Player Tracking in American Football: Spatio-temporal Modeling of Defensive Players' Intent

Karl Pazdernik

September 23, 2017

Player Value

Which players are the most valuable?

Player Value

Which players are the most valuable?

- Check the stats! I AM!

Player Value

Which players are the most valuable?

- Check the rings! I AM!

Player Tracking Motivation

We need to look beyond the box score.

- Which RB is more valuable?
  1. The one who leads the NFL in rushing yards?
  2. The one who rushes for 200 yards fewer against predominantly 8-man boxes with poor run-blocking?

- Which WR is more valuable?
  1. The one who leads his team in receiving yards?
  2. His teammate who has 100 fewer yards receiving while always matched up against the #1 CB with S help?

Player Tracking Motivation

We need to look beyond the box score.

- Which RB is more valuable?
    1. The one who leads the NFL in rushing yards?
    2. The one who rushes for 200 yards fewer against predominantly 8-man boxes with poor run-blocking?

- Which WR is more valuable?
    1. The one who leads his team in receiving yards?
    2. His teammate who has 100 fewer yards receiving while always matched up against the #1 CB with S help?

- The QB is always the most important...

Player Tracking Motivation

What makes an offense/defense successful?

Player Tracking Motivation

What makes an offense/defense successful?

- Individuals on offense gain separation.
- The defense collectively and continuously covers all offensive players.

Player Tracking Motivation

What makes an offense/defense successful?

- Individuals on offense gain separation.
- The defense collectively and continuously covers all offensive players.

To understand value, we need to model the complex dynamics of defensive coverage.

- Simultaneously determine **who** is following whom and **how** they are being followed.
- Uncover a defender's **intent** and **effectiveness**.

Player Tracking in Other Sports

Advancements in optical tracking and RFID technology have created rich spatio-temporal datasets for various sports and spurred numerous research efforts.

- Basketball
  - xyresearch.com (Alexander D'Amour, Alexander Franks, Andrew Miller, Dan Cervone, Luke Bornn, and Kirk Goldsberry)
  - Annals of Statistics, JASA, MIT Sloan Sports Analytics Conference, etc.
  - Franks et al (2015) "Characterizing the spatial structure of defensive skill in professional basketball." The Annals of Applied Statistics, 9(1), 94-121.
- Soccer
  - Bojinov and Bornn (2016) "The Pressing Game: Optimal Defensive Disruption in Soccer." MIT SSAC
  - Horton et al (2015) "Automated Classification of Passing in Football." Pacific-Asia Conference on Knowledge Discovery and Data Mining, 319-330.
  - Kim et al (2011) "Spatial and spatiotemporal analysis of soccer." 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 385-388.
  - Richly et al (2017) "Utilizing artificial neural networks to detect compound events in spatio-temporal soccer data."

Player Tracking in American Football

Although the National Football League records spatio-temporal data, it is not openly available. Despite this limitation, there has been some progress.

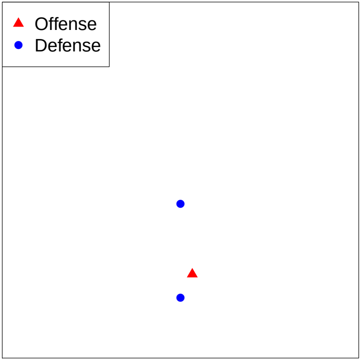
- Hochstedler (2016) “Finding the open receiver: A quantitative geospatial analysis of quarterback decision-making.” MIT Sloan Sports Analytics Conference.
- Hochstedler and Gagnon (2017) “American Football Route Identification Using Supervised Machine Learning.” MIT Sloan Sports Analytics Conference.

Yet to determine individual matchups and coverage types. Yet to characterize the spatial structure.

Location

Location provides a partial explanation.

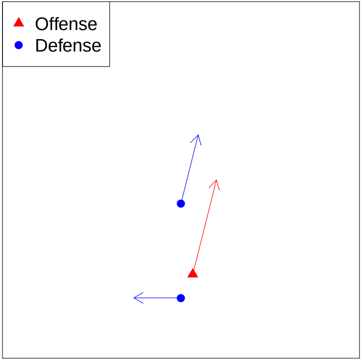
- Defenders are more likely to be "tracking" the closest offensive player.

