Team Control Number

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

24791

Problem Chosen

B

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

2014 Mathematical Contest in Modeling (MCM) Summary Sheet
(Attach a copy of this page to each copy of your solution paper.)

Abstract

Our research goal is to construct a coach evaluation model which could find the top 5 all time coaches among all possible sports. Specifically, we will first build our model based on college football, basketball and baseball coaches data. During this process, Satty's Analytics Hierachy Process will be used intensively with modifications. Next, we use some reliable online ranking sources to validate the correctness of output. By setting up an innovative algorithm, we can readily train our model until the evaluation function converges to its maximum. Afterward, we furthermore introduce non-quantitative factors into our model. Finally, we discuss the potential problems and those procedures that can be improved.

Discovering the Best Coaches Ever
A Modern Approach Of Analytics Hierachy Process

February 11, 2014

Abstract

Our research goal is to construct a coach evaluation model which could find the top 5 all time coaches among all possible sports. Specifically, we will first build our model based on college football, basketball and baseball coaches data. During this process, Satty's Analytics Hierachy Process will be used intensively with modifications. Next, we use some reliable online ranking sources to validate the correctness of output. By setting up an innovative algorithm, we can readily train our model until the evaluation function converges to its maximum. Afterward, we furthermore introduce non-quantitative factors into our model. Finally, we discuss the potential problems and those procedures that can be improved.

Key Words: Ranking fitness; Analytics Hierachy Process; Two level reciprocal matrix

1 Introduction

College sport has been a heated topic nationwide for a long time, and obviously the college coaches play a significant role in determining the performance of a team. Besides the famous athletes, the best coaches also get much attention from sport enthusiasts and even the media.

When the magazine, Sports Illustrated, is trying to find the Best All Time College Coach for the last century, we may be wondering about many problems, like how to build the appropriate criteria to judge a good coach, or how to treat the senior coaches, who we are unfamiliar with. Moreover, the time witnessed huge development of some sports with growing popularity and influence, such as football and basketball, of which there are new rules and top events. Thus, it is not fair to select the outstanding coaches by using univerisal standards. In this paper, our most important job is about building a relatively comprehensive mathematical model to diminish the prejudice of gender as well as limitation due to the one-hundred-year time horizon.

Our study is based on real people and facts, so we construct the research method considering the quality and quantity of the available data from online database. Though there are many factors affecting the results, we only choose the data that can be quantified easily. Moreover, further techniques beyond quantitative methods will be given at last.

2 Background

According to NCAA Sports Sponsorship and Participation Rates Report, we can find that football, basketball and baseball are the three most popular sports on the college level, which involve great participation and attention across the nation. During the sport season, even the Division III schools treat the game as a big event. Furthermore, all these three sports boast a long history, and have grown into a mature stage.

Football, basketball and baseball all call for teamwork. Unlike individual sports, athletes' physical ability and talent matter, but there are more other factors affecting results of team sports. Therefore, the coaching job exerts a significant influence on the performance. The coach's leadership can make difference in motivating the athletes, enhance the cooperation and guide the training. How to look for the best coaches is a question worthy of discussing.

If you search the Internet at present, it is not hard to get a good number of rankings about college coaches. Then, you may get to know a vivid story about a legendary coach. These rankings are from media, used to cater for the mass and far away from being objective. So some of the rankings have a serious problem of being subjective and irrational. What is more, we are looking for best coaches for a long time period, some rankings did not cover elder-generation coaches, while some previous rankings excluded the recent famous coaches. Thus, it is important to be prudent when referring to these materials.

As for the research field, we can hardly get academic articles about mathematical methods in evaluating coaches. However, there are adequate researches in the field of scoring other than human career achievements, things like vehicles, productivity levels, service qualities and etc. Marcel Bouman, Ton van der Wiele (1992) offered a method to measure the service quality in the car industry. Tian-Shyug Lee and I-Fei Chen (2005) constructed a two-stage hybrid credit scoring model using multivariate adaptive regression splines. The most-frequently-used measure in ranking is linear or other revised regression methods. But these econometric approaches are impractical in our modelling. Unlike the thesis mentioned above, it is so difficult to find an unbiased variable on the left hand side of the regression equation, due to the limitation of time and information. Since the lack of a reliable dependent variable is a general problem in coach ranking, some researchers turn to judge the career achievements of a coach on non-quantitative properties. Cliff Mallett and Jean Côté(2006) analyzed the psychological factors beyond winning and losing in their paper, which focused more on athletes' feedback and satisfaction. They collected first-hand data through delivering and collecting questionnaires, but this method is apparently impractical to us.

3 Assumptions

1. The conference to which each team belongs will be ignored due to the high circulation of coaches among teams of different conferences and frequent change of relative competitiveness among conferences.

2. Gender of the coaches is excluded from the ranking of best coaches.

3. Only coaches who have ever taught in the Division I can be viewed as a potential "All Time Best Coach" candidate.

4. Only coaches of man sports are chosen in our model.

5. A potential best coach is always facing convex indifference curves with fame and income on x and y axis respectively. He or she will always FIRSTLY choose the indifference curve that is farthest from origin and then make trade-off between fame and income.

6. We choose years as a Division I coach, overall number of games, overall game winning percentage, overall number of participating in top-level games and top-level game winning percentage as variables in our quantitative analysis, when all these five variables are available[1].

7. The five variables mentioned above can more or less reflect two more general characteristics of a coach, which are experience and efficiency.

8. Two basic criteria for a good ranking model are that absolute position of a coach is more critical than relative position; the accuracy of the bottom part of the ranking has to be sacrificed to guarantee the accuracy of the upper parts.

9. Comparing our model with the most reliable and popular one is a good and efficient approach to test its goodness of fit.

10. Saaty's Analytic Hierarchy Process is a proper approach for a ranking model based on the available data assumption above (Assumption 6) .

11. The initial relative importance we teammates assigned in the AHP model is close to the actual relative importance.[2]

12. The judging standards of younger-generation coaches and elder-generation coaches will be different in order to avoid the time horizon problem. The adjustment and the cut-off line between young and elder coaches varies among the sporting items (football, basketball and baseball).

13. According to the Rule of Large Number, our accessible data for each variable follows a normal distribution.

[1]Some of these five variables are not available due to the time horizon problem.

[2]Close indicates that the relative importance between two variables we gave is at most 1 levels away from the actual relative importance.

14. Only quantitative variables are included in our ranking model, so factors beyond the years and outcomes of matches are ignored in the modelling part.

15. The rankings given in the websites are partially reliable. This indicates that the rankings can be used to refine or test the goodness of our model, but none of the rankings is the real truth.

