

Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights

Joel Brooks Matthew Kerr John Guttag

Massachusetts Institute of Technology



Motivation

- Teams this year in La Liga are spending over \$3 billion on their players
- How do teams evaluate a player's worth?
- Quantitative metrics can help justify subjective evaluations, and provide new insight into non-obvious player contributions



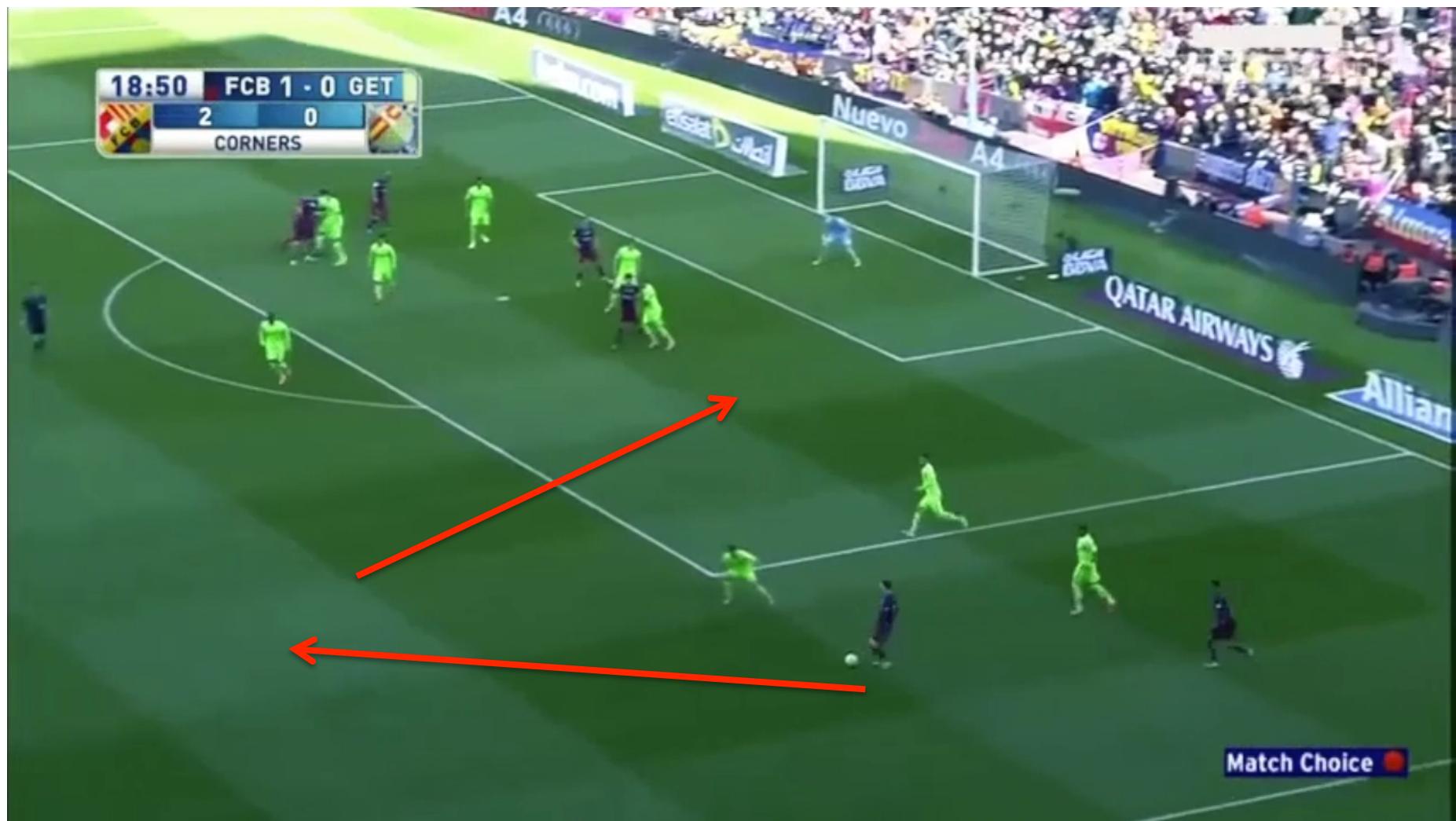
Image source: http://espn.go.com/sports/soccer/news/_/id/5580467/european-football-eating-itself

Passing Contributions on Offense

- Passing strategy is a key component of overall offensive success
- We focused specifically on the contribution of specific passes on offense
- Passes tell you a lot about how much an individual player contributes on offense outside of goals and assists

81:09 FCB 6 - 0 GET

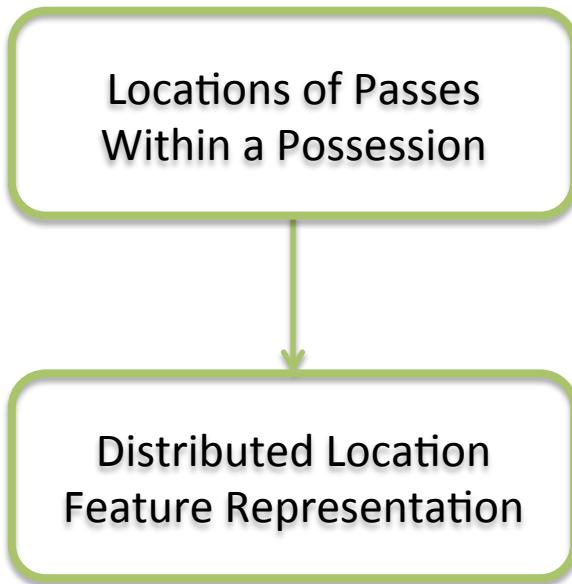
Match Choice



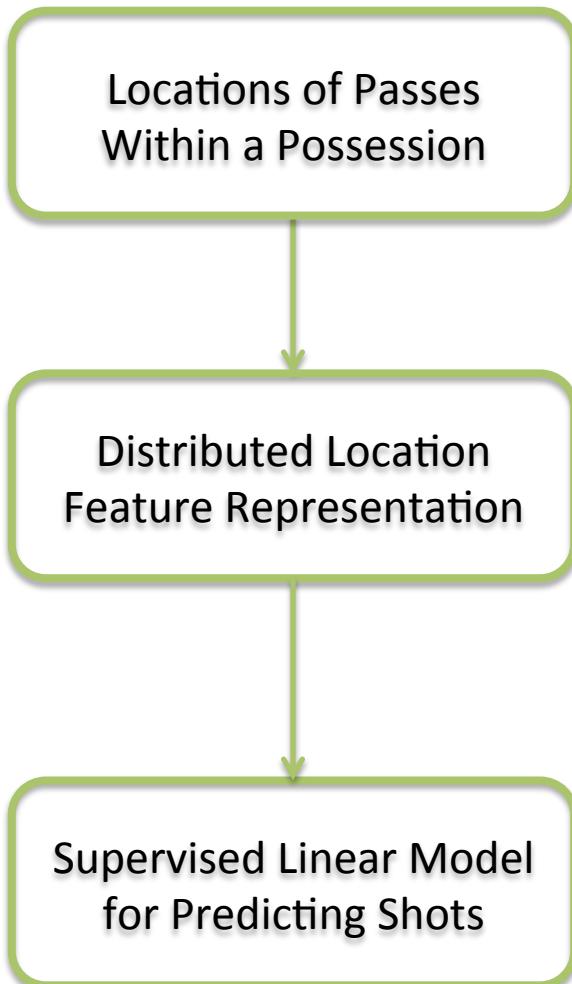
How to rank players from passes?

Locations of Passes
Within a Possession

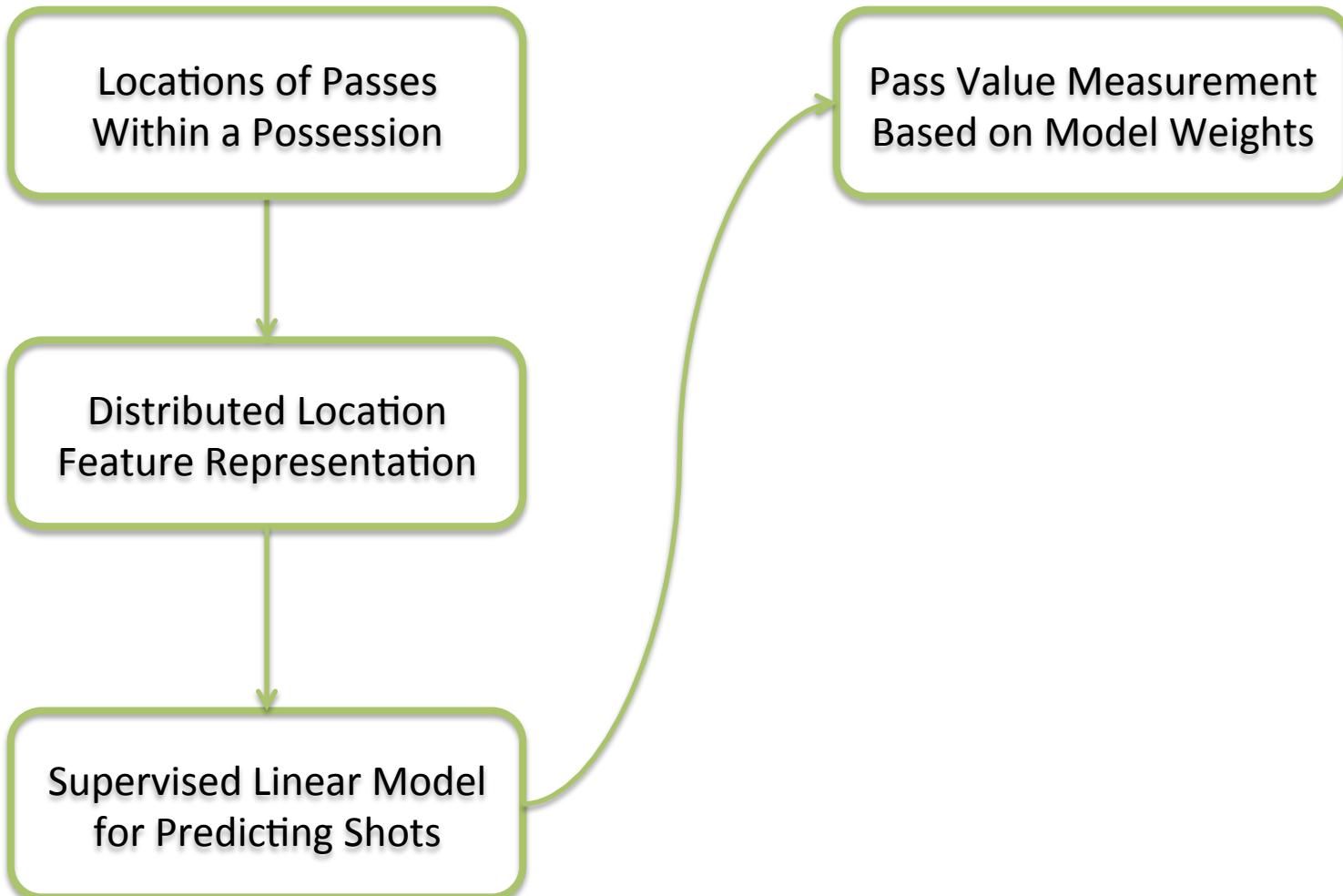
How to rank players from passes?



