STATS 50

Final Project

Tony Wu

NFL Power Rankings through Bradley-Terry

In this project, I apply the Bradley-Terry model to the NFL using results from the 208 games that have been played up through Week 14. I've developed two models, one purely based on the win/loss outcome of games and another that takes into account the point differentials as well, to rank all 32 teams. I will then compare my BT rankings with some of the online NFL rankings by the big news outlets. Finally, I will use each of my models to predict which teams will make it into the playoffs.

Before jumping into the modeling, I first examined if the data met the assumptions for using a Bradley-Terry model. The BT model is used for situations of repeated pairwise comparisons, which seems to apply perfectly to the NFL. However, if the teams in the NFL can be grouped into two groups, A and B, such that there are no intergroup pairings in the first 14 weeks, then the basis for using Bradley-Terry is flawed. The key assumption is that, "In every possible partition of the individuals into nonempty sets, some individual in the second set beats some individual in the first set at least once." (Hunter) I had to check if 14 weeks' worth of games met that criteria. Each team will always play the other three teams in its division twice every season (and with three games to go, most teams have already played at least one game against every divisional rival). One division from each conference, and the remaining teams in its conference that finished in the same place in their own divisions round out the sixteen game NFL regular season. As shown by Fig. 1, every possible partition of divisions into two groups satisfies the key assumption. I also checked grouping all the k^{th} place division winners from last year together but due to how scheduling works out, every grouping of this kind also satisfies the key assumption. After exhausting other checks, I conclude that the results from the 14 weeks of NFL games satisfies the requirements for using the Bradley-Terry model.

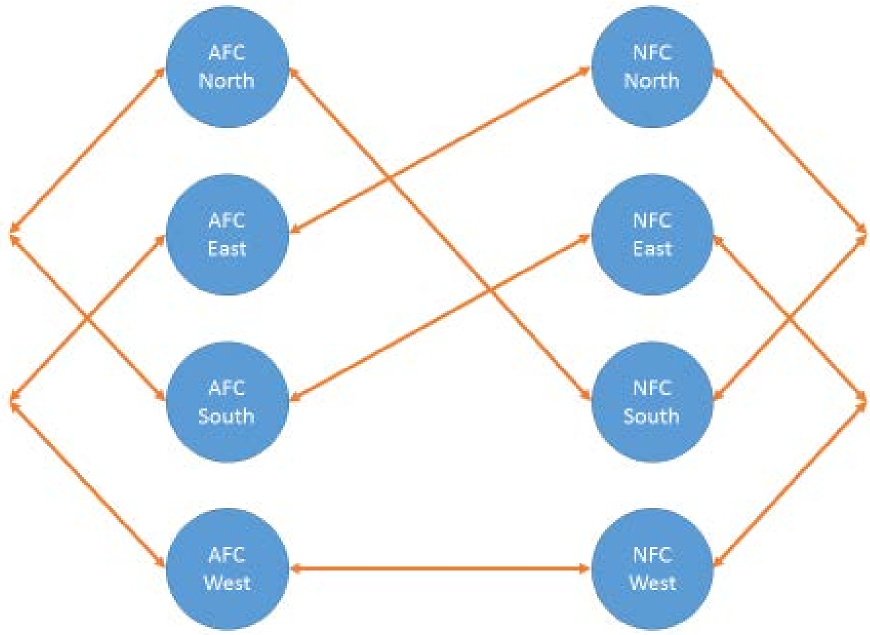

Figure 1- NFL interdivision schedule for 2014

I tabulated the data from this season into a 32 x 32 table of all the teams listed alphabetically (except SEA is before SF because of an error, but it was propagated consistently so it was not necessary to correct in the order). A team's wins are recorded in the row corresponding to that team and its losses are recorded in the column corresponding to that team. For example, the 41-10 win by the Panthers over the Saints in week 14 is recorded in the 5th row (Carolina) and the 20th column (New Orleans). There are two tables derived from the data. The first is purely win/loss so there's a 1 in (5,20) (and a different 1 in (20,5) for the Saint's victory over the Panthers in week 9). The second table incorporates the point differential of a game into the entry. However, the raw point difference is too irregular for Bradley-Terry so I divided it by 8 and rounded up to record the number of possessions a team won by (so the 31 point win is a 4 at (5,20) in the second table). All data recorded in included in the spreadsheet attachment for the project.