4 Data Manipulation

1. Only mainstream sports are considered: football, basketball and baseball.
   - High popularity. As is discussed in previous section, football, basketball and baseball have dominated the college sports nationally for a long time. Many games are played with statistics carefully recorded during the past decades. Therefore, data of coaches in these sports is much more accessible. Based on a sufficiently large sample, we could build our model with high credibility.
   - Long history. Compared with other sports, these three enjoy a long history. In fact it is the foundation of NCAA that promotes the development of collegiate sports in U.S.

2. Only man sports data is collected for our research.
   - Football and baseball are man only. Though girls also participate in these two sports on campus, official games are not held regularly at conference or national level.
   - Basketball is female friendly. However, the salary gap between male and female sports is huge. According to Bray's data(2003), average salaries for a head coach in men basketball team is twice as their colleagues who teach a women basketball team. Moreover, the history of women basketball is relatively short.

3. Only Division I coaches' data will be applied in modelling process.
   - There is a huge gap in salaries and requirements between coaches of different divisions. Our study will ignore coaches from Division II and III, and choose only data of Division I coaches to simplify the analysis. According the data from Bureau of Labor Statistics, the top 10% people earn $65,910 annually, while the bottom 10% of salaries are only $17,210. The great compensation for Division I coaches demonstrates the recognition and expectation from the college and society.
   - Coaching a Division I team brings the coach both fame and wealth. Besides higher salary, Division I coaching jobs also give coaches self-satisfaction. As we have mentioned in Assumption 5, even if best coaches face offer from Division II schools with wonderful compensation, they will take the factor, fame, into consideration and maximum their total utility.
   - College championship is a winner-take-all labor market. Those high-ranking teams take all the prize money and attentions, while other teams gain little. Because Division I teams usually have a large budget, and the salary that an outstanding coach takes has to match his marginal benefits tha the can bring to the team. So only the teams that can affords to the compensation of first-level coaches have the opportunity to employ them. This market structure leads to the phenomenon that best coaches concentrate on only Division I teams.

4. Where do we get reference sources?

- We choose reference ranking prudently. There are many online sources available with various information qualities. Thus, we choose Bleacher Report, Sportsonearth, Fox Sports and official data from NCAA. These sources are all based on click rates in Alexa and rankings of Google to ensure the data's profession and credibility.

5. The reasons and methods for filtering data.

- Filtering data can reduce the complexity of our analysis work. Our data are all from Sports Reference Data. Before the step of filtering, we have several thousands of coaches for each sport. To reduce the workload and simplify our study, filtering is necessarily.
- Our basic criteria are set under the consideration of both efficiency and experience. For instance, we find that all top coaches on famous rankings, like Bleacher Report and Orlando Sentinel, are rich in experience. Therefore, we set minimum years of coaching experience or minimum number of participating in important games as one of the filtering standards. The concrete steps for filtering are shown below:
  - football
    We only choose football coaches with more than 10 years' experience. Since our target is to discover the best coaches during the past century, and young coaches on the rising stage have not accumulated enough honor, the experience matters in this case.
  - basketball
    Step 1
    We use five variables when evaluating a team's performance, including CREG (number of Regular Season Conference Championships won), CTRN (number of Conference Tournament Championships won), NCAA (number of NCAA Tournament appearances), FF (number of NCAA Final Four appearances) and NC (number of NCAA Tournament championships won). Before filtering, there are around 3,700 coaches' data.
    Step 2
    If any of CREG, CTRN, NCAA, FF and NC is nonzero in a coach's record, this coach can be reserved for next step. Otherwise, the data of that coach would be dropped.
    Step 3
    If any of those five numbers is not available, we will judge by winning percentage and games. If the winning percentage is lower than 60% or number of overall games is less than 50, the coach's data will be deleted. After this step, there are 122 old teams left.
    Step 4
    If the five numbers are applicable and the sum of the five numbers is ≥ 11, the data of this coach will be reserved.

    After the steps mentioned above, only 326 basketball coaches' data in total remain for further study.
  - baseball
    Step 1
    Like how we deal with the data in football and basketball, we only choose baseball coaches with more than 10 years' experience.
    Step 2
    With more than 1000 coaches remained, we restrict the requirement on total winning percentage. Only if the coach has a winning percentage higher than 0.6, his profile remained.

So after these two steps, the number of coaches remained is 387.

- The reasons and methods for adding back data
  - football

    Step 1

    Coaches on Bleacher Report that are deleted due to shorter than 10 years career or out of Sport Reference Database are added.

    Step 2

    Those coaches with outstanding performance in bowl games are added back.

    After 2 steps above, we finally get data of 377 football coaches for further study.
  - basketball

    Compared with the reference ranking, we miss 5 prestigious coaches in earlier time. Thus we add those 5 coaches back to our dataset.
  - baseball

    Like steps above, we compare our ranking pool with the reference ranking and add 3 back. The final ranking pool has 390 candidates.
- Data normalization

  Our aim of normalization is to eliminate the effects of dimension in our data. According to Assumption 13, our data for each variable all follow a normal distribution. Thus, we can use the formula below to normalize data based on Law of Large Number.

  $$z = \frac{x - \mu}{\sigma}$$

  Moreover, we mentioned that the judging standards of younger-generation coaches and elder-generation coaches are different in Assumption 12. To avoid time horizon problem, we need to set the cut-off line to distinguish the elder-generation coaches. The cut-off line we set for football is 1932. Because before 1932, there was only one bowl game, while two more bowl games were held after that time. Thus, we divide the whole coaches into two groups by comparing their mean of years of coaching career to 1932. Then, the two groups of data should be calculated using the formula respectively to get them normalized. Finally, the two groups of data can be used together for further study. The data of basketball and baseball coaches can be treated in the similar manner, and the cut-off year of basketball and baseball are 1939 and 1950, respectively.

5 Constructing model

5.1 Our problem

After we normalized the data, we have all the variables without dimensions. The next crucial problem we face at this moment is how to determine the weights of each variable, so that we can multiple the data matrix with the weights to give a quantitative score for each person. The most apparent approach to find weights of each variable is linear regression. However, we find that there is no proper variables which can be used as the independent variables. Salaries for the top coaches could be an unbiased indicator to reveal the real abilities of the coaches, but the purchasing power fluctuates greatly in the time horizon. More importantly, the salaries of top head coaches are highly correlated with the popularity of the game. Take football as an example, the wage of a top coach isn100,000 USD, which is barely double of the average household income at that time. However, a top head coach at present earns at least 600,000 USD, which is ten times more than the average US household incomes[3]. As is shown, the problem of time horizon is extremely severe if we take coaches income as the dependent variable. Other than income, ranking could be a potential dependent variable if we use the Tobit regression. However, if we use one ranking source as the dependent variable, then our regression outcome will be ultimately close to that ranking, in other words, it is over-fitted.