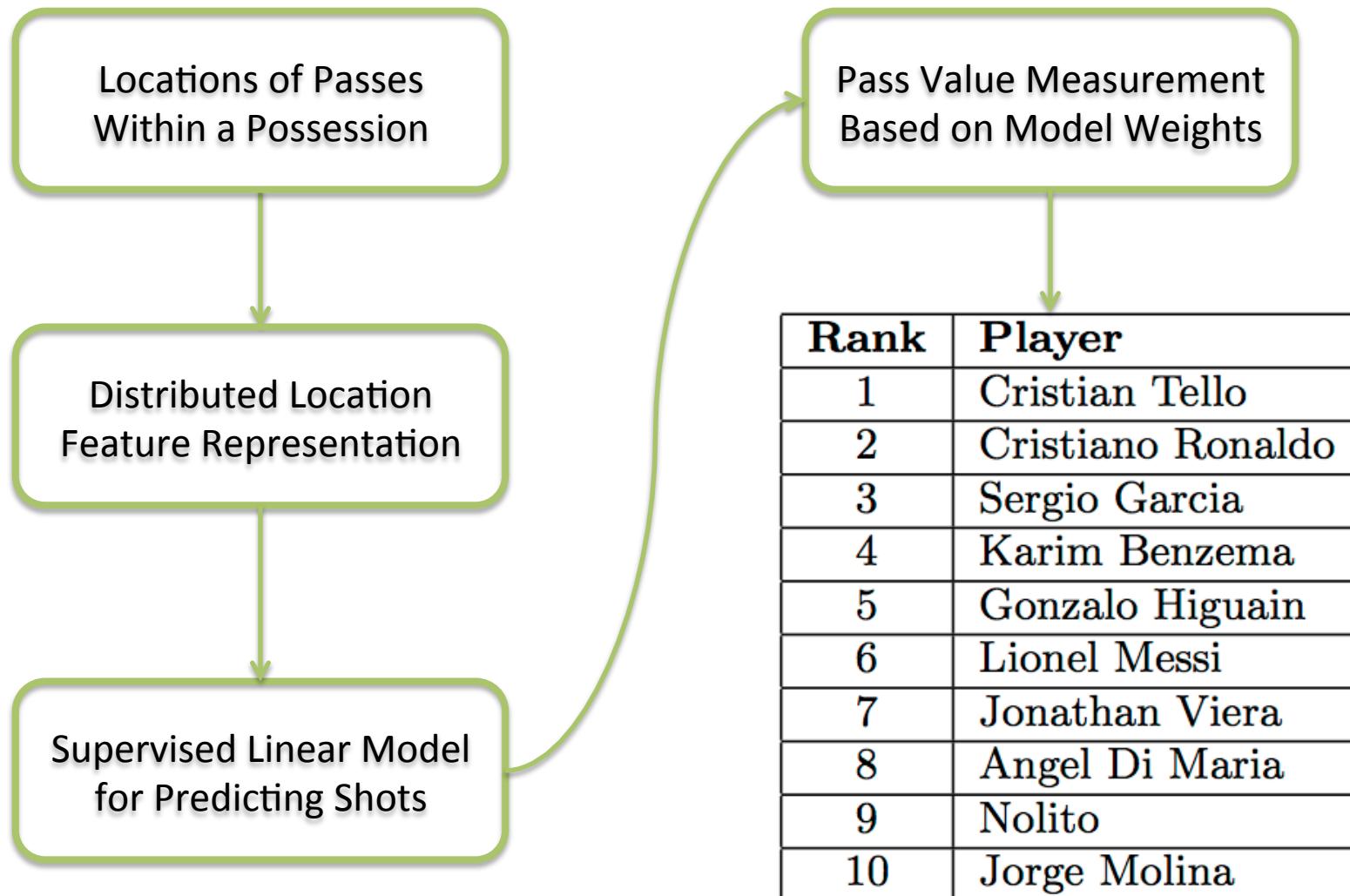
How to rank players from passes?



How to rank players from passes?



How to rank players from passes?



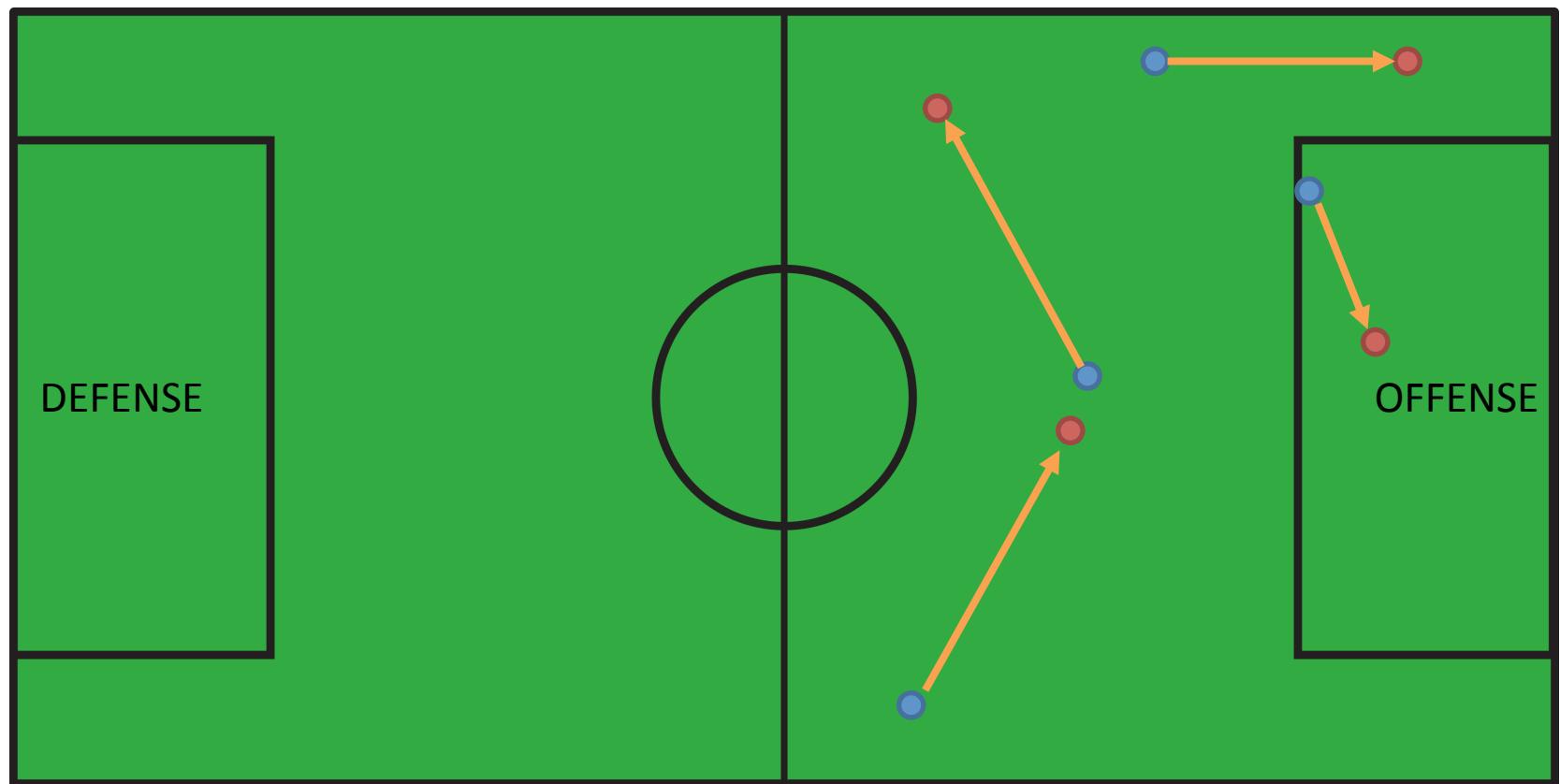
The data

- (x, y) coordinates of all pass origins and destinations from the 2012-2013 La Liga season
- > 300,000 passes
- 380 games
- > 500 players

