control can make a play last longer and give the QB more time
4. Implement a self-updating Elo score between specific pass blocker vs. pass rusher that could attribute a score to a team of defensive or offensive linemen

Next Steps