Another approach to decide the weight is the Analytic Hierarchy Process. Since ranking is viewed as a multi-decision-making process, the Analytic Hierarchy Process can be a potential solution if all other requirements are satisfied. But the core approach of the Analytic Hierarchy Process is to simplify the complicated criteria into more simplified alternatives. In our database, we only have five variables that can be applied. So apparently, the Analytic Hierarchy Process is excluded from our potential modelling method pool.

Take all our problems into consideration, we turn to the method of Analytic Hierarchy Process. The greatest advantage of AHP over other methods is that we do not need to find out dependent variables before giving weights to the independent variables. Although the biggest disadvantage of AHP is subjectiveness when we give the relative importance, we have come up with solutions to greatly diminish problems of subjectiveness. Thus, the method of AHP becomes our priority.

5.2 Introduction to AHP

5.2.1 AHP history and usage

Born in 1970s, Analytic Hierarchy Process created by Thomas L. Saaty attracts tons of attention from decision makers, creating numerous literature discussing its usage and possible extension. The practical nature of the method, suitable for solving complicated and elusive decision problems, has led to applications in highly diverse fields like economics, sociology, politics and so on.

5.2.2 Solving a decision making problem using AHP

Using AHP in solving a decision problem involves four steps (Johnson,1980):

1. Setting up the decision hierarchy by breaking down the decision problem into a hierarchy of interrelated decision elements.
2. Collecting input data by pairwise comparisons of decision elements.
3. Using the eigenvalue method to estimate the relative weights of decision elements.
4. Aggregating the relative weights of decision elements to arrive at a set of ratings for the decision alternatives (or outcomes).

5.2.3 Mathematical Primer

5.2.3.1 The positive reciprocal matrix In AHP, we have to assign a matrix profiling the relative importance of each pair of variables. Following is the definition of positive reciprocal matrix(Shunsuke, 1997).

[3]The source of top head coach income is USToday, the average household income is from bls.gov, US Bureau of Labor Statistics.

1. A is said to be positive provided that $a_{i,j} > 0$ for all $i, j = 1, 2, \cdots, n$.
2. A is said to be reciprocal provided that $a_{i,j} = 1/a_{j,i}$ for all $i, j = 1, 2, \cdots, n$.

Hence, we define the positive reciprocal matrix as:

$$A = \begin{pmatrix} w_1/w_1 & w_1/w_2 & w_1/w_3 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & w_3/w_3 & \cdots & w_2/w_n \\ w_3/w_1 & w_3/w_2 & w_3/w_3 & \cdots & w_1/w_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ w_n/w_1 & w_n/w_2 & w_n/w_3 & \cdots & w_n/w_n \\ & & & & \end{pmatrix}$$

Intuitively, the reciprocal matrix always has rank 1, and there exists a vector w such that:

$$A \cdot w = n \cdot w$$

where $w = (w_1, w_2, \ldots, w_n)^T$ is the vector of actual relative weights, and n is the number of variables. In algebra, n and w are called the eigenvalue and corresponding eigenvector of matrix A. By Saaty's eigenvector method, the priority vector(weight) will be derived.

5.2.3.2 Checking for consistancy Before introducing the priority vector, we must check the consistency of our settings in order to make sure its validness. AHP posits that the evaluator does not know the actual w, which produces inconsistency over different evaluators. Mathematical induction and proof of this issue are described intensively in voluminous body of literature, so here we only give the final formula for checking consistency.

(a) The consistent index(C.I.) is defined as:

$$C.I. = \frac{\lambda_{max} - n}{n - 1}$$

(b) Compute the Consistency Ratio(C.R.).

$$C.R. = \frac{C.I.}{R.I.}$$

In this formula, R.I. is known as the average random consistency index. Professor Saaty computed the R.I. for $n = 1, \cdots, 9$ shown in the following table:

n	1	2	3	4	5	6	7	8	9
R.I.	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

5.2.3.3 Deriving the priority factor Having a positive reciprocal matrix, our final step is to determine the priority factor. For simplicity, suppose we have a 3 by 3 reciprocal matrix called A:

$$\begin{array}{cc} & \begin{array}{ccc} f_1 & f_2 & f_3 \end{array} \\ \text{factors} & \\ A = \begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} & \begin{pmatrix} 1 & 1/3 & 5 \\ 3 & 1 & 7 \\ 1/5 & 1/7 & 1 \end{pmatrix} \end{array}$$

Aggregating entries by columns we get a vector $s = (21/5, 31/21, 13)^T$. Then we divide each entry by the sum of its column values, i.e. dividing a_{21} by 31/21.

$$A = \begin{array}{c c} \text{factors} & \begin{array}{ccc} f_1 & f_2 & f_3 \end{array} \\ \begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} & \begin{pmatrix} 5/21 & 7/12 & 5/13 \\ 15/21 & 21/31 & 7/13 \\ 1/21 & 3/31 & 1/13 \end{pmatrix} \end{array}$$

Finally, the priority vector w can be derived by averaging across each row.

$$w = \frac{1}{3} \begin{pmatrix} 5/21 + 7/12 + 5/13 \\ 15/21 + 21/31 + 7/13 \\ 1/21 + 3/31 + 1/13 \end{pmatrix} = \begin{pmatrix} 0.2828 \\ 0.6434 \\ 0.0738 \end{pmatrix}$$

Our mathematical background is enough to proceed.

5.3 Model Specification

5.3.1 Our hierarchy

Through the Analytic Hierarchy Process, we are trying to make the decision of choosing the best five coaches ever in the century. Based on the available data of five variables[4] we have, the decision we actually are making is finding out the relative importance between these five variables. According to Assumption 7, all the five variables mentioned above can more or less reflect two more general characteristics of a coach, which are experience and efficiency. So it will be pretty tricky if we simply make the decision of relative importance without noticing whether our relative decision is experience-oriented or efficiency-oriented.

As is said in the introduction, AHP method can be defined as a semi-supervised approach on making decisions. How much extent we can diminish the subjectives will highly affect the accuracy of our final ranking. Take variables years and overall winning percentage as an example. Obviously, both of them are crucial in deciding the century best coaches, and thus make it a tricky problem when giving the relative importance between them if we do not properly divide the characteristics of experience and efficiency. However, if we separate the properties of experience and efficiency, the problem will become much more straightforward. From the aspect of experience, the years as a head coach is much more important than winning and losing, but from the aspect of efficiency, the impact of working time becomes nontrivial, and winning percentage will be highly cited.

In this way, if we can accurately choose the weight between experience and efficiency, then we can greatly reduce the influence of subjectivity. According to Assumption 11, our weights are close to the real weights. Moreover, all the weights will go through the data training part, so the weights we give will not vary greatly from the actual ones. As is shown in the hierachical graph, the ultimate target we need to get is to find out the best coach of all time, which is shown in level 1. Then we separate the standards into two generalized categories, experience and efficiency. In the level 3 part, all the five concrete variables are listed, and subordinate lines[5] link the experience and efficiency with all these five variables, meaning that these five variables have both experience and efficiency properties. And finally, in level 4, all the candidates in our potential century best coaches pool are judged by these five variables, as the subordinate lines show.

