CoachRank

Akilesh Potti, Dominick Twitty, Wenhai Yang

February 10, 2014

Contents

1 The Problem

Sports Illustrated, a magazine for sports enthusiasts, is looking for the "best all time college coach" male or female for the previous century. We now need to build a mathematical model to choose the best college coach or coaches (past or present) from among either male or female coaches in such sports as college hockey or field hockey, football, baseball or softball, basketball, or soccer. We need to take into consideration the time horizon, discuss how our model can be applied in general across both genders and all possible sports, and finally present our model's top 5 coaches in each of 3 different sports.

2 Model Assumptions

1. The idea of "best all time college coach" is subjective in nature, and to assess our model we have to focus more on qualitative assessment and justification, instead of a quantitative measure.

2. A coach's rank is determined and only determined by his or her achievement in the sport he or she coaches in, and unrelated to other factors in his personal life.

3 Data Collection

The accuracy and relevance of our results depended heavily on the quality and volume of data we could collect. That said, data collection was perhaps the most difficult part of our modeling process. We found that data on college sports is much harder to come by than data on professional sports. Within the scope of our search, less popular sports like Hockey simply did not have easily accessible data, especially from before 2001. We often found that when a site stated they had statistics on file, this meant that they had scanned copies of the original documents, which we could not parse.

For data that was available, we had to write web scrapers to collect the human-readable data. This made collecting data on the largest scales time-prohibitive, even with generous caching and multithreading. Finally, we had to be mindful of the data use policies of our sources, and ensure that our collection methods used as little bandwidth as possible.

We managed to collect a fair amount of data for Football and Basketball, the most popular college sports. All the data for these two sports was taken from Sports-Reference. For Basketball, we collected the career statistics for all coaches listed, back to 1895, for a total of over 3500 coaches. These include total wins and losses, conference championships and appearances, and NCAA tournament statistics. We also collected the outcomes and point data for every NCAA tournament game. Aforementioned limitations prevented us from collecting outcome data for every regular season game.

We collected a similar scope of data for Football. We acquired career wins, losses, and bowl game statistics for every Football coach listed, back to 1877, for a total of over 2000 coaches. We also collected the outcomes and point data for every bowl game listed. Once again, time and data policies prevented us from collecting data on every game.

Finding data for a Baseball was extremely difficult. We were only able to collect career statistics from less than 100 coaches, each with over 1000 wins. That data came from a Wikipedia article, which was backed by the NCAA's coaching records. We did not have the ability to parse the data from the original source, as it was in PDF format. Despite the huge gap in data volume, we chose Baseball because we could not find historical data on any level for the other sports listed.

4 Coaches vs Teams

One of the key assumptions we make in our approach is that in general, "best" coaches are largely independent of the team that they coach. This is a bold claim, but an important one as if this were true it allows us to simply analyze the coaches rather than the specific teams they coached and the players another way. Said another way, the mathematical models would be far more complex if the coaches and teams were tied together in terms of determining who is the "best coach". We discuss the methodology used for testing this assumption for basketball and for conserving space we don't discuss it for fotoball and baseball but test this assumption by constructing a data set that contains 3513 basketball coaches with all of the teams they have ever coached. For each $Coach, Team$ value we have the percentage of games that the team won with that coach. From this dataset we know that if the coach has achieved approximately the same win percentage then we have strong reason to believe that the coaches and teams are independent enough to only construct the mathematical model around coach data without delving into player and team specific data. To test this formally, we use the Pearson chi squared statistical test to test if there is a discrete uniform distribution across the win percentage across all the teams coached. Out of the 3513 basketball coaches, 755 of them had coached more than one team in their career. Of these coaches, 87.6% of them had a discrete uniform distribution across the number of win percentages, which largely says that we are fine in constructing mathematical models with just the coach data and not diving into the player data. We repeated similar statistical tests for football and baseball and found similar results.

5 Models

5.1 Simple Win & Loss Models

We would like to start with the simplest model as our baseline, and decide where and how we can improve.

The most indicative measure of a coach's capability is the number of games she has won and the number of games she has lost, and this information is also easy to obtain. Therefore we start by considering different algorithms to rank coaches based on their number of wins and number of loss.

5.1.1 Sort by Win Percentage

The simplest model is to compute the win percentage. Let $V = \{v_1, ...v_n\}$ denotes the set of coaches, $w(v_i)$ returns the number of games v_i has won and $l(v_i)$ returns the number of games v_i has lost. This model can be represented with the function $r_1 : V \to \mathbb{R}$:

$$r_1(v_i) = \frac{w(v_i)}{l(v_i) + w(v_i)}$$

This method has obvious flaws: it doesn't take into account the number of games played by the coaches: a coach who has only played 4 games and won 3 of them doesn't not mean she is better than a coach who has played 100 games and won 70 of them, even though the first coach has better win percentage.

5.1.2 Sort by Net Wins

To take into account the number of games played, we can rank coaches by net wins, $r_2 : V \to \mathbb{R}$:

$$r_2(v_i) = w(v_i) - l(v_i)$$

In this model, the more net wins you have, the higher you have, therefore combining both win-percentage and number of games played. Someone might ask: is a coach who has won 10 games and lost 10 games the same as a coach who has won 100 and lost 100?