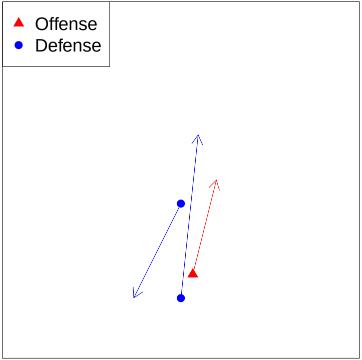


Location

Trajectory helps to complete the picture.

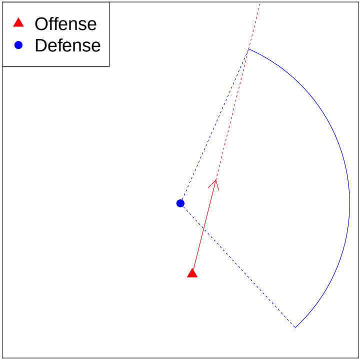
- Need to consider the direction and magnitude of all trajectories.





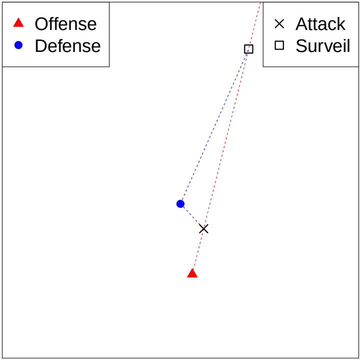
Angles of Intersection

- Given location/speed/direction for an offensive player and location/speed for a defensive player, we can determine where the players' paths may cross.
  - Assume speed and direction are constant for the specified time interval



Angles of Intersection

- Consider two motion types:
  - Attack: rushing the passer, closing in on ball carrier, anticipating a pass, catching up to an open receiver
  - Surveil: zone defense, safety help, off-coverage



Hidden Markov Model

- Since locations and trajectories are easily measured but player assignments are not, this lends itself to a state-space model with latent variables.
- As in Franks et al (2015), we use a Hidden Markov Model (HMM)
  - Conditional independence allows for simple/efficient parameter estimation and confidence interval computation (MLE)
  - HMM construct includes assignment probabilities
  - Can easily be expanded to include multiple states per offensive player (attack and surveil)
- Note: numeric corrections are necessary to account for sidelines and slow or stationary defenders.

Transition Probabilities

- $I_{tjk\ell}$ = latent indicator variable that defender j is following offensive player k with motion type ℓ at time t
- k^* = offensive player at time $t-1$
- ℓ^* = player motion type at time $t-1$

$$P(I_{tjk\ell} = 1 \mid I_{(t-1)jk^*\ell^*} = 1) = \begin{cases} pq & k = k^*, \ell = \ell^* \\ p(1-q) & k = k^*, \ell \neq \ell^* \\ c(1-p)q & k \neq k^*, \ell = \ell^* \\ c(1-p)(1-q) & k \neq k^*, \ell \neq \ell^* \end{cases}$$

- p = probability that a defender maintains coverage on the same opponent
- q = probability that a defender maintains current coverage type (attack or surveil)
- c = normalizing constant ($\frac{1}{5}$ assuming 5 eligible receivers and a quarterback)

Statistical Model

- $\boldsymbol{D}_{tj}$ = defender j's location at time t
- $\boldsymbol{O}_{tk}$ = offensive player k's location at time t
- θ_{tj} = angle of trajectory of defender j at time t
- $\theta_{tjk\ell}$ = optimal angle for defender j to take which will intersect offensive player k using a motion type ℓ, given vectors at time t

$$\begin{aligned} \boldsymbol{D}_{tj} \mid I_{tjk\ell} = 1 &\sim \text{MVN}(\boldsymbol{O}_{tk} + \boldsymbol{\mu}_j(t), \boldsymbol{\Sigma}_\ell) \\ \theta_{tj} \mid \boldsymbol{D}_{tj}, I_{tjk\ell} = 1 &\sim \text{vonMises}(\theta_{tjk\ell} + \phi, \eta_\ell) \end{aligned}$$

- Parameters = $\{\boldsymbol{\mu}_j(t), \boldsymbol{\Sigma}_\ell, \phi, \eta_\ell, p, q : 1 \le j \le 11, 1 \le \ell \le 2\}$

Tracking Summary Statistics

- Offensive Attention (player):
  - Average number of defensive players tracking any offensive player

$$OA_k = \frac{1}{T}\sum_{tj\ell} P(I_{tjk\ell} = 1 \mid \boldsymbol{D}, \boldsymbol{\theta}) \tag{1}$$

Let $\hat{k}$ be the offensive player who eventually receives the ball and $\hat{t}$ be the time point at which that is decided.