[4]Or three variables, due to the time horizon problem.

[5]The lines connect the upper level attributes and the lower level attributes.

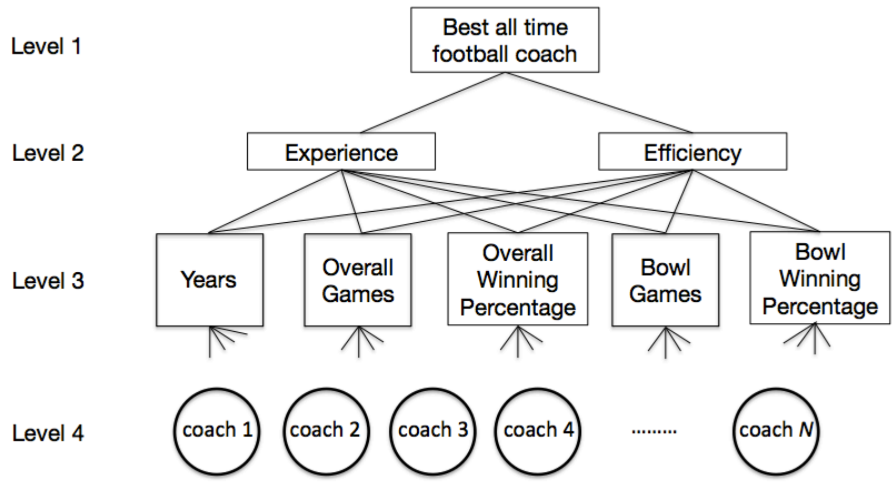


Figure 1: Hierarchy graph

5.3.2 Constructing relative importance matrices

$L_1 = (a_{ij})$ is called the first level relative importance matrix(mathematically defined as reciprocal matrix), $L_2^{exp} = (b_{ij})$ is called the second level reciprocal matrix with respect to experience, and $L_2^{eff} = (c_{ij})$ is called the second level reciprocal matrix with respect to efficiency.

$$L_1 = (a_{ij}) = \begin{array}{c} \text{factor} \\ experience \\ efficiency \end{array} \begin{array}{c} \begin{array}{cc} experience & efficiency \end{array} \\ \begin{pmatrix} 1 & exp/eff \\ eff/exp & 1 \end{pmatrix} \end{array}$$

The first level relative importance matrix is a 2 by 2 matrix with generalized categories Experience and Efficiency on the column and row. The relative importance of one variable itself is 1, while Exp/Eff means the how much efficiency is more important than experience. Say, if one thinks that efficiency is twice more important than experience, then the upper-right entry will be filled as 1/2. And symmetrically, the lower-left entry will be filled as 2. Mathematically speaking, the diagonally symmetric entries are reciprocal.
Our initial L_1 matrix along with its priority vector is shown in following table:

L_1	experience	efficiency	priority
experience	1	1/2	1/3
efficiency	2	1	2/3

$L_2^{exp} = (b_{ij})$ and $L_2^{eff} = (c_{ij})$ are pretty much the same with L_1, only with a higher complexity, both 5 by 5. To avoid redundancy, we will not explain them entry by entry.

L_2^{exp}	years	overall games	overall pct	bowl games	bowl pct	priority
years	1	3	9	4	9	0.4790
overall games	1/3	1	7	3	8	0.2703
overall pct	1/9	1/7	1	1/7	1	0.0377
bowl games	1/4	1/3	7	1	7	0.1762
bowl pct	1/9	1/8	1	1/7	1	0.0368

L_2^{eff}	years	overall games	overall pct	bowl games	bowl pct	priority
years	1	1/4	1/9	1/5	1/9	0.0326
overall games	4	1	1/4	1/5	1/7	0.0749
overall pct	9	4	1	2	1	0.3204
bowl games	5	5	1/2	1	1/3	0.1857
bowl pct	9	7	1	3	1	0.3867

5.3.3 Finding the overall factor weights

In previous section, we have already calculated two level 2 priority vectors with respect to experience and efficiency. Applying these two vectors to relative weights of the first level, we get our overall factor weights as following:

$$w^* = (1/3, 2/3) \begin{pmatrix} 0.4790 & 0.2703 & 0.0377 & 0.1762 & 0.0369 \\ 0.0326 & 0.0749 & 0.3204 & 0.1857 & 0.3867 \end{pmatrix}$$

$$= (0.1814, 0.1401, 0.2261, 0.1825, 0.2699)$$

5.3.4 Data Training

Through the priority vectors we can have the weights for all the five vectors. Then we can give a comprehensive score for all the coaches in our candidate pool. With the score of each candidate, the next step will be giving the initial ranking for all the candidates which rely simply on the numerical score of the candidates.

Since the score and the ranking relies much on the initial level of importance we gave in the relative importance matrix, so the ranking only reflects our own favors. In order to be more neutral and cover the favor of a more generalized population, we need to train our data with other famous rankings given in the websites. Also, according to Assumption 15, the rankings on the Internet is only partly reliable, so we use the ranking in 3 websites for training and testing the goodness of our model. The reliability of the websites is ranked by the number of daily UV (unique visitors) of the websites available on Alexa.com. The most viewed one will be used as goodness testing while the other two will be used as data training. This procedure will be discussed in detail in the goodness of fit part.

5.3.4.1 Goodness of Fit According to Assumption 8, Two basic criteria for a good ranking model are that absolute position of a coach is more critical than relative position; the accuracy of the bottom part of the ranking has to be sacrificed to guarantee the accuracy of the upper parts. So when we are trying to test how close our ranking modelling is to the other two rankings given in the websites, we will pay more attention to the upper part of the ranking and to the absolute rankings. Thus, we construct a scoring equation trying to find out how many coaches out of 5, exist both in the top five of our ranking and one of their rankings, how many out of 10 exist both in the top ten of our and one of their rankings, what if out of 20, 30 or 50? So through this step, we can have five percentages, then we will add up these five percentages to get the score of similarity. This means that two rankings are identical if we ignore the minor relative rankings inside top 5, 10, 20, 30 and 50. If the score is 5 while two rankings are completely different if the score we give at this step is 0.

