International Journal of Engineering and Management Research e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-10, Issue-2 (April 2020)
www.ijemr.net https://doi.org/10.31033/ijemr.10.2.10

The Problem of the Chinese Basketball Association Competing for the Championship

He Zhang
Student, Department of Chemistry, Yanbian University, CHINA

Corresponding Author: 850455757@qq.com

ABSTRACT

In China, basketball is a very popular sport and is loved by the people. The Chinese Basketball Association is China's highest-level basketball league. While watching the game, more and more people want to predict the outcome of the game. There are currently 14 teams participating in the game. According to the rules of the match, the game is divided into two stages: regular season and playoffs. Each game must have a victory, and each team has a fixed number. This article is based on about 100 historical score data from each team for four years, By building a mathematical model, Analyze and calculate the winning probability of the 14 teams. And give a qualitative analysis of the level of the 14 teams in the CBA league.

Keywords— Winning Probability Prediction, Time Series Model, Fuzzy Comprehensive Evaluation, Basketball League

I. INTRODUCTION

We process the comprehensive score data of each team's live performance in each game. First of all, divide these data into four groups, each of which is defined as the competition they participated in a year, and these four groups have different weight ratios. The fourth year has the largest weight. Then use matlab to compare the data between the two teams in order. If the value of the team's data is greater than the other pair, record it as 1, Then add up and divide by the number of historical score data to calculate four sets of probabilities. Each group is multiplied by their corresponding weight to obtain the team's probability of winning in the regular season. According to the rules of the playoffs, use the permutation and combination method to calculate the probability of winning each team. This paper establishes a time series model, and calculates four sets of winning probabilities based on historical score data and arranges them in chronological order to form a time series to infer the winning probabilities of each team in the next year.

This paper establishes a first-level fuzzy comprehensive evaluation model for the three variables: the degree of stability, the level of on-site performance, and the average level. Divide the data into four groups, perform group analysis and calculation, and calculate its horizontal range into five comment sets excellent, good, average, poor, and very poor. At the same time, the corresponding weights of each factor are determined and a fuzzy relation matrix from factor set to comment set is established. Finally, the comment with the largest value is taken as the comprehensive evaluation result.

II. TEAM WIN PROBABILITY ANALYSIS MODEL

According to the CBA regular season game rules, fourteen teams play against each other, and each team has to play 26 games. About 100 historical scoring data of each team is divided into 4 groups, each group is one year, and the results of the first 26 groups are the results of the most recent year. The fourth year is the most reliable, and the weighting ratio should be the largest. Use matlab to compare and analyze the historical score data of each group with each other, if

$$a_i > b_j \quad (i, j = 1, 2, 3 \ldots 100) \tag{1}$$

The data x_{ij} is recorded as 1, if

$$a_i < b_j \quad (i, j = 1, 2, 3 \ldots 100) \tag{2}$$

The data x_{ij} is recorded as 0.

$$P_t = \sum \frac{e_{ij}}{\frac{26}{13}} \quad (t = 1,2,3,4) \tag{3}$$

Formula (3) is the probability that Team A will win against other teams in the past four years. The weights for the first to fourth years are set to be 0.1, 0.2, 0.3, and 0.4. The total probability of team A winning in the regular season is

$$P = 0.1P_1 + 0.2P_2 + 0.3P_3 + 0.4P_4 \tag{4}$$

Use this method to calculate the total probability of other teams winning in the regular season.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-10, Issue-2 (April 2020)
https://doi.org/10.31033/ijemr.10.2.10

TABLE I
REGULAR SEASON WIN PROBABILITY

Team	A	B	C	D	E	F	G
Win rate (%)	29.68	55.28	57.17	51.27	34.72	43.17	39.26
Team	H	I	J	K	L	M	N
Win rate (%)	41.15	36.12	41.88	44.62	54.96	36.58	59.92

From this group of data, we can see the probability of each team winning in the regular season, sort the winning probability, and obtain the ranking of each team in the regular season as N C B L D K F J H G M I E A, and verify the accuracy of the ranking of this group below.