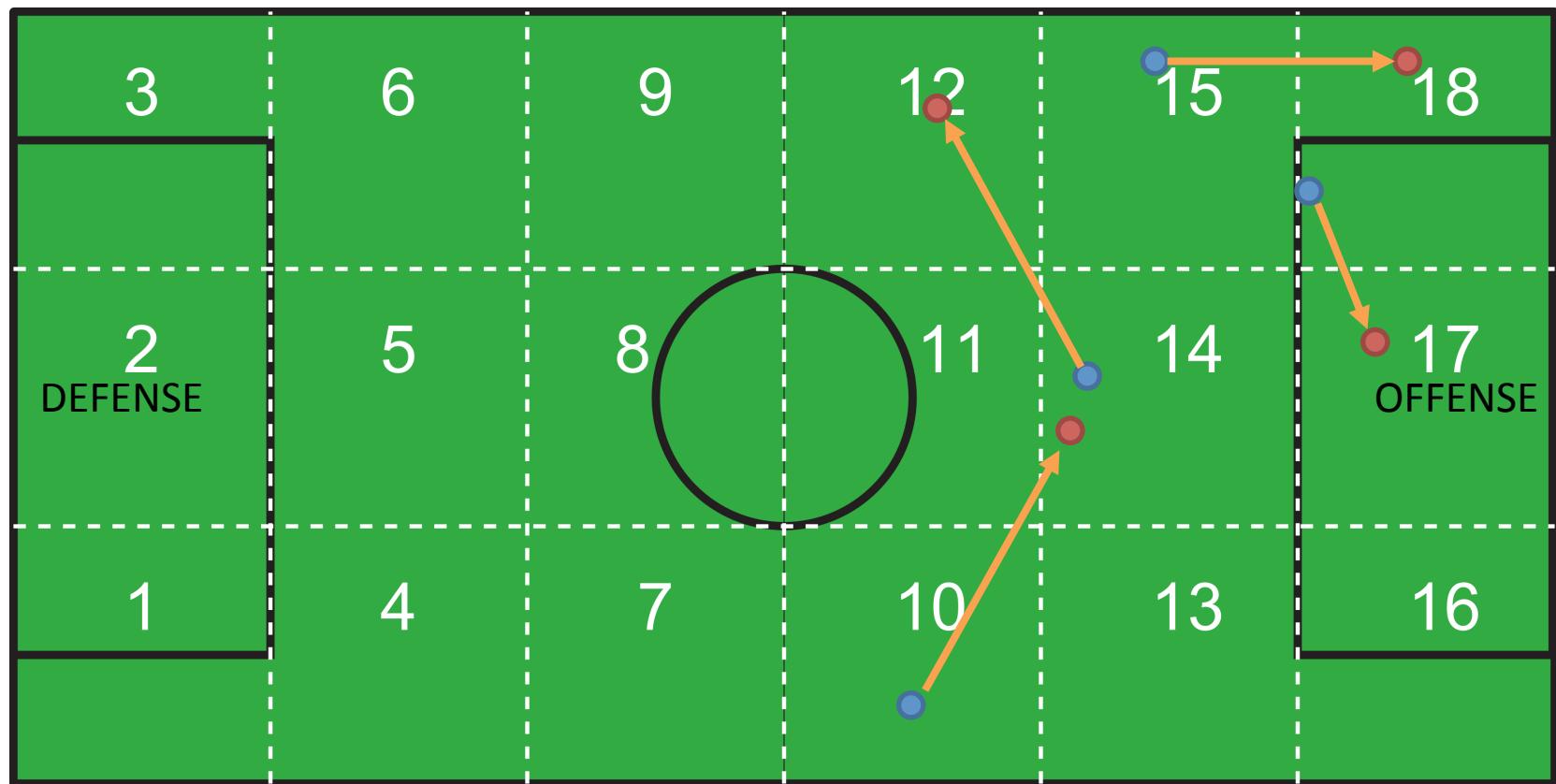


Image source: <https://flic.kr/p/nvVaHM>

Pass Location Representation



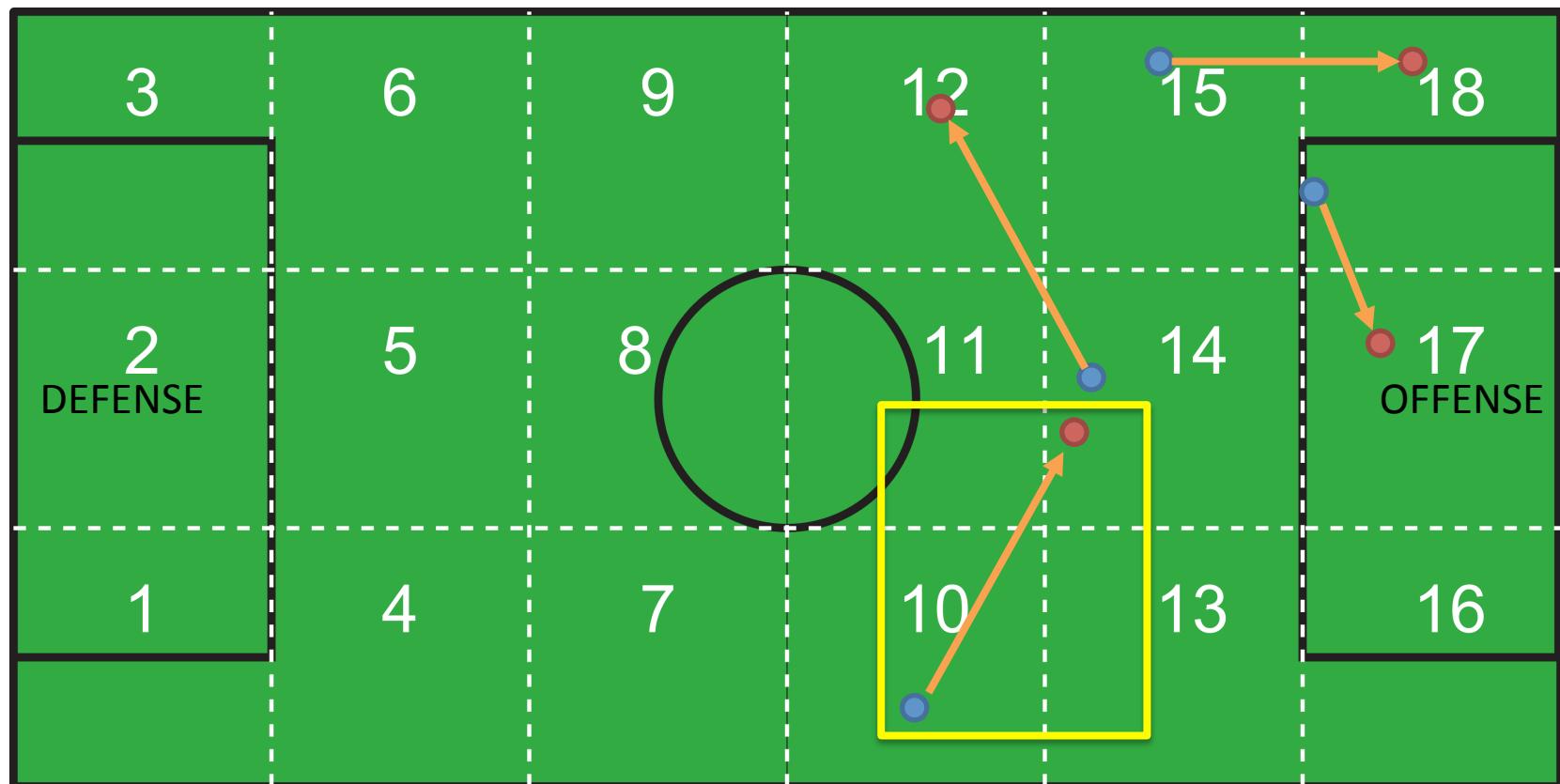
Pass Location Representation



Sparse Pass Location Representation

Origin (zone 10) = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]

