NFL Direction-Oriented Rushing Off-Def Plus-Minus

ID: 62839

In football, rushing is an action of advancing the ball forward by running with it, instead of passing. Rush offense refers to how well a team is able to gain yardage by running the ball, whereas rush defense refers to how well a defense team is able to prevent the rushers from gaining yardage.

In this report, we use the (regularized) normal Rasch model to estimate the offensive and defensive skills of individual teams in rushing. The advantage of using the Rasch model is that we can consider offense and defense separately, thereby taking into account opponent team's strength of defense when estimating a team's offense.

I have obtained the dataset from *NFLsavant*.com, which includes play-by-play data for every game in the NFL 2015-16 season. Similar data can be obtained from websites such as *FootballOutsiders*.com and *Pro-Football-Reference*.com. The dataset contains individual statistic collected in a single play, such as gained yardage, pass type, down, yard line, quarter, and indicators for pass/rush play.

1 The Rasch Model

In the Rasch model, we hope to estimate the number of yards gained by the rusher on the offensive team, depending on various pairs of offense vs. defense teams. In addition to the identity of teams on offense/defense, we must include other significant factors that could help decide the outcome. For example, we include the number of yards to go (`ToGo`) for first down, since the rushing team may only be attempting to get first down, rather than maximizing the gained yardage.

Furthermore, we include the rush directions chosen by the offensive team. This is significant because a team's defense skills may not necessarily be uniform throughout the defensive line – i.e. the left side of the scrimmage may be weaker than the right side. It may be the best of interest for an offensive team to predict which direction would help gain the most yardage against a particular team on defense, in order to boost their yardage.

For possession $i = 1, ..., n$:

- O_i = identity of offensive team on i^{th} possession. ← Offensive Adversary
- D_i = identity of defensive team on i^{th} possession. ← Defensive Adversary
- L_i = number of yards to go for first down
- For $j = 1, ..., 7$: (Directions: {`LE`, `LT`, `LG`, `C`, `RG`, `RT`, `RE`})
  - $R_{O_i j}$ = indicator of whether team O_i chose j^{th} direction on i^{th} possession
  - $R_{D_i j}$ = indicator of whether offense chose j^{th} direction on i^{th} possession against team D_i
- Y_i = number of yards gained by rusher on i^{th} possession

$$E[Y_i] = \alpha + \beta_{O_i} + \delta_{D_i} + \sum_{j=1}^{7} \rho_{R_{O_i j}} + \sum_{j=1}^{7} \tau_{R_{D_i j}} + \epsilon L_i \tag{1}$$

2 Results

The expected yardage per rush attempt for a league-average offense team (with 10 yards to go for 1st down) is **4.6 yards**. Our $\epsilon = 0.12$, meaning 10 yards to first down predicts about one extra rush yard than 1 yard to first down. The following shows the top 5 offense teams and top 5 defense teams according to our model:

Top 5 Offense	$E[Y\|L = 10]$	Top 5 Defense	$E[Y\|L = 10]$
Seattle Seahawks	5.2 yds	Denver Broncos	4.0 yds
Buffalo Bills	5.2 yds	Seattle Seahawks	4.1 yds
Tampa Bay Buccaneers	5.1 yds	New York Jets	4.3 yds
Pittsburgh Steelers	5.1 yds	Caroline Panthers	4.3 yds
Kansas City Chiefs	5.0 yds	Pittsburgh Steelers	4.4 yds

