Team Control Number

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

24270

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

Problem Chosen

B

2014 Mathematical Contest in Modeling (MCM) Summary Sheet

Summary

In order to estimate the excellence of different sports coaches and to give a ranking result, two distinct models are developed. The first model is a comprehensive evaluation method. And the second model is a ranking algorithm analogous to the *Journal Influence Algorithm***.**

In the first model, we take into account a variety of metrics, and divide them into two categories: *Objective Metrics* and *Subjective Metrics*. In the *Objective Metrics*, we consider four factors, the number of wins, winning percentage, champions and final fours. All these factors have contributions to the excellence of a coach. We deem that the total number of games in a year could affect the number of wins, and the unevenness of team quality could affect the winning percentage. By employing statistical regression method to process collected data, we establish two functions of influence coefficient to eliminate the discrepancy caused by the two kinds of effect. In the *Subjective Metrics*: we consider two factors, media popularity and tenure. We employ *Fuzzy Analysis Method* to quantify these two subjective factors. We further incorporate *Analytic Hierarchy Process* (AHP) and *Gray Relational Analysis Grade Method* (GRAP) to determine the weight allocation to different metrics. The final ranking gives a comprehensive result by weighing results returned by these two methods. Using data from *Sports Reference* and other websites, the rankings in basketball, football and baseball accord with previous media commentaries.

In the second model, we deem that the excellence of a certain coach can be reflected from the media impact over the span of history and that the interactions between two coaches can reflect the disparity of skill level between them. We use search results returned by *Google* to quantify the impact of one coach on another. Based on the search results, we build a *cross-reference matrix* to represent relationships between coaches. In view that the different time periods that two coaches were in may largely affect the interaction between them, and the personal reputation may influence the number of search results, we develop a weight function of two variables to compensate the influence of time and to rule out the redundant information.

In consideration of the similarity between personal influence and journal influence, we refer to the *Journal Influence Algorithm* introduced by *Eigenfactor* and establish a new ranking algorithm. The basic idea of the algorithm is subtle: using weight function to modify the *cross-reference matrix*, and taking into consideration of individual influence, the algorithm gives an evaluation vector to rank different coaches. To test the validity of this algorithm, we apply the algorithm into basketball, football and baseball. The algorithm gives a result that is similar to the result obtained in the first model. The ranking also agrees with previous media commentaries. Furthermore, by slightly adjusting the coefficients, we can apply the algorithm into various sports.

"Dream Team" of College Coaches

Team 24270

Contents

1. Introduction

1.1. Restatement of the Problem

Sports, by definition, is all forms of usually competitive physical activity which aim to use physical ability while providing entertainment to participants and spectators[1]. No wonder the word "sports" gives us a first impression of fierce competition, agitated spectators, sweating on the running track, combined with a joy of victory. It is the uncertainty that makes the sports game so intriguing. However, where there is competition, there will always be victory, defeat, and ranking. Loyal sport fans could debate day and night over the question who is the best player or coach. These debates have called forth a need for certain criterion of sports coaches and players. The criterion has to be: (1) all-encompassing to take into consideration a variety of factors; (2) applicable to various sports; (3) robust enough to remain unaffected by fluctuation.

1.2. Model Overview

- **Model Ⅰ**

The evaluation method in Model Ⅰ is based on a comprehensive method sophistically combining Analytic Hierarchy Process(AHP) and Gray Relational Analysis Grade Method(GRAP). In the evaluation process, we take into consideration the influence of time horizon, and incorporate Fuzzy Analysis Method, which make it feasible to compare diverse factors on the same level. The ranking results in three different sports accord with previous media report, which attest the validity of this method.

- **Model Ⅱ**

In model II, we assume that the excellence of a certain coach can be reflected from the media impact over the span of history and thus can be gauged by the impact on another coach within or without the same period of time. We use Google search results to quantify the impact of one coach on another. The relationship between coaches can be established as a cross-reference matrix. By further taking into account the influence of time, influence of reputation, and a modification to rule out the redundant information, we obtain a final evaluation vector. The final ranking result is roughly approximate to the result in model I. To sum up, we only need the search results returned by Google search engine to estimate the excellence of certain coach with high accuracy.

2. Assumptions

- **We assume that the competition rules of each sport do not change.**
  Although sports are developing, we do not take into account of time in the competition rules in order to compare the coaches of different years more fairly.

- **We neglect tied competitions since they have the same effect on the two compared teams.**

- **We only take the Division I into consideration.**

Competitions are divided into three parts: Division I, II and III according to the level of sport strengths of different colleges. Since Division I always concludes top coaches, we only take Division I into consideration.

- **The selected data are valid.**
- **Additional assumptions are made to simplify analysis for individual sections. These assumptions will be discussed at the appropriate sections.**

3. Model Ⅰ

3.1. Additional assumptions

- **The evaluation system includes two parts: Objective Metrics(OM) and Subjective Metrics (SM).**
- **We assume that OM include four specific indexes: the total number of wins, the winning-percentage, the number of final fours and the number of champions.**
- **Tenure and media popularity are considered in SM.**
  In the subjective metrics of ranking coaches, some factors are hard to investigate qualitatively and quantitatively due to lacking data, such as, his or her influence to players, range of knowledge, studying ability, team spirits, searching talents, acting in competitions, salary and so on. Therefore, we neglect these indexes in SM.
- **Time only makes a difference in the total number of wins, and the winning percentage.**
  In fact, the numbers of final fours and champions have no effect on the other two in OM, since the number of teams which are able to enter into final fours and even achieve champions is fixed. And we neglect the influence of time on media popularity in order to simplify the model.

3.2. Notations

Table 1: Notations and Descriptions

Notations	Descriptions
S_i	Evaluation object
x_j	Evaluation index
n	The number of evaluation objects
m	The number of evaluation indexes
$\mathbf{x}, \mathbf{x}', \mathbf{x}'', \mathbf{x}^*$	Evaluation index matrix
t	Time
p_i , q_i	Influence coefficients of time
$W(t)$	The total number of competitions in t
$s(t)$	The standard deviation of all winning-percentage in t
M_j	Maximum of x_{ij}
m_j	Minimum of x_{ij}

Notations	Descriptions
$f(x)$	Subordinate function
$\mathbf{A}$	Pairwise comparison matrix
λ	The largest eigenvalue
$\mathbf{w}$	Weight vector
CI	Consistency index
RI	Random consistency index
CR	Consistency ratio
$\mathbf{B}$	Evaluation vector of AHP
$r\left(x_i^{(0)}, x_i^{(j)}\right)$	Grey relational coefficient
$\Delta_{ji}^{(0)}$	Absolute difference
$\Delta_{\min}$	Minimum difference
$\Delta_{\max}$	Maximum difference
$\mathbf{r}$	Relation degree vector
$\mathbf{C}$	Evaluation vector of Grey Relation Degree
α , β	Partial coefficient
$\mathbf{U}$	Ultimate evaluation vector

3.3. Evaluation System

We define n as the number of evaluation objects, and S_1, S_2,…, S_n (n>1) are the evaluation objects. m is the number of evaluation indexes, and x_1, x_2,…, x_m are the evaluation indexes. Evaluation index vector is

$$\mathbf{x}=\left[x_1,x_2,\ldots,x_m\right]^T (m>1).$$

The total evaluation indexes include OM: the total number of wins, the winning-percentage(pct.), the number of final fours and the number of champions and SM: tenure and media popularity. So $m=6$,

$$\mathbf{x}=\left[\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4,\mathbf{x}_5,\mathbf{x}_6\right]^T$$

Where:

- $\mathbf{x}_1$ — the total number of wins vector.
- $\mathbf{x}_2$ — the winning-percentage vector.
- $\mathbf{x}_3$ — the number of final fours vector.
- $\mathbf{x}_4$ — the number of champions vector.
- $\mathbf{x}_5$ — tenure vector.
- $\mathbf{x}_6$ — media popularity vector.