After charting the data, I used the cyclic algorithm described in Hunter's paper to compute $\boldsymbol{\gamma}$, the 32 by 1 vector of $\gamma_i's$ for the 32 teams. My initial guess was each $\gamma_i = \frac{1}{32}$. The (k+1)th iteration was computed by:

$$\gamma_i^{(k+1)} = W_i \left[\sum_{j \neq i} \frac{N_{ij}}{\gamma_i^{(k)} + \gamma_j^{(k)}} \right]^{-1}$$

Where W_i is the total number of wins by team i, and N_{ij} is total number of times teams i and j played each other. For the point differential BT model, W_i is the total number of possessions team i won games by. The total number of possessions team i lost by is not subtracted out of W_i because they are included in the $W_j's$ of the teams that beat team i. After the (k+1)th iteration is completed, $\boldsymbol{\gamma}$, is renormalized before proceeding. When the $\gamma_i's$ of an iteration do not change by more than 10^{-4} from the previous iteration, I stopped the code and my last iteration became the MLE. Figures 2a and 2b plot the BT model rankings for win/loss and point differential.

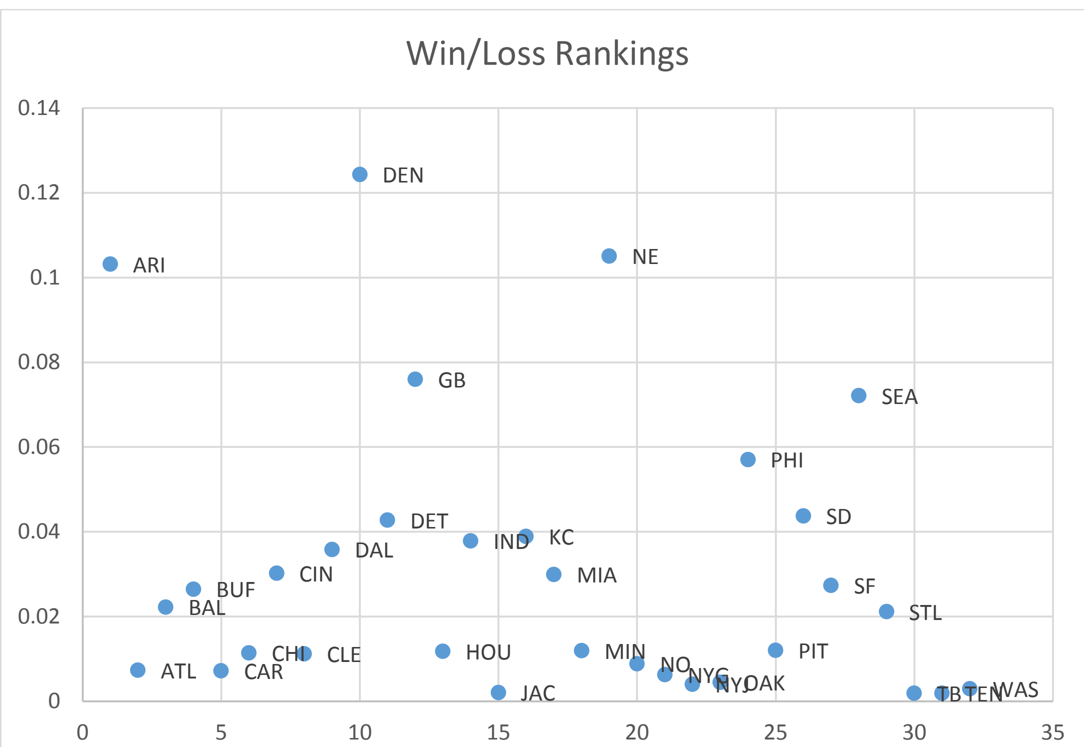

Figure 2a - BT model rankings based on teams' win/loss. The y axis is the team's γ_i