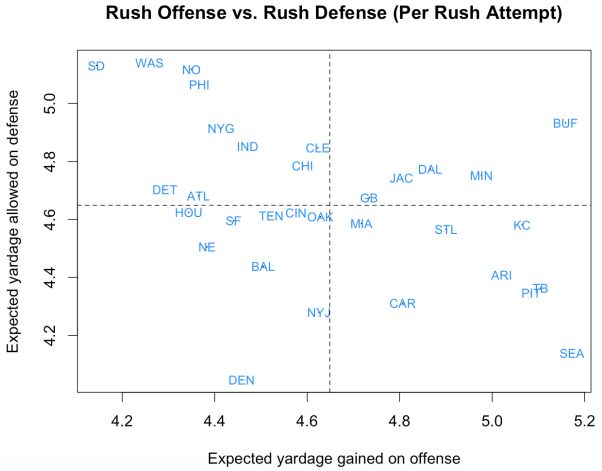

Figure 1: Average rush-offense plotted against rush-defense. Higher on x-axis correponds to better rush offense, and lower on y-axis corresponds to better rush defense. Seattle has both great rush offense and defense, whereas San Diego has both poor rush offense and defense.

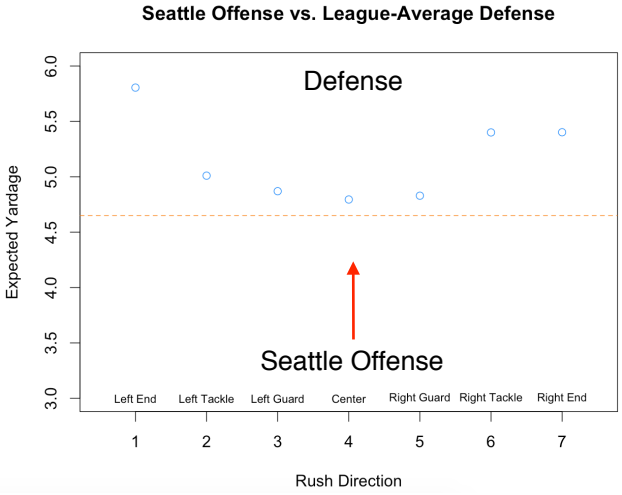

Figure 2: Rush-direction results for the Seahawks' offense against a league-average team. Orange dotted line shows the league average yardage. Note that the Seahawks rush offense is more successful away from the center than towards the center, giving them approx. one extra yard.

3 Analysis

In addition to each team having different levels of rush offense and defense, we have better insight about how a rush offense/defense can most benefit by analyzing rush directions. For example, when the best rush offense in the league (Seahawks) faces the worst rush defense in the league (Redskins), the Seahawks are estimated to gain 5.6 yards, which is 1.0 yard more than average. Furthermore, when Seahawks utilizes its best rush direction against the Redskins' worst rush defense direction at `RT`, then the Seahawks would gain about **6.1 yards**; on the other hand, Seahawks are estimated to have the poorest rush yardage at `RE`, where the Seahawks would gain only 4.9 yards. Even though `RT` and `RE` are close positions, wisely choosing the rushing direction will maximize yardage.

As an another example, consider the league's poorest rush offense (Chargers) playing the league's best rush offense (Broncos). The Chargers offense, when attempting to rush towards its weakest position at Broncos' strongest defense position at `LE`, would only gain about **3.1 rush yards**, which is 1.5 yards fewer than average. After all, since the Broncos won the Super Bowl this year, it's not surprising that the Broncos have the toughest rush defense in the league!