5.1.3 Sort by Squared Win & Loss Ratio

Another model can be represented with the function $r_3 : V \to \mathbb{R}$:

$$r_3(v_i) = w(v_i)\frac{w(v_i)}{l(v_i) + w(v_i)}$$

One can then object, this model, as well as the previous one, will clearly have a bias towards coaches who have coached longer, since an increase in by $w(v_i)$ will have a larger effect than a higher win percentage.

The three models above all share the same fundamental flaw: they are both arbitrary and subjective. We are simply picking some heuristics and choose among them the one that we think make the most sense, when we see a problem, we adjust the formula to compensate for it. But we whink we can do better:

5.2 Non-connected Model

We instead consider a more sophisticated approach: the Wilson score confidence interval. For each coach, we consider the event that the coach will win when playing any other coach. The observed binomial proportion is the fraction of games the coach wins. Using the Wilson score, we can determine a confidence interval for the proportion of wins. We take the lower bound of this interval and use that as our coach score.

We found that this model produced unsatisfactory result when comparing very weak coaches to very strong coaches. For example, it will favor a coach with a (15-1) record to one with a (876-190) record.

We found that inserting a filtering step dramatically improved results. For Basketball, we only considered coaches with at least one NCAA appearance and above-average win ratio and number of games. We filtered Football coaches similarly, except we only considered coaches with at least one bowl win.

The strengths of this model include

- Easy to understand and implement
- Works on the most easily accessible data (career totals)
- Does not rely on tuned parameters

The weaknesses of this model include

- Does not use connectivity data between coaches
- Relies on ad-hoc heuristics to remove outliers

However, we can still do better by incorporating more data into our model.

5.3 Graphical Model

It comes very natural for us to consider the problem as a graph. The nodes of the graph are the coaches. An edge between two nodes denotes the relation between two coaches, such as the game result of them playing against each other.

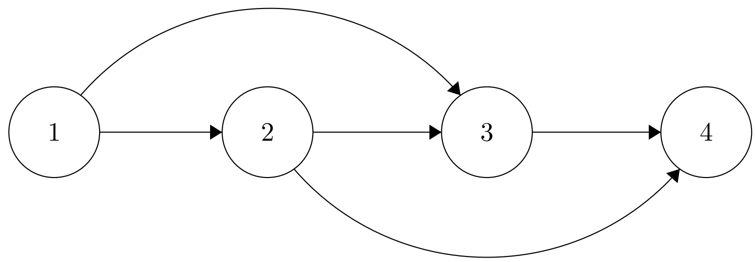
This idea of representing the problem as a graph is a very good step forward because it allow us to go beyond the simple statistics like win/loss ratio of a single coach, and gain more insight from the data.

Another advantage of this model is that it can aggregate data from different time periods. For example if coach A beats coach B and then retired in 1990, and coach B beat coach C later in 1995, then in the graph we will have a way to compare coach A and coach C because there is a path from A to C through B, even though they never played against each other.

5.3.1 Topological Ordering

A very first idea that comes to mind is Topological Ordering. In this sub-model, there exist an edge from i to j if and only if among all the games i played against j, j beat i more often than i beat j (net-loss). Then if there exists a Topological Ordering of the graph, we have a ranking of the coaches based on their rival history.

Consider the following example. Coach 1 has been beaten by both 2 and 3, coach 2 has been beaten by coach 3 and 4, coach 3 has been beaten by coach 4, and coach 4 is never beaten by anyone. We have a topological ordering in the graph: $\{1, 2, 3, 4\}$, and it is natural to conclude that coach 4 is the best coach among them. (For now, when we talk A is beaten by B, we mean net-loss, which means A is beaten by B more often than B is beaten by A. We will introduce more sophisticated measures later).



However, one natural limitation of this model is the following theorem:

Theorem 5.1 *A graph G has a topological ordering if and only if it is a DAG (directed acyclic graph).*

Proof See Section 3.6 of Algorithms Design by Kleinberg and Tardos.

To solve this issue, we think of two algorithms. One is to reduce the originial graph into a DAG: keep removing the most "unimportant" edges untils the graph is acyclic. By most "unimportant" we then have to find a way to measure the importance of an edge. In this model, the weight (or importance) an edge by the net-loss it is associated with:

$$w(e_{i,j}) = \text{number of times j beats i} - \text{number of times j beats i}$$

The algorithm then works as following:

while *G is not DAG* **do**
  $e_{min} = \operatorname*{argmin}_{e \in E} w(e)$;
  Remove $e_m in$ from G
end

However, this algorithm has apparent flaws: suppose even though coach A and coach B only encounter each other once, however it is in the SuperBowl, while the matchup between coach A and coach C is quite often but only in friendly games or early-season games and therefore the result is less important and less indicative of their comparative abilities. Therefore using the net-loss as the only measurement of the importance of an edge can possibly remove the important edge and keep the unimportant ones. We can of course tweak the edge-weight by introducing the importance of a game, the time in both coaches' career when the match happens, etc., but simply adding heuristics seem not like the best approach to the problem.

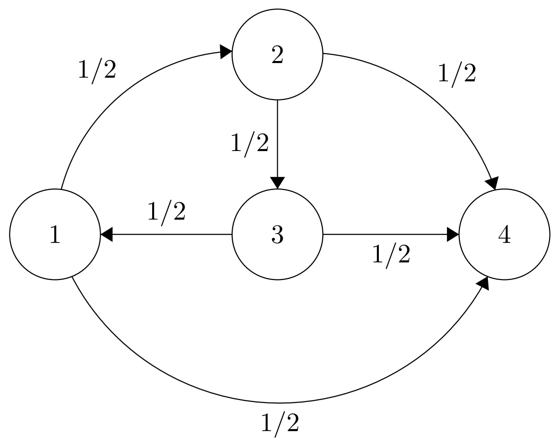
The second algorithm requires a certain "leap of faith" by introducing randomness into our model:

5.3.2 Markov Chain (Voting Model)

To solve the problem of Topogical Ordering, instead of trying to create a DAG to fit the algorithm, we decide adjust the mechanism behind ranking using Topological Ordering.