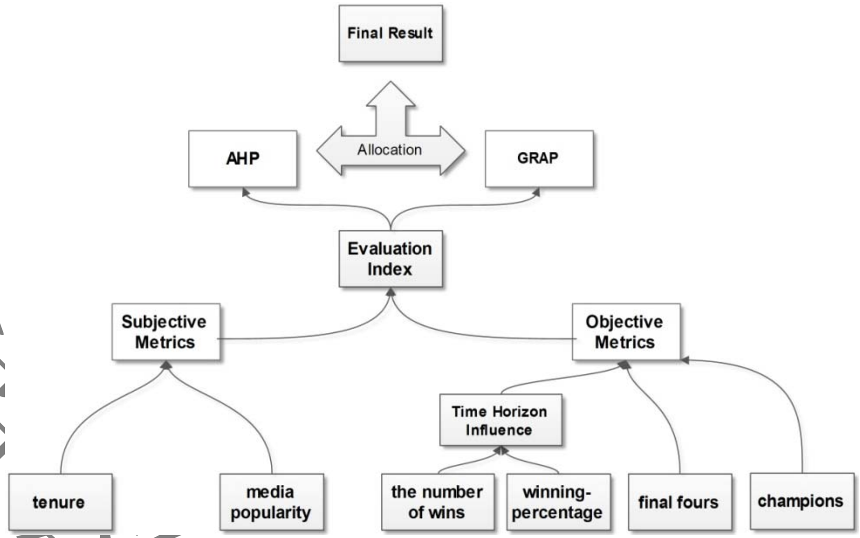

Figure 1: Flow chart of model I

Undoubtedly, time plays an important role in evaluating top coaches. According to the assumptions, time only makes a difference in the total number of wins, the winning-percentage.

3.3.1. The influence of time on the total number of wins

With the development of sports, the competition is getting relatively fiercer than ever, which means the disparity between teams become wider. The total number of games also increases with time going on. Therefore, when evaluating coaches in the previous century, the later certain coach begin his coaching career, the more likely he will get more wins. So we should put less weight on the coaches active in a later time period. And we can get a fairer evaluation of coaches within different time periods.

In order to compensate the influence of t, we establish Influence Coefficients of Time (ICT) $p_i\left(i=1,2,\ldots,n\right)$. We assume that the total number of competitions in t is $W\left(t\right)$. $W\left(t\right)$ can be obtained by statistical regression and simulating and curve fitting of selected data. So we define:

$$p_i = \frac{1}{W\left(t_{mi}\right)}$$

where t_{mi} is the middle year of tenure of S_i. And then $x_{1i}^{'} = x_{1i} \cdot p_i\left(i=1,2,\ldots,n\right)$.

3.3.2. The influence of time on the winning-percentage

As for the winning-percentage, sports were underdeveloped at an earlier time, and the quality disparity between teams is comparatively narrow. Therefore, the standard deviation of winning-percentage of each coach is closer to zero. Thus we should put less weight on the coaches active in a "mediocre" time period. We define ICT here as q_i (i=1,2,…,n), we assume that the standard deviation of all winning-percentage in t is $s\left(t\right)$. $s\left(t\right)$ can be obtained by statistical regression and simulating and curve fitting of selected data. So we define:

$$q_i = \frac{1}{s(t_{mi})}$$

and $x'_{2i} = x_{2i} \cdot q_i \left(i = 1,2,\ldots,n\right)$.

3.3.3. Fuzzy Analysis

As for SM indexes, we assume that they can be divided into five levels: " Excellent, Very Good, Good, Not Good, Bad". And we correspond the five levels into 5,4,3,2,1 successively For continuous quantification, we assume:

- As for "Excellent", we suppose $f(5) = 1$.
- As for "Very Good", $f(3) = 0.7$.
- As for "Bad", $f(1) = 0.1$.

We employ partial large Cauchy distribution and the logarithmic function as the subordinate function[2]:

$$f(x) = \begin{cases} \left[1 + a(x-b)^{-2}\right]^{-1}, 1 \le x \le 3 \\ c\ln x + d, \quad 3 \le x \le 5 \end{cases}$$

where *a*, *b*, *c*, *d* stands for undetermined constants. We use the initial conditions above to define their values. And solution of the subordinate function(**Figure 2**) is:

$$f(x) = \begin{cases} \left[1 + 2.8049(x-0.4417)^{-2}\right]^{-1}, 1 \le x \le 3 \\ 0.5873\ln x + 0.0548, \quad 3 \le x \le 5 \end{cases} \tag{1}$$

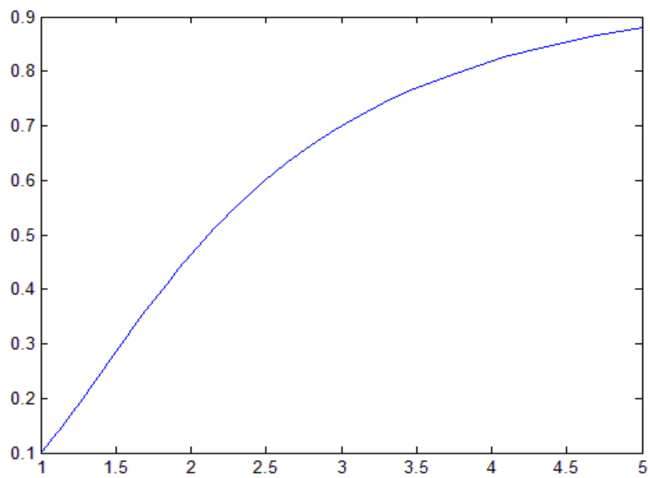

Figure 2: Trend of $f(x)$

Media popularity is measured by the number of search results via Google. The impact of duplication of names can be neglected by means of adding search keywords in order to rule out the redundant information.

We map $\mathbf{x_j}$ (j=5,6) into interval [1,5], through function (1),we can obtain:

$$x_{ji}^{'} = f\left(1+\frac{4\left(x_{ji}-m_j\right)}{M_j-m_j}\right)\left(i=1,2,\ldots,n, j=5,6\right) \tag{2}$$

where $M_j = \max_{1\le i\le n}\left\{x_{ij}\right\}, m_j = \min_{1\le i\le n}\left\{x_{ij}\right\} \left(j=5,6\right)$.

As for $\mathbf{x_3}$ and $\mathbf{x_4}$, we define that $\mathbf{x_3}' = \mathbf{x_3}$, $\mathbf{x_4}' = \mathbf{x_4}$.

we use $\mathbf{x_j^{'}}\left(j=1,2,\ldots,6\right)$ to proceed the following calculation.

3.3.4. Nondimensionalization process

We employ extreme difference method to nondimensionalize the different indexes so that we can compare them[2] on the same level. The method is as follows:

$$x_{ji}^{"} = \frac{x_{ji}^{'}-m_j}{M_j-m_j} \tag{3}$$

and $\mathbf{x}_j^{"} = \left[x_{j1}^{"}, x_{j2}^{"}, x_{j3}^{"}, \ldots, x_{jn}^{"}\right]^T \left(j=1,2,\ldots,6\right)$

where $M_j = \max_{1\le i\le n}\left\{x_{ij}^{'}\right\}, m_j = \min_{1\le i\le n}\left\{x_{ij}^{'}\right\} \left(j=1,2,\ldots,6\right)$.

and then we obtain the final evaluation index matrix:

$$\mathbf{x}^{*} = \left[\mathbf{x_1^{"}}, \mathbf{x_2^{"}}, \mathbf{x_3^{"}}, \mathbf{x_4^{"}}, \mathbf{x_5^{"}}, \mathbf{x_6^{"}}\right]^T \tag{4}$$

3.3.5. Final result

By using AHP as the subjective evaluation method and GRAP as the objective method, the final represents a comprehensive evaluation combined the merits of these two methods.

