ASSIGNMENT 1 : Ranking NFL 2007 teams using Markov Chain

Name: Kevin Kasundra

UIN: 327000151

"On my honor, as an Aggie, I have neither given nor received unauthorized aid on this assignment."

Introduction

The objective of this assignment is to rank the 32 teams who played the National Football League (NFL) under either National Football Conference (NFC) or American Football Conference (AFC), with 16 teams in each conference. Each team plays 16 regular season games, thus teams do not play all other teams during a single regular season. The focus of this assignment is the 2007 NFL regular season, which infers that Wildcard, Division and Superbowl are excluded while ranking the teams.

Methodology

The steps in which we would rank the teams is given below:

1. Abstract the required data provided from 'https://www.pro-football-reference.com' into an excel.
2. Analyse the data in excel and remove the columns which are not required by us to rank the teams, for example Week number, Day, Date, Time, Boxscore, etc.
3. Import the excel into R-Studio
4. Assign Teams to different states varying from 1 to 32.
5. Make a Transition Matrix based on the criteria you want to rank the teams
6. Normalize the rows of the matrix
7. Find Stationary distributions of different teams/states and export the data into excel.
8. Line teams in decreasing order of stationary distribution values and assign corresponding rank from 1 to 32.

- State assigned to each team as per step 4 of methodology

State	Team Name	State	Team Name
1	Indianapolis Colts	17	Tennessee Titans
2	New Orleans Saints	18	Jacksonville Jaguars
3	Carolina Panthers	19	Houston Texans
4	St. Louis Rams	20	Kansas City Chiefs
5	Minnesota Vikings	21	Detroit Lions
6	Atlanta Falcons	22	Oakland Raiders
7	Denver Broncos	23	San Diego Chargers
8	Buffalo Bills	24	Chicago Bears
9	Green Bay Packers	25	Seattle Seahawks
10	Philadelphia Eagles	26	Tampa Bay Buccaneers
11	Washington Redskins	27	Dallas Cowboys
12	Miami Dolphins	28	New York Giants
13	Pittsburgh Steelers	29	Cincinnati Bengals
14	Cleveland Browns	30	Baltimore Ravens
15	New England Patriots	31	San Francisco 49ers
16	New York Jets	32	Arizona Cardinals

TYPE: 1 Matrix by assigning W/L scores individually

The main concept of the method is that each individual competition between two teams (alternatives) results in the losing alternative voting for the winning alternative. These collection of votes will populate a square matrix that represents the head to head competitions between all the alternatives.

We develop the voting matrix (transition matrix) using the scores of each team in a game. Each time an alternative loses to another alternative, they will place a vote for that alternative in the matrix. We then normalize the rows of the voting matrix to develop a stochastic transition probability matrix. Furthermore we solve using either dominant eigenvector of this matrix or use power method to obtain the steady state probability. Finally arranging each state's steady probabilities in the descending order gives us the ranking of the teams (states) with first being the best among the set.

RANKING USING METHOD 1

RANK	STATE	TEAM NAME	PROBABILITY	RANK	STATE	TEAM NAME	PROBABILITY
1	15	New England Patriots	0.058812	**17**	24	Chicago Bears	0.030886
2	1	Indianapolis Colts	0.041671	**18**	32	Arizona Cardinals	0.030386
3	27	Dallas Cowboys	0.040701	**19**	13	Pittsburgh Steelers	0.029794
4	23	San Diego Chargers	0.039642	**20**	30	Baltimore Ravens	0.029008
5	9	Green Bay Packers	0.039377	**21**	29	Cincinnati Bengals	0.028882
6	10	Philadelphia Eagles	0.038392	**22**	22	Oakland Raiders	0.027074
7	18	Jacksonville Jaguars	0.03829	**23**	8	Buffalo Bills	0.025864
8	28	New York Giants	0.036431	**24**	26	Tampa Bay Buccaneers	0.025515
9	2	New Orleans Saints	0.034992	**25**	17	Tennessee Titans	0.024961
10	11	Washington Redskins	0.033845	**26**	16	New York Jets	0.023789
11	5	Minnesota Vikings	0.033046	**27**	12	Miami Dolphins	0.02365
12	19	Houston Texans	0.0323	**28**	6	Atlanta Falcons	0.02309
13	14	Cleveland Browns	0.032001	**29**	3	Carolina Panthers	0.022812
14	7	Denver Broncos	0.031744	**30**	4	St. Louis Rams	0.022113
15	25	Seattle Seahawks	0.03107	**31**	20	Kansas City Chiefs	0.021159
16	21	Detroit Lions	0.030915	**32**	31	San Francisco 49ers	0.017789