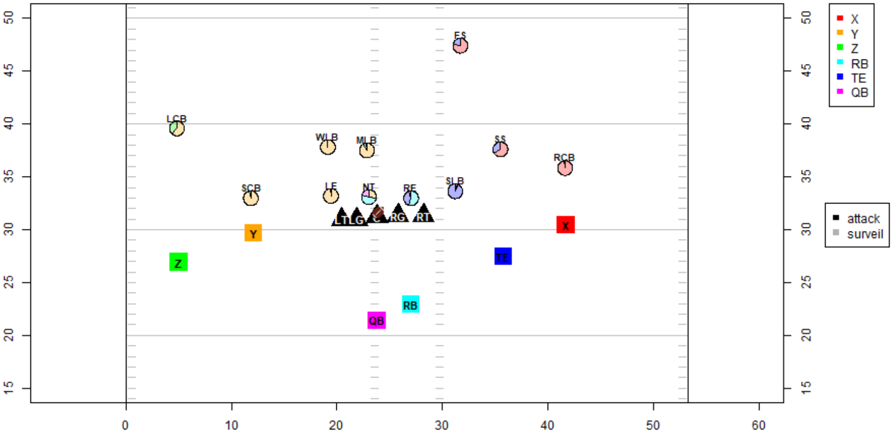
- Swarming Defense (team):
  - Average number of defensive players tracking the eventual ball carrier

$$SD = \frac{1}{T - \hat{t}}\sum_{t>\hat{t}}\sum_{j\ell} P(I_{tj\hat{k}\ell} = 1 \mid \boldsymbol{D}, \boldsymbol{\theta}) \tag{2}$$

- Defensive Instincts (player):
  - Switched to the eventual ball carrier.
  - Removes those covering the ball carrier by initial assignment.

$$DI_j = \min_t\{t - \hat{t} \mid \hat{k} \neq \underset{k}{\operatorname{argmax}}\sum_{\ell=1}^{2} P(I_{(t-1)jk\ell} = 1), \hat{k} = \underset{k}{\operatorname{argmax}}\sum_{\ell=1}^{2} P(I_{tj\hat{k}\ell} = 1)\} \tag{3}$$

Simulation - Sack



Simulation: Summary Statistics

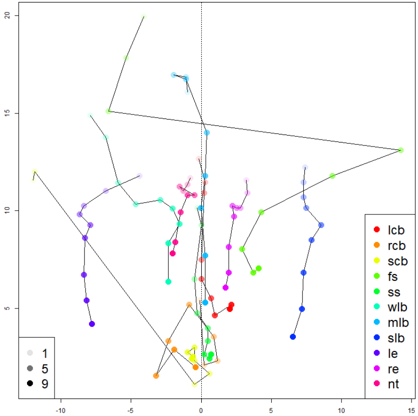
Table 1: Transition Probability Estimates

	Lower 95%	MLE	Upper 95%
p	0.978	1.000	
q	0.737	0.987	
pq	0.736	0.987	
$p(1-q)$		0.013	0.263

Table 2: Offensive Attention and Defensive Summary Statistics

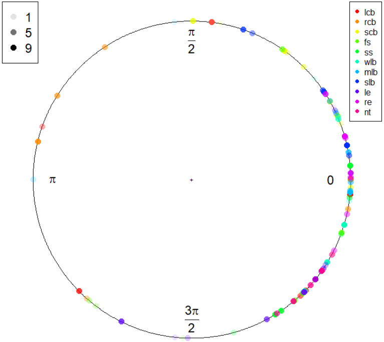
X	Y	Z	RB	TE	QB	SD	DI
1.00	1.43	1.00	1.00	1.35	5.22	5.13	NA

Simulation: Location Offset



Player specific offsets and $\mu_y \to 0$ as $t \to T$

Simulation: Angle Offset



RCB moves in the opposite direction almost half of the time.

Super Bowl LI

Super Bowl LI provided the greatest comeback in Super Bowl history.