The detail scoring equation is shown as below:

$$score = \frac{common_5}{5} + \frac{common_{10}}{10} + \frac{common_{20}}{20} + \frac{common_{30}}{30} + \frac{common_{50}}{50}$$

As is shown in the equations, the top five are redundantly calculated for five times, this is corresponded with the assumption that the upper parts of the ranking is more important than the lower parts. Also, we do not give any penalty to the relative ranking inside the top n, this is corresponded with our initial assumptions as well. After this step, we can have a score of similarity between our data and either of these two datasets, and in the next part, we are going to slowly release the degrees of freedoms in the reciprocal matrices to give more relative importance matrices that could be closer to the true one.

5.3.4.2 Enumerating possible reciprocal matrices As is fully discussed above, the most crucial disadvantage of the AHP process is that the initial relative importance level we gave in the reciprocal matrices could be too subjective to reflects the truth of the ranking. So in the step of enumerating possible reciprocal matrices, we want to find out all the potential reciprocal matrices that could possibly be the one that is the closest to the true reciprocal matrix. In this process, we use the most robust approach to find the best possible reciprocal matrices. That is, enumerating all the possible reciprocal matrices and use irritation to find out which matrix has the highest score for the two rankings mentioned above.

Apparently, the reciprocal matrices that maximize either of the two rankings can not be identical, we have to make a compromising between the two rankings. We use the number of daily UV when deciding the weights between these two websites. In this way, we are trying to find out the relative importance matrix that maximize a weighted score, the calculation of the score is shown as below:

$$score_{total} = \frac{UV_1}{UV_1 + UV_2} \cdot score_1 + \frac{UV_2}{UV_1 + UV_2} \cdot score_2$$

But we face a trade off there between the accuracy of the weights and the efficiency of calculation. Since the diagonally symmetric entries are reciprocal, and the values for the diagonal entries are all ones. So for an n by n reciprocal matrix, there are only $n(n-1)/2$ free variables. As is shown in the reciprocal matrices above, if we sum up the free variables in all these three matrices, we will have 21 free variables. And for each free variable, we can choose the value from $1/9$ to $1/2$ and from 1 to 9. So totally, if we contains all the possible reciprocal matrices, we will have 17^{21} times of iterations. A normal PC we can have access to run one iteration for at least 0.025 second, so the total time we need is more than decades. According to Assumption 11, the initial relative importance levels we teammates assigned in the AHP parameters are close to the actual relative importance levels.Thus, we set the value for each free variable can only have 3 possible answers, which are the initial level we gave in the entries, one level up, or one level down[6]. This way, the base of this exponential function is down from 17 to 3, and the efficiency will be increased greatly. However, $3^{21} \cdot 0.025$ still takes years to finish. So we still have to furthermore limit the degrees of freedoms in our reciprocal matrices.

It is obvious that if the number of free variables is zero, this means we take the initial scores we gave directly. And degree of freedom equals one means that only the relative importance in the L_1 matrix is changeable while all the relative importances entries in the

[6]Up means the closer to 9 while down means closer to $1/9$.

other two matrices remain the same as the initial matrices. We slowly relax the restrictions of the three matrices, with more and more relative importance parameters changeable. The procedure of furthermore free variables relaxation is a little tricky. When the free variables we relax are between 2 and 6, we just randomly choose the free variables, and it turns out that the outcome of the similarity score which is concretely discussed in the next subsection is pretty much the same. And when the number of free variables reaches 7, we assume that there are linear relationships between the entries. For simplicity, I will only explain the linear relationships we assumed when the number of free variables is 11.

$$\begin{cases} b_{13} = 9 \\ b_{15} = 9 \\ b_{24} = b_{14} - 1 \\ b_{25} = 8 \\ c_{13} = 1/9 \\ c_{15} = 1/9 \\ c_{24} = 1/5 \\ c_{25} = 1/7 \end{cases}$$

Here we explain the first equation. $b_{13} = 9$ means the relative importance of years over overall winning percentage with respect to experience factor is 9. The times of iterations is exponentially correlated with the number of free variables, if we relax one more entry into a free variable, then the times of iterations will be tripled.

5.3.4.3 Optimized model Each time we relax one more free variables and irritate to get our revised optimized model. Then we compare the ranking using our optimized model with the ranking we used for testing the goodness of fit (ranking from Bleacher Report) to get a similarity score between them. We find out that the score is rising with a higher number of free variables, and finally converges to a score about 3 for all three sports, as is shown in Figure 2.

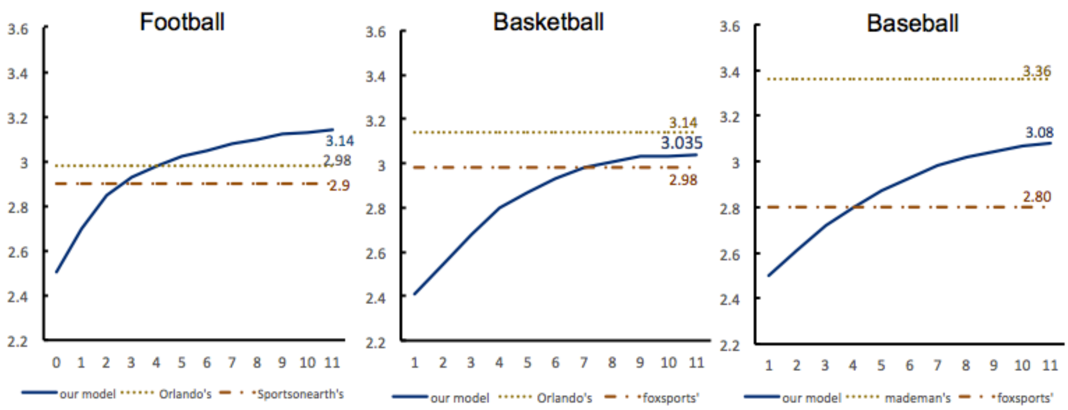


Figure 2: Training improves model performance significantly

For simplicity, we will only show the optimized reciprocal matrices for football, shown as below(optimized $L1$ is identical to initial $L1$):

$L_2^{exp}(optimized)$	years	overall games	overall pct	bowl games	bowl pct
years	1	3	9	5	9
overall games	1/3	1	7	4	8
overall pct	1/9	1/7	1	1/7	6
bowl games	1/5	1/4	7	1	7
bowl pct	1/9	1/8	1/6	1/7	1

$L_2^{eff}(optimized)$	years	overall games	overall pct	bowl games	bowl pct
years	1	1/4	1/9	1/6	1/9
overall games	4	1	1/8	1/6	1/7
overall pct	9	8	1	3	2
bowl games	6	6	1/3	1	1/4
bowl pct	9	7	1/2	4	1

The weights corresponding to the optimized reciprocal matrices for all three sports are shown as below:

weights	football	basketball new	basketball old	baseball new	baseball old
years	0.1477	0.182	0.2625	0.1808	0.2663
overall games	0.1172	0.1411	0.2471	0.1377	0.2386
overall pct	0.2791	0.3239	0.4904	0.3097	0.4951
top games	0.1654	0.1264		0.1657	
top games pct	0.2906	0.2266		0.2060	