- **Analytic Hierarchy Process[3] (AHP)**

By comparing the effect of two indexes $\mathbf{x_j^{'}}$, the weights of the two method $w\left(\mathbf{x_j^{'}}\right)$ (j=1,2,…,m)are given. Then we construct the pairwise comparison matrix $\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \frac{1}{3} & \frac{1}{7} & 5 & \frac{1}{2} \\ 1 & 1 & \frac{1}{3} & \frac{1}{5} & 5 & 3 \\ 3 & 3 & 1 & \frac{1}{6} & 5 & 3 \\ 7 & 5 & 6 & 1 & 7 & 5 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{7} & 1 & \frac{1}{3} \\ 2 & \frac{1}{3} & \frac{1}{3} & \frac{1}{5} & 3 & 1 \end{bmatrix}$$

We can obtain the largest eigenvalue of $\mathbf{A}$: λ=6.0496 and its weight vector :

$$\mathbf{w} = \left[0.1248, 0.1469, 0.4593, 0.8125, 0.0775, 0.2928\right]^T$$

After that, we must check the consistency of matrix **A**. The consistency index is calculated as follows:

$$CI = \frac{\lambda - n}{n-1} = 9.92 \times 10^{-3}$$

From **Table 2**, the random consistency index RI=1.24

Table 2: The Quantitative Values of RI[2]

n	1	2	3	4	5	6	7	8	9	10	11
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51

Then, we can obtain consistency ratio:

$$CR = \frac{CI}{RI} = 0.008 < 0.1$$

Therefore, we can safely draw the conclusion that the inconsistent degree of matrix **A** is in a tolerable range, and we can take its eigenvector as weight vector **w**[3].

We define **B** as the evaluation vector of AHP, and **B** can be calculated as follows:

$$\mathbf{B} = \mathbf{x}' \cdot \mathbf{w} \tag{5}$$

In evaluation vector, the greater B_i is, the higher ranking S_i is.

- **Gray Relational Analysis Grade Method[4] (GRAP)**

We use integral grey relational degree to analyze the metrics data. And we take the total number of wins as the reference sequence:

$$\mathbf{x}^{(0)} = \left\{ x_i^{(0)} \right\} (i = 1, 2, \ldots, n)$$

and then we can obtain the gray relational coefficient[4]:

$$r\left(x_i^{(0)}, x_i^{(j)}\right) = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{ji}^{(0)} + \rho\Delta_{\max}} (i = 1, 2, \ldots, n, j = 1, 2, \ldots, 6)$$

Where:

- $\Delta_{ji}^{(0)} = \left| x_i^{(0)} - x_i^{(j)} \right|$ —absolute difference.
- $\Delta_{\min} = \min\limits_j \min\limits_i \Delta_i^{(j)}$ —minimum difference of all indexes data.
- $\Delta_{\max} = \max\limits_j \max\limits_i \Delta_i^{(j)}$ —maximum difference of all indexes data.
- ρ—resolution ration.

For every coach S_i, we determine its weight as w_i, which should satisfy the requirements:

$$0 \le w_i \le 1, \quad \sum_{i=1}^{n} w_i = 1$$

After determining the weight, we can obtain the relational degree[4]:

$$r_j = r\left(x^{(0)}, x^{(j)}\right) = \sum_{i=1}^{n} w_i r\left(x_i^{(0)}, x_i^{(j)}\right) \tag{6}$$

And then we construct the relation degree vector $\mathbf{r}=[r_1,r_2,r_3,r_4,r_5,r_6]^{\mathrm{T}}$, where $r_1=1$.

We define **C** as the evaluation vector of AHP, and **C** can be calculated as follows::

$$\mathbf{C}=\mathbf{x}'\cdot\mathbf{w} \tag{7}$$

In evaluation vector, the greater C_i is, the higher ranking S_i is.

- **Combination of AHP and GRAP**

At first, we employ extreme difference method to nondimensionalize the two evaluation vector **B** and **C**. And then, we construct an ultimate evaluation vector:

$$\mathbf{U}=\alpha\mathbf{B}+\beta\mathbf{C} \tag{8}$$

where α, β respectively stands for the weight of AHP and GRAP, which should satisfy the requirements of $\alpha+\beta=1$.

Finally, we sort the value of U_i (i=1,2,…,n), and S_i that corresponds to the top 5 of U_i are top five coaches.

3.4. Solutions to Model Ⅰ

We choose three sports to verify our model and get the results, which include basketball, football and baseball.

3.4.1. Basketball

- **Searching and selecting data**

We search and select data through the Internet[5][6][7]. For example, first, we search 100 coaches and their evaluation index data. Secondly, we rank them by comprehensively considering the total number of wins and the winning-percentage, so we can get top 40 coaches. And then, we consider other metrics and rank top 20 coaches. Finally, the evaluation system is based on the selected 20 data. **Table A1** in **Appendix** show the selected coaches and their evaluation index data.

- **Determining the final evaluation index matrix**

At first, we determine vector **x**'. For $\mathbf{x_1}$, via the data in **Table A2**, we use $W(t_i)$=Num^2, where Num represents the total number of teams in t_i.

We utilize software MATLAB to plot the graph of $W^{(1)}(t)$ by simulating and curve fitting of data (**Figure 3**). So we can get p_i for each S_i, and then we obtain the vector $\mathbf{x_1}'$.

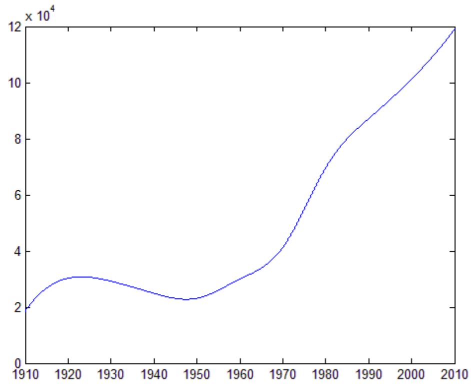

Figure 3: Trend of $W^{(1)}(x)$

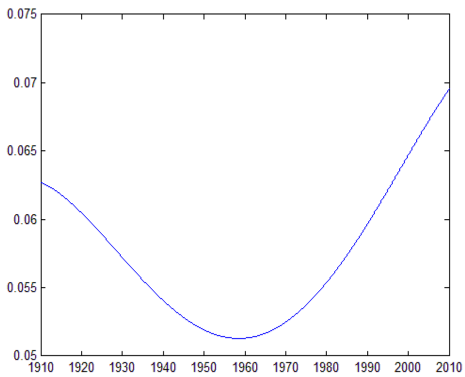

Figure 4: Trend of $s^{(1)}(x)$

For $\mathbf{x_2}$, via the data in **Table A3**, we also plot the graph of $s^{(1)}(t)$ by simulating and curve fitting of data (**Figure 4**). So we can get q_i for each S_i, and then we obtain the vector $\mathbf{x_2}'$.

For $\mathbf{x}_5$ and $\mathbf{x}_4$, from (1)(2), we can obtain $\mathbf{x}_5'$ and $\mathbf{x}_6'$.

Secondly, from (3), we can obtain $\mathbf{x}_j^{*}\ (j=1,2,\ldots,6)$. Finally, from (4), we can obtain $\mathbf{x}^*$.

We list the quantitative value of $\mathbf{x}^*$ in **Table A4**.

