International Journal of Innovation in Science and Mathematics
Volume 8, Issue 3, ISSN (Online): 2347–9051

A Study of CBA League Team's Winning Probability Based on "Grey Prediction"

Ning Kang [1] and Hui Xu [1*]
[1] Department of mathematics, school of science, yanbian university, Yanji 133002 P.R. China.
*Corresponding author email id: 601617267@qq.com

Date of publication (dd/mm/yyyy): 12/06/2020

Abstract **– Currently, the Chinese men's basketball league (CBA) is the highest level basketball league in China. In order to study the probability of each team winning the championship and its ranking, this paper uses the grey prediction model and the normal distribution model to analyze the data, and further uses the MATLAB software to calculate the probability of each team winning the championship. According to the data, among the 14 teams, the highest probability of winning the championship is 84.03%, which is N team. The team with the lowest probability of winning the championship is 19.00%, which is team A. Thus it can be seen that there is still a certain gap between the level of each team, but there is a corresponding space for progress, we can work together to promote the new development of China's sports development.**

Keywords **– CBA League Team, Grey Prediction Model, Optimization, Normal Distribution, Win Probability.**

I. INTRODUCTION

Chinese men's basketball professional league is the highest level of basketball league in China. In the past 20 years, it has become a professional brand with high international influence in China. CBA has made remarkable achievements in promoting and popularizing basketball, cultivating excellent basketball players, improving the level of Chinese basketball, enriching the people's entertainment life, and promoting the socialization and industrialization of Chinese basketball. Along with the attention of people from all walks of life at home and abroad, the overall level of the CBA team is also concerned. In order to study the winning probability of each team, to solve this problem by mathematical model, more accurate prediction can be obtained.

Article [1] studied the grey model GM (1, 1,) from the aspects of the generation and development of the model, modeling methods, application examples, application scope and model optimization. Article [2] puts forward the concept of grey system, and then studies some basic properties of grey parameters, grey matrix and grey system, and puts forward some corresponding theorems. In this paper [3], the traditional grey prediction method is improved and the grey prediction algebraic recursive equation is established to replace the original grey prediction differential equation or whitening equation. However, because the prediction accuracy is still low, the adjacent generation formula is improved, and the corresponding weight is determined according to the automatic optimization method. Aiming at the low accuracy of the traditional model prediction and the one-sided disadvantages of its solution optimization and polynomial fitting, an improved grey equal-dimensional dynamic prediction method based on the combination of solution optimization and polynomial fitting optimization is proposed in this paper [4]. In view of the large error of the grey model commonly used in prediction, the paper [5] gives the reason for the inconsistency between the basic form of the model (used for parameter estimation) and the structure of the whitening equation of the model (used for prediction), and further selects the three-sum method and three-point method to optimize the model and improve the accuracy. In this paper, the gray model was used to predict the 1-26 data selected from the 14 teams by MATLAB. Through the establishment of the model and the application of its characteristics, the comprehensive data of each team's on-

the-spot performance was predicted and further ranked. Then, according to the normal distribution principle, the winning rate of each team in rotation with other teams was calculated, and the winning rate of each team was weighted and averaged, and then the probability of each team winning the championship was comprehensively predicted. After that, according to the characteristics of grey prediction, the defects are found out, and the model is optimized to make it more in line with the prediction curve of the original data, so that people can understand the real level of each team more accurately and the team can make better self-adjustment to make progress.

II. Model for Predicting the Probability of Winning

1. *Data Acquisition and Preprocessing*

CBA basketball league to carry on the regular season and the playoffs two stages, the regular season refers to all participating teams to carry on the double cycle competition, the top eight of the regular season enters the playoffs, the playoffs adopt the cross knockout stage. According to this rule,14 teams in CBA basketball league need to play 26 games a year. And basketball games have the inertia of winning and losing, so the prediction with recent data is more persuasive. Assuming that the historical scores of each team are arranged in chronological order, so in the 100 historical data provided, we select 1-26 data to represent the on-the-spot performance of 26 games in nearly a year of each team.

