Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080302
ISSN 2371-8838 Vol. 8 Num. 3

Comprehensive evaluation model for athletes based on PageRank and complex networks

Yanjiong Zhu*

The School of Finance, Shanghai University of International Business and Economics, Shanghai, 201620, China
**Corresponding author: zyjzyj0927@163.com*

Keywords: Athletes, Evaluation, Complex Networks, PageRank

Abstract: A necessary condition for the accurate evaluation of the comprehensive strength and greatness of athletes in sports is the provision of objective and quantifiable criteria, which can reduce subjective bias and increase persuasiveness. This paper builds a model based on the characteristics of individual sports competitions, from which 'The Greatest Athlete of All Time' (The G.O.A.T.) is selected. Boxing was chosen as the object of study, and the model first collected relevant data on the BoxRec website, and built a complex network among boxers based on the relationship between opponents' fights against each other. With the support of a large amount of data, this paper uses the PageRank algorithm to score and rank the players according to the objectivity and practicality of the data, and obtain the 'greatest athlete of all time' in boxing. In order to extend the evaluation model to all individual sports, this paper subdivided the individual sports into direct and indirect athletics, and implemented differentiated evaluation. For indirect athletics, the indicators like 'relative score' and 'record keeping time', are added. The aim is to select the 'greatest athletes of all time' through comprehensive analysis.

1. Introduction

Athletes are sometimes referred to as 'The Greatest Athlete of All Time' (The G.O.A.T.) based on their past accomplishments or outstanding performances in individual or team sports. Athletes are sometimes referred to as 'The Greatest Athlete of All Time' (The G.O.A.T.) based on their past accomplishments or outstanding performances in individual or team sports. This designation is often reported in sports news. Athletes are sometimes referred to as 'The Greatest Athlete of All Time' (The G.O.A.T.) based on their past accomplishments or outstanding performances in individual or team sports. It is important to avoid subjective evaluations and maintain a clear, objective language. The text should also adhere to conventional structure and formatting features, while using precise word choice and avoiding biased language. Various factors affect the G.O.A.T. Therefore, enhancing the credibility of evaluation results requires setting objective and quantifiable criteria. This paper evaluates the strength of athletes, establishes a model based on objective data, and promotes the scientization and fairness of sports evaluation. Currently, there are limited studies on the 'comprehensive strength assessment of athletes', with most falling under the categories of medical rehabilitation and sports training. For instance, Du Huiling et al.'s study, 'The application

of MRI imaging in the evaluation of lumbar disc herniation in swimmers and the evaluation of therapeutic efficacy', is founded on the MRI-based study of sports rehabilitation [1]. There are studies that use deep learning and other methods to analyse athletes. As studied by Ghosh Indrajeet et al. in 'DeCoach: Deep Learning Based Coaching for Badminton Player Evaluation', they utilized deep learning to study the training of badminton players by evaluating their postures and stances to rank their performance [2]. This paper uses complex networks to analyse the strength of athletes, compensating for the effects of differences in the era in which they played. The PageRank algorithm is applied to synthesise connections between a large number of athletes over a long time span, resulting in the identification of 'the Greatest Athletes of All Time'.

2. A model for evaluating athletes using PageRank and complex networks

Data source: The data comes from the BoxRec website and contains the date of the match, the name of the opponent, and the win or loss situation. The specific data is shown in the table 1 below:

Table 1: Data Table for Boxers' Competition

	Date	Opponent 1	Opponent 2	W=1/L=0
0	1965/11/10	Ray Robinson	Joey Archer	0
1	1965/10/20	Ray Robinson	Rudolph Bent	1
2	1965/10/1	Ray Robinson	Peter Schmidt	1
3	1965/9/23	Ray Robinson	Harvey McCullough	1
4	1965/9/15	Ray Robinson	Neil Morrison	0
5	1965/8/10	Ray Robinson	Stan Harrington	0
…				
998	1916/7/28	Benny Leonard	Freddie Welsh	0
999	1916/6/23	Benny Leonard	Vic Moran	1
1000	1916/6/12	Benny Leonard	Johnny Dundee	0
1001	1916/5/1	Benny Leonard	Charley Thomas	1
1002	1916/4/20	Benny Leonard	Phil Bloom	1
1003	1916/3/31	Benny Leonard	Freddie Welsh	1
…				
5663	1985/7/11	Mike Tyson	John Alderson	1
5664	1985/6/20	Mike Tyson	Ricardo Spain	1
5665	1985/5/23	Mike Tyson	Don Halpin	1
5666	1985/4/10	Mike Tyson	Trent Singleton	1
5667	1985/3/6	Mike Tyson	Hector Mercedes	1