4 Appendix

4.1 Additional Figures

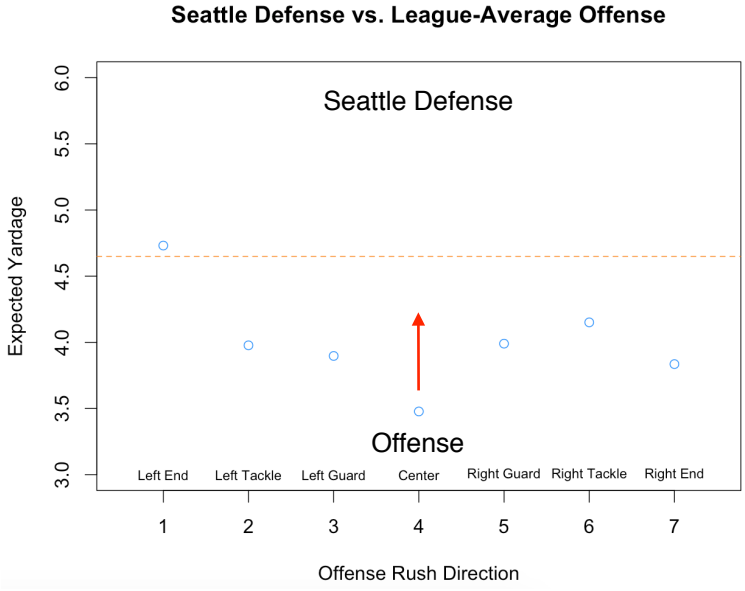

Figure 3: Rush-direction results for a league-average team against the Seahawks' defense (from defense point of view). Orange dotted line shows the league average yardage. Note that the Seahawks' weakest spot is the left-end, whereas its strongest defensive spot is closer to the center.

4.2 Overall Rush Offense Coefficients Ranking

```
   Off-SEA    Off-BUF     Off-TB    Off-PIT     Off-KC    Off-ARI    Off-MIN    Off-STL
 0.52368432 0.50980515 0.45571072 0.43559587 0.41594557 0.37162467 0.32786583 0.25123025

   Off-DAL    Off-CAR    Off-JAC     Off-GB    Off-MIA    Off-OAK    Off-NYJ    Off-CLE
 0.21710622 0.15767589 0.15521996 0.08519952 0.06828680 -0.02056526 -0.02324948 -0.02573651

   Off-CHI    Off-CIN    Off-TEN    Off-BAL    Off-IND    Off-DEN     Off-SF    Off-NYG
-0.05875664 -0.07282904 -0.12644443 -0.14401599 -0.17776539 -0.19073783 -0.20792840 -0.23505960

    Off-NE     Off-PHI    Off-ATL     Off-NO    Off-HOU    Off-DET    Off-WAS     Off-SD
-0.26557606 -0.28212433 -0.28383298 -0.29894858 -0.30516887 -0.35662417 -0.38961032 -0.50457897
```

4.3 Overall Rush Defense Coefficients Ranking

```
   Def-DEN    Def-SEA    Def-NYJ    Def-CAR    Def-PIT     Def-TB    Def-ARI    Def-BAL
-0.60202540 -0.51086271 -0.37021228 -0.33752742 -0.30306280 -0.28583022 -0.24067483 -0.21109561

    Def-NE     Def-STL     Def-KC    Def-MIA     Def-SF    Def-OAK    Def-TEN    Def-CIN
-0.14401654 -0.08312008 -0.06724460 -0.06279973 -0.05336218 -0.03924906 -0.03671143 -0.02518764

   Def-HOU     Def-GB    Def-ATL    Def-DET    Def-JAC    Def-MIN    Def-DAL    Def-CHI
-0.02436692 0.02624550 0.03275653 0.05419611 0.09418219 0.10388913 0.12542584 0.13634504

   Def-CLE    Def-IND    Def-NYG    Def-BUF    Def-PHI     Def-NO     Def-SD    Def-WAS
 0.19798279 0.20390159 0.26616354 0.28466206 0.41710238 0.46895691 0.48268671 0.49157197
```