- **Obtaining the result via ultimate evaluation vector**

At first, from (5), we use AHP and get $\mathbf{B}$. Secondly, we use GRAP and define that ρ=0.3 and w_i=0.05(i=1,2,…,n). From (6), we can get the relation degree vector **r**. And then, from (7), we can obtain **C**. Finally, from (8), by defining $\alpha=0.6, \beta=0.4$, we can obtain the ultimate evaluation vector:

$$\mathbf{U}=[0.293,0.288,0.468,0.241,0.422,0.160,0.521,0.868,0.713,0.311,0.481,0.836,0.168,0.998,\\ 0.130,0.318,0.243,0.482,1.761,0.138]^T$$

By sorting the value of U_i (i=1,2,…,n), we can obtain the ranking result of S_i. And the ranking vector is:

$$\mathbf{Rank}^{(1)}=[19,14,8,12,9,7,18,11,3,5,16,10,1,2,17,4,13,6,20,15]^T$$

Therefore, we list top five coaches of basketball in the previous century in **Table 3**:

Table 3: Top 5 Coaches of Basketball

No.1	No.2	No.3	No.4	No.5
S_{19}	S_{14}	S_8	S_{12}	S_9
John Wooden	Dean Smith	Mike Krzyzewski	Adolph Rupp	Bob Knight

This result is largely agreement with the widely accepted result[8][9].

3.4.2. Football

- **Searching and selecting data**

Like what we do in basketball, we search and select data through the Internet[5][10][11]. However, we calculate that the number of final fours is the sum number of times that teams can enter into Super Bowl.

- **Determining the final evaluation index matrix**

At first, we determine vector $\mathbf{x}'$. For $\mathbf{x}_1$, we use $W(t_i)$=Num^2, where Num represents the total number of teams in t_i.

We can obtain $W^{(2)}(t)$ by simulating and curve fitting of data. So we can get p_i for each S_i, and then we obtain the vector $\mathbf{x}_1'$.

For $\mathbf{x}_2$, we also obtain $s^{(1)}(t)$ by simulating and curve fitting of data. So we can get q_i and the vector $\mathbf{x}_2'$.

Finally, from (4), we can obtain $\mathbf{x}^*$. We list the quantitative value of $\mathbf{x}^*$.

- **Obtaining the result via ultimate evaluation vector**

Like what we do in Basketball, we can obtain the ultimate evaluation vector:

$$\mathbf{U}=[0.234,0.879,0.738,0.106,0.359,0.296,0.383,0.193,0.291,0.453,0.494,0.248,0.180,0.615,\\ 1.000,0.316,0.151,0.047,0.392,0.052]^T$$

By sorting the value of U_i (i=1,2,…,n), we can obtain the ranking result of S_i. And the ranking vector is:

$$\mathbf{Rank}^{(2)} = [15,2,3,14,11,10,19,7,5,16,6,9,12,1,8,13,17,4,20,18]^T$$

Therefore, we list top five coaches of basketball in the previous century in **Table 4**:

Table 4: Top 5 Coaches of Football

No.1	No.2	No.3	No.4	No.5
S_{15}	S_2	S_3	S_{14}	S_{11}
Joe Paterno	Bobby Bowden	Bear Bryant	Tom Osborne	Don James

This result is largely agreement with the widely accepted result[12].

3.4.3. Baseball

- **Searching and selecting data**

Like what we do in basketball, we search and select data through the Internet[13][14]. But in this sport, we assume that the number of final fours is the number of champions of NCAA competitions that teams can achieve. And we assume that the number of champions is the number of champions of National competitions that teams can get.

- **Determining the final evaluation index matrix**

At first, we determine $\mathbf{x}'$. For $\mathbf{x_1}$, due to scarcity of the data, we can only search a little information of several years[9]. We use $W(t_i)$=Num^2, where Num represents the total number of competitions of champion in t_i.

We can obtain $W^{(3)}(t)$ by simulating and curve fitting of data. So we can get p_i for each S_i, and then we obtain the vector $\mathbf{x_1}'$.

For $\mathbf{x_2}$, due to lacking the standard difference of winning-percentage in every ten year, we choose another approach to get $\mathbf{x_2}'$. Considering the influence of time, first, we employ extreme difference method to nondimensionalize $\mathbf{t_m}$ into $\mathbf{t_m}'$, where t_{mi} is the middle year of tenure of S_i. Then, we define

$$t_m'' = \frac{1}{t_m' + 5}$$

and then we define $x_{2i}' = x_{2i} \cdot t_{mi}''$. So from (3), we obtain $\mathbf{x}_2''$.

Finally, from (4), we can obtain $\mathbf{x}^*$. We list the quantitative value of $\mathbf{x}^*$.

- **Obtaining the result via ultimate evaluation vector**

Like what we do in Basketball, we can obtain the ultimate evaluation vector:

$$\mathbf{U} = [0.353,0.278,0.587,0.223,0.205,0.302,0.208,0.189,0.314,0.494,0.203,1.000,0.214,0,\\ 0.022,0.044,0.232,0.225,0.138,0.118]^T$$

By sorting the value of U_i (i=1,2,…,n), we can obtain the ranking result of S_i. And the ranking vector is:

$$\mathbf{Rank}^{(3)} = [12,3,10,1,9,6,2,17,18,4,13,7,5,11,8,19,20,16,15,14]^T$$

Therefore, we list top five coaches of basketball in the previous century in **Table 5**:

Table 5: Top 5 Coaches of Baseball

No.1	No.2	No.3	No.4	No.5
S_{15}	S_2	S_3	S_{14}	S_{11}
John Barry	Mike Martin	Rod Dedeaux	Augie Garrido	Jim Morris

This result is largely agreement with the widely accepted result[15].

3.4.4. Sensitivity analysis

By changing the weight of AHP and GRAP in equation (8), we analyze the changing result of basketball. For example, we define $\alpha = 0.5, \beta = 0.5$, and the result is listed in **Table 6**. The coaches who rank top 5 do not change:

Table 6: Top 5 Coaches of Basketball

No.1	No.2	No.3	No.4	No.5
S_{19}	S_{12}	S_{14}	S_8	S_9
John Wooden	Adolph Rupp	Dean Smith	Mike Krzyzewski	Bob Knight

When defined $\alpha = 0.4, \beta = 0.6$, the result changes, which is listed in **Table 7**. The coaches who rank top five change:

Table 7: Top 5 Coaches of Basketball

No.1	No.2	No.3	No.4	No.5
S_{19}	S_{12}	S_{14}	S_7	S_8
John Wooden	Adolph Rupp	Dean Smith	Hank Iba	Mike Krzyzewski

When defined $\alpha = 0.7, \beta = 0.3$, the result is listed in **Table 8**. The coaches who rank top five do not change:

Table 8: Top 5 Coaches of Basketball

No.1	No.2	No.3	No.4	No.5
S_{19}	S_{14}	S_{12}	S_8	S_9
John Wooden	Dean Smith	Adolph Rupp	Mike Krzyzewski	Bob Knight

As can be seen from above, when there is a slight change of weights, the result do not change. But with a relatively greater change, weights have an effect on the result.

4. Model Ⅱ

How could one's reputation affect another's? One way is to follow the implication in the saying: "You wouldn't mention A and B in the same breath." It means if the difference between two people is too wide, it would be unlikely for most of individuals to mention them in a same talk. The same holds true for the sports coaches. That means, if two coaches are absolutely not on the same level, more likely than not, there will be few reports on these two coaches. On the other hand, if two of them are top coaches, there will be a plethora of reports: such as *"The Greatest Coaches Ever" "Basketball Hall of Fame"*, on the two coaches. Informed by this natural law, we may find an innovative approach to estimate a coach's level of excellence and popularity. The working flow is shown as follows:

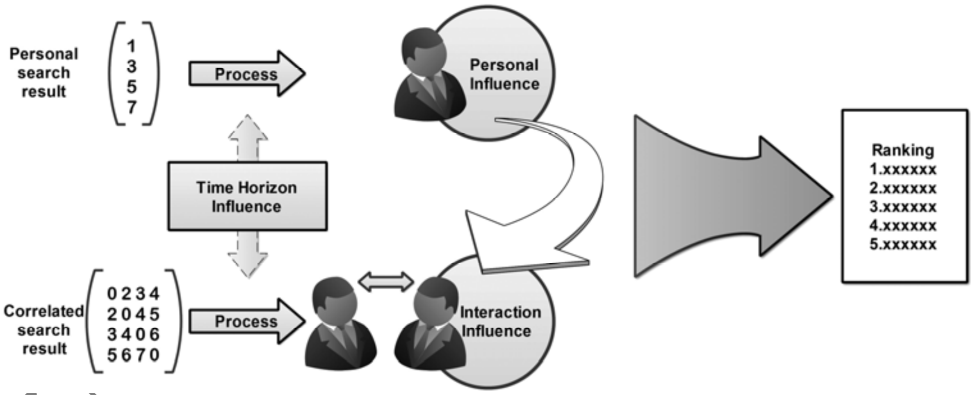

Figure 5: Flow chart of model II

4.1. Additional assumptions

- The excellence of a coach and association between two coaches can be accurately reflected by the mass media.
- The attention that the mass media have on certain coach is related to the search results on Google, in terms of number of pages, report orientation and report time.
- The media attention is related to time and the excellence of certain coach. The influence of time and excellence on the media attention remains unchanged to different kinds of people.