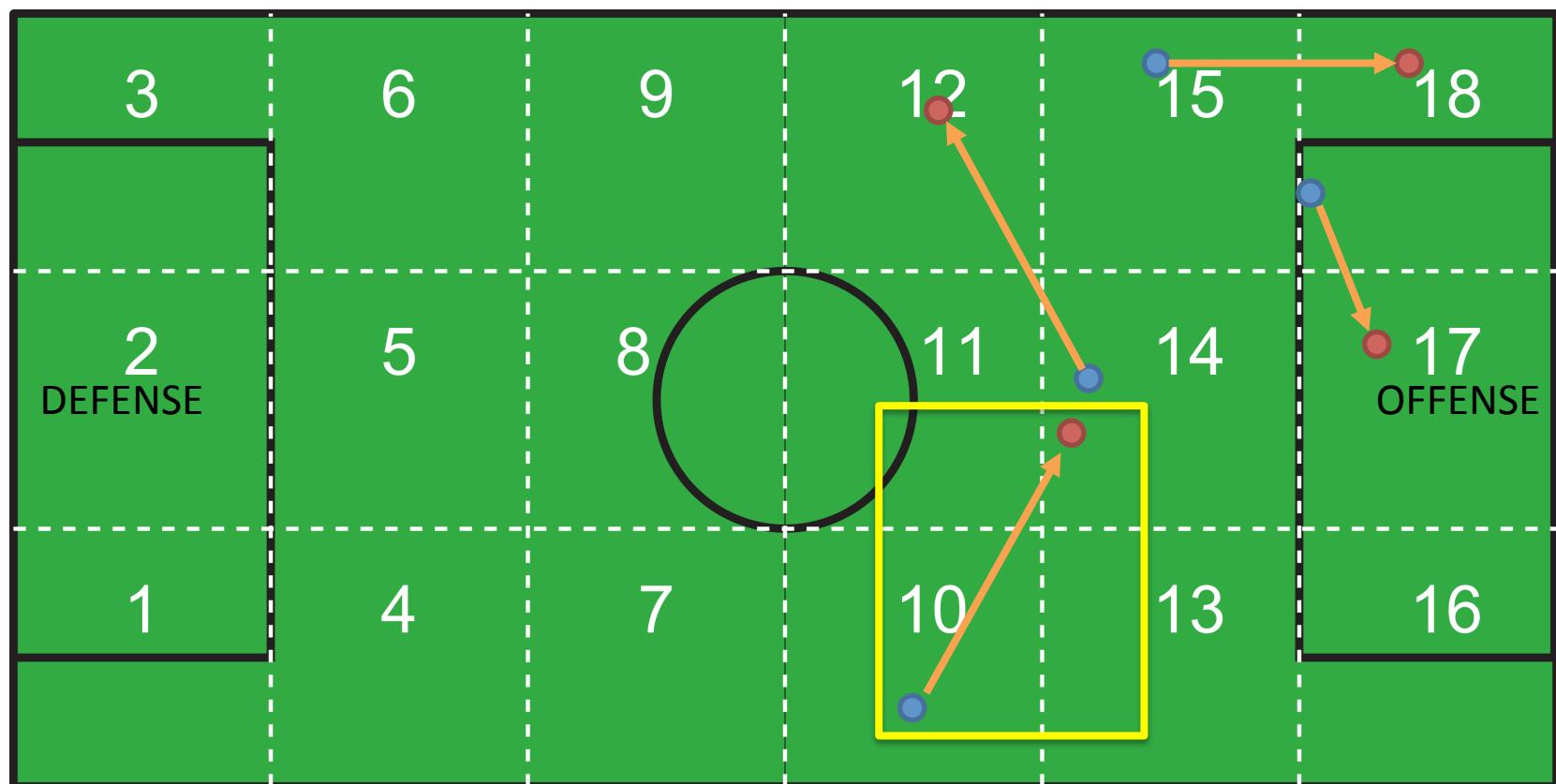
Destination (zone 14) = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]



Dense Pass Location Representation

Origin (zone 10) = [0, 0, 0, 0, 0, 0, 0.26, 0, 0, 0.74, 0, 0, 0, 0, 0, 0, 0]

Destination (zone 14) = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0.48, 0, 0, 0.52, 0, 0, 0, 0]



Pass Location Representation Formula

- Represent a pass location l as:

$$\mathbf{r}^l = [r_1^l \dots r_{18}^l]$$

- Each element of r is:

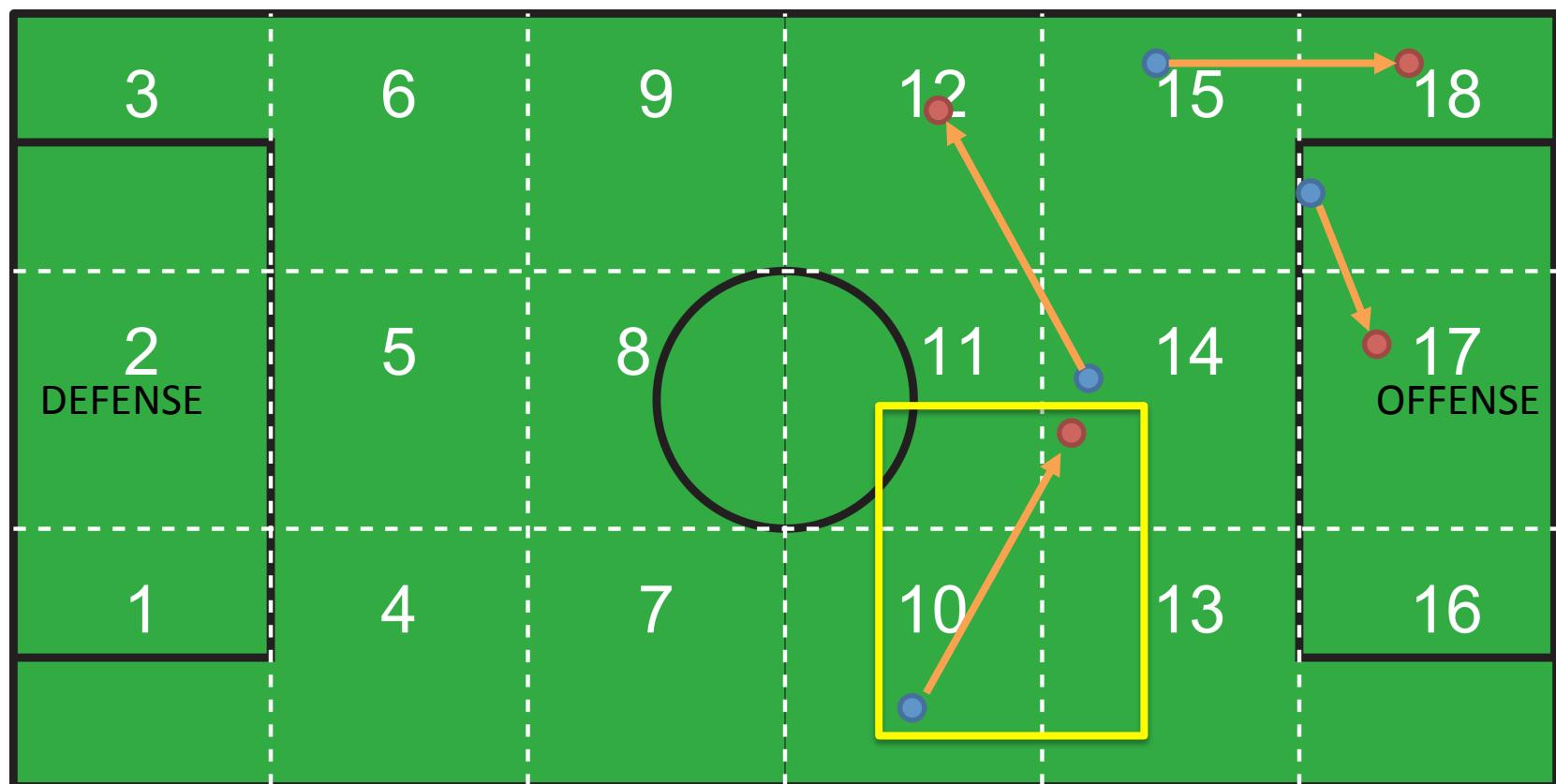
$$r_i^l = \frac{c_i}{\max(d(l, z_i), 1)}$$

- $d(l, z_i)$ is the Euclidean distance between l and the center of zone i
- c_i is an indicator variable that is 1 if i is one of the N closest zones, 0 otherwise
- In practice $N = 2$ seem to lead to the best results

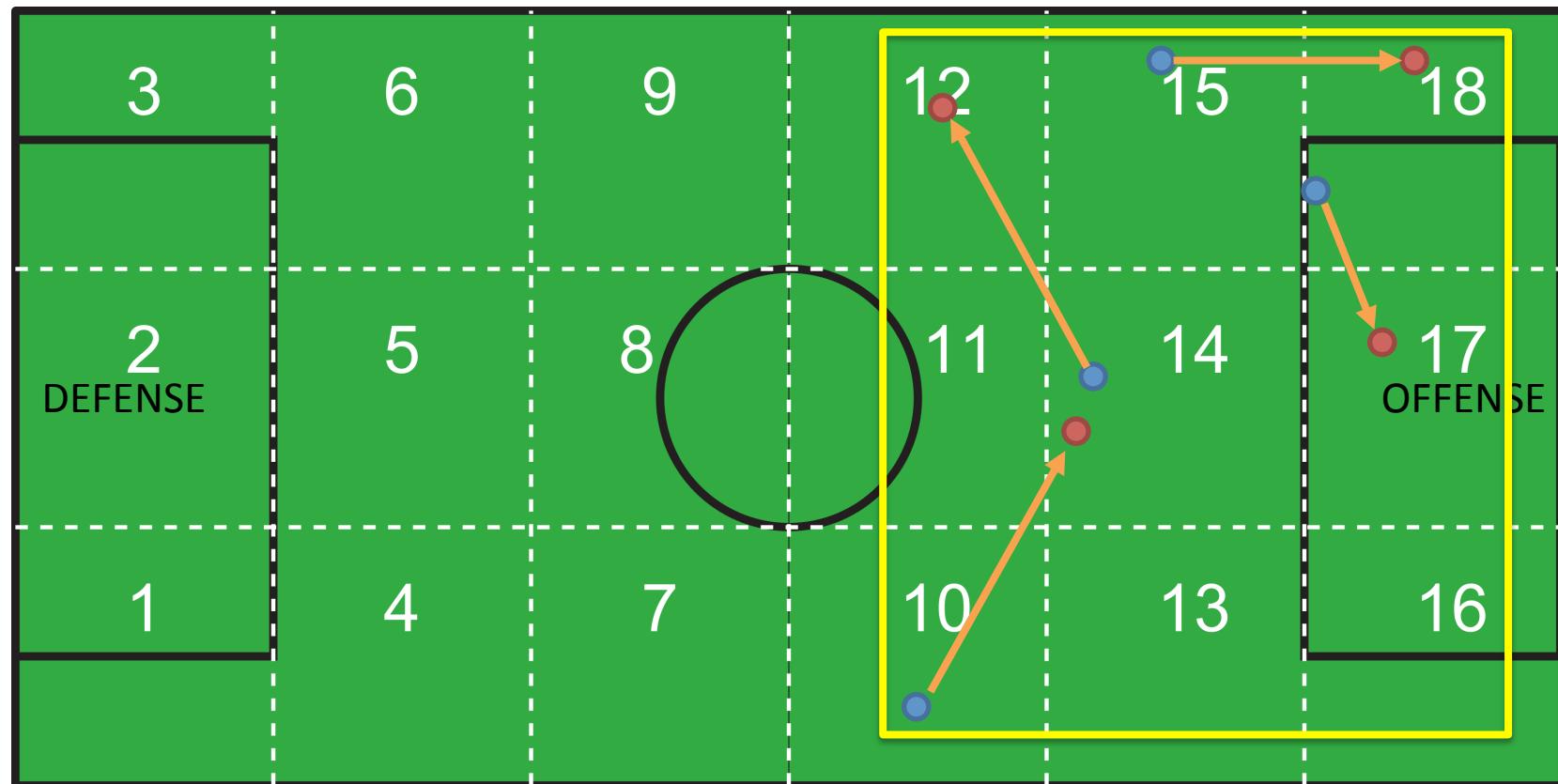
Dense Pass Location Representation

Origin (zone 10) = [0, 0, 0, 0, 0, 0, **0.26**, 0, 0, **0.74**, 0, 0, 0, 0, 0, 0, 0]