TYPE: 2 Random Walk Method

Here we consider the win and loss between various teams to construct the transition matrix. The voting matrix is made by the defeated team placing a vote for the winning team and this will be done for all individuals. There is always one winner per individual per group while there can be any number of losers. Just the number of wins is not important, but also wins against strong teams will result in larger impact on the rankings. The voting matrix can be represented as a Markov chain. If a random walk is taken along the Markov graph, the long run proportion of time spent at each state will be rating of that team strength. Before we find the long term steady state matrix, we check to see if any row has sum of elements equal to zero. Any undefeated team represents an absorbing state as there was no losses and hence no votes. Using the dangling node adjustment, we add value of '1/n', n = number of states to the row of absorbing state. Thus now the matrix is stochastic and solvable.

Finally the steady state probability of the transition matrix can be found out and arranging this probabilities in decreasing order gives us the strengths of each team in decreasing order (ranking).

RANKING USING METHOD 2

RANK	STATE	TEAM NAME	PROBABILITY	RANK	STATE	TEAM NAME	PROBABILITY
1	15	New England Patriots	0.079883	**17**	29	Cincinnati Bengals	0.027852
2	27	Dallas Cowboys	0.07411	**18**	14	Cleveland Browns	0.027777
3	28	New York Giants	0.05031	**19**	30	Baltimore Ravens	0.026334
4	23	San Diego Chargers	0.049362	**20**	20	Kansas City Chiefs	0.025409
5	11	Washington Redskins	0.048942	**21**	21	Detroit Lions	0.023812
6	32	Arizona Cardinals	0.04659	**22**	5	Minnesota Vikings	0.023666
7	18	Jacksonville Jaguars	0.045263	**23**	2	New Orleans Saints	0.023199
8	25	Seattle Seahawks	0.041907	**24**	13	Pittsburgh Steelers	0.022389
9	26	Tampa Bay Buccaneers	0.038658	**25**	3	Carolina Panthers	0.020908
10	17	Tennessee Titans	0.037153	**26**	6	Atlanta Falcons	0.016352
11	9	Green Bay Packers	0.034726	**27**	7	Denver Broncos	0.013631
12	19	Houston Texans	0.034359	**28**	22	Oakland Raiders	0.013046
13	24	Chicago Bears	0.033803	**29**	8	Buffalo Bills	0.010642
14	10	Philadelphia Eagles	0.031849	**30**	4	St. Louis Rams	0.007555
15	31	San Francisco 49ers	0.031205	**31**	16	New York Jets	0.006958
16	1	Indianapolis Colts	0.02856	**32**	12	Miami Dolphins	0.00379

TYPE: 3 Difference of W/L points method

Its core idea of this method is based on similar lines to the PageRank algorithm of Google. It's similar to the first method on the basis of the general steps. However instead of each team's score against the other for voting, we simply take the difference of the score and assign it to the winning team in the matrix. But after making initial matrix we normalize the matrix and multiply the new transition matrix by a constant alpha between zero and one. Moreover we add a personalization vector having size equal to transition matrix, with all entries being equal, positive and summing up to one. This is done to make the overall transition matrix irreducible.

Now we find the long term stationary probabilities of the states in the new matrix. Arranging these probabilities in decreasing order gives us the ranking of teams.