2.1 Athlete evaluation model using complex networks

When evaluating the 'greatest athletes of all time' in a particular sport, it is necessary to make longitudinal comparisons across a wide range of time periods. Since athletes' careers are sequential and there is often no direct competition between great athletes, this paper considers their indirect connections. Through the collection of games between different athletes in various time periods in history, a comprehensive and interconnected complex game network is established, and the athletes in various time periods are connected through the form of complex network, based on which to judge their superiority or inferiority.neural network is a multi-layer network with error reverse propagation, which is composed of input layer nodes, hidden layer nodes and output layer nodes.

This process has been reduced to an acceptable level of error to the network output, or to a predetermined number of learning times.

In sports competitions, the point system calculates a player's points based on three factors: the player's own strength, the game's outcome, and the opponent's strength. Points are added for winning and losing hand-to-hand combat. This paper refers to the relatively mature PageRank model to optimize the athlete evaluation model. The method of considering the opponent's point system and the complex network above is similar to the idea of the PageRank evaluation system.

PageRank is a method used by Google, the well-known web search engine, to evaluate the 'importance' or 'influence' of a web page. It is based on the traditional concept of citation analysis: when page A links to page B, page B is considered to have gained the value of A's contribution to it. The value of this contribution depends on the importance of webpage A itself. In other words, the more important webpage A is, the higher the contribution value of webpage B will be [3].

This paper applies the idea of determining the 'greatest athlete of all time' for a single sport by abstracting links between web pages as competitions between players. If athlete A has a link to B, then A loses a match to B, and B receives the points that A has contributed to him. The degree of impact of a match is considered by taking into account the relative power situation between A and B.

This paper proposes an athlete evaluation model based on complex network theory by combining the characteristics of a single sport with the concept of PageRank. The model is developed through the following process:

The competitions between athletes in individual sports can be represented as a directed graph, where each athlete is a vertex and their matches are arcs. When there are many athletes with races between them, this forms an athlete-race complex network [4].

The adjacency matrix, $B=(b_{ij})_{N\times N}$, corresponds to the links in the directed graph for subsequent calculations.If j beats i,then $b_{ij}=1$.If i beats j or if there is no match between i then j,then $b_{ij}=0$.

There are N athletes. The row sum of the adjacency matrix B, $r_i=\sum_{j=1}^{N} b_{ij}$, is the out-degree of the vertex i, which denotes the sum of the number of games lost by i. And the column sum $n_i=\sum_{i=1}^{N} b_{ij}$ of the adjacency matrix B is the in-degree of the vertex j, which denotes the sum of the number of games j won.

Suppose that the process of a match between an athlete and his opponent is independent of who he has competed with in the past, but is only related to the current opponent. Then the process of this match can be considered as a finite-state, discrete-time stochastic process, and therefore the law of the process of the match between athletes is described by a Markov chain. Define the matrix $A=(a_{ij})_{N\times N}$, where:

$$a_{ij}=\frac{1-d}{N}+d\frac{b_{ij}}{r_i} \tag{1}$$

Referring to the PageRank model, d is a model parameter, usually taken as $d=0.85$. A is the transfer probability matrix of Markov chain, and a_{ij} represents the probability of i losing to j in

the model. According to the basic properties of Markov chain, there exists $x=[x_1,\cdots,x_N]^T$ satisfies:

$$A^T x = x, \sum_{i=1}^{N} x_i = 1 \tag{2}$$

x denotes the probability distribution of each athlete's victory over other athletes when the number of transfers tends to be infinite, which is defined as an athlete's 'strength' in this paper. Then x satisfies the equation:

$$x_k = \sum_{i=1}^{N} a_{ik} x_i = (1-d) + d \sum_{i:b_{ik}} \frac{x_i}{r_i} \tag{3}$$

Athlete i has a 'strength' of x_i and has lost r_i matches, so divide his strength value x_i into r_i shares. Add each share to the athlete who beat him. x_k, this is the 'strength' of athlete k, i.e. the final value added to athlete k by the athletes he has beaten in all the matches.