Use the historical scoring data to find the average value x_i (i=1,2,3,4) of each team for each year, and use this set of data to find its weighted average

$$x = 0.1x_1 + 0.2x_2 + 0.3x_3 + 0.4x_4 \quad (5)$$

TABLE II
WEIGHTED AVERAGE OF HISTORICAL SCORE DATA

Team	A	B	C	D	E	F	G
average value	3.294	5.417	6.204	5.150	3.743	5.120	4.477
Team	H	I	J	K	L	M	N
average value	4.967	4.196	4.822	4.885	6.231	4.051	6.619

Combining these two sets of data, we roughly get the ranking of the regular season as N L C B D F H J K G I M E A. According to the calculation of the total probability of each team winning in the regular season, we know that each team's winning ratio accounts for the total winning ratio. The analogy is the average winning percentage of each team in the playoffs

TABLE III
PLAYOFF AVERAGE WIN RATE

Team	A	B	C	D	E	F	G
average win rate (%)	4.74	8.83	9.14	8.19	5.55	6.89	6.27
Team	H	I	J	K	L	M	N
average win rate (%)	6.57	5.77	6.69	7.13	8.78	5.84	9.58

Based on the above data, we know that the top four players in the regular season this year are N L C B. According to the CBA competition system, the top four players in the regular season go directly to the quarter-finals. Simulate average probability, using permutation and combination algorithm

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-10, Issue-2 (April 2020)
https://doi.org/10.31033/ijemr.10.2.10

$$P = C_n^k P^l (1-P)^{n-k} \quad (6)$$

The teams that have not entered the top four need to play three games and win two games. The average simulated probability of other teams is

$$P = C_n^k P^l (1-P)^{n-k} C_3^2 P^2 (1-P)^1 \quad (7)$$

The above is the formula of the average winning probability of each team. After calculating the proportion of each team, the average winning rate of each team is obtained as shown in Table IV.

TABLE IV
AVERAGE CHAMPIONSHIP RATE

Team	A	B	C	D	E	F	G
Average championship rate	0.02	19.98	21.91	0.99	0.05	0.18	0.13
Team	H	I	J	K	L	M	N
Average championship rate	0.19	0.039	0.22	3.48	19.12	0.07	33.51

Calculate using the average probability of winning the championship over the years and the following model.

III. TIME SERIES MODEL

Let the time series of group A be y1, y2, y3, y4, y5, a is the weighting factor. 0 < a < 1, take a = 0.2, 0.5 and 0.8 for calculation, Initial value

$$S_0^{(1)} = \frac{y_1 + y_2}{2} = \frac{29.29\% + 31.95\%}{2} = 30.62\% \quad (8)$$

$$\hat{y}_1 = S_0^{(1)} = 30.62\% \quad (9)$$

Predictive model

$$\hat{y}_{t+1} = S_t^{(t)} \quad (10)$$

that is

$$\hat{y}_{t+1} = a y_t + (1-a)\hat{y}_t \quad (11)$$

Calculate the predicted value of team A for each year

TABLE V
VARIABLE GROUP

T	1	2	3	4
team A win probability	29.29%	31.95%	28.11%	27.60%
Predictive value $\hat{y}_t$ (a=0.2)	30.35%	30.67%	30.15%	29.64%
Predictive value $\hat{y}_t$ (a=0.5)	29.96%	30.96%	29.54%	28.57%
Predictive value $\hat{y}_t$ (a=0.8)	29.56%	31.47%	28.78%	27.84%

Calculate the prediction standard error S at different weighting coefficients a, and select the value of a that minimizes S.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-10, Issue-2 (April 2020)
https://doi.org/10.31033/ijemr.10.2.10

TABLE VI
WEIGHT

A	0.2	0.5	0.8
S	1.669	1.238	0.450

calculations proves that: when a = 0.8, S is the smallest, therefore, a=0.8 is selected to predict the winning probability of team A this year.

$$\hat{y}_5 = a y_4 + (1-a)\hat{y}_4 \quad (12)$$

Use this method to calculate the probability that other teams will win next year.

TABLE VII
PROBABILITY OF WINNING NEXT YEAR

Team	A	B	C	D	E	F	G
Winning probability (%)	0.016	19.871	22.105	0.811	0.062	0.182	0.137
Team	H	I	J	K	L	M	N
Winning probability (%)	0.192	0.039	0.220	3.492	19.214	0.082	34.509

IV. FIRST-LEVEL FUZZY COMPREHENSIVE EVALUATION MODEL

The variance represents the stability of each team's strength, and the average value represents the average level of the team.

