Mathematical Modeling Project: BCS Computer Rankings

Landon Carter, Calvin Deng, Jason Liang

October 26, 2012

Abstract

In this paper, we introduce three methods to rank teams in a sparse schedule (meaning that the number of games played per team is small compared to the number of teams). One such example is the BCS system, in which 124 teams play an average of 12-13 games. The BCS rankings determine which universities will be allocated millions of dollars in revenue from profitable college football bowl games. Recent controversy has surrounded the algorithms used in part to determine the BCS rankings, citing their differences from human polls. We develop new metrics that include more factors, such as score and various offensive and defensive statistics, than computers currently take into account. Our use of more data to rank teams result in the generation of rankings that are more in agreement with human rankings.

1 Problem

College football rankings for the FBS (Football Bowl Subdivision) have been a subject of controversy since they were first used in the early 2000s. The currently used official poll is the BCS (Bowl Championship Subdivision) Ranking. This ranking system takes into account two human polls: the Coaches' Poll and the Harris Poll, and six different computer rankings.

One main difficulty with comparing teams under the BCS standings is that there are relatively few games compared to the number of teams. Contrast this with college basketball, where teams often play 2 to 3 times more games, and the NFL, where there are about as many games but a fourth as many teams. Indeed, different ranking systems will give different rankings for the teams in the BCS standings, and we aim to make a more accurate system ranking

All of the current computer algorithms are based on win-loss record and strength of schedule, which is determined by the win-loss record of a team's opponents. A few rankings take preseason projections into account, but decrease its importance as the season goes on. No computer ranking factors in points scored in order to discourage running up scores, and no ranking accounts for other statistics such as total offensive yards, sacks, etc. The difficulty with only ranking teams by win-loss record in college football is that very few games are played, so there is not a large sample size (as mentioned before).

For many years, critics have complained that the computers do not generate accurate rankings, citing the discrepancies between human and computer polls. We aim to develop new computer algorithms that take into account factors that the the ranking systems currently in use do not, in order to more closely mimic human polls.

For comparison with our calculated rankings, here are the BCS, Sagarin (a computer ranking), and Harris rankings, current as of Friday, October 26:

Ranking	BCS Team	W	L	Sagarin Team	W	L	Harris Team	W	L
1	Alabama	7	0	Alabama	7	0	Alabama	7	0
2	Florida	7	0	Florida	7	0	Oregon	7	0
3	Kansas St.	7	0	Oklahoma	5	1	Florida	7	0
4	Oregon	7	0	Oregon	7	0	Kansas St.	7	0
5	Notre Dame	7	0	Kansas St.	7	0	Notre Dame	7	0
6	LSU	7	1	Notre Dame	7	0	LSU	7	1
7	Oregon St.	6	0	Texas A&M	5	2	Oklahoma	5	1
8	Oklahoma	5	1	South Carolina	6	2	Oregon St.	6	0
9	USC	6	1	Texas Tech	6	1	USC	6	1
10	Georgia	6	1	Florida St.	7	1	Florida St.	7	1
11	Mississippi St.	7	0	LSU	7	1	Georgia	6	1
12	Florida St.	7	1	USC	6	1	Mississippi St.	7	0
13	South Carolina	6	2	Oregon St.	6	0	Clemson	7	1
14	Texas Tech	6	1	Stanford	5	2	Louisville	7	0
15	Rutgers	7	0	Arizona St.	5	2	Rutgers	7	0
16	Louisville	7	0	Texas	5	2	South Carolina	6	2
17	Stanford	5	2	Michigan	5	2	Texas Tech	6	1
18	Clemson	7	1	Ohio St.[1]	8	0	Stanford	5	2
19	West Virginia	5	2	TCU	5	2	Boise St.	6	1
20	Texas A&M	5	2	Oklahoma St.	4	2	Michigan	5	2
21	Boise St.	6	1	Georgia	6	1	Texas A&M	5	2
22	Michigan	5	2	Mississippi St.	7	0	West Virginia	5	2
23	Texas	5	2	Clemson	6	1	Ohio	7	0
24	Ohio	7	0	Arizona	4	3	Texas	5	2
25	Wisconsin	6	2	Rutgers	7	0	TCU	5	2

This problem is of vital importance because the final BCS Rankings help determine the participants in the five BCS Bowls: the National Championship, Rose, Orange, Sugar, and Fiesta. Teams that are selected to play in these games gain millions of dollars in revenue for their school, so fair allocation of these valuable spots is of prime importance in the intercollegiate athletics landscape.