2.2 The establishment of evaluation model for boxers

This paper takes boxing as an example and uses a complex network-based athlete evaluation model to select the 'Greatest Athletes of All Time'. The data is collected based on the ranking of the top 50 boxers of all time on Live About. These 50 boxers were used only as a reference for data collection to build a network of boxing matches in which the top boxers have fought almost all of the best boxers of all time, so that the network can be considered to contain all candidates for the 'Greatest Athletes of All Time'. All of the historical fight records for these 50 fighters were collected from BoxRec.com (Boxing Records Archive) and organized into a database as shown in Table 1 in the data source.

This more than 5500 data contains more than 3600 athletes across all weight classes, with a large sample size, wide coverage, and a time span of nearly 100 years, which should reflect the best athletes in boxing history.

The data was processed using a program and a directed network was generated as shown in Figures 1 and 2.

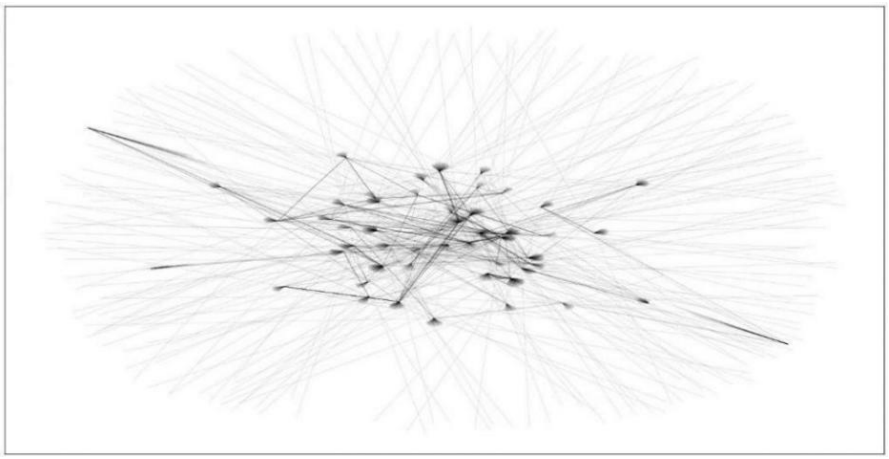

Figure 1: Complex network in its entirety (abstract)