Imagine instead of having us, the spectators doing the voting and the ordering, each coach will have a **vote** of a certain **size**, which she can divide up into pieces. She can then give the chunks of the vote to different coaches which she think is definitely better coach than she is. We assume the coaches to be rational: she doesn't take into account the opinions of amateurs and public media, and only give out votes to the one she truly believe is better than her (one possible way is only give out vote to the coaches who has beaten her). Then we will rank the coaches based on how many votes each coach eventually get. However since the coaches form a graph and there even might be cycles in the game, so at each timestep a coach can be both receiving and giving. What we are interested in is the distribution of the votes at some large time.

Consider the following example. There are four coaches: 1, 2, 3, 4. Coach 1 has been beaten by 1 and 4, coach 2 has been beaten by 3 and 4, coach 3 has been beaten by 4 and 1, coach 4 has never been beaten.



The Markov Chain above shows how the vote will be given out at each timestep: coach 1 will give out half of her vote to coach 2 and the other half to 3, coach 2 will give out half of her vote to coach 3 and the other half to coach 4, coach 3 will give out half of her vote to coach 1 and the other half to coach 4. Coach 4, since she never lost to anyone, will simply keep her vote to herself. We can therefore see this can be modeled as a Markov Chain, and the distribution of votes at some large time is the stationary probability distribution of this Markov Chain:

$$\pi = \lim_{n\to\infty} \phi_0 P^n \text{ for some initial distribution } \phi_0$$

where $\mathbf{P}$ is the probability transition matrix, with $\mathbf{P_{i,j}}$ equal to the proporption of her chunk coach i will give to j (can also be interpreted as the probability of going from i to j in the Markov Chain):

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We know that if the Markov Chain is irreducible and aperiodic, then such a stationary distribution exists, and is unique, and it is the left eigenvector of $\mathbf{P}$ for the eigenvalue 1.

Irreducible & Aperiodic

Apparently, the example we show doesn't satisfy the property. To make sure the matrix we encounter will be irreducible and aperiodic, we use a similar technique as the PageRank algorithm used by Google:

1. For dead-ends (coach 4 in the example), in each step she will split her vote into equal pieces and give one piece to every coach, (go to a random state in the Markov Chain)
2. For non-deadends (coach 1, 2, 3 in the example), in each step with he will first split her vote into two, with proportion α and $1-\alpha$. With the α proportion she did what she did before, spliting it and give the pieces to coaches who have beaten her, the remaining $1-\alpha$ she will split into equal pieces and give one piece to every coach.

Therefore, after adjusting for deadends in $\mathbf{P}$, the new probability transition matrix can be represented as:

$$\mathbf{D} = \alpha\mathbf{P} + (1-\alpha)\mathbf{U}$$

where $\mathbf{U}$ with all entries equal to $\frac{1}{n}$, n being the total number of coaches (or number of states in the Markov Chain).

With the new probability transition matrix, since all entries in $\mathbf{D}$ are positive, we know by Perron-Frobenius theorem that $\mathbf{D}$ is both irreducible and aperiodic, and that the stationary distribution π exists and is unique. With π we can find the coaches with the top 5 probability and they will be the best coaches "of all time" since they get the most proportion of votes from all coaches in the long-run.

Formal Representation

$$G(V, E) \text{ represents the graph of coaches}$$

$$V = \{v_1, v_2 ... v_n\} \text{ where } v_i \text{ is coach i}$$

$$(v_i, v_j) \in E: \text{ coach j beats coach i more often than coach i beats coach j}$$

$$\mathbf{P}(i,j) = \begin{cases} \dfrac{I\{(v_i, v_j) \in E\}}{d(v_i)}, & \text{if } d(v_i) \geq 1 \\ \dfrac{1}{n}, & \text{otherwise} \end{cases}$$

$$\mathbf{D}(i,j) = \alpha\mathbf{P}(i,j) + (1-\alpha)\frac{1}{n}$$

Efficient Implementation

Since the matrix $\mathbf{D}$ is huge and all entries are non zero, calculating its eigenvalues and eigenvectors directly can be very inefficient computationally. However $\mathbf{P}$ is a sparse matrix and most of its entries are zero. To make sure we will be able to produce the eigenvector that represents the stationary distribution efficiently, we use Power Method:

$v = (1/n, ..., 1/n)$;
while $\|\alpha v\mathbf{P} + (1-\alpha)(1, ..., 1) - v\|_2 < e^{-8}$ **do**
  $v = \alpha v\mathbf{P} + (1-\alpha)(1, ..., 1)$;
end
return v;

Since $v\mathbf{P}$ can be calculated using sparse matrix $\mathbf{P}$, this algorithm is much more computationally efficient than simply calculating the eigenvectors of $\mathbf{D}$. As α increases, the time it takes for v to converge is shorter, but the difference in the ranking might be less significant, therefore this is a tradeoff and we pick $\alpha = 0.95$.