2. *Establishment and Solution of Grey Prediction Model*

The simple principle of the prediction model like *GM* (1, 1) means that the first order differential equation is established and solved by using the cumulative technique to make the data have exponential law, and then the result is reduced and reduced to grey prediction value, so as to predict the future.

A. *Set the Known Original Reference Number as* $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$

The first cumulative number is listed as:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)) = (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2), \ldots, x^{(0)}(1) + \ldots + x^{(0)}(n)$$

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i),\ k = 1, 2, \ldots, n \tag{1}$$

An adjacent value in an original sequence called a sequence of numbers $x^{(0)}(k-1), x^{(0)}(k)\ x^{(0)}$

$x^{(0)}(k-1)\ x^{(0)}(k)$ For a posterior neighbor, for a constant $a \in [0, 1]$

$$Z^{(0)}(k) = ax^{(0)}(k) + (1-a)\,x^{(0)}(k-1),\ k = 2, 3, .., n$$

The resulting series $x^{(0)}$

α It is called the number of adjacent values generated under the weight of the sequence, and the weight is also called the coefficient of generation α.

In particular, when a coefficient is generated, it is called $\alpha = 0.5$,

$$Z^{(0)}(k) = 0.5x^{(0)}(k) + 0.5x^{(0)}(k-1),\ k = 2, 3, \ldots, n$$

Is the mean number, also known as the equivalent number of neighbors.

$x^{(1)}$ The mean generation sequence is:

$Z^{(1)} = (Z^{(1)}(2), Z^{(1)}(3), Z^{(1)}(4),\ldots, Z^{(1)}(n))$ (2)

$Z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2, 3,\ldots,n,$ of which (3)

B. *Establishment of Grey Differential Equations*

$x^{(1)}$ The gray derivative defined is

$D(k) = x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$ (4)

$Z^{(1)}(k) = \alpha x^{(1)}(k) + (1-\alpha)\, x^{(1)}(k-1)$ (5)

So the model of grey differential equation *GM* (1, 1) is defined as

$D(k) + \alpha z^{(1)}(k) = x^{(0)}(k) + \alpha z^{(1)}(k) = b, k = 2, 3,\ldots,n$ (6)

C. *Find Vectors* μ

Put the schedule $k = 2, 3,\ldots,n,$ in (6)

$$\begin{cases} x^{(0)}(2) + \alpha z^{(1)}(2) = b \\ x^{(0)}(3) + \alpha z^{(1)}(3) = b \\ \cdots \\ x^{(0)}(n) + \alpha z^{(1)}(n) = b \end{cases} \quad (7)$$

Introduction of matrix vector notation $\mu = \begin{bmatrix} a \\ b \end{bmatrix}$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (8)$$

So the model *GM* (1, 1) can be represented $Y = \mu B$

Now the problem comes down to the value of the coefficient of development α and the amount of grey b. We use the MATLAB to superimpose the first set of 26 data according to the above, and generate the accumulation matrix Y, and obtain the gray derivative, $x^{(0)}(k)$

$$\mu = \begin{bmatrix} \hat{\alpha} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y \quad (9)$$

On the basis of. We can predict the result of each team's next match, based on the grey prediction model, we used MATLAB to speculate on each team's next game.

Table 1. Forecasts of 14 team.

Team Name	N team	D team	C team	L team	B team	F team	K team
Projected	7.941	7.281	6.257	6.028	5.947	5.769	5.369
Ranking	1	2	3	4	5	6	7
Team Name	I team	H team	J team	M team	G team	E team	A team
Projected	5.083	4.65	4.517	4.134	4.044	3.409	2.824
Ranking	8	9	10	11	12	13	14

Clearly, we forecast results show with the result of above average μ normal distribution, according to the 3σ principles of normal distribution and the distribution of interval, we can order σ is 8/3, according to the forecast result, due to the number of games and game between them is unknown, so we, between each set of the team will play, in order to solve the forecast data of the first n of probability to win the championship, for example, we are going to this team and probability to find any other team win the game, this is the probability of solving the team beat all the team, due to this team and the prediction result of the ith the probability of winning that game with the team and the probability of the first j a different prediction results and has a regularity, we believe that this probability has an analytical weight, so the weighted average of the winning probability of all games of each team is the team's probability of winning the game.