4.2. Notations

Table 9: Notations and Descriptions

Notations	Descriptions
a_i	The number of search results of coach i
k, b	Coefficient of the function through linear regression
$\boldsymbol{t_i}$	Characteristic year of coach i
u	The number of search results about sports career
ICT	Influence coefficient of time
ICR	Influence coefficient of reputation
$\mathbf{I}$	Individual influence vector
$\mathbf{Z}$	Original cross-reference matrix
WF	Weight function
$\mathbf{W}$	Weighted cross-reference matrix
α, β	Partial coefficient

4.3. The Individual Influence Vector

By our hypotheses, the excellence of a coach can be accurately reflected by the mass media. There are several ways to evaluate the media attention on a celebrity. One of the most simple and direct way is to record the number of search results on Google. However, the search results can be influenced by a variety of factors, such as time periods, tenure, etc. By simulating and curve fitting of sorted data, we evaluate the impact of such factors separately. Finally, we obtain a normalized individual influence vector.

4.3.1. Original data

Here we define t_i as the **characteristic year**, the average of the year that the coach i start coaching and the year of his or her retirement. (If the coach i is still active, then t is the average of the year that the coach i start coaching and this year, that is, 2014)

The search results vector a is the original data we use to estimate the individual influence, where a_i is the number of search results of coach i. Particularly, the coaches here are sorted by characteristic year in a descend order. This can be a great convenience to our later discussion about time factor.

4.3.2. The influence coefficient of time

According to the growth law of web information[16], the information aiming at a certain field is similar to an exponent increase. To test this hypothesis and better apply it to sports, we entered the Google website. Using "1910 basketball", "1920 basketball"， and "1930 basketball" as the "exact keywords"[17] respectively. The numbers of search results are shown in **Table 10**.

Table 10: The Numbers of Search Results

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
Results	2750	5160	7440	11700	16200	26400	40200	25800	27000	67700	326000

Assuming that this is an exponential function: $y_1 = c \cdot e^{dt}$. We use the least squared method to obtain the unknown numbers in the function. See **Figure 6**.

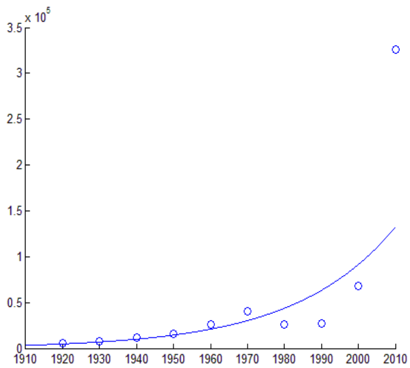

Figure 6: Trend of exponential function y_1

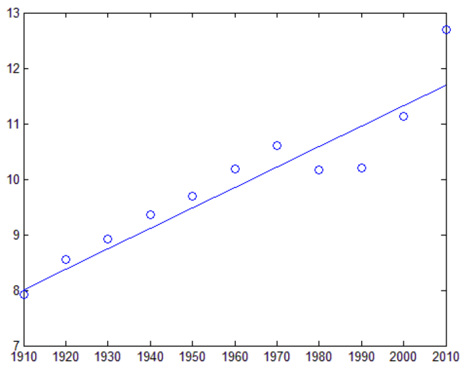

Figure 7: Trend of linear function y_2

The result gives a satisfying simulation to the numbers of search results. However, the distinction between 2000s and 1900s is too large. In our observation, the search results of coaches at different period of time is almost of the same magnitude of as each other. So, here we use the natural logarithm of the search results. Again we obtain a linear function y_2 as showed in **Figure 7**.

The difference between maximum and minimum is about half of the minimum value. This is a modest value that we can safely put into use to estimate *ICT*. Common sense told us that the greater number of total reports is, the more "valuable" the search result is, the greater weight the search result will get. So, we define **ICT** as

$$ICT_i\text{(influence coefficient of time)}=\frac{1}{kt_i+b}\left(i=1,2,\ldots,n\right)$$

where k and b are the unknown variables related to searching data.

4.3.3. The influence coefficient of reputation

As mentioned in the overview, by searching data with different methods and using diverse keywords, we observe that on the track of fame, the media will turn at first to the achievements of sports one has then to the other aspects in his or her life. Therefore, the media attention can be interpreted and quantified by using the overall search results of one coach and his or her triumph and achievement.

To extract the information about news reports on the triumph of a certain coach from the search engine, first, we use Wordnet[18] as our tool to obtain a host of synonyms of the word "winning", the result is:

"booming; flourishing; palmy; prospering; prosperous; roaring; thriving; in; made; no-hit; productive; self-made; sure-fire; triple-crown; victorious; successful"

Using these words as our "alternative keywords", we can obtain the numbers of winning search results u_i for coach i.

The result turns out to be an indication that the ratio between winning search results and overall search results is negatively correlated with the degree of reputation. That is, the higher reputation one coach gets, the more likely the mass media will concentrate on the other aspects of life of this coach. According to this rule, we establish a function of **ICR**:

$$ICR_i\text{(influence coefficient of reputation)}=1-\left(\frac{a_i}{u_i}\right)^2 \left(i=1,2,\ldots,n\right)$$

4.3.3. The individual influence vector

The individual influence vector **I** can be interpreted as the overall search results modified by ICT and ICR.

$$l_i=\frac{a_i \bullet ICT_i \bullet ICR_i}{\sum_{j=1}^{n} a_j \bullet ICT_j \bullet ICR_j}\left(i=1,2,\ldots,n\right) \tag{9}$$

The individual influence vector gives an accurate estimation of the media attention certain coach got. It is a normalized vector, so that we can conveniently put it into use in the later section.

4.4. The Cross-Reference Matrix

With the help of the Google search engine, the degree of correlation of two coaches can be measured by the number of search results using two names as "Citation Keywords"[17] simultaneously. And we define the original cross-reference matrix **Z**. The entries of the matrix is:

$$Z_{ij}=\text{number of search results number of coach } i \text{ and coach } j$$

Since exchange of the two names does not affect the result, the matrix **Z** is symmetrical. And we set all of the diagonal elements of this matrix to be 0.

4.4.1. The weight function

The elements of cross-reference matrix are influenced by the distinct period of time and reputation. However, things get a bit more complicated here: if two coaches exist in the same period of time, then there could be more reports on competitions they engaged. The competition reports are the redundant information that we want to avoid. What we are aiming to do is to evaluate a certain coach's impact over the span of the sports history and to rule out the redundant information. To do that, we assign a lower weight to the matrix element in which the two coaches is in the same period of time, and a higher weight to those in a different period of time.