Edge Weights

In the model we introduced above, in situation where $d(v_i) \geq 1$ (which is something we assume in the rest of this section), $\mathbf{P}(i,j) = \dfrac{I\{(v_i, v_j) \in E\}}{d(v_i)}$, where $I\{(v_i, v_j) \in E\}$ is simply an indicator variable indicating whether i has a net-loss against j. However there are a lot more information we can capture, if we think more about the vote coach i will give:

1. Game's importance: The more important the game is, the more effort both coaches will put in, and therefore the game result will be more convincing and informative, and more likely coach i will be to give the vote to coach j, had coach i lost.

2. Time in Career: If it is early in the career of coach i when he lost to coach j, she is going to put less weight than the game she lost later in her career, since it is typical for a coach to learn from failure early and become a better coach later. Therefore coach i will be more convinced to give vote to someone she lost to later in her career.

3. Game Score: A game lost very close is definitely less convincing than a game entirely dominated by opponent. Therefore the more score coach i lost, the more convinced she will be to give your vote to your opponent.

There are definitely more information that can be informative, however due to the difficulty of data-mining and significance comparison, we come up with the following updated formula to calculate the weight for edge weight:

Let T be the set of all games
$f : T \to \mathbb{R}$ map a game to the score difference of the game (positive winning score, 0 if draw)
$h : T \to \mathbb{I}$ returns the importance multiplier of a game, it is sport specific since different sport have different game structure.
β : the score difference adjustment, it is specific to a type of sport. For example, in basketball β is set to be smaller than in soccer since the score difference in basketball is much larger.

$$T_{i,j} = \{t \in T : \text{ coach i lost to coach j in } t\}$$
$$w(v_i, v_j) = \max(0, \sum_{t \in T_{i,j}} h(t)\log(1 + \beta f(t)) - \sum_{t \in T_{j,i}} h(t)\log(1 + \beta f(t)))$$
$$(v_i, v_j) \in E \text{ if } w(v_i, v_j) > 0$$

$$\mathbf{P}(i,j) = \begin{cases} \dfrac{w(v_i, v_j)}{\sum_{w(v_i,v_k)>0} w(v_i, v_k)}, & \text{if } d(v_i) \geq 1 \\ \dfrac{1}{n}, & \text{otherwise} \end{cases}$$

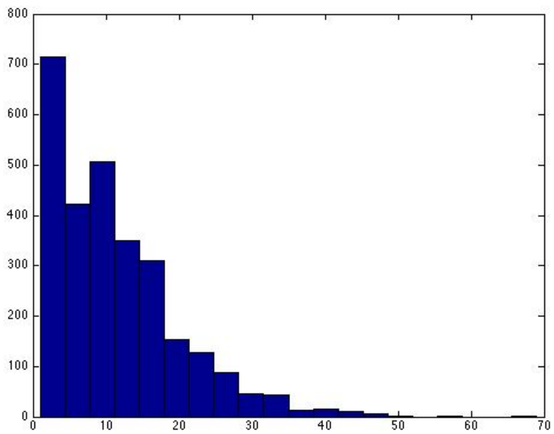
The intuition for this formula is really straightforward: for example, if a coach lost 30 points in Super Bowl, it definitely is going matter much more than a loss of 4 points in an early-season game. The weight for an edge is just a combination of the two factors: the importance of the game, and the score difference of the loss.

Parameter Estimation

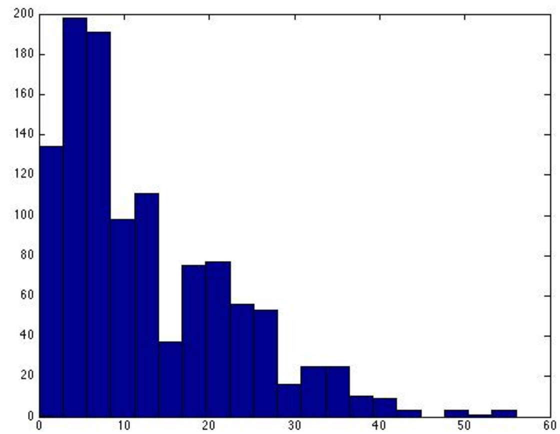
f : we approximate f it to be $\dfrac{1}{\text{maximum number of such game a team can play per season}}$. The reason for this approximation is clear, the less frequent a type of sub-game is, the more it is valued. In NCAA, a team 30 season games, 5 tournament games, and 1 championship game, therefore $h(\text{season}) = \dfrac{1}{30}$, $h(\text{tournament}) = \dfrac{1}{5}$, $h(\text{championship}) = 1$. Similarly, in College Football, $h(\text{season}) = \dfrac{1}{12}$, $h(\text{playoff}) = \dfrac{1}{2}$, $h(\text{championship}) = 1$

β : it is set to be the inverse of the medium score difference of a sport in the data. From the data, we calculated the median and set $\beta_{basketball} = \dfrac{1}{9}$, $\beta_{football} = \dfrac{1}{10}$. We do not use the mean because there are outliers in the score difference in its distribution, as we can see in the following graph:

(a) Distribution of Basketball Score Difference



(b) Distribution of Football Score Difference



There are a lot more possible variation with edge weights in the graphical model. However as more detailed information is added into the model, the less significant difference it is going to make, especially among the top results returned, where the differences between coaches are already quite large. Therefore due to time constraint the above model and method of parameter estimation is the final version we decided to use for our graphical approach. As we can see in later sections, the results are promising.