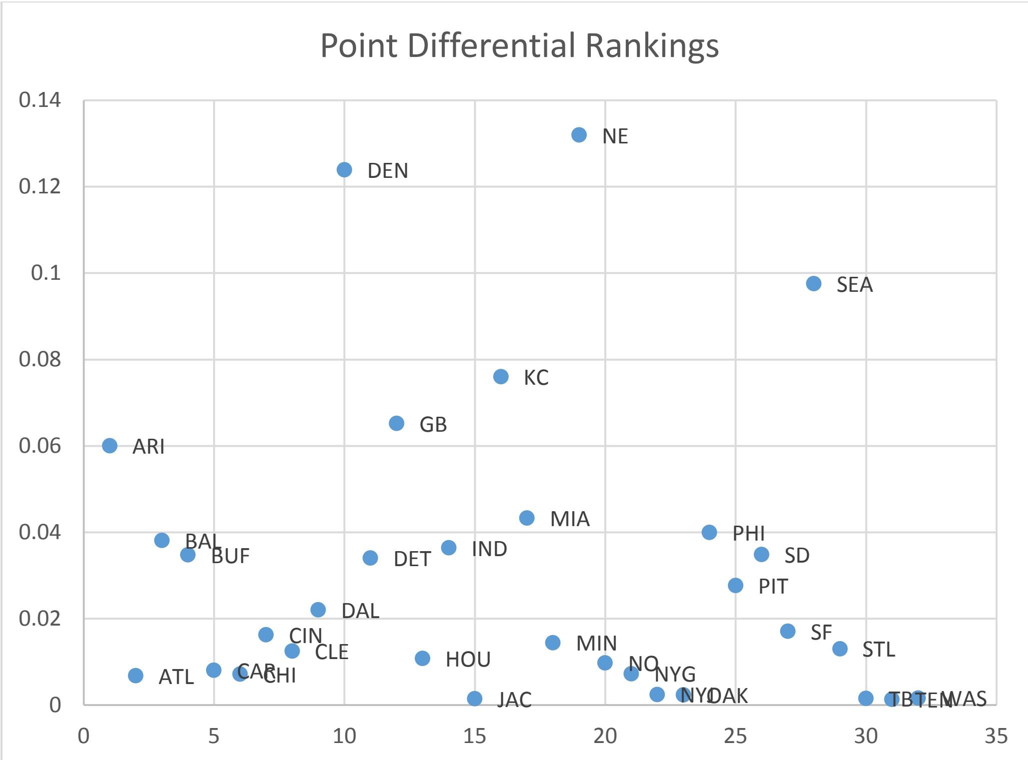

Figure 2b – BT model rankings based on teams' number of possessions won by. The y axis is the team's γ_i

The modeling results led to some interesting observations. Denver is rated higher than NE in the first model, even though NE won the head-to-head. This is because the opponents Denver defeated were stronger as a whole than the group NE beat. However, the 22 point beating NE put onto Denver was given more weight in the second model resulting in NE being ranked higher than Denver. However, the second model has KC (7-6) ranked higher than GB (10-3), due to the fact that KC beat NE by 27 so when NE moved up, it dragged KC with it. The graphs themselves are a little misleading, because the rankings themselves are not absolute. In the second model, KC has a $\gamma_i = 0.076$ while GB is at $\gamma_i = 0.065$, which translates to $P(KC\ beats\ GB) = \frac{0.076}{0.076+0.065} = 0.539$. If the public perception is to be relied upon, $P(KC\ beats\ GB) \approx 0.4$ could be plausible and not far from 0.539 after accounting for injuries, recent performances, etc. Most of the inconsistencies in the results of these two models can be explained by variability and small sample size (even with 208 games played, very rarely have the same two teams played more than once and only 197 out of a possible 496 matchups).