2 Assumptions

Assumption 1. There are two subdivisions of Division I college football: the FBS (Football Bowl Subdivision) and FCS (Football Championship Subdivision). For our purposes, we focused on FBS teams because only they are eligible for BCS bowls. However, FBS teams play FCS teams, so in order to simplify the ranking system we discounted all games played between FBS and FCS teams. This simplified the ranking but also introduced some inaccuracy into the standings.

Assumption 2. In Model 3, we assume that all statistics have some correlation with the result of a game. This is justified because the large amount of data that we have minimizes random correlations.

[1] Ohio St. is not eligible to be ranked in the BCS poll due to violations the program has incurred. However, we have chosen to show it in our rankings and models for completeness.

3 Definitions

Before proceeding, we provide a few definitions.

Definition 1. We define G_t to be the number of games team t has played.

Definition 2. We define O_t to be the set of all opponents of team t.

Definition 3. We define W_t to be the number of wins team t has, and L_t to be the number of losses team t has.

Definition 4. We define $i \succ j$ to be team i beating team j, and $i \prec j$ to be team i losing to team j.

4 Models

4.1 Model 1: Iterated Strength of Schedule

In this most basic model, we develop a self-consistent ranking based on win-loss record. In this metric, we take into account both win-loss record and strength of schedule. Thus we have our metric:

$$R_t = \frac{W_t + \sum_{o \in O} R_o}{2G_t}$$

We iterate this metric, reassigning team ratings until we reach a stable point. Once the metric has stabilized, the teams are ranked in descending order. The result is a value between 0 and 1 that shows both how many wins each team has and the strength of the teams each team has played.

The primary purpose of Model 1 is as a stepping stone for a later Model (Model 3)

Here is the current ranking for Model 1:

Ranking	Team	W	L	Rating
1	Notre Dame	7	0	0.7747
2	Florida	7	0	0.7717
3	Kansas St.	7	0	0.7655
4	Oregon St.	6	0	0.7573
5	Oregon	7	0	0.7543
6	Alabama	7	0	0.7521
7	Ohio St.	8	0	0.7381
8	Rutgers	7	0	0.7202
9	Mississippi St.	7	0	0.7140
10	Texas Tech	6	1	0.7009
11	Louisville	7	0	0.6970
12	LSU	7	1	0.6873
13	Ohio	7	0	0.6778
14	Oklahoma	5	1	0.6762
15	USC	6	1	0.6582
16	Boise St.	6	1	0.6564
17	Toledo	7	1	0.6558
18	Florida St.	7	1	0.6533
19	Stanford	5	2	0.6498
20	Georgia	6	1	0.6475
21	Clemson	6	1	0.6386
22	South Carolina	6	2	0.6377
23	Tulsa	7	1	0.6364
24	Louisiana Tech	6	1	0.6309
25	Cincinnati	5	1	0.6274

4.2 Model 2: Probabilistic

In our next approach, we assign a power rating to teams with the goal of maximizing the probability that the current situation happens. We assume that if team 1 has a rating of r_1 and team 2 has a rating of r_2, then team 1 wins with probability $P(r_1, r_2) = \frac{e^{r_1}}{e^{r_1}+e^{r_2}} = \frac{1}{1+e^{r_2-r_1}}$. In addition, we assume that the teams follow a normal distribution, $N(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{\sigma^2}}$. Thus the probability distribution given a set of ratings $r_1, r_2, \ldots, r_N$ and games $g_1 = (w_1, l_2), \ldots, g_m = (w_m, l_m)$ (where $i, j \in \{1, 2, \ldots, N\}$) is proportional to

$$p(r_1, r_2, \ldots, r_n) \sim f(r_1, r_2, \ldots, r_n) = e^{\frac{\sum_{i=1}^{N} r_i^2}{\sigma^2}} \prod_{i=0}^{m} \frac{1}{1 + e^{r_{l_i} - r_{w_i}}}$$

This model aims to maximize the above quantity.