5.4 Algorithm in pseudocode

Require: optimized ranking model

Input: 3 ranking datasets (2 for training, 1 for testing)

  5 variables describing each coach

  Initial *L1* and two *L2* reciprocal matrices

1: **for** $var_1^{free}-1,\ var_1^{free},\ var_1^{free}+1$ **do**

  ……

2: **for** $var_n^{free}-1,\ var_n^{free},\ var_n^{free}+1$ **do**

3:   $currentParameter := \{var_1^{free}\cdots var_n^{free}; var_1^{fixed}\cdots var_{21-n}^{fixed}\}$

4:   Compute eigenvalues and eigenvectors of *L2* matrices

5:   **if** $C.R.(L_2^{exp})>0.1$ or $C.R.(L_2^{eff})>0.1$ **do**

6:    **continue**

7:   **end if**

8:   Combine two priority vectors to get the overall weights

9:   Score all coaches by overall weights

10:   Rank all coaches by their cuurent score

11:   Evaluate the goodness of fit between current model version with reference rank

12:   **if** $currentScore > bestScore$ **then**

13:    $bestScore := currentScore$

14:    $bestParameters := currentParameters$

15:   **end if**

16: **end for**

  ……

17: **end for**

18: Construct the optimized model using *bestParameters*

5.5 Adaptation for basketball and baseball coaches

In the database for basketball and baseball coaches, elder generation coaches do not have records for the top tournaments as the tournaments started later than their main coaching careers, so we only use variables years as a Division I coach, overall number of games, overall game winning percentage and overall number of participating in top-level games as the parameters. Plus, we will have minor adjustments to the steps for processing the football data.

For step 4 in the pseudocode, we will extract the upper-left 3 by 3 submatrices from both $L2$ matrices to calculate the factor weights with respect to elder generation coaches. In this way, we can get the score for younger generation coaches with the weights from the complete 5 by 5 $L2$ matrices, and 3 by 3 submatrices for elder generation coaches. And the final step for adjustment is to combine the score of two generation coaches together. Other procedures are identical with the algorithm processing football data.

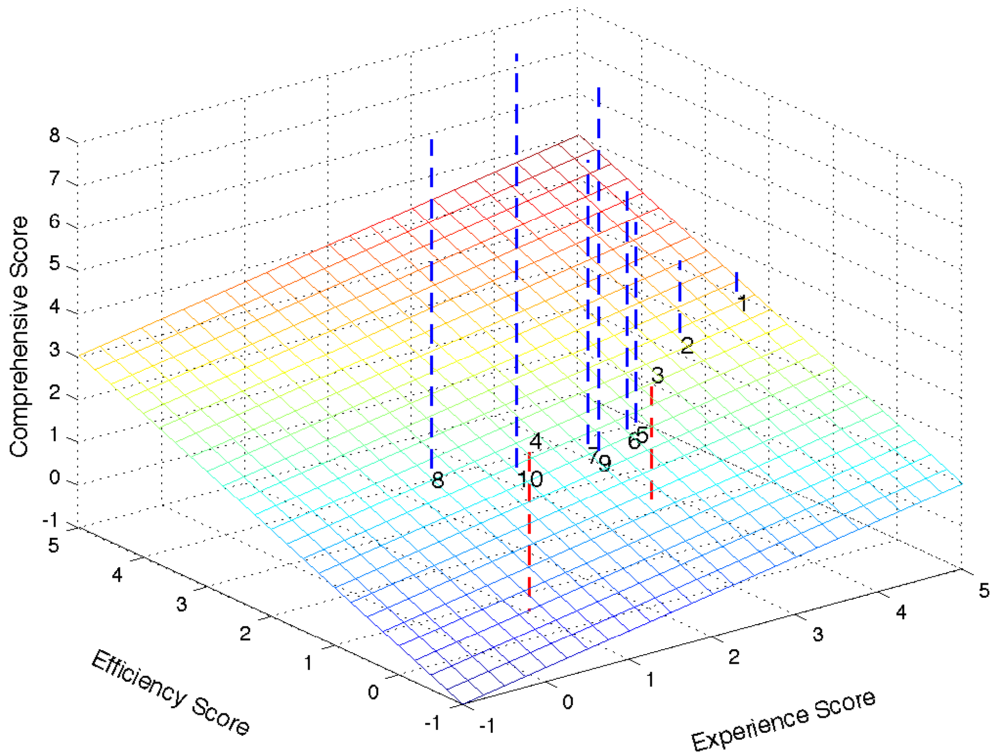


Figure 3: Fitted model against reference ranking

5.6 Visualization

In this part we want to report the goodness of fit between our optimized model and the testing ranking from Bleacher Report. In Figure 3, the rainbow-colored grids represent the overall scoring plane, with score of experience, efficiency, comprehensive score on x, y and z axis respectively. The numbers shown on the surface are the top ten football coaches in our ranking, with dash lines indicating how high we are over or underestimated, red for underestimated and blue for overestimated.

6 Model results

After the model construction part, we find out the weights for the five variables in all three sports. So it is very straightforward to give a final ranking for all the three sports. In this part, we will report the top ten coaches in our ranking for all three sports. It is worth noticing that the top ten ranking we give in this part is purely quantitative data oriented, which means that the ranking will not capture the achievements for one coach that beyond the five variables we measure above. And that is why we choose ten coaches for each sport. We will furthermore refine our ranking using factors beyond winning and losing in the next section.

The ranking of the three sports are shown as below:

rank	football	basketball	baseball
1	Joe Paterno	John Wooden	Rod Dedeaux
2	Bobby Bowden	Mike Krzyzewski	Skip Bertman
3	Bear Bryant	Dean Smith	Augie Garrido
4	Tom Obsorne	Adolph Rupp	Bobby Winkles
5	Pop Warner	Bob Knight	Dick Siebert
6	Lou Holtz	Rick Pitino	Ray Tanner
7	Mack Brown	Roy Williams	Jerry Kindall
8	LaVell Edwards	Billy Lush	Cliff Gustafson
9	Fielding Yost	Denny Crum	Wayne Graham
10	Howard Jones	Otto Rittler	Bibb Falk

7 Non-quantitative factors

Through building a mathematical model, we get the numerical results presenting top 10 coaches of each sport in last section. However, it is difficult to explain why some coaches are better than the others. In fact, the ten coaches are almost identical considering their accomplishments and performances. Each of them is outstanding enough for our model to implement perfect evaluation using only quantitative variables. So introducing other non-quantitative dimensions will greatly improve our evaluation.