First, let t_i and t_j to be the characteristic year of two coaches, we want to construct a weight function related to t_i and t_j. Because we previously set the diagonal elements to be 0, the weight on the diagonal can be zeros. Assuming that the average term of office is 2σ, according to the 2σ principle[3], there is little likelihood that the two coaches can encounter each other in the sport games. Then the corresponding weight can be approximate to be 1 outside the interval 2σ. By carefully weighing the pros and cons of diverse types of function, we establish the original weight function as

$$\frac{1}{\sqrt{2\pi}\sigma}(1-e^{-\frac{(t_i-t_j)^2}{2\sigma^2}})$$

Additionally, to take into consideration of *ICT*, we simply transform the one variable function IIT to two variable function multiplied by the original weight function. The final weight function *WF* is:

$$WF=\frac{1}{\frac{1}{2}kt_i+\frac{1}{2}kt_j+b}\times\frac{1}{\sqrt{2\pi}\sigma}(1-e^{-\frac{(t_i-t_j)^2}{2\sigma^2}})\left(i,j=1,2,\ldots,n\right) \tag{10}$$

From (10), the values of the function in intervals (1910, 2010), (1910, 2010) are drawn in 3D graph as showed in **Figure 8** and **Figure 9**.

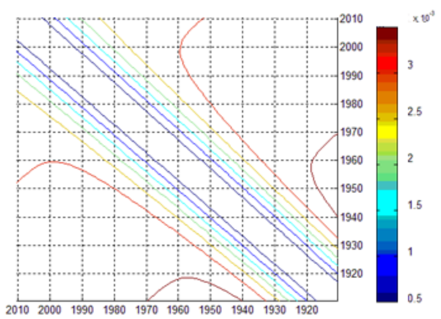

Figure 8: Contour plot of weight function

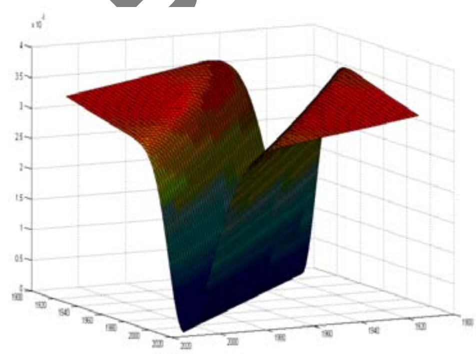

Figure 9: 3D graphic of weight function

As can be seen from the graphs above, as the characteristic year increases, the weight decreases linearly. And with the gap between two characteristic years increase, the weight increases somehow similar to a normal distribution curve.

4.4.2. The final cross-reference matrix

The weighted cross-reference matrix is defined as the original cross-reference matrix multiplied by the function value of weight function, via:

$$W_{ij} = Z_{ij} \cdot WF(t_i, t_j)$$

Normalizing the weighted cross-reference matrix by column sums, we obtain the final cross-reference matrix **N**.

$$N_{ij} = \frac{W_{ij}}{\sum_{k=1}^{n} W_{ij}} \tag{11}$$

4.5. The Evaluation Vector

So far we have obtained the individual influence vector and the cross-reference matrix. Each of them partly reflect the impact that a certain coach has over the course of sports history. Furthermore, if one is related to a very influential coach, then this relationship can be more "valuable" to him. This law is far better than merely counting the correlated search results. Following this law, we sophistically combine the Article Influence Algorithm[19] with our data.

Following the algorithm described in Eigenfactor Article Influence Algorithm, a new matrix is calculated as:

$$S = \alpha N + \beta l \cdot e^T \tag{12}$$

Where α and β are partial coefficients, $\alpha + \beta = 1$, and e^T is a row vector of 1's. We find out the greatest eigenvalue of matrix **S**, and the corresponding eigenvector is defined as the evaluation vector **v**. Through the interpretation given by Eigenfactor, [19]the values of each element of the leading vector represent an average fraction of time spent on each article[19]. However, the situation here is a little different in three respects: (1) the matrix **N** is a symmetrical matrix; (2) The individual influence vector here is the corresponding individual search results divided by the sum of search results; (3) the relative magnitude of α and β cannot be determined by previous data.

Finally, the evaluation vector is sorted in descending order. The location of element is the ranking of corresponding coach i.

4.6. Solutions to Model II

4.6.1. Basketball

- **Searching and selecting data**

First, we choose 100 coaches and their evaluation index data as our candidate pool. Secondly, we rank them by comprehensively considering the total number of wins and the winning-percentage. In this way, we get top 20 coaches from the candidates. The evaluation system is based on the selected 20 data.

Through Google search results, we obtain the original search result vector **a**(the third column in **Table A5**) and the original cross-reference matrix **Z**(**Table A6**).

- **The influence coefficient of time**

We use linear regression method, and the result turns out to be: $k = 0.0368, b = -62.2719$.

$$\text{Therefore},\quad \text{ICT}=\frac{1}{0.0368t_i-62.2719}.$$

- **The individual influence vector**

Using software Microsoft Excel, we can easily get the individual influence vector from the data shown in **Table A5**.

- **Weight function**

Using the linear regression result in the previous section, from (10), the weight function can be written as:

$$WF=\frac{1}{\frac{1}{2}\times 0.0368t_i+\frac{1}{2}\times 0.0368t_j+b}\times\frac{1}{\sqrt{2\pi}\cdot 12.5}(1-e^{-\frac{(t_i-t_j)^2}{2\times 12.5^2}})$$

- **The final cross-reference matrix**

By following several steps of simple Matlab matrix operation, the final cross-reference matrix is listed in **Table A7**.

- **Evaluation vector**

In the equation $S=\alpha N+\beta l\cdot e^T$, by following the recommendation from Eigenfactor, we first set the variable α and β to be 0.8 and 0.2 separately. The elements of the leading vector are all negative. Then we change α and β to be 0.85 and 0.15, the result turns out to be satisfying: the greatest eigenvalue and elements in the leading vector are all positive. After that, we alter the value of α and β, the result changed in a manageable range. The top five coaches are:

1. John Wooden 2. Dean Smith 3. Mike Krzyzewski 4. John Thompson 5. Rick Pitino.

Three of the top five coaches are also in the top five list of Model I.

4.6.2. Football

The data selection and processing method are completely in consistent with the method used in 4.6.1 section. α and β remains to be 0.85 and 0.15. The top five coaches are:

1. Frank Thomas 2. Joe Paterno 3. Urban Meyer 4. Don James 5. Bear Bryantlisted

Three of the top five coaches are also in the top five list of Model I.

4.6.3. Baseball

Like how we process the data in basketball, the final top five coaches are:

1.John Barry 2. Jim Morris 3. Gary Ward 4. Gene Stephenson 5. Rod Dedeaux

Again, three of the top five coaches are also in the top five list of Model I.

4.6.4. Sensitivity analysis

In the process of calculating evaluation vector, the α and β are defined to be 0.85 and 0.15. But it is not often the case: in various sports, the partial coefficient can change slightly.

We only take basketball for example here. First, we set the partial coefficient α and β to be 0.8 and 0.2, the amount of change(in percentage) in the returned evaluation vector is showed in **Table 14**:

Table 14: The Amount of Change

Number	1	2	3	4	5	6	7	8	9	10
Change (in %)	-3.73	-3.06	-1.31	-7.50	-5.61	-1.66	-2.13	-4.97	-1.48	-5.80
Number	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Change (in %)	-6.65	12.05	-7.09	-6.45	-5.73	15.23	-1.69	-6.25	14.74	-3.35

The ranking list of top five is:

1. John Wooden 2.Dean Smith 3.Mike Krzyzewski 4.Bob Knight 5.Rick Pitino

This list is closer to the list obtained in Model I(**Table 3**), in which four of the coaches are the same.

Secondly, we set the partial coefficient α and β to be 0.9 and 0.1, the returned evaluation vector changed in a moderate range. See **Table 16**.

Table 16: The Amount of Change

Number	1	2	3	4	5	6	7	8	9	10
Change (in %)	3.03	2.12	0.29	7.23	5.21	0.5	1.3	4.38	0.99	4.99
Number	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Change (in %)	6.3	-12.74	6.77	6.03	5.4	-16.02	1.45	5.64	-16.05	2.29

The ranking list of top 5 in this case is:

1. John Wooden 2.Mike Krzyzewski 3.Dean Smith 4.John Thompson 5.Rick Pitino

Three of the coaches are listed in Model I.

