Post Graduate Department of Statistics, University Of Delhi

Application of Linear Systems in Sports Rankings

Linear Algebra Project Dissertation

Guide: Prof Garima Priyadarshini

Saikat Ghosh (M.Sc. Statistics)

15

Contents:

Introduction :

Rankings are important for many different applications in a wide variety of fields; in essence, any time there are multiple options to choose from, it is generally useful to have a systematic way of deciding which choice is the best. Whether considering colleges, cars or web pages, people value well-designed ranking systems that help them to make informed decisions. Ranking systems are especially important within the world of sports, where team rankings can make or break a season or a career.

A natural way to compare two athletic teams exists already; they play against each other, and their relative order is determined by the game's outcome. However, given the variety of complicated factors that can decide who wins or loses, these one-to-one comparisons can't reliably be extended to give a complete ordering of a set of many teams -- that is, beating one team doesn't in turn guarantee superiority over every team that they have already beaten.

Some method is needed to take the basic information provided by individual games and consolidate it into a measure of overall strength.

How to rank

• Vote (possibly by "experts")

• AP poll

• The tournament selection "seeds"

• Use Math to process data and try to come up with a rating that reflects the results of games played before the tournament

A "fair" ranking system should

• utilize simple mathematical techniques

• be available for verifying by non-specialists

• use win–loss information only (or limit score margins)

• produce reasonable and unbiased results.

Simple Win Percentage Approach:

Calculation: # of wins/# of games (this is the winning percentage) The team with the highest winning percentage is #1 and so on. Problems in it is a team that plays lots of weak teams may have a higher winning percentage than a team that plays lots of strong teams.

An Example:

teams	Spurs	Man U	Chelsea	Liverpool
Spurs	x	LW	W	L
Man U	WL	x	WW	L
Chelsea	L	LL	x	L
Liverpool	W	W	W	x

We calculate winning percentage for each team = #wins/#games

Spurs had 2 wins and 2 losses so their winning percentage was =2/4=0.5

Man Utd had 3 wins and 1 loss so their winning percentage was =3/5=0.6

Chelsea had 0 win and 4losses so their winning percentage was =0/4=0.0

Liverpool had 3 wins 0 losses so their winning percentage was =3/3=1.0

Results:

• So based on this criteria Liverpool would be #1 (the most likely to win)

• Man Utd would be #2, Spurs would be #3 and Chelsea would be #4.

• And according to these rankings if Liverpool played Chelsea then Liverpool should win and if Man Utd played Spurs, Man Utd should win.

• Most people would agree that based on this data Liverpool should be strongly favored over Chelsea. However if we look at Man Utd and Spurs it appears that the main reason Man Utd was ranked above Spurs was that Man Utd played Chelsea twice and Spurs only played Chelsea once.

Fate of Fixtures

• One method to try to get a “better ranking” is to try to incorporate the idea of “Fixture Toughness”. So instead of “all fixtures are equal" we will try to incorporate a method to count a win over a strong team as weighing more than a win over a weak team. So a team that plays lots of strong teams may be ranked higher than a team with a higher winning percentage that plays lots of weak teams.

Vectors for our example

• Spurs=team1, Man Utd. = team2, Chelsea = team3, Liverpool = team4 and the first coordinate of each vector corresponds to Spurs and so on.

• “w” is the wins vector so w = [2, 3, 0, 3]'

• “l” is the losses vector so l = [2, 2, 4, 0]'

• Consistency check: total wins should equal total losses since each game has a winner and loser and these games only involved the 4 teams.

• $t = w + l = [4,5,4,3]'$

This gives the total number of games played by each team.

When we calculated the winning percentage $r_i = w_i / t_i$, we solved 4 equations that were not interconnected.

Adjusted winning percentage

• At the start of the season each team's winning percentage would be 0/0 which is undefined.

So we will assign a rating to each team which is **(1+#wins) / (2+#games)**

So $\mathbf{r_i = (1+w_i)/(2+t_i)}$ and each team starts with a rating of .5, (So all teams start off the same) and since in each game the denominator increases by 1, the losers rating will decrease and the winners rating will increase after each game.

• Since the average of all ratings should be ~ .5 .The total of the ratings for all the teams played so far by team i should be roughly .5*(number of games) throughout the season.

Relative Rating

• Now when we use winning percentage, i.e. each rating only depends on wins and losses

$$(2 + t_i)\, r_i = 1 + w_i$$

Or

$$(2 + w_i + l_i)\, r_i = 1 + w_i$$

We can replace w_i by $(w_i/2 + w_i/2)$

And now we add $(l_i/2 - l_i/2) = 0$, and rearrange so we have

$$w_i = \tfrac{1}{2}(w_i - l_i) + \tfrac{1}{2}(w_i + l_i)$$

Now, The term $\tfrac{1}{2}(w_i + l_i) = \tfrac{1}{2}(t_i)$

And we will approximate this by the sum of the ratings of the teams that team i

played.(If team i played team 4 times then we will have $r_j + r_j + r_j + r_j$ in the sum)

This is the idea of **Colley Model** which brings the rating of the opponents into the

Equation and if team 'i' plays strong opponents the sum of rankings will be larger than if they play weak opponents. These rankings are called Colley Rankings.

For the above example,

- For r1 we have $(2+4)r1= 1 + ½ (2 – 2) + 2r2 +r3 +r4$
- For r2 we have $(2+5)r2= 1 + ½ (3 – 2) + 2r1 +2r3 +r4$
- For r3 we have $(2+4)r3= 1 + ½ (0 – 4) + r1 +2r2 +r4$
- For r4 we have $(2+3)r4= 1 + ½ (3 – 0) + r1+r2 +r3$
- 4 equations in the 4 unknowns r1,r2,r3,r4 ;

In matrix vector form Ax = b (Colley System of Equations).