Destination (zone 14) = [0, 0, 0, 0, 0, 0, 0, 0, 0, **0.48**, 0, 0, **0.52**, 0, 0, 0, 0, 0]



How do we represent the location of a collection of passes?

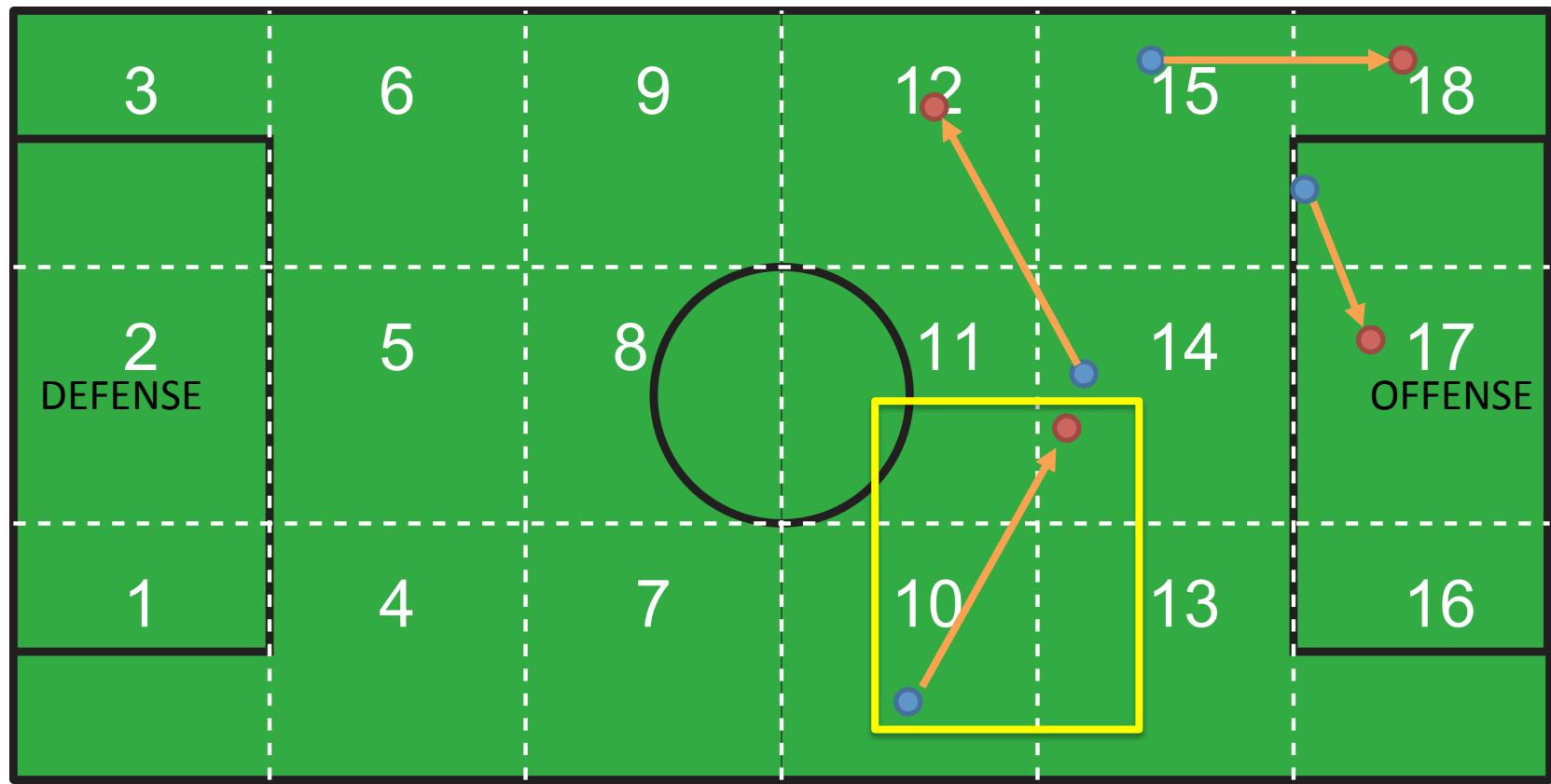


Feature Vector for a Possession

- For each pass in the possession with an origin I_o and destination I_d :

Feature Vector for a Possession

- For each pass in the possession with an origin I_o and destination I_d :
 1. Compute \mathbf{r}^{lo} and \mathbf{r}^{ld} , the vector representations of the pass origin and destination



Feature Vector for a Possession

- For each pass in the possession with an origin I_o and destination I_d :
 1. Compute \mathbf{r}^{lo} and \mathbf{r}^{ld} , the vector representations of the pass origin and destination

$$\begin{array}{lcl} \text{Origin (zone 10)} & = [0, 0, 0, 0, 0, 0, \\ & \textcolor{blue}{0.26}, 0, 0, \textcolor{blue}{0.74}, 0, 0, 0, 0, 0, 0, 0] \end{array}$$

$$\begin{array}{lcl} \text{Destination (zone 14)} & = [0, 0, 0, 0, 0, 0, 0, \\ & 0, 0, 0, \textcolor{red}{0.48}, 0, 0, \textcolor{red}{0.52}, 0, 0, 0, 0] \end{array}$$

Feature Vector for a Possession

- For each pass in the possession with an origin I_o and destination I_d :
 1. Compute \mathbf{r}^{lo} and \mathbf{r}^{ld} , the vector representations of the pass origin and destination
 2. Compute the matrix $\mathbf{R}^{lod} = \mathbf{r}^{lo} \otimes \mathbf{r}^{ld}$, the outer product of the origin and destination representations

$$\begin{array}{lcl} \text{Origin (zone 10)} & = [0, 0, 0, 0, 0, 0, \\ & \textcolor{blue}{0.26}, 0, 0, \textcolor{blue}{0.74}, 0, 0, 0, 0, 0, 0, 0] \end{array}$$

$$\begin{array}{lcl} \text{Destination (zone 14)} & = [0, 0, 0, 0, 0, 0, 0, \\ & 0, 0, 0, \textcolor{red}{0.48}, 0, 0, \textcolor{red}{0.52}, 0, 0, 0, 0] \end{array}$$

Feature Vector for a Possession

- For each pass in the possession with an origin l_o and destination l_d :
 - Compute \mathbf{r}^{lo} and \mathbf{r}^{ld} , the vector representations of the pass origin and destination
 - Compute the matrix $\mathbf{R}^{lod} = \mathbf{r}^{lo} \otimes \mathbf{r}^{ld}$, the outer product of the origin and destination representations

$$\begin{array}{ll} \text{Origin (zone 10)} & = [0, 0, 0, 0, 0, 0, \\ \textcolor{blue}{0.26}, 0, 0, \textcolor{blue}{0.74}, 0, 0, 0, 0, 0, 0, 0] \end{array}$$

$$\begin{array}{ll} \text{Destination (zone 14)} & = [0, 0, 0, 0, 0, 0, 0, \\ 0, 0, 0, \textcolor{red}{0.48}, 0, 0, \textcolor{red}{0.52}, 0, 0, 0, 0] \end{array}$$

Origin-Destination Outer Product (10, 14) =

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	.32	0	0

⋮

...