RANKING USING METHOD 3

RANK	STATE	TEAM NAME	PROBABILITY	RANK	STATE	TEAM NAME	PROBABILITY
1	32	Arizona Cardinals	0.07639	**17**	9	Green Bay Packers	0.025262
2	15	New England Patriots	0.075324	**18**	5	Minnesota Vikings	0.022115
3	27	Dallas Cowboys	0.073966	**19**	2	New Orleans Saints	0.021354
4	25	Seattle Seahawks	0.06392	**20**	14	Cleveland Browns	0.020605
5	28	New York Giants	0.061516	**21**	21	Detroit Lions	0.019465
6	31	San Francisco 49ers	0.054538	**22**	17	Tennessee Titans	0.019044
7	23	San Diego Chargers	0.047032	**23**	1	Indianapolis Colts	0.018178
8	29	Cincinnati Bengals	0.043806	**24**	3	Carolina Panthers	0.017776
9	30	Baltimore Ravens	0.042915	**25**	10	Philadelphia Eagles	0.012929
10	26	Tampa Bay Buccaneers	0.03751	**26**	22	Oakland Raiders	0.011114
11	13	Pittsburgh Steelers	0.036675	**27**	7	Denver Broncos	0.008269
12	11	Washington Redskins	0.035759	**28**	6	Atlanta Falcons	0.007663
13	24	Chicago Bears	0.033279	**29**	16	New York Jets	0.007547
14	19	Houston Texans	0.03244	**30**	8	Buffalo Bills	0.006496
15	18	Jacksonville Jaguars	0.031722	**31**	4	St. Louis Rams	0.00428
16	20	Kansas City Chiefs	0.027134	**32**	12	Miami Dolphins	0.003978

Summary

Comparison of Different methods used is given in the table below

State	Name	W/L Score	Random Walk (W/L)	Difference in W/L Score
1	Indianapolis Colts	2	16	23
2	New Orleans Saints	9	23	19
3	Carolina Panthers	29	25	24
4	St. Louis Rams	30	30	31
5	Minnesota Vikings	11	22	18
6	Atlanta Falcons	28	26	28
7	Denver Broncos	14	27	27
8	Buffalo Bills	23	29	30
9	Green Bay Packers	5	11	17
10	Philadelphia Eagles	6	14	25
11	Washington Redskins	10	5	12
12	Miami Dolphins	27	32	32
13	Pittsburgh Steelers	19	24	11
14	Cleveland Browns	13	18	20
15	New England Patriots	1	1	2
16	New York Jets	26	31	29
17	Tennessee Titans	25	10	22
18	Jacksonville Jaguars	7	7	15
19	Houston Texans	12	12	14
20	Kansas City Chiefs	31	20	16
21	Detroit Lions	16	21	21
22	Oakland Raiders	22	28	26
23	San Diego Chargers	4	4	7
24	Chicago Bears	17	13	13
25	Seattle Seahawks	15	8	4
26	Tampa Bay	24	9	10
27	Dallas Cowboys	3	2	3
28	New York Giants	8	3	5
29	Cincinnati Bengals	21	17	8
30	Baltimore Ravens	20	19	9
31	San Francisco 49ers	32	15	6
32	Arizona Cardinals	18	6	1

- The highest team ranked is New England Patriots for Method 1&2 while it is Arizona Cardinals for Method 3.
- The rankings are almost similar to the actual rankings of the NFL 2007 regular season.
- It is observed that Method 1 and 2 are very sensitive to tail end values, meaning an upset match can have a considerable impact on the ranking.
- Each method has its pros and cons and no method perfectly implies to real life scenario.

Reference

1. A. Y. Govan and C. D. Meyer, “Ranking national football league teams using Google’s pagerank,” in AA Markov Anniversary Meeting, 2006.
2. B. Vaziri, S. Dabadghao, Y. Yih, and T. I. J. J. o. t. O. R. S. Mortin, “Properties of sports ranking methods,” vol. 69, no. 5, pp. 776-787, 2018.
3. T. Callahan, P. J. Mucha, and M. A> J. T. A. M. M. Porter, “Random walker ranking for NCAA division IA football,” vol. 114, no.9, pp. 761-777, 2007.