As can be seen from above, with a slight change in the partial coefficient, the result changed in a controllable range and remain close to the result in Model 1. The lager β is, the more weight it is attributed to individual influence and vice versa.

5. Applicability

- **Genders**

As for gender factors, we can employ two models to evaluate coaches with different genders due to the same evaluation indexes. Through the data collected by Internet[20], we use and modify model I to obtain the women basketball result. See **Table 16**.

Table 16: Top 5 Coaches of Women Basketball

No.1	No.2	No.3	No.4	No.5
Pat Summitt	Tara VanDerveer	Barbara Stevens	C. Vivian Stringer	Geno Auriemma

The ranking list of top coaches is agreement with the data from Internet.

- **Other Sports**

As for other sports，like Track and Field Events or Swimming Events, these evaluation systems of two models can be applied very well. For model I, we only need to change evaluation indexes and modify the data, while the analytical method is the same. For model II, because we do not need to consider other evaluation indexes but the data of search result, we also evaluate and rank top coaches of other sports conveniently.

6. Strengths and Limitations

6.1. Model Ⅰ

- **Strengths**
  - — **High accuracy**. We combine objective and subjective indexes in model Ⅰ, and meanwhile, we employ subjective and objective methods to evaluate the rank of coaches in the previous century. Therefore, this evaluation system has a high accuracy and is accordance with reality.
  - — **Extendibility**. With so many sports and fields, this model extracts the common features of different sports so that it can be adapted to a large range of sports and fields via comparing the same metrics.
  - — **Easy to understand**. This model is succinct and clear, which can be easily understood.
- **Limitations**
  - — **The lack of data**. When a game is not hot, the data are difficult to find, which may result in the lack of data.
  - — **Difficult to determine weights**. In order to obtain an accurate result, this model needs to determine weights carefully. However, it is difficult to choose appropriate values of weights.

6.2. Model II

- **Strengths**
  - — **Efficient**. Once the model has available data, the final result can be obtained efficiently, which means that it do not rely largely on the work of human.
  - — **The lack of data do not occur**. Since the data of model is based on the searching number of Google, the needed data is always sufficient.
  - — **Simple**. Information of coaches do not need to be specific.
- **Limitations**
  - — **The influence of media**. Since this model is based on the degree of media attention, it do not take comprehensive factors in consideration.

7. Conclusions

In model I, we have defined evaluation indexes and determined the weights of their importance. By searching and selecting needed data, we obtained the final amount of evaluation of each coach in a certain sport. Finally, we compared the evaluation amounts and listed top five coaches, which is accordant with widely accepted result. And in the sensitivity analysis, the result is satisfying and of robustness. Further work of model I should choose more evaluation indexes to rank coaches in a more fairly way.

In model II, the algorithm offers us a convenient way to evaluate the excellence of certain coach, without any need for detailed information. That is, by a series of repetitive search on the Internet, one can get a grasp of rankings in any given sports. Because of the simplicity of the search job, a program can be developed to perform the rote tasks. Given that the algorithm does not need any details, the method can be applied to any fields that need a ranking system. However, the partial coefficient in this algorithm must be determined on a mass of trail and error adjustments, which we cannot adequately explain here.

8. The Article for *Sports Illustrated*

"*Dream Team*" of College Coaches

Where there is a competition, there will always be victory, defeat, and ranking. Loyal sport fans could debate day and night over the question who is the best coach. However, people are born to be biased to their own preference. Without any objective analysis, the ranking is dubious at best. Nevertheless, after a reference to various data about college coaches from *Sports Reference*[5], we can safely put forward our unbiased evaluation. Here are 5 greatest college coaches over the course of sports history in basketball, football and baseball, as shown below:

Sport	No.1	No.2	No.3	No.4	No.5
Basketball	John Wooden	Dean Smith	Mike Krzyzewski	Adolph Rupp	Bob Knight
Football	Joe Paterno	Bobby Bowden	Bear Bryant	Tom Osborne	Don James
Baseball	John Barry	Mike Martin	Rod Dedeaux	Augie Garrido	Jim Morris

There is a dream team of college coaches in every sport fan's heart. People tend to have different criteria for "excellence". Some people would agree that trophies are the touchstones for a great coach, while others may prefer winning percentage. In view that there are a host of criteria, we consider objective and subjective factors together with different weight degrees to obtain our final result. We have incorporate total wins, the winning percentage, champions, final fours, tenure and their media influence as our basic criterion. With this method, we select five greatest college basketball coaches including *John Wooden, Dean Smith, Mike Krzyzewski, Adolph Rupp, and Bob Knight*. Amazingly, the results obtained from our comprehensive analysis models are quite similar to ideal ones hold by the majority[9].

John Wooden is thought to be one of the best college basketball coaches, or even the best coach ever. During 12 tenure years at UCLA, he won 10 national championships. The win-loss record is 664-162 and winning-percentage is 0.804, the highest in NCAA history[21]. The 27 tenure years are considered to be a glorious history of UCLA basketball team, which put the name John Wooden into spotlights. Take various factors into consideration, there is no doubt that Wooden was the greatest college basketball coach. We appreciate the honors of Wooden, furthermore, we praise his personality and contributions he had made to basketball[20]. We can tell that, a great coach can not only help his own team become better, but also promote the development of basketball.

Whenever people speak of great coaches, they often compare them with other great coaches. So we can abandon time limits to compare coaches with previous coaches, and in the same era, we can also compare any two of coaches. The relation between two great coaches can make up the network of relationships. Here we employ the data of search results from media to consider this network of relationships when two coaches were mentioned together. We also consider the effect of the tenure periods to evaluate the historical status of coaches.

We find that through using two methods to find top five coaches, we get quite similar results.

Similarly, we can employ these methods into college football and baseball coaches, though their data to evaluate coaches are not the same as basketball. For instance, their games have ties while basketball only has wins and loses. But our methods can also be used into these two sports and get satisfying results.

Maybe you want to know whether women coaches can be evaluated by our methods, we can confirm that they also fit the methods. The result can tell everything. By our methods, the ranking of top five women basketball coaches is: *Pat Summitt, Tara VanDerveer, Barbara Stevens, C. Vivian Stringer, Geno Auriemma*[20]. We find that the result is quite similar with public ideal ones. So the gender factor dose not make a difference in our methods.

College sports are various, can we put these methods into use in other sports such as hockey, track and field? We give out these two methods and they fit almost overall sports by collecting the data.

In brief, these two methods to evaluate college coaches have a high applicability towards different sports and genders. Whether you have enough data or not, you will find ways to construct your *"Dream Team"* of college coaches.

References

[1] Sport. Wikipedia.[2014-2-7].http://en.wikipedia.org/wiki/Sports.
[2] Baidu Library.The evaluation model of mathematical modeling method[EB/OL]. [2014-2-7].
http://wenku.baidu.com/link?url=jt3jrHz6x47Q0TGQsbx3Cb0BHnSTXGDWyxXPPCCZ7GwHSZC6ZbTHgTrSO_Ee4uGoItw1h8EPNKD1pEiYewEXE0u6Vi0mWIYQSn-gVt6ERdm.
[3] Qiyuan Jiang, Jinxing Xie, Jun Ye.Mathematical Modeling[M].4th ed.Higher Education Press, 2011 :249-257.
[4] Baidu Library. Grey Relational Degree[EB/OL].[2014-2-7].
http://wenku.baidu.com/view/ aba402ccba0d4a7302763a83.html.
[5] Sports Reference.[2014-2-7]. http://www.sports-reference.com/.
[6] Sports Reference. College Basketball. Coaches.[2014-2-7].
http://www.sports-reference.com/ cbb/coaches/.
[7] Sports Reference. College Basketball. Seasons.[2014-2-7].
http://www.sports-reference.com/cbb/seasones/.
[8] Men's College Basketball.[2014-2-7].
http://insider.espn.go.com/ncb/insider/columns/story?columnist=bilas_jay&id=2423815.
[9] 10 Greatest Coaches in NCAA Basketball History.[2014-2-7].
http://bleacherreport.com/articles/1341064-10-greatest-coaches-in-ncaa-basketball-history/page/6.
[10] Sports Reference. College Football. Coaches. [2014-2-7].
http://www.sports-reference.com/cfb/coaches/.
[11] Sports Reference. College Football. Seasons. [2014-2-7].
http://www.sports-reference.com/cfb/years/.
[12] Top Ten College Football Coaches of All-Time. [2014-2-7].
http://www.barrystickets.com/blog/top-ten-college-football-coaches-of-all-time/.