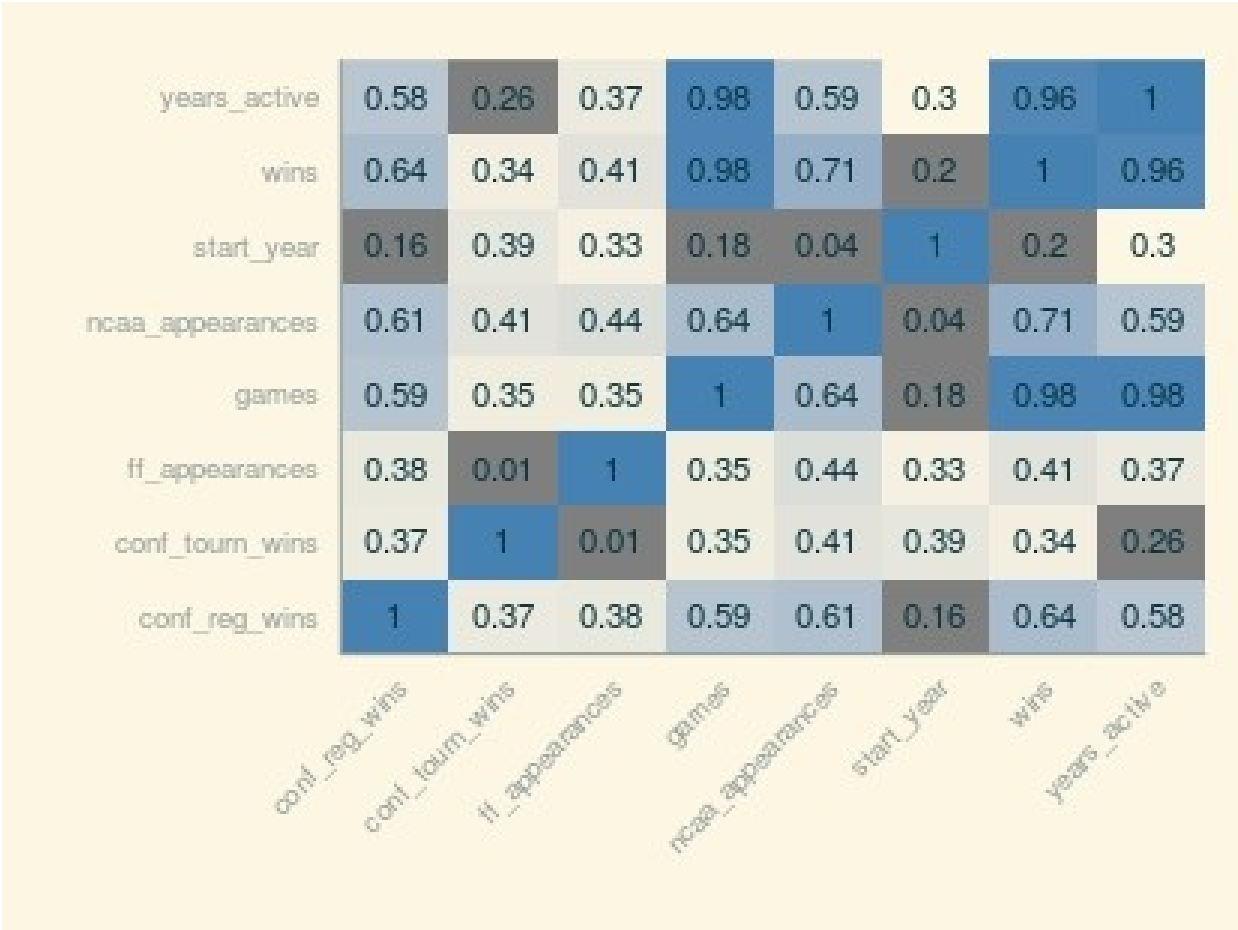
5.4 Machine Learning Model

So far we have algorithms that produce the top 5 "best coaches" according to historical data, but we coudl also interpret "best" as who has the best chance of winning in the future, or is overall "best" looking forward. Additionally, it is difficult to assess the CoachRank graphical model or more important figure out what attributes of a coach are the most important for defining "best coach". There are a lot of spors fans who may not understand the whole graph model, but can understand certain ket attributed of coaches. This leads us to the use of machine learning methods for tackling this unsupervised learning problem- we found significant difficulty finding structure within the massive amounts of sports data, but we realized that the CoachRank graphical model posed above can bring *structure* to formulate a supervised learning problem. Furthermore, we can use variable selection methods to extract out which are the most important attributes for deciding who the best coach is.

Feature Correlation

Before we start to use machine learning methods, it's important to get a feel for the data to make sure the models we produce are intuitively soundsl. After the initial data munging phase, we realized tht a lot of the attributes (now on refered to as features) were correlated, which made sense because our data sets had wins, losses, total games played, etc. To make it the most interpretable model and simple, we removed features that had a high degree of correlation. Below is the Pearson correlation matrix for basketball:

Figure 2: Pearson Correlation matrix for basketabll features



From the above matrix, we can clearly see that there are some attributes (wins, losses, games, etc.) that are colored dark blue, which we removed moving forward to create a compact set of features that explained what constituted as the "best coach". There are a wealth of supervised machine learning methods that we can use, but we have two defined goals: figure out what features makle up the best coach and assess how well the graphical model works. To best achieve these goals, we decided to use multiple linear regression to figure out how each feature correlates with the coach's CoachRank value. This is also easy to explain to sports fans.