5 R Code

```
library(glmnet)
library(Matrix)

########## PROCESSING ##########
data = read.csv('pbp-2015.csv')
data = data[data$IsRush == 1,]
data$RB.Description = str_extract(data$Description, "-[A-Z]+.[A-Z]+ ")
data$RB = str_extract(data$RB.Description, "[A-Z]+\\.[A-Z]+")
players = sort(unique(data$RB))

# Extract each team occurence from the data.
teams = unlist(data[,c("OffenseTeam", "DefenseTeam")])

# How many players are present in the data?
num.players = length(players)

# How many teams?
unique.teams = sort(unique(teams))
num.teams = length(unique(teams))

# Label each player occurence as offense or defense.
prefix = rep(c('Off', 'Def'), each = nrow(data))
tag1 = paste(prefix, teams, sep='-')
tag2 = paste('OffRush', data$OffenseTeam, data$RushDirection, sep='-')
tag3 = paste('DefRush', data$DefenseTeam, data$RushDirection, sep='-')

tag.factor = as.factor(c(tag1, tag2, tag3))

# Define the row and column of X in which to put each 1.
i = rep(1:nrow(data), 4)
j = as.numeric(tag.factor)

# Construct X as a sparse matrix.
X = as.matrix(sparseMatrix(i, j))
colnames(X) = sort(unique(tag.factor))
X = cbind(X, data$ToGo)
colnames(X)[ncol(X)] = "ToGo"
X = as(X, "sparseMatrix")

# Create y, the vector storing the number of yards gained on rush attempt.
y = data$Yards

########## ANALYSIS ##########

# Use cv.glmnet to fit a regularized normal Rasch model.
lambda = exp(seq(-10, 0, length = 100))
rasch = cv.glmnet(X, y, alpha = 0, standardize = FALSE, lambda = lambda)
plot(rasch)

# Extract and label the fitted regression coefficients.
coef = coef(rasch, s = 'lambda.min')[, 1]
alpha = coef[1]
togo = coef[length(coef)]

# Best offense in the NFL
sort(coef[select=grep("Off-", names(coef))], decreasing=T)

# Best defense in the NFL
sort(coef[select=grep("Def-", names(coef))], decreasing=F)

# Plot for Overall Offensive-Defensive Skills
average = alpha + togo*10
offense = coef[select=grep("Off-", names(coef))] + alpha + togo*10
defense = coef[select=grep("Def-", names(coef))] + alpha + togo*10
plot(offense, defense, cex=0.1, main="Rush Offense vs. Rush Defense (Per Rush Attempt)",
```

```
      xlab="Expected yardage gained on offense", ylab="Expected yardage allowed on defense")
text(offense, defense, labels=sort(unique.teams), cex = 0.8, col = 'dodgerblue')
abline(h = average, lty=2)
abline(v = average, lty=2)

# Stats for the Seattle Seahawks
coef[select=grep("SEA",names(coef))]

# Best rush directions on offense for the Seahawks.
sort(coef[select=grep("OffRush-SEA", names(coef))], decreasing=T)

# Best rush directions when rushing against the Seahawks defense.
sort(coef[select=grep("DefRush-SEA", names(coef))], decreasing=T)

# Coefficients for all rush directions when Seahawks are on offense.
positions = c("OffRush-SEA-LEFT END", "OffRush-SEA-LEFT TACKLE", "OffRush-SEA-LEFT GUARD",
              "OffRush-SEA-CENTER",
              "OffRush-SEA-RIGHT GUARD", "OffRush-SEA-RIGHT TACKLE", "OffRush-SEA-RIGHT END"
                  )
# positions = c("DefRush-SEA-LEFT END", "DefRush-SEA-LEFT TACKLE", "DefRush-SEA-LEFT GUARD",
#               "DefRush-SEA-CENTER",
#               "DefRush-SEA-RIGHT GUARD", "DefRush-SEA-RIGHT TACKLE", "DefRush-SEA-RIGHT
    END")
coefs = coef[positions]
yardages = alpha + coefs + togo*10 + coef["Off-SEA"]

# Plot Expected Yardage for different rush directions for Seahawks offense.
plot(yardages,
      xlab = "Offense Rush Direction", ylab="Expected Yardage", main="Seattle Offense vs.
          League-Average Defense",
      xlim=c(0.5,7.5), ylim=c(3, 6), cex=1, col='dodgerblue')
abline(average, 0, col='darkorange', lty=2)
labels = c("Left End", "Left Tackle", "Left Guard", "Center", "Right Guard", "Right Tackle",
      "Right End")
text(1:7, rep(3, 7), labels = labels, cex = 0.7, col = 'black')
```

hw4-code.R