[13] NCAA. College Baseball. Championship History. [2014-2-7].
http://www.ncaa.com/history/baseball/d1.
[14] NCAA. College Baseball. Records. [2014-2-7].
http://fs.ncaa.org/Docs/stats/SB_Records/2011/coaches.pdf.
[15]10 Best College Baseball Coaches Ever. [2014-2-7].
http://www.mademan.com/mm/10-best-college-baseball-coaches-ever.html.
[16] Jingsong Yan.Study on the Growth Law of Web Information[D].Lanzhou university, 2006.
[17]Machang Zhao.Secret of Google-Interpretation of PageRank.[2014-2-8].
http://www. itlearner.com/ good/pagerank_cn.htm.
[18] Princeton. Wordnet Search.[2014-2-8]. http://wordnetweb.princeton.edu/ perl/webwn.
[19] Eigenfactor. Eigenfactor Method Overview.[2014-2-8].
http://www.eigenfactor.org/ methods.pdf.
[20] List of college women's basketball coaches with 600 wins.[2014-2-8].
http://en.wikipedia.org/wiki/List_of_college_womens_basketball_coaches_with_600_wins.
[21] John Wooden. [2014-2-8].
http://www.sports-reference.com/cbb/coaches/john-wooden-1.html.

Appendix

Table A1: Selected 20 Basketball Coaches and Their Metrics Data

	Name	win-lose	years	pct.	champions	final fours	season	Google
S_1	Phog Allen	719-259	48	0.735	1	3	1905	182000
S_2	Jim Boeheim	942-314	38	0.75	1	4	1976	918000
S_3	Jim Calhoun	877-382	40	0.697	3	4	1972	2660000
S_4	John Calipari	585-171	22	0.774	1	4	1988	2470000
S_5	Denny Crum	675-295	30	0.696	2	6	1971	426000
S_6	Don Haskins	719-353	38	0.671	1	1	1961	1540000
S_7	Hank Iba	752-333	40	0.693	2	4	1929	5020000
S_8	Mike Krzyzewski	975-302	39	0.764	4	11	1975	1430000
S_9	Bob Knight	899-374	42	0.706	3	5	1965	34000000
S_{10}	Lute Olson	776-285	34	0.731	1	5	1973	298000
S_{11}	Rick Pitino	681-239	28	0.74	2	7	1978	1880000
S_{12}	Adolph Rupp	876-190	41	0.822	4	6	1930	375000
S_{13}	Nolan Richardson	509-207	22	0.711	1	3	1980	1020000
S_{14}	Dean Smith	879-254	36	0.776	2	11	1961	60100000
S_{15}	Bill Self	524-169	21	0.756	1	2	1993	2390000
S_{16}	Jerry Tarkanian	761-202	30	0.79	1	4	1969	1150000
S_{17}	John Thompson	596-239	27	0.714	1	3	1972	4210000
S_{18}	Roy Williams	715-187	26	0.793	2	7	1988	2140000
S_{19}	John Wooden	664-162	29	0.804	10	12	1946	39700000
S_{20}	Gary Williams	668-380	33	0.637	1	2	1978	2130000

Table A2: The Number of Teams in Every Ten Year

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
Num	136	174	171	158	152	173	203	264	295	318	345

Table A3: The Standard Difference of Winning-percentage in Every Ten Year

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
s	0.069	0.063	0.068	0.050	0.055	0.034	0.075	0.040	0.075	0.043	0.068

Table A4: The final evaluation index matrix $\mathbf{x}^*$ of Basketball

	x_1''	x_2''	x_3''	x_4''	x_5''	x_6''
S_1	0.592	0.284	0.182	0.000	1.000	0.055
S_2	0.157	0.389	0.273	0.000	0.794	0.083
S_3	0.149	0.246	0.273	0.222	0.841	0.145
S_4	0.030	0.415	0.273	0.000	0.136	0.138
S_5	0.103	0.331	0.455	0.111	0.552	0.064
S_6	0.165	0.322	0.000	0.000	0.794	0.106
S_7	0.854	0.476	0.273	0.110	0.841	0.219
S_8	0.170	0.445	0.909	0.333	0.818	0.102
S_9	0.187	0.368	0.364	0.222	0.885	0.748
S_{10}	0.122	0.397	0.364	0.000	0.685	0.059
S_{11}	0.082	0.399	0.545	0.111	0.473	0.118
S_{12}	1.000	1.000	0.455	0.333	0.864	0.062
S_{13}	0.026	0.310	0.182	0.000	0.136	0.087
S_{14}	0.251	0.736	0.909	0.111	0.742	1.000
S_{15}	0.000	0.300	0.091	0.000	0.055	0.136
S_{16}	0.147	0.710	0.273	0.000	0.552	0.092
S_{17}	0.076	0.405	0.182	0.000	0.428	0.195
S_{18}	0.063	0.453	0.545	0.111	0.380	0.127
S_{19}	0.520	0.995	1.000	1.000	0.514	0.813
S_{20}	0.069	0.000	0.091	0.000	0.655	0.127

Table A5: Overall search results and winning search results of basketball

Year	Name	Overall	Winning
2003.5	Bill Self	2390000	256000
2001	Roy Williams	2140000	883000
1999	John Calipari	2470000	578000
1995	Jim Boeheim	918000	254000
1994.5	Mike Krzyzewski	1430000	476000
1994.5	Gary Williams	2130000	1470000
1992	Jim Calhoun	2660000	136,000
1992	Rick Pitino	1880000	671,000
1991	Nolan Richardson	1020000	31,500

1990	Lute Olson	298000	275,000
1986	Denny Crum	426000	31,500
1986	Bob Knight	34000000	215,000
1985.5	John Thompson	4210000	1,160,000
1984	Jerry Tarkanian	1150000	59,000
1980	Don Haskins	1540000	313,000
1979	Dean Smith	60100000	397,000
1960.5	John Wooden	39700000	1,010,000
1950.5	Adolph Rupp	375000	94,000
1949	Hank Iba	5020000	31,500
1929	Phog Allen	182000	32,200

Table A6: Parts of Original Cross-Reference Matrix of Basketball

	Bill Self	Roy Williams	John Calipari	Jim Boeheim	Mike Krzyzewski
Bill Self	0	92900	97100	282000	271000
Roy Williams	92900	0	111000	86300	111000
John Calipari	97100	111000	0	81400	69700
Jim Boeheim	282000	86300	81400	0	140000
Mike Krzyzewski	271000	111000	69700	140000	0

Table A7: Parts of Final Cross-Reference Matrix of Basketball

	Bill Self	Roy Williams	John Calipari	Jim Boeheim	Mike Krzyzewski
Bill Self	0	0.002142	0.00986	0.096381	0.0426
Roy Williams	0.001882	0	0.002294	0.015612	0.009703
John Calipari	0.006253	0.001656	0	0.006777	0.003033
Jim Boeheim	0.060131	0.011086	0.006667	0	7.82E-05
Mike Krzyzewski	0.063972	0.016584	0.007182	0.000188	0