BT W/L	BT PtDiff	ESPN	NFL.com
DEN	NE	GB	GB
NE	DEN	NE	NE
ARI	SEA	DEN	DEN
GB	KC	SEA	SEA
SEA	GB	ARI	IND
PHI	ARI	IND	ARI
SD	MIA	PHI	PHI
DET	PHI	DAL	DAL
KC	BAL	DET	DET
IND	IND	PIT	PIT
DAL	SD	SD	KC
CIN	BUF	BAL	CIN
MIA	DET	CIN	SD
SF	PIT	KC	BAL
BUF	DAL	MIA	MIA
BAL	SF	SF	HOU
STL	CIN	BUF	BUF
PIT	MIN	HOU	STL
MIN	STL	CLE	SF
HOU	CLE	STL	CLE
CHI	HOU	ATL	MIN
CLE	NO	MIN	ATL
NO	CAR	NO	CAR
ATL	NYG	CAR	NO
CAR	CHI	CHI	NYG
NYG	ATL	NYG	CHI
OAK	NYJ	WAS	TB
NYJ	OAK	NYJ	WAS
WAS	WAS	OAK	OAK
JAC	TB	TB	JAC
TEN	JAC	JAC	NYJ
TB	TEN	TEN	TEN

Comparing the power rankings from the two BT models with those made by ESPN and NFL.com, it seems that the model gets the “echelons” of teams is pretty consistent. That is, the elite teams from BT are the same as the ones in the online rankings, though not necessarily in the same order. One of the major factors creating this difference is the lack of bias towards more recent games in the BT models. ARI’s drop-off and SEA’s recent rise are not accounted for very well in the model. Also, key injuries and other factors that aren’t captured by the W’s and L’s are not present in the BT models. Though, who’s to say the ESPN analysts aren’t putting too much stock into the “eye test” for teams ignoring the hard data?

Another interesting application for the BT models is predicting the playoffs. Using the computed $\boldsymbol{\gamma}$, probabilities for each playoff-hopeful team winning its last three games can be estimated. The best way to simulate the playoff scenarios would be doing Monte Carlo simulations using the win probabilities predicted by the Bradley-Terry models. However, I went with a simpler approach by calculating the expected number of wins for each team over the last 3 weeks to create the following end of season standings.

W/L	Predictions			Pt. Diff
E(wins)		NFC		E(wins)
12.36	GB	1	GB	12.29
12.21	ARI	2	ARI	11.98
11.47	PHI	3	PHI	11.45
6.81	NO	4	NO	7.02
10.93	DET	5	SEA	11.35
10.91	SEA	6	DET	10.87
		AFC		
12.54	NE	1	DEN	12.65
12.51	DEN	2	NE	12.53
11.23	IND	3	IND	11.36
10.23	BAL	4	BAL	10.49
9.13	KC	5	PIT	9.7
8.64	CIN	6	KC	9.39

Note that tiebreakers were not computed (for example DEN finishing as the 1 seed over NE even though NE owns the tiebreaker) because a higher expected wins number can easily turn into a 1 win advantage (so DEN could end up 13-3 to NE's 12-4 and the head-to-head won't matter). It's interesting to note that the γ_i for DEN was higher than NE in the W/L side and lower than NE in the Pt. Diff side, yet the standings show the opposite. This meant that the changes to the $\gamma_i's$ of DEN and NE's opponents also shifted by amounts that gave DEN more expected wins than NE.

The results of the Bradley-Terry modeling seem to be pretty satisfactory, considering only a fraction of the pertinent data on NFL games was used. The logical next step would be taking other tangible factors into account. Home field advantage and turnover luck are two such factors. Even with some important pieces left out, the BT model seemed to work very well given that it only uses a fraction of what others consider when ranking NFL teams. With further refinement, a Bradley-Terry model would seem to be exceptional way to rank NFL teams and make predictions with. Or you could just get rich in Vegas.

References

Hunter, D. R. (2004). “MM Algorithms for Generalized Bradley-Terry Models”. *The Annuals of Statistics,* Vol. 32, No. 1, 384-406. http://sites.stat.psu.edu/~dhunter/papers/bt.pdf

Tierney, Luke. (1997). “Fitting a Bradley-Terry Model”. http://homepage.divms.uiowa.edu/~luke/xls/glim/glim/node8.html

http://www.pro-football-reference.com/