Multiple Linear Regression

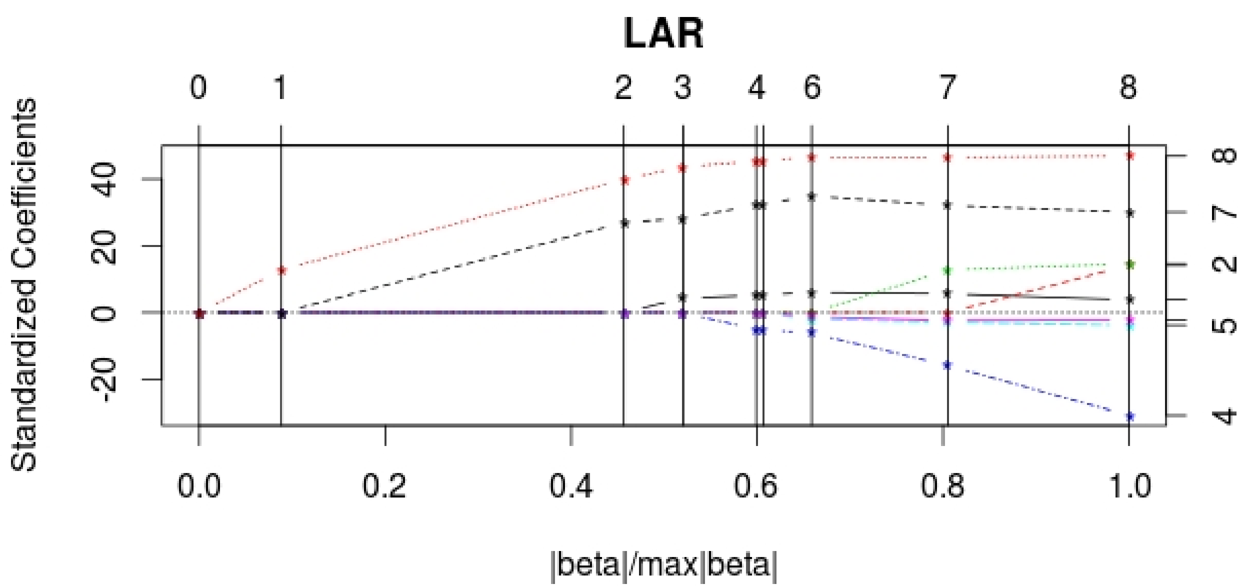
Multiple Linear Regression is a supervised machine learning method that predicts a real valued output given a set of features. More formally, the multiple linear regression function is:

$$E[CoachRankValue|X] = \alpha + \sum_{i=0}^{n} X_i$$

That is, we are finding the conditional expectation of the CoachRank value of a coach given set of features X is the set of features. Since we would not only like to predict CoachRank values for coaches but also individually analyze the features, we handled multicollinearity by using the Pearson correlation matrix above to ensure none of the features were correlated above a minimum threshold of 0.8. We also looked at the distribution of the features to make sure the homoscedasiticity, weak exogeneity, and independence assumptions of linear regression held true. Below are the results of the multiple linear regression model.

We notice that the features that are the most signficicant are $ncca-appearances$ and $ff-appearances$, which intuitively make sense and are used in the construction of the CoachRank graphical model. Additioanlly, the sign of the features make sense as the more final four and NCAA appearances a coach has the more popular or "good" they are likely to be. What's not intuitive is the real-valued coefficients of the features, and the number of features that we have: does one feature "predict" the CoachRank value of the coach better than other features? That is commonly referred to as variable selection, where there exist a wealth of methods. We use least-angle regression (LARS) for computing which features best correlate with the CoachRank. In a nutshell, LARS intializes all the coefficients of the features at zero and take the largest step in the direction of the most correlated vairbale with CoachRank, until some other feature is more correlated, at which point LARS proceeds in a equiangular direction to both features, hence the name "least-angle". At the end of the day, LARS provides us which features are the most "important" in determining the CoachRank values. The plot is shown below:

Figure 3: LARS of Basketball Features



As expected, the $ncca-appearances$ and $ff-appearances$ enter first and are thus deemed to be the most "important". Now it remains how to make the coefficients of the features more interpretable in a way that sports fans can understand it. Fortunately, intereptable models in supervised learning are a growing area of research within predictive modeling and machine learning. In the next section we discuss a natural extension to the multiple linear regression model discussed above known as Supersparse Linear Integer Models, which one of the group members is currently working on with Professor Cynthia Rudin at MIT, so are able to use the code although it's not publicly released yet.

Supersparse Linear Integer Model

Supersparse Linear Integer Models (SLIM) create predictive scoring systems that are both practical and interpretable. They are widely refered to as intrepretable models because they require users to perform only a few operations to make a prediction. They further restrict the coefficients to a certain set of integral values so that we can better understand the effect of the features on the predicted value. SLIM is formulated as a mixed integer program with an objective function that minimizes 0-1 training loss and L_0 and L_1 norm to ensure both accuracy, sparsity, and interpretability respectively. Although this is a computationally hard problem (NP-Hard), due the size of our data set and CPLEX solver from IBM we are able to obtain results within a reasonable amount of time. Additionally, as this is a classification model, we had to have a surjective mapping of our real valued CoachRank to discrete buckets (< 0.01, > 0.02, etc). Since we had a imbalanced data set from this, that is the number of coaches falling into each bucket varied widely, we used the imbalanced formulation of SLIM. After doing so, we obtained results that, as expected, are close to the results of our multiple linear regression model but now have integral values for the features.