To do this, note that for each variable r_k,

$$f = \left(e^{-\frac{\sum_{i \neq k} r_i^2}{\sigma^2}} \prod_{k \notin g_i} \frac{1}{1 + e^{r_{l_i} - r_{w_i}}} \right) \left(e^{-\frac{r_k^2}{\sigma^2}} \prod_{k \succ i} \frac{1}{1 + e^{r_i - r_k}} \prod_{k \prec i} \frac{1}{1 + e^{r_k - r_i}} \right)$$

Here the first product is independent of the variable r_k. Thus we have

$$\frac{\partial f}{\partial r_k} = \left(e^{-\frac{\sum_{i \neq k} r_i^2}{\sigma^2}} \prod_{k \notin g_i} \frac{1}{1+e^{r_{l_i}-r_{w_i}}} \right) \frac{\partial}{\partial r_k} \left(e^{-\frac{r_k^2}{\sigma^2}} \prod_{k \succ i} \frac{1}{1+e^{r_i - r_k}} \prod_{k \prec i} \frac{1}{1+e^{r_k - r_i}} \right)$$

However,

$$\begin{aligned}
\frac{\frac{\partial}{\partial r_k}\left(e^{-\frac{r_i^2}{\sigma^2}} \prod_{k \succ i} \frac{1}{1+e^{r_i - r_k}} \prod_{k \prec i} \frac{1}{1+e^{r_k - r_i}} \right)}{e^{-\frac{r_i^2}{\sigma^2}} \prod_{k \succ i} \frac{1}{1+e^{r_i - r_k}} \prod_{k \prec i} \frac{1}{1+e^{r_k - r_i}}} &= \frac{\frac{\partial}{\partial r_k} e^{-\frac{r_k^2}{\sigma^2}}}{e^{-\frac{r_k^2}{\sigma^2}}} - \sum_{k \succ i} \frac{\frac{\partial}{\partial r_k}(1+e^{r_i - r_k})}{1+e^{r_i - r_k}} - \sum_{k \prec i} \frac{\frac{\partial}{\partial r_k}(1+e^{r_k - r_i})}{1+e^{r_k - r_i}} \\
&= -\frac{2}{\sigma^2} r_k + \sum_{k \succ i} \frac{e^{r_i - r_k}}{1+e^{r_i - r_k}} - \sum_{k \prec i} \frac{e^{r_k - r_i}}{1+e^{r_k - r_i}} \\
&= -\frac{2}{\sigma^2} r_k + \sum_{k \succ i} \frac{1}{1+e^{r_k - r_i}} - \sum_{k \prec i} \frac{1}{1+e^{r_i - r_k}}
\end{aligned}$$

Since at every local maximum of f, we must have $\frac{\partial f}{\partial r_k} = 0$ for all k, by the above equation we get

$$r_k = \frac{\sigma^2}{2} \left(\sum_{k \succ i} \frac{1}{1+e^{r_k - r_i}} - \sum_{k \prec i} \frac{1}{1+e^{r_i - r_k}} \right)$$

for all $1 \leq k \leq N$ at a local maximum.

We determine values for the power ratings $r_1, r_2, \ldots, r_N$ by using repeated iteration with binary search on a computer and the fact that $-\frac{2}{\sigma^2} r_k + \sum_{k \succ i} \frac{1}{1+e^{r_k - r_i}} - \sum_{k \prec i} \frac{1}{1+e^{r_i - r_k}}$ is decreasing in r_k. Sorting these power ratings in order gives us our desired ranking. We get consistent numbers every time we run our program, so we assume (but as of now cannot prove) that the above system has only one solution.