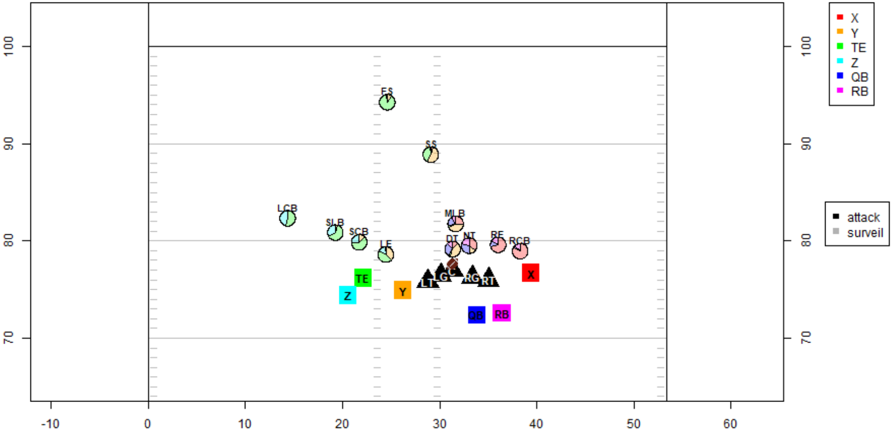
- NE 3 - ATL 28, 8:31 remaining in the 3rd quarter
- NE 34 - ATL 28, final score

What information can we extract using player tracking data from the most influential plays?

- Difference in win probabilities (WP) relative to NE provided by `nflscrapR` (Horowitz & Yurko)

ATL Int-to-TD: WP ≈ -10%

NE 3 - ATL 21, Qtr 2, 2:36



ATL Int-to-TD: Summary Statistics

Table 3: Transition Probability Estimates

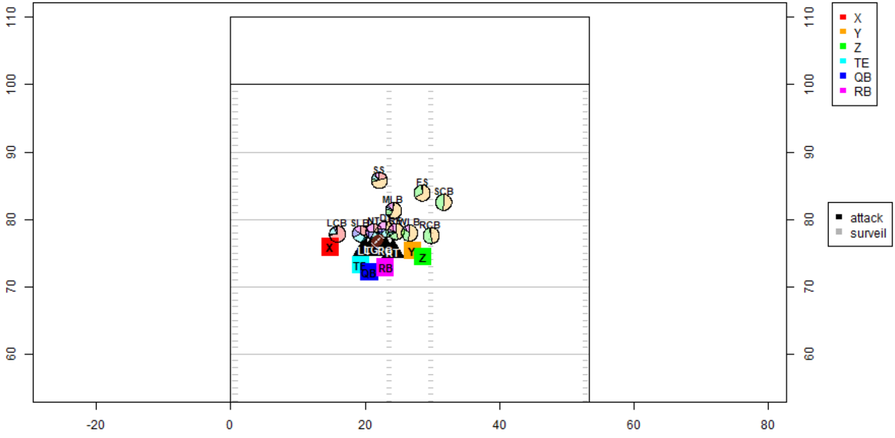
	Lower 95%	MLE	Upper 95%
p	0.853	0.993	
q	0.685	0.969	
pq	0.649	0.962	
$p(1-q)$		0.031	0.313