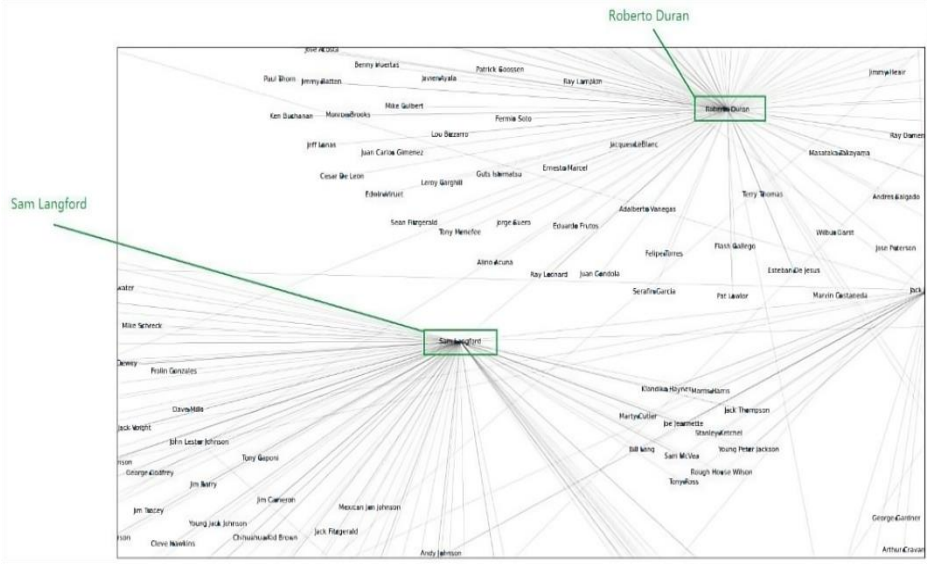

Figure 2: Complex network localization

The adjacency matrix $B = (b_{ij})_{N \times N}$, derived from the directed network, is shown in Figure 3.

	0	1	2	3	4	5	6	7	8	9	...	3618	3619	3620	3621	3622	3623	3624	3625	3626	3627
0	0	1	0	0	0	1	1	1	1	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3623	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3624	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3625	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3626	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3627	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

3628 rows x 3628 columns

Figure 3: Adjacency matrix

Based on the adjacency matrix, the following steps are performed:

1) Calculate the r_i, n_i values of the adjacency matrix,

2) The probability transfer matrix A is obtained from $a_{ij} = \frac{1-d}{N} + d\frac{b_{ij}}{r_i}$,

3) Calculate the player's x_k value based on $x_k = \sum_{i=1}^{N} a_{ik} x_i$,

4) Multiple iterations are performed to obtain a smooth distribution of Markov chains $x = [x_1, \cdots, x_N]^T$, which is the player's strength value.

The results of the calculations were taken from the top 15, as shown in Table 2.

Table 2: Calculation results of boxers' strength

serial number	Contestants	Score
1	Carlos_Monzon	0.85161
2	Marvin_Hagler	0.73098
3	Sam_Langford	0.64908
4	Archie_Moore	0.49355
5	Ted_Kid_Lewis	0.48892
6	Joe_Gans	0.44306
7	Willie_Pep	0.43769
8	Henry_Armstrong	0.41479
9	Ray_Robinson	0.4089
10	Tony_Canzoneri	0.38321
11	Roberto_Duran	0.35179
12	Bernard_Hopkins	0.34206
13	Ray_Leonard	0.31986
14	Benny_Leonard	0.31772
15	Julio_Cesar_Chavez	0.31719

From the results shown in Table 2, Carlos Monzon is the ‘greatest athlete of all time’ in the sport of boxing in history.

3. Results

3.1 Analysis of calculation results

In this article, we found some of the official media rankings of the ‘50 Greatest Boxers of All Time’ for comparison, as shown in Table 3.

Table 3: Comparison of calculation results and media ranking

serial number	Calculate Ranking	Media Ranking
1	Carlos_Monzon	Muhammad Ali
2	Marvin_Hagler	Sugar Ray Robinson
3	Sam_Langford	Jack Dempsey
4	Archie_Moore	Sam Langford
5	Ted_Kid_Lewis	Roberto Duran

As can be seen from the table, there is not much difference between the rankings obtained in this paper and the rankings given by the media, considering that the media has a certain degree of subjective opinions and introduces more external criteria for judgment. The rankings in this paper are based on the collected database, which reflects the rationality of the model to a certain extent.

3.2 Extension of the model

The above complex network-based athlete evaluation model is based on the direct competition between players to calculate the adjacency matrix, and in order to generalize it to any single sport, this paper divides the single sport into two different categories:

1) Direct competition category, the players directly between the game, divided into winners and losers. Such as badminton, tennis, boxing, etc.

2) Indirect sports category, the players through the score of high and low to distinguish winners and losers. Such as high jump, gymnastics, diving and so on.

For the first category: according to the model, it is only necessary to collect enough race data to build a directed network among the players to be able to use the same method to abstract the race results into mathematical concepts and substitute them into the model to calculate the 'greatest athlete of all time'.