Using the above method, we use and rank the results according to the order of the ranking of the forecast results. The results of the forecast results and the winning rate are as follows:

Table 2. The odds of winning 14 teams in rotation.

Team	N	D	C	L	B	F	K	H	I	J	M	G	E	A
N		0.4023	0.2639	0.2366	0.2273	0.2077	0.1674	0.1419	0.1086	0.0996	0.0767	0.072	0.0446	0.0275
D	0.5977		0.3505	0.3192	0.3084	0.2854	0.2367	0.2049	0.1619	0.15	0.119	0.1124	0.0733	0.0473
C	0.7361	0.6495		0.4658	0.4537	0.4274	0.3696	0.3299	0.2734	0.257	0.213	0.2033	0.1428	0.099
L	0.7634	0.6808	0.5342		0.4879	0.4613	0.4024	0.3615	0.3027	0.2855	0.2388	0.2284	0.163	0.1148
B	0.7727	0.6916	0.5463	0.5121		0.4734	0.4142	0.373	0.3134	0.2959	0.2483	0.2377	0.1706	0.1208
F	0.7923	0.7146	0.5726	0.5387	0.5266		0.4404	0.3985	0.3374	0.3194	0.2699	0.2589	0.1881	0.1347
K	0.8326	0.7633	0.6304	0.5976	0.5858	0.5596		0.4573	0.3937	0.3747	0.3216	0.3096	0.2312	0.1699
H	0.8581	0.7951	0.6701	0.6385	0.627	0.6015	0.5427		0.4355	0.416	0.361	0.3484	0.2651	0.1985
I	0.8914	0.8381	0.7266	0.6973	0.6866	0.6626	0.6063	0.5645		0.4801	0.4233	0.4101	0.3208	0.2468
J	0.9004	0.85	0.743	0.7145	0.7041	0.6806	0.6253	0.584	0.5199		0.4429	0.4296	0.3389	0.2628
M	0.9233	0.881	0.787	0.7612	0.7517	0.7301	0.6784	0.639	0.5767	0.5571		0.4865	0.3929	0.3116
G	0.928	0.8876	0.7967	0.7716	0.7623	0.7411	0.6904	0.6516	0.5899	0.5704	0.5135		0.4059	0.3237
E	0.9554	0.9267	0.8572	0.837	0.8294	0.8119	0.7688	0.7349	0.6792	0.6611	0.6017	0.5941		0.4132
A	0.9725	0.9527	0.901	0.8852	0.8792	0.8653	0.8301	0.8015	0.7532	0.7372	0.7792	0.6763	0.5869	
Wining probability	**0.8403**	**0.7718**	**0.6446**	**0.6135**	**0.6023**	**0.5775**	**0.521**	**0.4802**	**0.4189**	**0.4003**	**0.3545**	**0.3359**	**0.2557**	**0.19**

The table 2 shows that if the two sets of predicting the achievement gap, the greater the probability of the high performance team wins, the greater the as predicted results in solving n of probability to win the championship, the team ranking and forecast the probability of winning a match for the P_{na}, however in the first b and prediction results of the probability of winning that game for P_{nb}, if $a < b$ is $P_{na} < P_{nb,}$ meaning that P is a weight here, as a result, we solve every team all game winning probability weighted average of the winning probability of the team.