In this model, a major consideration is the value of σ. We notice that a larger σ places greater emphasis on strength of schedule, while a smaller σ places smaller emphasis on win-loss record. Here is the correlation constant between win-loss record and rating for the teams for various values of σ:

σ	r^2	σ	r^2
0.5	0.9567	5.0	0.7418
1.0	0.9309	5.5	0.7301
1.5	0.8955	6.0	0.7196
2.0	0.8619	6.5	0.7102
2.5	0.8331	7.0	0.7017
3.0	0.8087	7.5	0.6940
3.5	0.7881	8.0	0.6869
4.0	0.7704	8.5	0.6804
4.5	0.7552	9.0	0.6743

Here is the ranking we get when we set $\sigma = 5.0$:

Ranking	Team	W	L	Rating
1	Florida	7	0	5.2064
2	Kansas St.	7	0	4.8286
3	Notre Dame	7	0	4.2823
4	Oregon St.	6	0	3.9919
5	LSU	7	1	3.9754
6	Oregon	7	0	3.9673
7	Alabama	7	0	3.9516
8	Oklahoma	5	1	3.7047
9	Ohio St.	8	0	3.5572
10	Texas Tech	6	1	3.1994
11	Mississippi St.	7	0	2.9953
12	South Carolina	6	2	2.9910
13	Ohio	7	0	2.7098
14	Rutgers	7	0	2.6562
15	Stanford	5	2	2.6073
16	Louisville	7	0	2.6071
17	Texas A&M	5	2	2.4770
18	West Virginia	5	2	2.4670
19	Georgia	6	1	2.4268
20	Arizona	4	3	2.3326
21	USC	6	1	2.1561
22	Michigan	5	2	2.0829
23	Texas	5	2	2.0568
24	Toledo	7	1	1.8079
25	Nebraska	5	2	1.7007

Note that like in Model 1 and BCS computers, this ranking only considers wins and losses.

4.3 Model 3: Analytic Hierarchy Process

We take into account other statistics besides score and win-loss record in this model. This will allow computer and human polls to be more in agreement with each other, since human perception of teams is not only based on their win-loss record and strength of schedule. For example, the consensus number one team throughout the season, according to human rankings, has been Alabama, yet it has never occupied the top spot in the average computer rankings. The high placement of Alabama in human rankings is mainly due to the perceived strength of its defense, which is the best in the country by nearly every statistical measure.

In order to account for offensive and defensive statistics when evaluating teams, we use an analytic hierarchy process to produce an accurate ranking. This technique has been analyzed and has been found to be statistically justified [1]. To produce pairwise comparison matrices, we first analyzed 65 different game statistics to determine correlations between specific statistics and winning games [2].

Let p_i be the probability that in a game, the team with the more desirable value of statistic i wins the game.

Let g_j be the modified percentage of games played in which team j has a more desirable value of a statistic than their opponent according to Laplace's rule of succession (that is, $g_j = \frac{w_j+0.5t_j+1}{w_j+l_j+t_j+2}$, where w_j, l_j, t_j are the wins, losses and ties in each category for team j)

We then used the method from Model 1 to calculate new ratings g'_j for each statistic.

In the pairwise comparison matrix for criteria, $a_{ij} = \frac{p_i - 0.5}{p_j - 0.5}$ for criteria i and j. For the pairwise comparison matrices comparing the alternatives, $b_{ij} = \frac{g'_i}{g'_j}$.

The pairwise comparison matrices are then processed using standard methods to calculate priorities, then teams are ranked with larger priorities corresponding to higher ranks.

Here are the top 25 teams under the new calculated rankings:

Ranking	Team	W	L	Rating
1	Alabama	7	0	0.05452
2	Oregon	7	0	0.04991
3	Kansas St.	7	0	0.04777
4	Notre Dame	7	0	0.04766
5	Florida	7	0	0.04718
6	LSU	7	1	0.04516
7	South Carolina	6	2	0.04455
8	Ohio St.	8	0	0.04399
9	Oregon St.	6	0	0.04398
10	Florida St.	7	1	0.04387
11	USC	6	1	0.04327
12	Mississippi St.	7	0	0.04249
13	Oklahoma	5	1	0.04188
14	Ohio	7	0	0.04179
15	Rutgers	7	0	0.04166
16	Arizona St.	5	2	0.04118
17	Stanford	5	2	0.04092
18	Texas A&M	5	2	0.04034
19	Texas Tech	6	1	0.04025
20	Wisconsin	6	2	0.03998
21	Boise St.	6	1	0.03973
22	Louisiana Tech	6	1	0.03915
23	Michigan	5	2	0.03848
24	Tulsa	7	1	0.03845
25	Toledo	7	1	0.03835

5 Results

Here are the rankings that we determined from Models 2 and 3, as compared to the BCS rankings (Model 1 is omitted as it is primarily used as a stepping stone to Model 3).

