The Immaculate Reception

Dimensionality-Reduced Receiver Route Optimization



Problem

How can we optimize routes so that we can increase expected yardage in any situation?

Shape Based Clustering



Turn x,y coordinates of every player at every moment into usable receiver routes.





Combine situational data with route information to predict Yards and EPA.



First Two Attempts

Time series clustering and auto-encoding routes worked, but it didn't give us the quality of insights we were hoping for.





Examples of auto-encoded routes

Time series clusters for one game

Shape-Based Clustering



Shape-Based Clustering: Example Routes

10 Yard Crossing Route



RB Out Route

Shape Based Clustering

Odell Beckham



Rob Gronkowski



Ezekiel Elliott





Odell Beckham's 5 most common routes





0.10

Percent Run

0.15

0.20

0.05

0.05

0.00

Route Grou

C

Route Group

Curl In (Left Side, 10Yds)

Robb Gronkowski's 6 most common routes

Double Model Approach



Likelihood of Completion

Accuracy 71%

AUC .75



Yards Gained Given Completion

Cor .51 RMSE 10.0

Situational Variables

- Seconds Remaining in Game
- Yard Line
- Down and Distance
- Score Difference
- Offensive Formation
- # of Pass Rushers
- Quarterback

Engineered Variables

- The routes run on the play
- Position (WR,TE,etc...) of the player running the route

Important Variables

Routes are much more important than the Quarterback at predicting play success

Completion % Important Vars

- Yard Line
- Seconds Remaining in Game
- Score Difference
- Number of Pass Rushers
- Route Groups
- ... x65!
- Matt Ryan

Yards Given Completion Important Vars

- Yard Line
- Score Difference
- Seconds Remaining in Game
- Route Groups
- ... x16
- EJ Manuel

Importance of Routes on Predicted Yards

Broncos vs Bills



Ravens vs Raiders



Optimizing Routes to Improve Yardage



Change TE (82) from "blocking" to "hitch"

Predicted Yards +1.65

4.73 → 6.38



Run Short 5 Yard Hitches



Good Routes >>>> Good Quarterbacks

Top 20 Most Important Factors by Model



Thank You

Jake Flancer - *jflancer@wharton.upenn.edu, @jakef1873*

Jack Soslow - jsoslow2@gmail.com, @jack_soslow

Andrew Castle - *castla@wharton.upenn.edu, @AndrewCastle510*

Eric Dong - *ericdong@seas.upenn.edu*

Special thanks to Professor Abraham J. Wyner of the Wharton School for his advice and assistance, and to Michael Lopez and Jay Reid.

References:

- 1. Keogh, Eamonn, and Jessica Lin. "Clustering of time-series subsequences is meaningless: implications for previous and future research." Knowledge and information systems 8.2 (2005): 154-177.
- 2. Steinbach, Michael, Levent Ertöz, and Vipin Kumar. "The challenges of clustering high dimensional data." New directions in statistical physics. Springer, Berlin, Heidelberg, 2004. 273-309.
- 3. Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." Proceedings of the 25th international conference on Machine learning. ACM, 2008.



Appendix

First Two Modeling Attempts

Modeling helps us optimize routes for any situation as well as tells us what variables matter when predicting yards gained



Variable Importance Plot for XGboost

Neural Net Diagnostic Charts

Model One - Completion % Anova

Analysis of Variance Table

Response: completion_indicator

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
offenseFormation	6	1.48	0.2469	1.2740	0.265630	
defendersInTheBox	1	0.15	0.1531	0.7901	0.374121	
numberOfPassRushers	1	9.68	9.6813	49.9471	1.856e-12	***
scoreDiff	1	1.78	1.7795	9.1807	0.002462	**
yardline_100	1	5.90	5.8980	30.4289	3.683e-08	***
<pre>game_seconds_remaining</pre>	1	0.22	0.2163	1.1161	0.290832	
passer_player_name	40	12.73	0.3181	1.6414	0.006787	**
Route_Group.x	29	88.79	3.0619	15.7965	< 2.2e-16	***
Route_Group.y	29	34.68	1.1959	6.1699	< 2.2e-16	***
Route_Group.x.x	29	13.82	0.4764	2.4580	2.289e-05	***
Route_Group.y.y	29	23.97	0.8266	4.2646	2.080e-13	***
Route_Group	29	16.55	0.5708	2.9446	2.266e-07	***
PositionAbbr.x	2	0.08	0.0410	0.2115	0.809387	
PositionAbbr.y	2	0.35	0.1764	0.9103	0.402494	
PositionAbbr.x.x	2	0.00	0.0017	0.0089	0.991151	
PositionAbbr.y.y	2	0.48	0.2420	1.2486	0.287030	
PositionAbbr	2	0.95	0.4773	2.4626	0.085346	
Residuals	3984	772.22	0.1938			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model Two - Yards Given Completion Anova

Analysis of Variance Table

Response: PlayResult

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
offenseFormation	5	1597	319.34	4.4148	0.0005266	***
defendersInTheBox	1	29	29.43	0.4068	0.5236475	
numberOfPassRushers	1	798	798.13	11.0338	0.0009080	***
scoreDiff	1	158	157.50	2.1774	0.1401812	
yardline_100	1	2934	2934.46	40.5677	2.266e-10	***
<pre>game_seconds_remaining</pre>	1	94	93.80	1.2967	0.2549294	
passer_player_name	40	6201	155.03	2.1432	4.501e-05	***
Route_Group.x	29	30797	1061.98	14.6814	< 2.2e-16	***
Route_Group.y	29	11730	404.50	5.5920	< 2.2e-16	***
Route_Group.x.x	29	5177	178.53	2.4681	2.260e-05	***
Route_Group.y.y	29	7968	274.76	3.7985	4.380e-11	***
Route_Group	29	6389	220.30	3.0456	9.664e-08	***
PositionAbbr.x	2	26	13.23	0.1829	0.8328851	
PositionAbbr.y	2	64	31.96	0.4419	0.6428831	
PositionAbbr.x.x	2	108	53.96	0.7459	0.4744109	
PositionAbbr.y.y	2	122	60.89	0.8418	0.4310547	
PositionAbbr	2	128	63.97	0.8844	0.4131048	
Residuals	2415	174689	72.33			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Starting Positions of Receivers



Important Variables

Routes are much more important than the Quarterback at predicting play success

yardline 100 yardline 100 scoreDiff scoreDiff game seconds remaining game seconds remaining Route Group.x24 numberOfPassRushers Route Group.y9 Route Group.x4 Route_Group.y.y24 defendersInTheBox Route Group24 Route_Group.y.y4 defendersInTheBox Route Group4 numberOfPassRushers Route Group.y7 Route Group.x9 Route Group24 Route Group9 Route Group.y.y8 Route Group.x.x9 Route Group.y.y3 Route Group.x11 Route Group.y.y13 Route Group26 Route Group.y24 Route Group.y11 Route_Group.y.y9 Route Group.x.x24 Route Group.y11 Route Group25 Route Group.y9 Route Group.x6 Route Group.x9 offenseFormationSHOTGUN Route Group9 Route Group.y.y25 Route Group6 0.00 0.02 0.04 0.06 0.08 0.10 0.00 0.02 0.04 0.06 0.08

Yards Given Completion Important Vars

Completion % Important Vars

Final Double Model Diagnostics