Feature Vector for a Possession

- For each pass in the possession with an origin l_o and destination l_d :
 - Compute \mathbf{r}^{lo} and \mathbf{r}^{ld} , the vector representations of the pass origin and destination
 - Compute the matrix $\mathbf{R}^{lod} = \mathbf{r}^{lo} \otimes \mathbf{r}^{ld}$, the outer product of the origin and destination representations

$$\begin{array}{ll} \text{Origin (zone 10)} & = [0, 0, 0, 0, 0, 0, \\ & \textcolor{blue}{0.26}, 0, 0, \textcolor{blue}{0.74}, 0, 0, 0, 0, 0, 0] \end{array}$$

$$\begin{array}{ll} \text{Destination (zone 14)} & = [0, 0, 0, 0, 0, 0, 0, \\ & 0, 0, 0, \textcolor{red}{0.48}, 0, 0, \textcolor{red}{0.52}, 0, 0, 0, 0] \end{array}$$

Origin-Destination Outer Product (10, 14) =

$$\mathbf{R}(7, 11) = \textcolor{blue}{0.26} * \textcolor{red}{0.48} = \textcolor{blue}{0.12}$$

$$\mathbf{R}(7, 14) = \textcolor{blue}{0.26} * \textcolor{red}{0.52} = \textcolor{blue}{0.14}$$

$$\mathbf{R}(10, 11) = \textcolor{blue}{0.74} * \textcolor{red}{0.48} = \textcolor{blue}{0.36}$$

$$\mathbf{R}(10, 14) = \textcolor{blue}{0.74} * \textcolor{red}{0.52} = \textcolor{blue}{0.38}$$

Feature Vector for a Possession

- For each pass in the possession with an origin I_o and destination I_d :
 - Compute \mathbf{r}^{lo} and \mathbf{r}^{ld} , the vector representations of the pass origin and destination
 - Compute the matrix $\mathbf{R}^{lod} = \mathbf{r}^{lo} \otimes \mathbf{r}^{ld}$, the outer product of the origin and destination representations
 - Construct the feature vector as:
 $[\mathbf{r}^{lo}, \mathbf{r}^{ld}, \text{flatten}(\mathbf{R}^{lod})]$

Origin (zone 10) = [0, 0, 0, 0, 0, 0,
0.26, 0, 0, **0.74**, 0, 0, 0, 0, 0, 0, 0]

Destination (zone 14) = [0, 0, 0, 0, 0, 0, 0,
0, 0, 0, **0.48**, 0, 0, **0.52**, 0, 0, 0, 0]

Origin-Destination Outer Product (10, 14) =

$$\mathbf{R}(7, 11) = \mathbf{0.26} * \mathbf{0.48} = \mathbf{0.12}$$

$$\mathbf{R}(7, 14) = \mathbf{0.26} * \mathbf{0.52} = \mathbf{0.14}$$

$$\mathbf{R}(10, 11) = \mathbf{0.74} * \mathbf{0.48} = \mathbf{0.36}$$

$$\mathbf{R}(10, 14) = \mathbf{0.74} * \mathbf{0.52} = \mathbf{0.38}$$

Feature Vector for a Possession

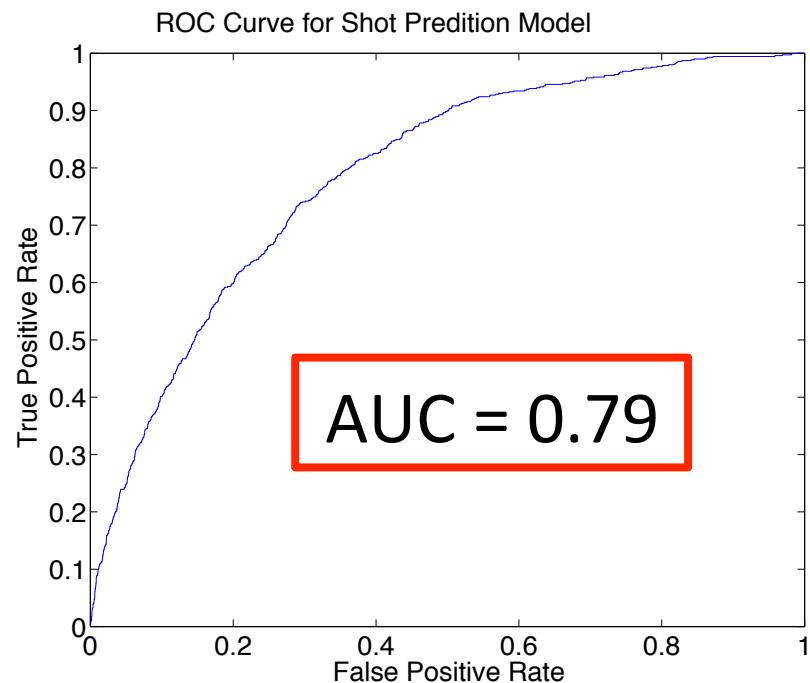
- 18 origin + 18 destination + 324 origin-destination pair features = **360 features**
- The feature vector for a possession is the average of the feature vectors for each individual pass
- Each feature vector is assigned a label:
 - +1 if the possession ended in a shot
 - -1 otherwise

Experimental Overview

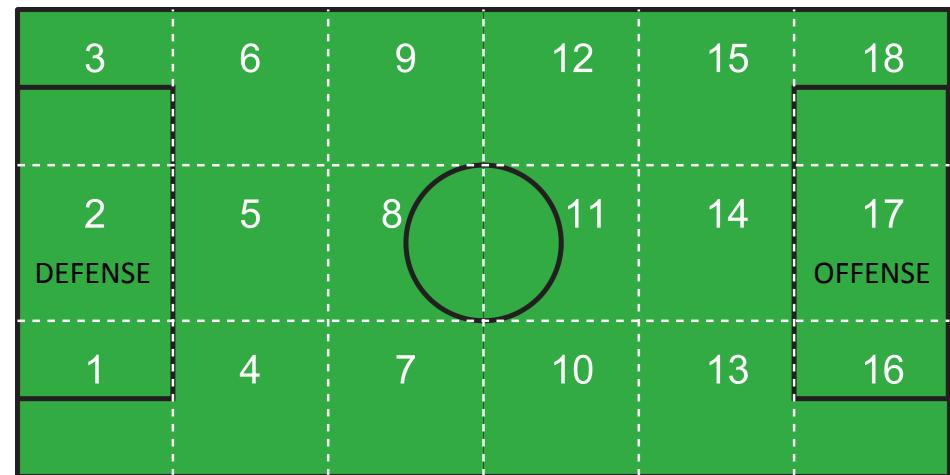
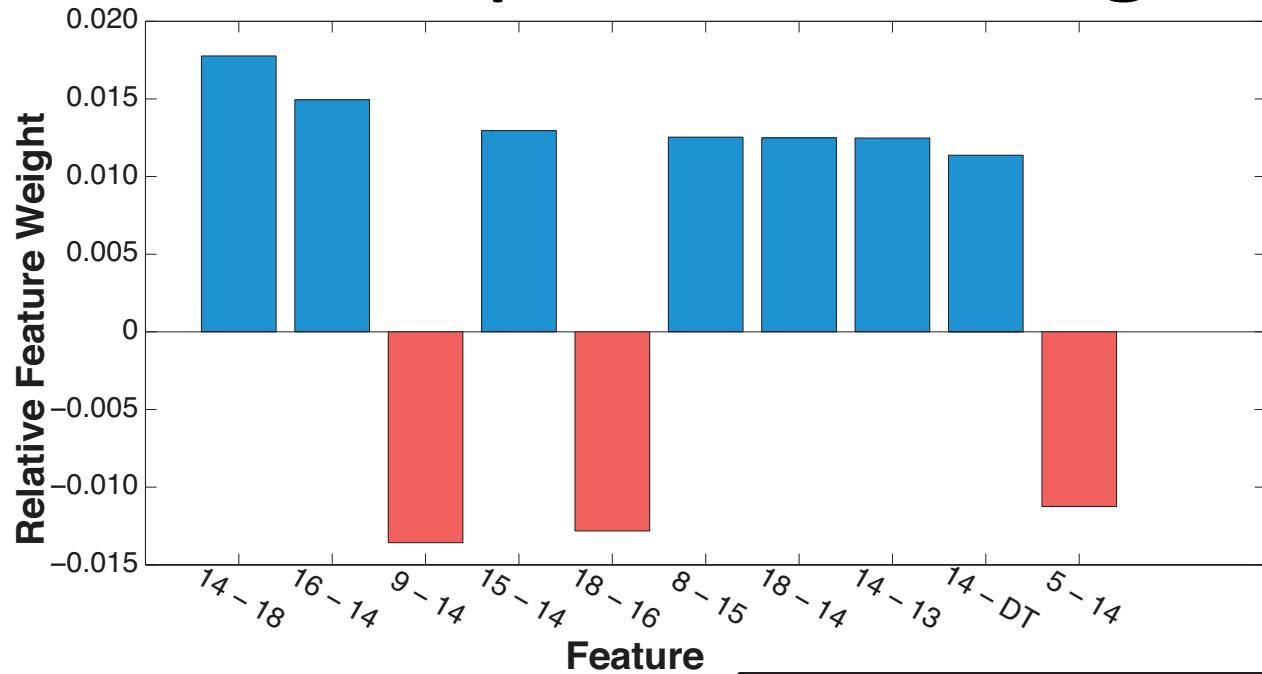
- Used the first 80% of games in the 2012-2013 season as a training set
- Evaluated the model on the final 20%
- Trained a L2-regularized SVM model finding the w that minimizes:

$$\frac{1}{2}w^T w + C_k \sum_{i=1}^l (\max(0, 1 - y_i w^T x_i))$$

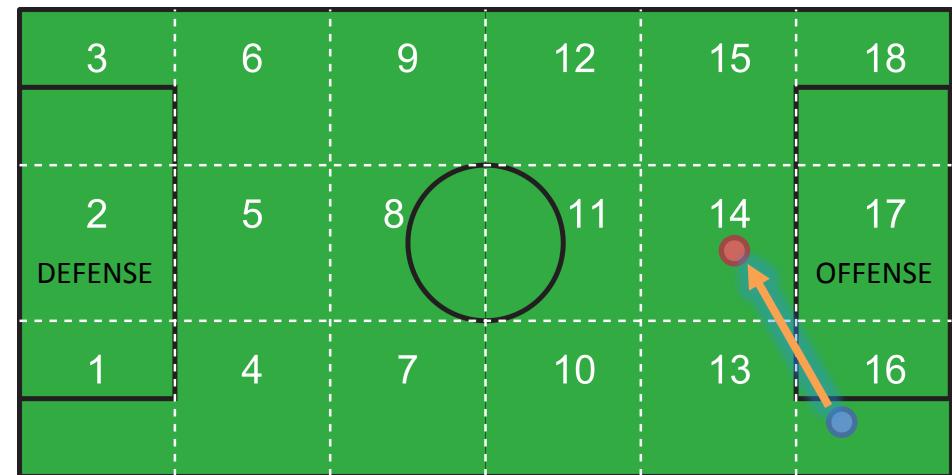
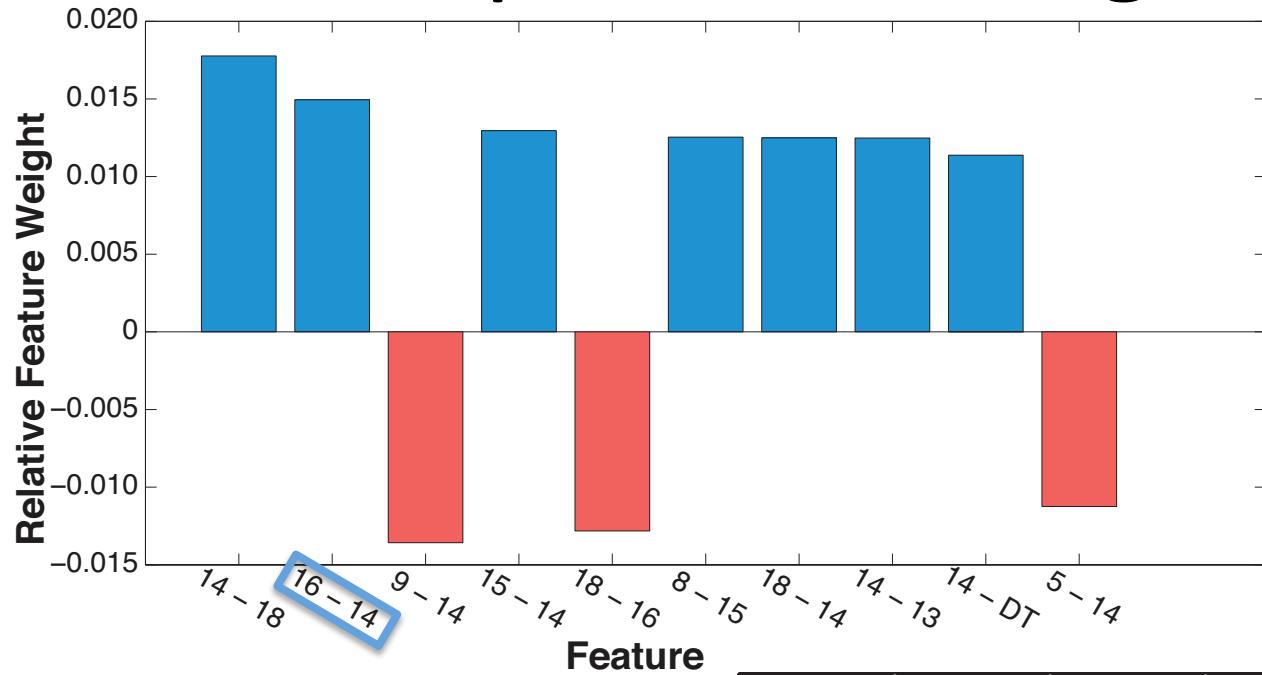
- Class specific cost parameters C_k chosen with 5-fold cross validation



Top Model Weights



Top Model Weights



18:50 FCB 1 - 0 GET
2 0 CORNERS



Nuevo

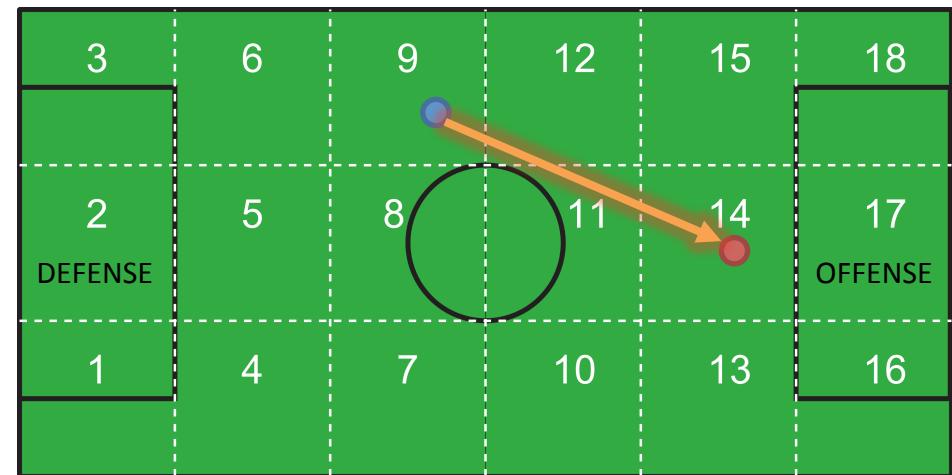
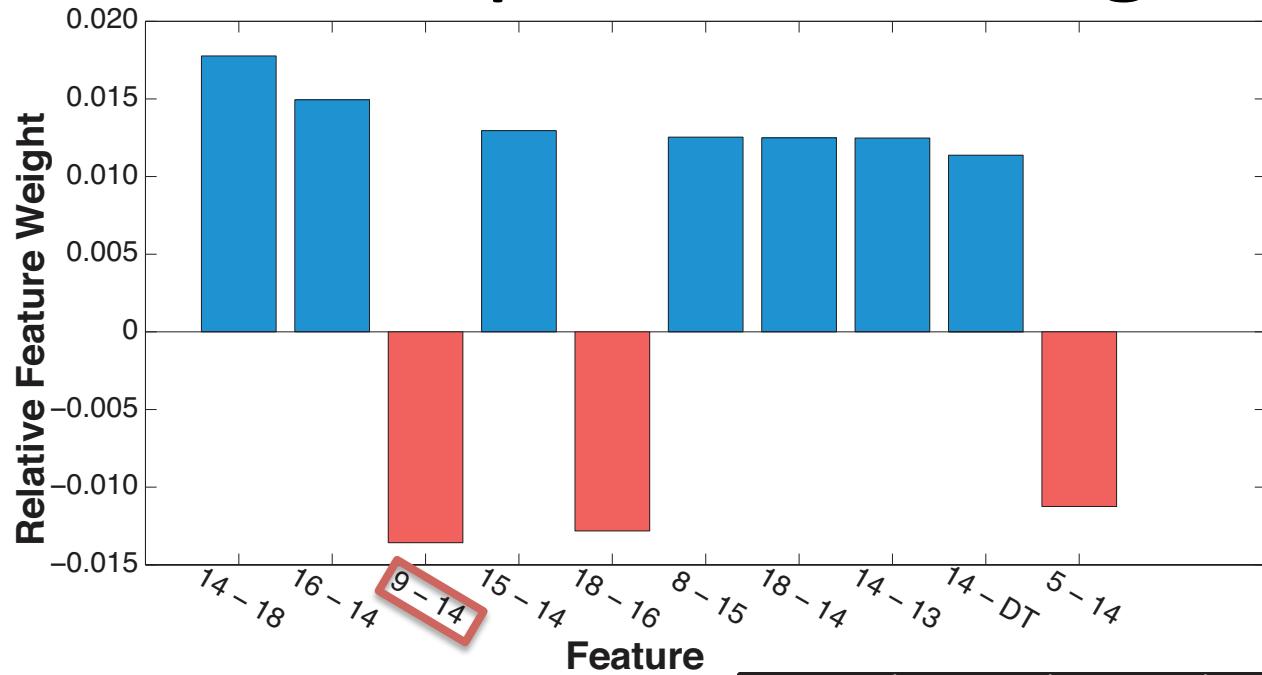
A4

QATAR AIRWAYS

Allian

Match Choice

Top Model Weights



Model Weights → Pass Value Metric

- Weights provide a conceptual map to which locations lead to shots
- Each pass has three relevant model weights
 - Origin
 - Destination
 - Origin-Destination pair

Pass Shot Value (PSV)

- Pass Shot Value (PSV) is computed for a pass with origin in zone i and a destination in zone j as:

$$\text{PSV}(i, j) = w_i^o + w_j^d + w_{ij}^{od}$$

- Sum of the model weights for the corresponding origin, destination, and origin-destination pair, respectively

Pass Shot Value (PSV)