Putting all the unknowns on the left and we have,

- $6r1 - 2r2 - r3 - r4 = 1+0 = 1$
- $-2r1 + 7r2 - 2r3 - r4 = 1 + ½ (1) = 3/2$
- $-r1 - 2r2 + 6r3 - r4 = 1 + ½ (0-4) = -3$
- $-r1 - r2 – r3 -5r4 = 1+ ½ (3-0) = 5/2$

With x= [r1 , r2, r3, r4]'

b = [1 , 3/2, -3, 5/2]'

and the coefficient matrix A is

$$\begin{pmatrix} 6 & -2 & -1 & -1 \\ -2 & 7 & -2 & -1 \\ -1 & -2 & 6 & -1 \\ -1 & -1 & -1 & 5 \end{pmatrix}$$

Construction Of A :

•The entry in row i and column j is the negative of the # of times team i played team j. So the entry in row j and column i should be the same number since the number of times team j played team i is the same as the number of times team i played team j. So the matrix A is symmetric.

- In order to construct the matrix we only need to know how many times team i played team j to get the i,j entry for i ≠ j and the total number of games played by team i to get the diagonal entry i,i . (the i,i entries are 2 +the number of games played by team i)

- Forming the augmented coefficient matrix [A b] and solving gives

$$\begin{pmatrix} 1.0000 & 0 & 0 & 0 & 0.5040 \\ 0 & 1.000 & 0 & 0 & 0.5278 \\ 0 & 0 & 1.000 & 0 & 0.2183 \\ 0 & 0 & 0 & 1.000 & 0.7500 \end{pmatrix}$$

So Spurs have a Colley Rating of 0.5040 (#3), Man Utd have 0.5278(#2), Chelsea have 0.2183(#4) and Liverpool Fc have 0.75(#1).

Generalization of the Model:

- If we had the data for N teams then we can define

N dimensional vectors w and L (w is wins and L is losses)

So t = w+L

- We will call the N dimensional vector [r1, r2,…,rN]’ = R

the ratings vector.

- We will define an N dimensional column vector ones = [1,1,…,1]’

Matrices

- We can define a matrix G as follows

G(i,j) = # of times team i played team j when $i \neq j$

0 when $i = j$

- Define a matrix T by

T(i,j) = ti for i = j

= 0 for $i \neq j$

- Note that we can find G and T from a fixture of games.

Matrix Algebra

• we can write the equations

(2I +T)*R = ones + 1/2*(W-L) +G*R

This gives us,

(2I+T)*R –G*R = (2I +T –G)*R = ones +.5*(W-L)

• So in our example A = 2I + T – G

x = R

b = ones+ 0.5(W-L)

Advantages

• We can see a that it is true for any number of teams that and we can expand the system size with any size system to accommodate our needs.

• We can find if there is always a unique solution to our system of equations regardless of the number of teams involved.

Applications and Inputs

Suppose we rank all Division I men's basketball teams in the NCAA.

That's almost 350 teams playing each other in about 5,000 games.

Once we solve the linear system, we get a rating for every team, and the higher-rated team, again, is predicted to win. But that means all would have the same prediction because the Colley method has been used.

This is where we gave our inputs.

We wanted to weight the games. We can weight by various things, whether the home team won, if they were on a winning streak. We're going to weight on time with the intent that a team with momentum might be better prepared to win in the tournament.

The main diagonal of the main matrix A from our system Ax =b changes. We had 2-plus total games going down that diagonal. Now, total games are simply total weighted games.

So if a team might have played 2 games in the second half of the season and 3 games in the first half, in that case, it played 2 plus 3 half games, or 3.5 total games.

In the same way, we look at that right-hand side vector, which had (ones + 1/2 wins minus losses).Now it’s simply weighted wins and weighted losses. So a team might have 2 wins, 1/2 a win, 1 loss, and another 1/2 of a loss. So in total, 2.5 wins and 1.5 losses ie (2.5- 1.5) = 1 as my wins minus losses.

After integrating the weights that, we can still form Ax equal b, And solve the linear system to find the ratings.

And Beyond

Now in our quest for extending the Colley Method with our ideas, we came across the works of Kenneth Massey. A few years ago he also tried to give a new look to the Colley the method and came up with the idea of integrating scores, not just wins and losses.

So if team i beats team j by 7 points,
we have the equation $r_i - r_j = 7$ for that game.
This system will probably not be consistent so we have to solve it by using least squares. This model can be weighted. Like away wins can be weighted more than home wins.
The Massey Method, though complex, was also a pleasant find for us and it was inspiring to see how he gave a totally new look to the problem at hand by capping and incorporating scores.

Acknowledgement:

- We, the presenters of this project express our deepest thanks to the department of Statistics, Delhi University, for giving us the opportunity to work with this theoretical project.
- We thank all the professors of the Statistics department, Delhi University, for their support and for helping us in moving ahead with this project.

Bibliography and Reference:

- W. N. Colley. Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained. http://www.colleyrankings.com/matrate.pdf,
- K. Massey. The college football ranking comparison. http://www.mratings.com/cf/compare.htm
- Matrix-based Methods for College Football Rankings Vladimir Boginski , Sergiy Butenko and Panos M. Pardalos
- DavidsonX: D003x.1 Applications of Linear Algebra (Part 1) by Tim Chartier
- Wikipedia
- MIT Open Course Ware.