Code in R

```
library(readxl)
library(MASS)
library(xlsx)
nfldata <- read_excel("nfldata.xlsx")
View(nfldata)
mat<- as.data.frame(nfldata)
summary(mat)
#-----------------------------------------------------#
## TYPE 1: Matrix by assigning W/L scores individually

#Transition Matrix
P1=matrix(data = 0, nrow = 32, ncol = 32)
for (i in 1:256) {
  k=mat$Winner[i]
  for (j in 1:256) {
    l=mat$Loser[j]
    if(i==j){
      P1[k,l]=mat$PL[i]
      P1[l,k]=mat$PW[i]
    }
  }
}
Pfinal_1 <- P1
write.xlsx(Pfinal_1, "TransitionMat1.xlsx")

#Stochastic Matrix using Normalization of rows
for(i in 1:32){
  for (j in 1:32) {
    Pfinal_1[i,j] = P1[i,j]/sum(P1[i,1:32])
  }
}
write.xlsx(Pfinal_1, "NormalizedMat1.xlsx")

#Steady State Matrix
r=eigen(Pfinal_1)
rvec=r$vectors
# left eigenvectors are the inverse of the right eigenvectors
lvec=ginv(r$vectors)
# The eigenvalues
lam<-r$values
# Two ways of checking the spectral decomposition:
# Standard definition
rvec%*%diag(lam)%*%ginv(rvec)
Pinf_1 <- Re(lvec[1,]/sum(lvec[1,]))
Pinf_1
write.xlsx(Pinf_1, "Ranking1.xlsx")

#-----------------------------------------------------#
## TYPE 2: Random Walk Method

#Transition Matrix
P2=matrix(data = 0, nrow = 32, ncol = 32)
for (i in 1:256) {
  k=mat$Winner[i]
  for (j in 1:256) {
    l=mat$Loser[j]
    if(i==j){
      P2[l,k]=1
    }
```

```
  }
}
Pfinal_2 <- P2
for(i in 1:32){
if(sum(Pfinal_2[i,])==0){
  Pfinal_2[i,]=1/32
   }
}
write.xlsx(Pfinal_2, "TransitionMat2.xlsx")

#Stochastic Matrix
for(i in 1:32){
  for (j in 1:32) {
    Pfinal_2[i,j] = Pfinal_2[i,j]/sum(Pfinal_2[i,1:32])
  }
}
write.xlsx(Pfinal_2, "NormalizedMat2.xlsx")

#Steady State Matrix
r=eigen(Pfinal_2)
rvec=r$vectors
lvec=ginv(r$vectors)
lam<-r$values
rvec%*%diag(lam)%*%ginv(rvec)
Pinf_2<-Re(lvec[1,]/sum(lvec[1,]))
Pinf_2
write.xlsx(Pinf_2, "Ranking2.xlsx")
#-------------------------------------------------------#
## TYPE 3: Difference of W/L points method

#Transition Matrix
P3 <- matrix(data = 0, nrow = 32, ncol = 32)
for (i in 1:256) {
  k=mat$Winner[i]
  for (j in 1:256) {
    l=mat$Loser[j]
    if(i==j){
      P3[l,k]=0.85*mat$diff[i]
    }
  }
}
S <- matrix(data=0.15, nrow=32, ncol=32)
Pfinal3 <- P3+S

#Stochastic Matrix
for(i in 1:32){
  for (j in 1:32) {
    Pfinal3[i,j] = Pfinal3[i,j]/sum(Pfinal3[i,1:32])
  }
}
write.xlsx(Pfinal3, "NormalizedMat3.xlsx")

#Steady State Matrix
r=eigen(Pfinal3)
rvec=r$vectors
lvec=ginv(r$vectors)
lam<-r$values
rvec%*%diag(lam)%*%ginv(rvec)
Pinf_3<-Re(lvec[1,]/sum(lvec[1,]))
Pinf_3
write.xlsx(Pinf_3, "Ranking3.xlsx")
#########################################################################
```