For the second category, since there is no direct competition between the players, it is not possible to simply build a directed network with the results. However, this type of competition is used in high-level international events (e.g., the Olympics), where a winner is determined by a knockout tournament [5]. Therefore, eliminating an opponent to advance to the next round can be considered as 'beating' the opponent and can be analyzed using the same algorithm as in direct competition sports. In addition, it should be noted that in many cases the greatness of a person is evaluated by his or her athletic performance in a particular sport, as well as by the number of points and times he or she breaks and holds a record. To account for this characteristic, it is planned to add 'relative score', 'record time', 'record score', 'improvement over previous score' to the original algorithm to reflect the characteristics of these indirect athletic competitions in the hope of analyzing and comparing to select the greatest athletes. The flow chart is shown in Figure 4.

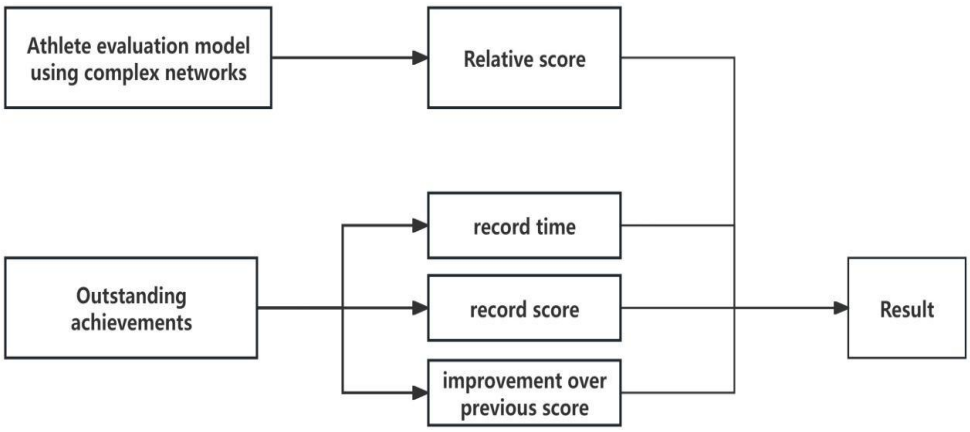

Figure 4: Flowchart of any single sport evaluation model

4. Conclusions

In this paper, the complex network is used to analyze the strength of an athlete, which compensates the influence of the difference of the era in which the athlete lives. At the same time, the PageRank algorithm is applied, and a comprehensive evaluation model of athletes based on PageRank and complex networks is proposed. Boxing is also selected as a sport, and the found data are calculated and compared with the ranking of official media. The results show that the rankings obtained in this paper are not much different from the rankings given by the media, and the rankings in this paper are based on the collected data to reduce the influence of subjective factors. Therefore, the comprehensive evaluation model of athletes based on PageRank and complex network can accurately and effectively evaluate the comprehensive strength and greatness of athletes.

However, in the actual evaluation of the comprehensive strength of athletes, it will be affected by many factors. Therefore, when evaluating the comprehensive strength of athletes in other sports, a large amount of data is needed to support the model and calculation. Meanwhile, this paper has only studied a single sport using boxing as an example, which has fewer influencing factors. More research is needed on how to introduce more influencing factors so that this comprehensive strength assessment system can cover team sports.

References

[1] Du Huiling Liu Changbian, MRI imaging in the assessment of lumbar disc herniation in swimmers and evaluation of therapeutic efficacy [J]. Imaging Science and Photochemistry, 2022, 40(6): 1581-1585.
[2] Ghosh I, Ramamurthy S R, Chakma A, et al. DeCoach: Deep Learning-based Coaching for Badminton Player Assessment [J]. Pervasive Mob. Comput. 2022, 83. 101608.
[3] Li G L , Li H , Wang Y R ,et al.The Solution to Node Importance in Complex Networks Based on PageRank Algorithm[C]//International Conference on Frontiers of Manufacturing Science and Measuring Technology.2014.
[4] Li Xiaolong. Research on node ranking algorithm in big data of table tennis tournament based on complex network [J]. Journal of Zhejiang Institute of Commerce and Industry, 2020, 19(1):6.
[5] Li Mengchu. Analysis of new standards and methods of sports event classification [J]. Contemporary Sports Science and Technology, 2014(13): 170-170.