Features	Coefficients
Adolph Rupp	
John Wooden	
Roy Williams	
Jerry Tarkania	
Dean Smith	

Tying it together

Okay, so we have these neat supervised machine learning methods that complement the complicated graphical CoachRank algorithm that we discussed earlier- so what? Well, the purpose of the machine learning methods was two fold: one to extract features from the CoachRank graphical model to actually determine how it's determining what coaches are "best" and two to understand how these features interact with each other and affect the CoachRank value of the coach in a practical and interpretable way, both of which were done through the multiple linear regression model and supersparse linear integer model.

6 Results & Validation

In this section we are just giving the results of the top-5 coaches for 3 sports. The 3 sports we choose are Male College Basketball, Male College Football, and Male College Baseball. We will start with the result using the Non-connected Model.

6.1 Result of Non-connected Model

Male College Basketball

Coach Name	Wins	Loss	Wilson Score
Adolph Rupp	876	190	0.8210
John Wooden	664	162	0.8030
Roy Williams	715	187	0.7918
Jerry Tarkania	761	202	0.7894
Dean Smith	879	254	0.7750

Male College Football

Coach Name	Wins	Loss	Wilson Score
Tom Osborne	255	49	0.8374
Bear Bryant	323	85	0.7904
Bo Schembechler	234	65	0.7811
Wallace Wade	171	49	0.7755
Woody Hayes	205	61	0.7691

Male College Baseball

Coach Name	Wins	Loss	Wilson Score
Ed Cheff	1705	430	0.7980
Mike Martin	1771	611	0.7429
Gene Stephenson	1837	675	0.7307
Augie Garrido	1874	871	0.6821
Gordie Gillespie	1893	952	0.6648

Since this model only takes into account a coach's career win and loss, the result will differ a lot from the graphical model, which capture more information:

6.2 Result of Graphical Model (Voting Model)

Using the Markov Chain model with the edge weight function that takes game's importance, and score difference into account, we computed the following result and showed the top five coaches for Male College Football and Male College Basketball (due to time-constraint, we were not able to find game data for College Male Baseball to form a graph):

Male College Football

The graph $G(V, E)$ for Male College Football have 529 nodes, 1032 edges, and 27 weakly connected component. The fact that there is no one single connected component is possibly due to the small size of our dataset. The top five coaches are:

Coach Name	Vote %
Joe Paterno	0.02421
Mack Brown	0.01728
Bear Bryant	0.01663
Lloyd Carr	0.01491
Pete Carroll	0.01338

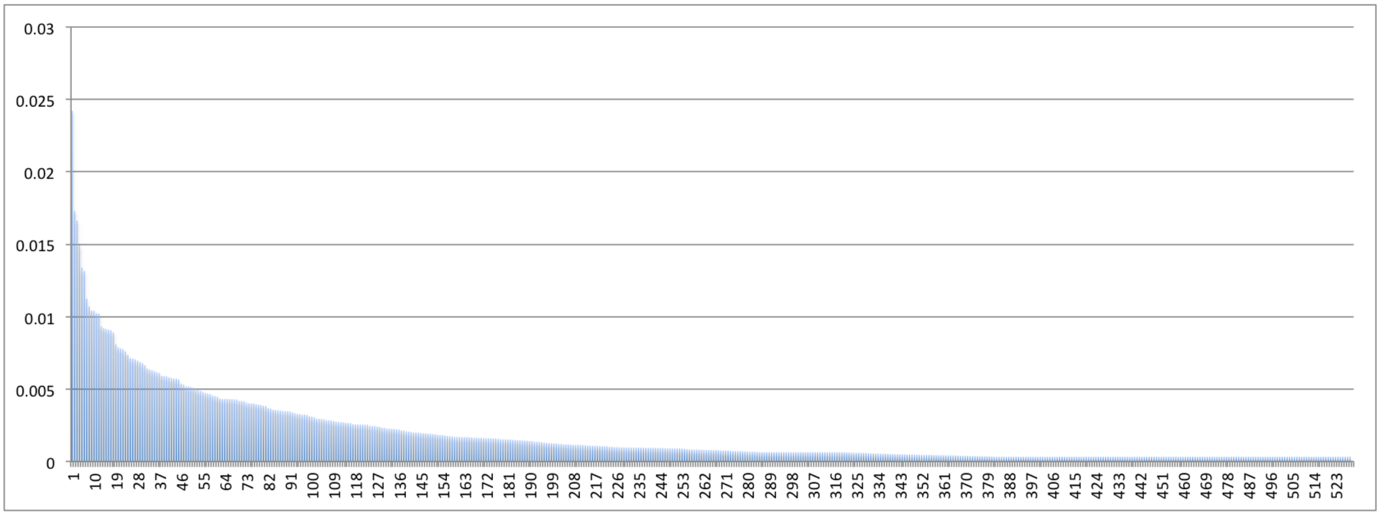
Male College Basketball

The graph $G(V, E)$ for Male College Basketball have 763 nodes, 2582 edges, and 1 weakly connected component. The fact that G is weakly connected is really useful in that it allow us to compare every two coaches in the graph, even though they are playing in different time horizon. The top five coaches are:

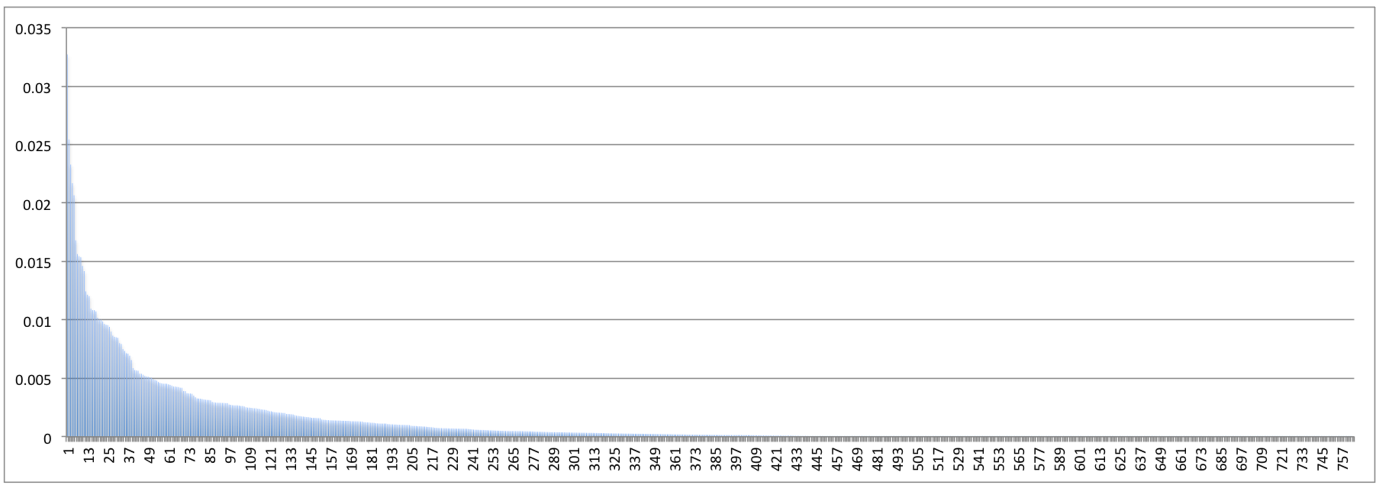
Coach Name	Vote %
Mike Krzyzewski	0.03272
Dean Smith	0.02544
Roy Williams	0.02329
John Wooden	0.02170
Rick Pitino	0.02066

If we plot the Vote % of all the coaches from high to low, we can get the following histogram:

(a) Male College Football Vote %



(b) Male College Basketball Vote %



Comparing the two graphs, we can see some clear similarities and differences.

1. The curve drops off very quickly, which means there are large differences between coaches with the top votes.

2. A small proportion of coaches hold the majority of the votes. In Male College Basketball, 5.6% of the coaches hold 50% of the votes. In Male College Football, 12.4% of the coaches hold 50% of the votes.

3. The graph for football have fatter tails than the graph for basketball, indicating that the vote are more spread out. This is a valid result since there are much more uncertainty in football than in basketball.

Assessment

Due to the subjective nature of this problem, assessment of the result can be trickly. By comparing both of our results with public polls and professional sports media, and looking at the achievements of coaches returned, we conclude that the result is consistent with public opinions.

Robustness

Varying parameters α from 0.75 to 0.95, and scaling β upwards and downwards by 10%, the top 5 coaches returned doesn't have much difference, just with minor rank movement among the top 10 coaches. This is in part due to the quick drop-off rate of the curve above, and the large score differences between top coaches. Therefore, we conclude that the result of the graphical model is valid and it is robust to change in parameters.

7 Strengths and Weaknesses

7.1 Strengths

- The graphical model allow us to be more objective in our ranking algorithm since we can understand it as the coaches voting among themselves based on their game history, instead of arbitrary tweaking of heuristics.
- We take into account not only the career data of a coach, but also the game result, game score, and the game importance of the games they play against each other.
- We have efficient implementation using power method to calculate the stationary distribution of the Markov Chain, and the results show clear differentiation between coaches, especially high-ranking ones.

7.2 Weaknesses

- Due to the limited amount of data we could collect, we could not consider coaches not in our dataset. For example, John Gagliardi, the coach with the most wins in college Football history, was not in our data set because he competed in the NAIA and NCAA Division III leagues.
- Our graphical models used only postseason games as input. We justify this heuristically by saying that only skilled coaches will be play in the postseason. However, this produces somewhat sparse graphs, especially compared to those we could have generated if we were able to collect data on every game.
- Our data for Baseball does not include any data concerning connections between coaches, therefore we could not use our graph model to choose the best baseball coach. Furthermore, we were unable to collect championship-specific data for Baseball coaches.
- Assessment is also difficult since ranking is in itself subjective. Our methods of assessment are limited to public polls online, professional sport media, coaches' achievements, and cross-validation between our different models.

8 Conclusions

Todo

9 Future work

- Obtain more game data and coach career information, and more information about the coaches to train a better model.
- Come up with an objective mechanism to assess the result of our ranking, such as incorporating more public sources.

10 Bibliography

References

[1] Pagerank Algorithm, http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf

[2] A stochastic optimization model for real-time ambulance redeployment, http://www.sciencedirect.com/science/article/pii/S0305054813000385

[3] Ithaca Demographics, http://www.ci.ithaca.ny.us/maps/index.cfm

[4] Cornell Demographics, http://www.cornell.edu/about/facts/stats.cfm

[5] Ithaca College, http://www.ithaca.edu/admission/facts/