7.1 Moral issue

Coaches may be judged by more than the factors listed above, and moral issue is a big determining factor difficult to quantify, which can totally change a coach's evaluation. Normally, good coaches tend to be aggressive, which make the situation full of criticism and controversy, but a best coach should at least obey the basic moral principles. To introduce this factor into our study, we will delete the coaches on our final list with moral problems regardless of their achivements.

In last section, we get top 10 coaches of each sport. Then, we go over their career life to check whether there is a scandal that may affect our results. For example, one of the most famous football coaches, Joe Paterno, involved in Penn State child sex abuse scandal and was dismissed by the team. So he would not appear in our finalist. Another example, Adolph Rupp, a basketball coach is excluded from the rankings of best coaches due to 1951 point shaving scandal.

7.2 Contribution to development of sport

We have emphasized that our rankings is made based on a long time horizon. Under the background that all three sports have experienced sufficient development, though data have been divided into two groups according to cut-off year when normalizing, we still believe that this treatment will ignore some important factors.So in this subsection, we will focus on an factor beyond the winning and losing of the games, the contribution to the development of sport.

In order to enhance the strength of elder-generation coaches, we reserve one place in the finalist to award those coaches with great lifetime achievements. For football, Pop Warner, ranking No.5 in our optimized model is reserved, because he devised the dominant offensive systems used over the first half of the 20th century, and he is remembered for having given his name to one of the country's major football organizations for young boys, the Pop Warner Youth Football League, in 1934. For basketball, Dean Smith, ranking No.3 in our model, is reserved due to his contribution in diversified fields. For baseball, Dick Siebert, is reserved because he was the first baseman and succeeded in continuing his career as a coach.

7.3 Popularity

Other than the factors we mentioned above, public acceptance and satisfaction are also important standard for making our finalist. This factor reflects how influential the coaches are. Here, we use statistical data from Google Trends to represent the current social influence of one coach. After deleting coaches with moral issues and excluding the reserved coach, the four most popular of our candidates will be finalised. Plus the reserved coach, we can finally give result of five coaches.

The graph below demonstrates the popularity of five football coaches. Four lines (Bobby Bowden, Bear Bryant, Tom Osborne and Lon Holtz, all in our finalist), are obviously higher than the purple one (Fielding Yost, ranked No.9 in our model).

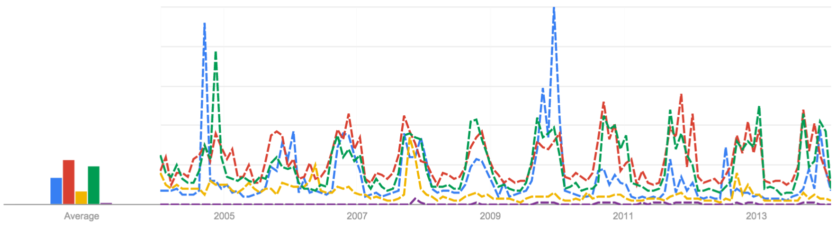


Figure 4: Popularity trends of top football coaches

8 Conclusion

We will present the finalist of the five All Time Best College Coaches for football, basketball and baseball, after filtering in Section 7.

football	basketball	baseball
Pop Warner	Dean Smith	Rod Dedeaux
Bobby Bowden	Bob Knight	Skip Bertman
Bear Bryant	Mike Krzyzewski	Augie Garrido
Tom Obsorne	Rick Pitino	Wayne Graham
Lou Holtz	John Wooden	Ray Tanner

Our quantitative model is based on AHP, combined with statistical and mathematical knowledge we acquire from the undergraduate curriculum, which has given special attention to address time horizon and gender problems. Meanwhile, non-quantitative factors are considered carefully in order to eliminate limitation of evaluation dimensions in quantitative model.

In short words, we design and construct a straightforward but effective model, and succeed in finding the best college coaches. The result we present can somehow evaluate coaches objectively and comprehensively. In addition, we attach an article for Sports Illustrated to the end of our paper, and we hope you will enjoy it.

9 Strength and weakness

9.1 Our strength

- Pros from AHP. Since our model is based on Saaty's Analytic Hierarchy Process, we have naturally inherited most of its advantages. Moreover, when Assumption 7 holds, evaluator can easily judge the relative importance of two variables from only one perspective (experience or efficiency) each time. And that is the main reason why we define our model in two levels. This stratification will effectively eliminate evaluator's subjectivity when setting up the initial reciprocal matrix. Taking computational efficiency into consideration, our model requires the searching radius be relatively limited. The split of experience and efficiency makes it possible that Assumption 11(initial solution is close to optimal solution) will not be destroyed.
- Benefits from training. In our training process, we supervise our original model with two or three online rankings other than the final test ranking. In other words, we prepare two less reliable rankings to train and another(most authoritative) ranking to test. As the figure above shows, our training process will improve the fitness score from 2.50 to 3.14, which can be interpreted as 22 percent increase in model's accuracy.

9.2 Our weakness

Though we have spent a long time discussing how excellent our model is, several weak points and deficiencies cannot be denied. Here are some important ones:

- Local optima dilemma. Obviously some of our model assumptions are too strong. For instance, our final conclusion will be no longer valid once Assumption 11 does not hold. The time complexity of a full exhaustive searching is $O(n^{21})$. We have to make the trade-off between the breadth of search and the amount of running time. Decreasing the number of free variables from 21 to 16, a full searching process can be finished in 10 minutes. Plus, our parameter selection is discrete, so the corresponding manifold is not smooth along with our objective function non-derivative. Without the help of traditional derivative methods, our algorithm design for finding maximum is extremely difficult.
- Our model may be oversimplified. There exists several factors that contribute to polish our data. The conference to which our best coaches are affiliated is ignored and the competition intensity varies greatly among different conferences.
- Female coaches might be underestimated. Though we tried to evaluate all college sports coaches beyond their gender, few female coaches qualify to final top 100. Women coaches may be well educated and have their own theory of training a baseball or football game, but the lack of real-game experience makes them less respected by players. No one would ever deny the indispensable role of female coaches, especially those in women sport teams.
- Data distribution may be skewed or abnormal. Even if we tried to collect our raw dataset fairly large, requiring all variables to be standard normal is a strong but impractical argument (Assumption 13). Our scoring function weights variables by their relative importance with regard to the best coach problem rather than their distributions. More statistical issues should be exposed and clarified, with our expectation on building a better ranking model.
- Our scoring formula can be improved. As a reminder, our scoring formula calculates the sum of accuracy percentage by top 5, 10, 20, 30 and 50 coaches, using an authoritative ranking as reference. This formula is sensitive to threshold values like 6, 11, 21,

31 and so on. When our algorithm iterates to a model version predicting one coach as 21st but his(her) reference ranking is in top 20, this formula underestimates the accuracy of this iteration. In a word, our scoring function is innovative, effective but not enough, a model using more satisfactory evaluation method is expected.