Table 4: Offensive Attention and Defensive Summary Statistics

X	Y	TE	Z	QB	RB	SD	DI_{LCB}
0.05	1.82	0.95	2.55	4.17	1.47	2.40	-0.8s

Matt Ryan 4th qtr Sack: WP ≈ +3%

NE 20 - ATL 28, Qtr 4, 3:56



Matt Ryan 4th qtr Sack: Summary Statistics

Table 5: Transition Probability Estimates

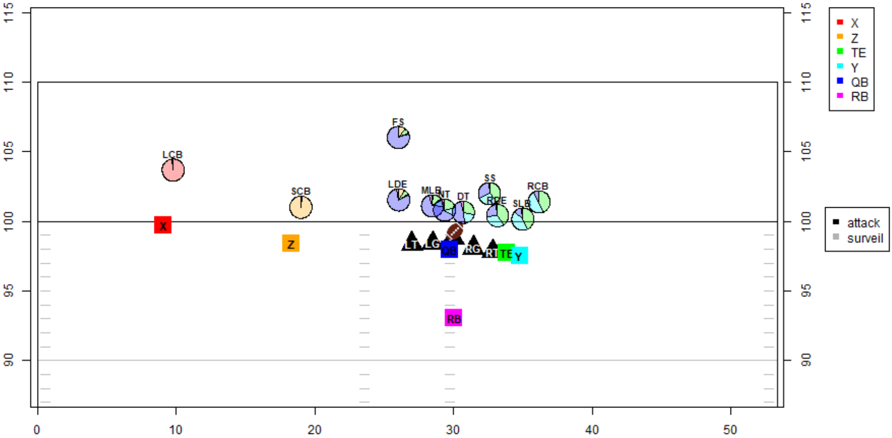
	Lower 95%	MLE	Upper 95%
p	0.918	0.995	
q	0.922	0.995	
pq	0.884	0.990	
$p(1-q)$		0.005	0.078

Table 6: Offensive Attention and Defensive Summary Statistics

X	Y	Z	TE	QB	RB	SD	DI_{MLB}
1.85	0.27	2.04	1.29	4.20	1.34	4.23	1.4s

James White 4th qtr TD: WP ≈ +23%

NE 20 - ATL 28, Qtr 4, 1:00



James White 4th qtr TD: Summary Statistics

Table 7: Transition Probability Estimates

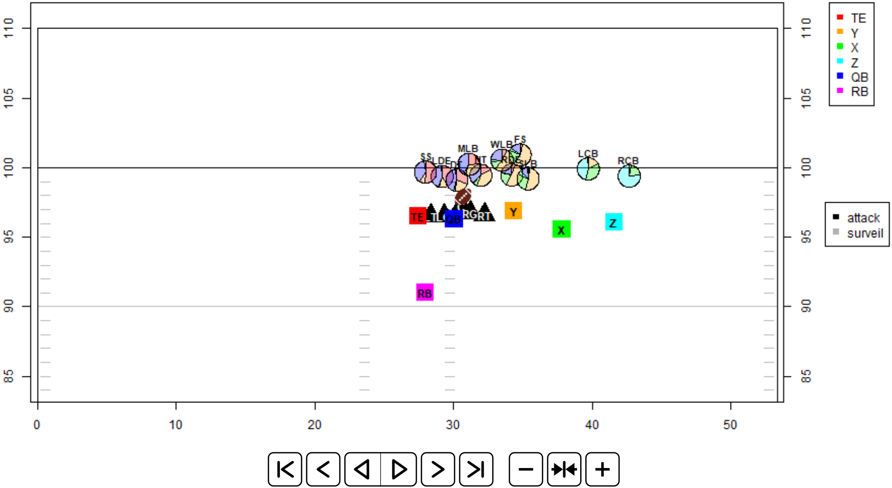
	Lower 95%	MLE	Upper 95%
p	0.842	0.988	
q	0.889	0.994	
pq	0.804	0.981	
$p(1-q)$		0.006	0.109

Table 8: Offensive Attention and Defensive Summary Statistics

X	Z	TE	Y	QB	RB	SD	DI
0.67	1.15	1.44	5.00	2.74	0.00	0.00	NA

James White OT TD: WP ≈ +14%

NE 28 - ATL 28, OT, 1:00



James White OT TD: Summary Statistics

Table 9: Transition Probability Estimates

	Lower 95%	MLE	Upper 95%
p	0.852	0.986	
q	0.856	0.987	
pq	0.788	0.973	
$p(1-q)$		0.013	0.142

Table 10: Offensive Attention and Defensive Summary Statistics

TE	Y	X	Z	QB	RB	SD	DI_{LCB}
0.82	2.91	3.52	0.50	1.33	1.92	2.28	2.6s

Conclusion

- Player tracking data for American football provides a path towards improved metrics and a better understanding of the complex motion on the field.

- Our methodology allows us to
  - Estimate player tracking probabilities
  - Differentiate between an attacking or surveillance motion
  - Quantify individual offensive impact
  - Quantify team hustle on defense
  - Assess individual defensive instincts

- Challenges include data veracity and limited volume

Thank you

- Thank you, Mark Glickman, Scott Evans, and Harvard for hosting!
- Thank you to Jacques Kvam and Vanessa Pazdernik for support and aiding in the data collection process.

For more information, visit **DeepFootball.com**

Questions?