- Pass Shot Value (PSV) is computed for a pass with origin in zone i and a destination in zone j as:

$$\text{PSV}(i, j) = w_i^o + w_j^d + w_{ij}^{od}$$

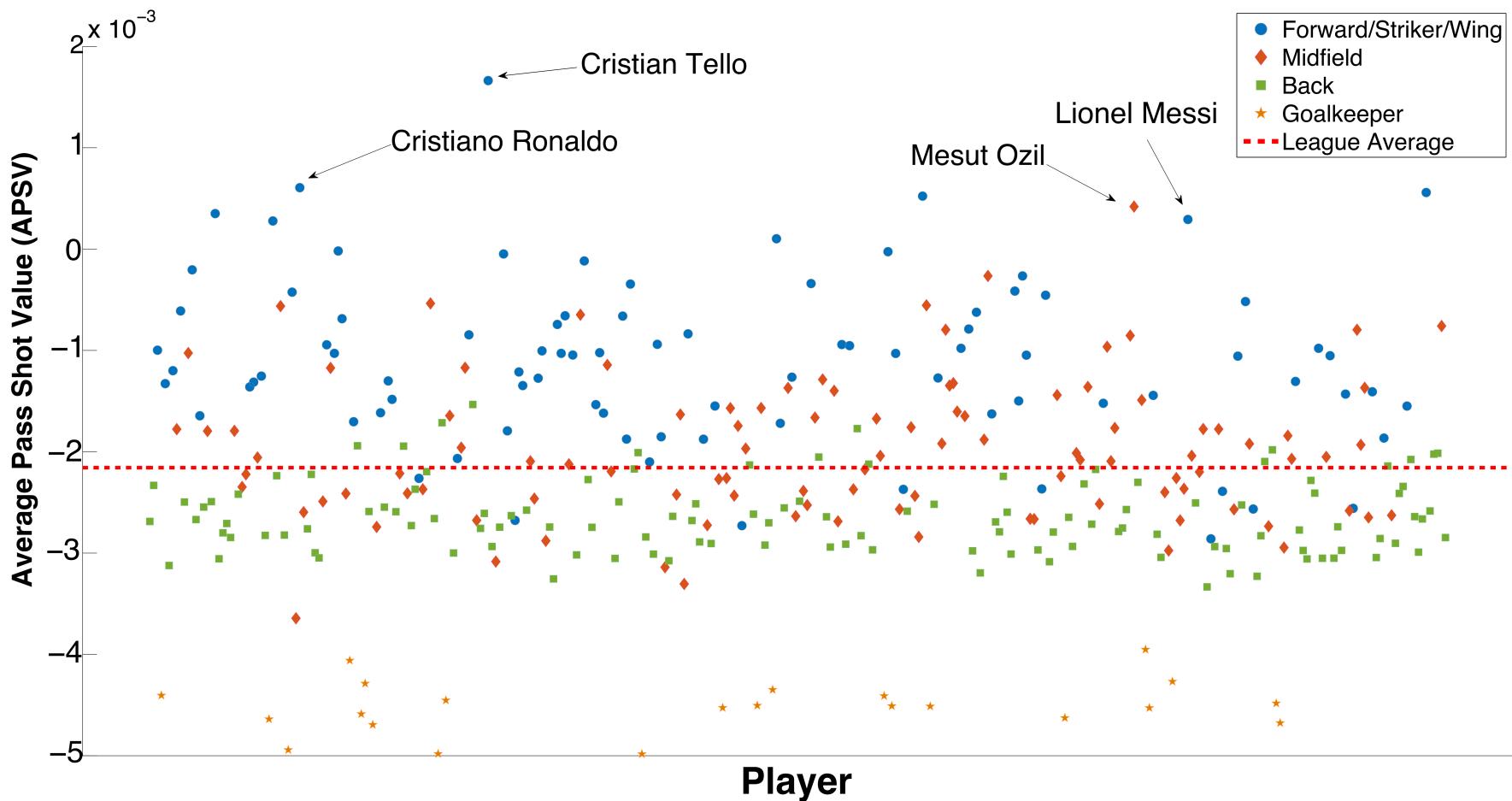
- Sum of the model weights for the corresponding origin, destination, and origin-destination pair, respectively
- e.g.:

$$\text{PSV}(16, 14) = w_{16}^o + w_{14}^d + w_{16,14}^{od} = 0.002 + 0.011 + 0.015 = 0.028$$

PSV as a Player Metric

- For every player, compute the PSV for every completed pass in which they were the distributor
- Average these values over the entire course of the season
- Limited analysis to players with > 200 completed passes
 - ~350 players

Average PSV for La Liga 2012-2013



Top Players by Average PSV

Rank	Player
1	Cristian Tello
2	Cristiano Ronaldo
3	Sergio Garcia
4	Karim Benzema
5	Gonzalo Higuain
6	Lionel Messi
7	Jonathan Viera
8	Angel Di Maria
9	Nolito
10	Jorge Molina

Offense

Top Players by Average PSV

Rank	Player
1	Cristian Tello
2	Cristiano Ronaldo
3	Sergio Garcia
4	Karim Benzema
5	Gonzalo Higuain
6	Lionel Messi
7	Jonathan Viera
8	Angel Di Maria
9	Nolito
10	Jorge Molina

Winner Ballon D'or 2013

Runner-up Ballon D'or 2013

Offense

Top Players by Average PSV

Rank	Player
1	Cristian Tello
2	Cristiano Ronaldo
3	Sergio Garcia
4	Karim Benzema
5	Gonzalo Higuain
6	Lionel Messi
7	Jonathan Viera
8	Angel Di Maria
9	Nolito
10	Jorge Molina

Offense

Rank	Player
1	Mesut Ozil
2	Diego Buonanotte
3	Kaka
4	Emiliano Armenteros
5	Miguel de las Cuevas
6	Julio Baptista
7	Arda Turan
8	Jose Barkero
9	Cesc Fabregas
10	Ever Banega

Midfield

Top Players by Average PSV

Rank	Player
1	Cristian Tello
2	Cristiano Ronaldo
3	Sergio Garcia
4	Karim Benzema
5	Gonzalo Higuain
6	Lionel Messi
7	Jonathan Viera
8	Angel Di Maria
9	Nolito
10	Jorge Molina

Offense

Rank	Player
1	Mesut Ozil
2	Diego Buonanotte
3	Kaka
4	Emiliano Armenteros
5	Miguel de las Cuevas
6	Julio Baptista
7	Arda Turan
8	Jose Barkero
9	Cesc Fabregas
10	Ever Banega

Midfield

Rank	Player
1	Marcelo Vieira
2	Eliseu
3	Dani Alves
4	Filipe Luis
5	Aly Cissokho
6	Diego Colotto
7	Martin Demichelis
8	Nacho Monreal
9	Sergio Sangez
10	Lolo

Defense

Top Players by Average PSV

Rank	Player
1	Cristian Tello
2	Cristiano Ronaldo
3	Sergio Garcia
4	Karim Benzema
5	Gonzalo Higuain
6	Lionel Messi
7	Jonathan Viera
8	Angel Di Maria
9	Nolito
10	Jorge Molina

Offense

Rank	Player
1	Mesut Ozil
2	Diego Buonanotte
3	Kaka
4	Emiliano Armenteros
5	Miguel de las Cuevas
6	Julio Baptista
7	Arda Turan
8	Jose Barkero
9	Cesc Fabregas
10	Ever Banega

Midfield

Rank	Player
1	Marcelo Vieira
2	Eliseu
3	Dani Alves
4	Filipe Luis
5	Aly Cissokho
6	Diego Colotto
7	Martin Demichelis
8	Nacho Monreal
9	Sergio Sangez
10	Lolo

Defense

Rank	Player	Goals
1	Lionel Messi	46
2	Cristiano Ronaldo	34
3	Radamel Falcao	28
4	Alvaro Negredo	25
4	Roberto Soldado	24
6	Ruben Castro	18
6	Piti	18
8	Gonzalo Higuain	16
9	Carlos Vela	14
9	Helder Postiga	14
9	Artiz Aduriz	14

Top Goal Scorers

Rank	Player	Assists
1	Andres Iniesta	16
2	Mesut Ozil	13
3	Lionel Messi	12
4	Karim Benzema	11
4	Cesc Fabregas	11
6	Cristiano Ronaldo	10
6	Ivan Rakitic	10
8	Ibai Gomez	9
8	Carlos Vela	9
8	Koke	9
8	Alexis Sanchez	9

Top Players by Assists

Top Players by Average PSV

Rank	Player
1	Cristian Tello
2	Cristiano Ronaldo
3	Sergio Garcia
4	Karim Benzema
5	Gonzalo Higuain
6	Lionel Messi
7	Jonathan Viera
8	Angel Di Maria
9	Nolito
10	Jorge Molina

Offense

Rank	Player
1	Mesut Ozil
2	Diego Buonanotte
3	Kaka
4	Emiliano Armenteros
5	Miguel de las Cuevas
6	Julio Baptista
7	Arda Turan
8	Jose Barkero
9	Cesc Fabregas
10	Ever Banega

Midfield

Rank	Player
1	Marcelo Vieira
2	Eliseu
3	Dani Alves
4	Filipe Luis
5	Aly Cissokho
6	Diego Colotto
7	Martin Demichelis
8	Nacho Monreal
9	Sergio Sangez
10	Lolo

Defense

Rank	Player	Goals
1	Lionel Messi	46
2	Cristiano Ronaldo	34
3	Radamel Falcao	28
4	Alvaro Negredo	25
4	Roberto Soldado	24
6	Ruben Castro	18
6	Piti	18
8	Gonzalo Higuain	16
9	Carlos Vela	14
9	Helder Postiga	14
9	Artiz Aduriz	14

Correlation:
 $\rho = 0.27, p < 0.05$

Top Goal Scorers

Rank	Player	Assists
1	Andres Iniesta	16
2	Mesut Ozil	13
3	Lionel Messi	12
4	Karim Benzema	11
4	Cesc Fabregas	11
6	Cristiano Ronaldo	10
6	Ivan Rakitic	10
8	Ibai Gomez	9
8	Carlos Vela	9
8	Koke	9
8	Alexis Sanchez	9

Top Players by Assists

Conclusion

- The locations of passes can predict whether a possession ends in a shot
- The relationship between pass location and shots can be used to understand the offensive value of individual passes
- Average PSV separates players by position, and seems to correlate well with offensive ability within each position
- Almost every other popular sport is collecting locations of events, so a similar methodology could be applied

Conclusion

- The locations of passes can predict whether a possession ends in a shot
- The relationship between pass location and shots can be used to understand the offensive value of individual passes
- Average PSV separates players by position, and seems to correlate well with offensive ability within each position
- Almost every other popular sport is collecting locations of events, so a similar methodology could be applied

Questions?