Ranking	BCS Team	W	L	Probabilistic Team	W	L	AHP Team	W	L
1	Alabama	7	0	Florida	7	0	Alabama	7	0
2	Florida	7	0	Kansas St.	7	0	Oregon	7	0
3	Kansas St.	7	0	Notre Dame	7	0	Kansas St.	7	0
4	Oregon	7	0	Oregon St.	6	0	Notre Dame	7	0
5	Notre Dame	7	0	LSU	7	1	Florida	7	0
6	LSU	7	1	Oregon	7	0	LSU	7	1
7	Oregon St.	6	0	Alabama	7	0	South Carolina	6	2
8	Oklahoma	5	1	Oklahoma	5	1	Ohio St.	8	0
9	USC	6	1	Ohio St.	8	0	Oregon St.	6	0
10	Georgia	6	1	Texas Tech	6	1	Florida St.	7	1
11	Mississippi St.	7	0	Mississippi St.	7	0	USC	6	1
12	Florida St.	7	1	South Carolina	6	2	Mississippi St.	7	0
13	South Carolina	6	2	Ohio	7	0	Oklahoma	5	1
14	Texas Tech	6	1	Rutgers	7	0	Ohio	7	0
15	Rutgers	7	0	Stanford	5	2	Rutgers	7	0
16	Louisville	7	0	Louisville	7	0	Arizona St.	5	2
17	Stanford	5	2	Texas A&M	5	2	Stanford	5	2
18	Clemson	7	1	West Virginia	5	2	Texas A&M	5	2
19	West Virginia	5	2	Georgia	6	1	Texas Tech	6	1
20	Texas A&M	5	2	Arizona	4	3	Wisconsin.	6	2
21	Boise St.	6	1	USC	6	1	Boise St.	6	1
22	Michigan	5	2	Michigan	5	2	Louisiana Tech	6	1
23	Texas	5	2	Texas	5	2	Michigan	5	2
24	Ohio	7	0	Toledo	7	1	Tulsa	7	1
25	Wisconsin	6	2	Nebraska	5	2	Toledo	7	1

6 Analysis of Solution

There are two methods by which we can analyze our rankings: first, by comparing to other ranking methods; and second, by predicting the results of games. By comparison, we notice that our two win/loss models closely resemble the rankings presented by other computer ranking systems. We also notice that the results of Model 3 are similar to both human polls used to determine the BCS rankings. This indicates that the model is a better approximation of the true rankings, because we are able to factor in as many or more statistics than humans can, then rank the teams in a mathematically justifiable and exact way. This is the ideal all computer ranking systems hope to achieve: to be able to rank teams using as many or more statistics as humans, without any of bias or emotion clouding the judgement.

7 Future Work

Our current models all work in a so-called "black and white" mode, where all wins and losses are the same, regardless of how lopsided the result is. One way to account for large disparities in score would be to give each win or loss a weight based on how lopsided the win is.

One other way our model could be improved is to extend the probabilistic model (Model 2) to Model 3, which currently implements Model 1 when making modified ratings for each category.

8 Conclusions

In this paper, we have introduced several methods to rank teams in a league with a sparse schedule.

9 References

[1] Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process (1994). Pittsburgh: RWS. ISBN 0-9620317-6-3. "A thorough exposition of the theoretical aspects of AHP., Thomas L. Saaty".

[2] "Split Statistics." Collegefootballdata.org, 14 Oct. 2012. Web. 26 Oct. 2012.

<http://www.cfbstats.com/blog/college-football-data/>.