3. *Model Optimization*

Because this model only shows the winning situation of each team of these 14 teams, and the error caused by the CBA system, this model does not make an analysis. However, due to the large fluctuation of some data involved in the topic, there may be a distortion problem in the grey prediction, that is, the deviation of the prediction. In order to solve this problem, we make a preliminary improvement method to the model:

Rewrite the one-time accumulation formula (1) of the model *GM* (1, 1) to:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) = \sum_{i=1}^{k-1} x^{(0)}(i) + x^{(0)}(k) = x^{(1)}(k-1) + x^{(0)}(k) \tag{10}$$

And from formula (6) of the grey differential equation of the model *GM* (1, 1), it can be obtained:

$$x^{(0)}(k) + \frac{\alpha}{2}\left(x^{(1)}(k) + x^{(1)}(k-1)\right) = b = \left(1 + \frac{\alpha}{2}\right) x^{(0)}(k) + \alpha x^{(1)}(k-1) = b \tag{11}$$

$$c = 1 + \frac{\alpha}{2} \tag{12}$$

From formula (8), the following grey prediction iterative equations can be obtained:

$$x^{(0)}(n) = \frac{(c-\alpha)^{n-2}}{c^{n-1}}\left(-\alpha x^{(0)}(1) + \mu\right) \tag{13}$$

Substitute the values of parameters *a*, *b* and *c* into the above equation to get:

$$\hat{x}^{(0)}(n) = \frac{2(2-\hat{\alpha})^{(n-2)}}{(\hat{\alpha}+2)^{(n-1)}}\left(\hat{b} - \hat{\alpha}x^{(0)}(1)\right) \tag{14}$$

The algebraic recursive equation of grey prediction is derived from equation (11) so that the prediction results are only relevant $x^{(0)}(1)$. And the improvement of the model has some smoothness, which avoids the problem of prediction distortion of the original model. When we use the improved model to predict, we get the best prediction curve which is more suitable to the original data.

III. Conclusion

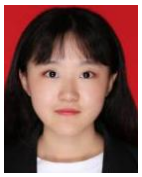
This paper uses the grey prediction model to build and solve the problem, and predicts the results of each team in the next game. Combined with the 3σ principle of the normal distribution model, we finally get the probability of fourteen teams winning the championship and their ranking. It can be seen that there is still a certain gap in the overall strength of all participating teams in CBA, which indicates that the teams with a high winning rate generally maintain a good state in a period of time, the coach makes excellent game strategies, the players cooperate well with each other, and the comprehensive strength steadily improves. Conversely, the team with a lower winning percentage may be in a period of adjustment, in a further exploration of their own strategy. By solving the team wins and rank, every team can make a better comparison and his team, so as to analyze self shortcomings, to improve the level of the team, for all the teams have progress, this means that raising the level of competition in the domestic CBA, gradually to the international level, it is also an embodiment of national soft power.

References

[1] Zhang yongbo. Research on prediction model based on grey system theory [D]. Harbin engineering university, 2005.
[2] Deng J l. grey control system [J]. Journal of huazhong institute of technology, 1982 (03) : 9-18. (in Chinese)
[3] Zhang Jinxiu, Xu Bingzhen. GM (1,1) Grey Prediction Method Improvement [J]. Statistics and Decision Making, 2016(11): 16-18.
[4] Li Mengwan, Sha Xiuyan. Improvement and application of GM (1, 1) Grey prediction model [J]. Computer engineering and applicati-

-ons, 2016, 52 (04): 24-30.
[5] Cheng Maolin. Grey model GM (1, 1) direct optimization modeling [J]. Journal of Suzhou university of science and technology (Natural Science Edition), 2018, 35(02): 6-10.
[6] Cui zhiquan. Principle and application of gaussian distribution [J]. China external education, 2018 (16) : 63-64.

AUTHOR'S PROFILE



First Author
Ning Kang, 1999, Student study in the department of mathematics in Yanbian University (Yanji, China).

Second Author
Xu Hui, Associate professor and master tutor of mathematics department, School of science, Yanbian University. Main research directions: mathematics education technology, mathematical modeling, intelligent algorithm (Yanji, China).