Variance formula

$$\sigma^2 = \frac{\sum (x - \mu)^2}{n} \quad (13)$$

Average formula

$$X = \frac{\sum_{i=1}^{n} x_i}{n} \quad (14)$$

TABLE VIII
VARIANCE AND AVERAGE VALUE

team	average value	variance
A	3.18364	8.99023
B	5.31673	6.33259
C	5.94717	10.43111
D	5.00924	11.12333
E	3.78569	2.87088
F	4.89017	7.55718

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-10, Issue-2 (April 2020)
https://doi.org/10.31033/ijemr.10.2.10

G	4.43262	7.77291
H	5.04875	8.34586
I	4.15575	3.84332
J	4.78661	5.77152
K	4.82817	6.66793
L	6.31773	5.81866
M	4.12178	4.05495
N	6.45025	4.72911

A first-level fuzzy comprehensive evaluation model was established to qualitatively analyze the level of the 14 teams in the CBA league.

4.1 Determine factor set

First of all, a comprehensive assessment of the stability of the fourteen teams' performance, on-field performance and average level. These factors constitute the evaluation index system for evaluating the level of fourteen teams, that is, the factor set

$$U = U_1 + U_2 + U_3 \tag{15}$$

4.2 Determining the review set

$$V = \{A\upsilon_1, B\upsilon_2, C\upsilon_3, D\upsilon_4, E\upsilon_5\} \tag{16}$$

4.3 Determine the weight of each factor

$$A = [0.35, 0.4, 0.25] \tag{17}$$

4.4 Determine the fuzzy comprehensive evaluation matrix

About 1400 historical rating data of fourteen teams are roughly divided into four groups, and the variance of each group is calculated. 2.82 to 2.89 is excellent, 2.74 to 2.82 is good, 2.67 to 2.74 is average, 2.60 to 2.67 is poor, and 2.52 to 2.60 is very poor.

$$R_1 = [0.25, 0.25, 0.25, 0, 0.25] \tag{18}$$

For the performance level of the 14 teams, we sorted about 1,400 historical score data of the 14 teams in descending order. 13.17-17.64 is excellent, 8.71-13.17 is good, 4.24-8.71 is normal, -0.23-4.24 is poor, and -4.70 to -0.23 is very poor.

$$R_2 = [0.008, 0.07, 0.59, 0.308, 0.024] \tag{19}$$

Similarly, the historical score data of the fourteen teams are divided into four groups, and the average value of each group is calculated. 4.50-5.08 is excellent, 3.92-4.50 is good, 3.34-3.92 is normal, 2.76-3.34 is poor, and 2.18-2.76 is very poor.

$$R_3 = [0.75, 0, 0, 0, 0.25] \tag{20}$$

It is thought that the i-th row constitutes a fuzzy relation matrix from the factor set U to the comment set V

$$R = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0 & 0.25 \\ 0.01 & 0.07 & 0.59 & 0.31 & 0.02 \\ 0.75 & 0 & 0 & 0 & 0.25 \end{bmatrix} \tag{21}$$

Perform matrix composition operations

$$B = A * R = [0.35, 0.4, 0.25] \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0 & 0.25 \\ 0.01 & 0.07 & 0.59 & 0.31 & 0.02 \\ 0.75 & 0 & 0 & 0 & 0.25 \end{bmatrix}$$

$$= [0.28, 0.12, 0.32, 0.12, 0.16] \tag{22}$$

The comment with the highest value is taken as the result of comprehensive evaluation, and the result is "average".

V. CONCLUSION

The model considers the format of the competition and clearly reflects the rankings. The fuzzy comprehensive evaluation model fully analyzes from all aspects to comprehensively obtain the CBA league strength. At the same time, it uses weight analysis to increase the accuracy.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-10, Issue-2 (April 2020)
https://doi.org/10.31033/ijemr.10.2.10

REFERENCES

[1] Sima Kui & Sun Zhaoliang. (2018). *Mathematical modeling algorithms and applications*. China: National Defense Industry Press.

[2] Ceng Yuhua, Yang Xuxin, & Cheng Xiayan. (2009). Construction of NBA schedule analysis and evaluation mathematical model. *Journal of Chongqing University of Science and Technology (Natural Science Edition)*, *11*(02), 157-162.

[3] Wang Jijian. (2014). Application of fuzzy comprehensive evaluation method in NBA schedule Evaluation. *Bulletin of Science and Technology, 30*(01), 24-30+38

[4] Xin Chi & Zhao Xueqing. (2012). Analysis and prediction of sports performance based on time series analysis. *Technology Wind, 16*, 185-186.