- Strong assumptions revisited. As described in previous sections, all assumptions explained explicitly function as cornerstones for this model. Without any one of them, the whole pyramid collapses. Relaxing some of these assumptions makes modelling process easier, the result however will be unpredictable.

10 Summary

Evaluating is a complicated process involving various factors and subjective ideas. To understand the problem and clarify the objective of our study is the key step to start. Before constructing model, we set up to learn the background knowledge, discover available sources online and collect related data.

Afterwards, we choose the core method for our model based on all resources that we are able to utilize. The property of semi-supervised is the greatest advantage Analytic Hierarchy Process has over other methods, meaning that we do not need to find out a dependent variable before we decide our initial weights. After building a theoretical framework by dividing our problem into subproblems, we can easily give a better general categorical-oriented relative importance and construct a more unbiased two level reciprocal matrices, and initial weights and ranking afterwards. Based on two popular rankings available on the Internet, we train our model to make our criteria closer to public sense by optimizing relative importance matrices and weights. Based on the optimized outcomes, we can give the top ten ranking considering only quantitative database.

All discussed above is more about quantitative method, but actually there are some non-quantitative factors affecting the evaluation of coaches. In Section 7, we give special attention to moral issue, special contribution to sports and popularity, which are also standards to get the finalist of five best coaches in each sport. We also attach a two-page article publishable for Sports Illustrated, which is non-technical and sport-fans-oriented.

References

[1] Ben-Israel, Adi, Aharon Ben-Tal, and Sanjo Zlobec. Optimality in nonlinear programming: a feasible directions approach. New York: Wiley, 1981.

[2] Bouman, Marcel, and Ton Van der Wiele. "Measuring service quality in the car service industry: building and testing an instrument." International Journal of Service Industry Management 3.4 (1992): 4-16.

[3] Gorney, Blaine, and R. Gary Ness. "Evaluation dimensions for full-time head coaches at NCAA division II institutions." Journal of Personnel Evaluation in Education 14.1 (2000): 47-65.

[4] Grant, Randy R., John Leadley, and Zenon X. Zygmont. The economics of intercollegiate sports. World Scientific, 2008.

[5] Harker, Patrick T. "Alternative modes of questioning in the analytic hierarchy process." Mathematical Modelling 9.3 (1987): 353-360.

[6] Johnson, C. R., & Hihn, J. M. (1980). Prioritization of research projects. Energy Storage: The Transactions of the First International Assembly on Energy Storage, 562-566.

[7] Lee, Tian-Shyug, and I-Fei Chen. "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines."Expert Systems with Applications 28.4 (2005): 743-752.

[8] Mallett, Cliff, and Jean Côté. "Beyond winning and losing: Guidelines for evaluating high performance coaches." Sport Psychologist 20.2 (2006): 213.

[9] Martinez, David M. "Study of the key determining factors for the NCAA Division I head strength and conditioning coach." The Journal of Strength & Conditioning Research 18.1 (2004): 5-18.

[10] National Collegiate Athletic Association. "822004-05 NCAA Sports Sponsorship and Participation Rates Report." Indianapolis, IN: National Collegiate Athletic Association (2006).

[11] Randy R. Grant, John Leadley, Zenon X. Zygmont. The Economics of Intercollegiate Sports.The Economics of Intercollegiate Sports(2008):221

[12] Saaty, Thomas L. "How to make a decision: the analytic hierarchy process."European journal of operational research 48.1 (1990): 9-26.

[13] Saaty, Thomas L. "How to make a decision: the analytic hierarchy process."Interfaces 24.6 (1994): 19-43.

[14] Shiraishi, Shunsuke, Tsuneshi Obata, and Motomasa Daigo. "Properties of a positive reciprocal matrix and their application to AHP." Journal of the Operations Research Society of Japan-Keiei Kagaku 41.3 (1998): 404-414.

[15] Vaidya, Omkarprasad S., and Sushil Kumar. "Analytic hierarchy process: An overview of applications." European Journal of operational research 169.1 (2006): 1-29.

[16] Zahedi, Fatemeh. "The analytic hierarchy processa survey of the method and its applications." interfaces 16.4 (1986): 96-108.

Best College Coaches Ever

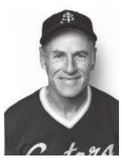
Mining the Century Best College Coaches in Football, Basketball and Baseball

Ask 30 people to name the 10 best college baseball coaches ever and you'll get 30 different answers. The NCAA has seen its fair share of exceptional, record-breaking coaches, and every fan has his favorite. Recently, we launched an activity to find the Best All Time College Coach, and our editors would like to share the final rankings with you now.

The aim of our activity is trying to discover the best coaches for the last century. As football, basketball and baseball all developed rapidly, a growing number of top events were generated, like bowl games for football, which attracted great attention and participation. Therefore, the elder-generation coaches before 1940s faced totally different social background and competitive structure.

Though it is difficult to answer the question "what makes a good coach"? We can easily list many factors that should be taken into consideration.

In our ranking, we have taken different treatments to younger and elder generation coaches in order to be fair with time horizontal problem in the mathematical model. Second, we collected data about coaches, and scored their performance under a comprehensive framework of priorities from the perspective of efficiency and experience. Candidates' information from professional database for scoring includes length of coaching career, number of games and winning percentage. We gave special attention to those coaches, who had made strong contribution to sport promotion and growth, and other factors hard to quantify. For instance, the moral issues and scandals may lower the ranking position.

FOOTBALL	BASKETBALL	BASEBALL
		
 1 **Pop Warner**	 1 **Dean Smith**	 1 **Rod Dedeaux**
• As a coach in both football and baseball • Championship in 4 National, 1 SIAA and 3 PCC • Amos Alonzo Stagg Award • Dominant offensive systems used for half the century	• Overall Records: 879-254 in 40-year coaching tenure • One of the highest winning rates in history • Gold Medal of Men's Basketball in 1976 Olympics • Ran a clean program with high graduation rates	• 11 national titles including a 5 straight titles during the 45 seasons coaching at USC • High recognition among greatest records in sports' amateur history • Names as the "Coach of the Century"
 2 **Bobby Bowden**	 2 **John Wooden**	 2 **Ray Tanner**
 3 **Bear Bryant**	 3 **Mike Krzyzewski**	 3 **Wayne Graham**
 4 **Tom Osborne**	 4 **Bob Knight**	 4 **Skip Bertman**
 5 **Lou Holtz**	 5 **Rick Pitino**	 5 **Augie Garrido**

The coaches deserve to be *All Time College Coach*, and they have spared no efforts to promote the development of sports, and realize dream of every team. Then even changed the history.

However, many college coaches cannot appear in our table. We should be proud of great era, as there were so many masters in the past century.

Let us remember every glorious moment of them to express our appreciation and respect.