ATLANTIS PRESS

Advances in Engineering Research, volume 118

2nd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2017)

Olympic medal prediction

Wei Li

Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China;

liwei5240@qq.com

Keywords: Verhulst model; Correlation , the Olympic Games medal standing.

Abstract: In order to determine the Olympic metal standing, this paper proposes a Verhulst model and a Grey Correlation Calculation model.. First, we select the Chinese perfermance from 1984 to 2016 as the research object. After data preprocessing for eliminating the influence of the host effect,we propose a Verhulst model and calculate the predicted values. Next, a Grey Correlation Calculation model is adopted to to estimate the relation between the nationnal comprehensive strength and the forecast results.

1. Introduction

General specification for bathtub:1700mm in length,800mm in width,700mm in height. General capacity is 230L to 320L.When people is in the bathtub, the water should drown the people's shoulder. If the bathtub is too small, in which people would be not comfortable. If the bathtub is too big, there is a sense of floating instability. The height of the overflow gate determines the height of the water capacity.

There exists no doubt in the significance of Olympic Games is the world. The medal standings not only represents a national level of competitive sports, but also reflects a country's competitive and social comprehensive national strength. Olympic medal table is a hot topic of concern and the Olympic Games performance prediction is the hotspot of national sports.

The prediction of Olympic medals has attracted scholars from all walks of life. This research has been heating up, and in the short term, there are many kinds of prediction models for different disciplines. Theoretical methods and forecasting models have been developed and perfected rapidly.

At home and abroad , forecasting methods of the Olympic Games performance are mainly time series model , empirical model based on the principle of econometrics and neural network model. Johnson and Ali[1] mainly utilize economic methods to conduct research. Bernard and Busse[2] used the Cobb-Douglas production function to discuss the distribution of medal, and established a regression prediction model to make the prediction more accurate. Kuper and Sterken, proposed a empirical model combined Olympic participation and performance prediction. Matros and Namoro[3] adopt game theory to build a predictive model explaining why countries with roughly the same population and GDP have a large gap in Olympic performance. Oyebanke and Oyeyinka, established multiple linear regression equation and artificial neural network model, to predict the Olympic Games results.

2. Basic Prediction Model

In this section, we establish a Verhulst model[4] based on historical performance. According to the topic, we select the Chinese team as the research object. Taking into account the Chinese team return to the Olympic Games in 1984, we selected the Chinese performance from 1984 to 2016 for the original data. To make the results more accurate and reliable, we choose the gray system theory to solve the problem causing by small samples.

2.1 Data Preprocessing

According to the visual representation of the data, as can be seen in Figure 1, the number of medals fluctuates greatly in 2008, resulting in data is not smooth. According to some scholarly research, we speculate that the host effect has an effect on the outcome. In order to make the basic prediction model more accurate, we will not discuss the host effect temporarily. So do smoothing processing to the point. We get

$$x_7^{(0)}=\frac{1}{2}(x_6^{(0)}+x_5^{(0)})=75 \tag{0}$$

Figure 1 Number of Medal Data of Chinese Team from 1984 to 2016

2.2 Prediction

The original data sequence $x^{(0)}$ is

$$x^{(0)}=\left(32,28,54,50,59,63,75,88,70\right). \tag{1}$$

After 1-IAGO, we get

$$x^{(1)}=\left(32,60,114,164,223,286,361,449,519\right). \tag{2}$$

Mean generating sequence is as follows:

$$\begin{aligned} z^{(1)}&=\left(z_2^{(1)},\ z_3^{(1)},\cdots,\ z_9^{(1)}\right)\\ &=\left(46,87,139,193.5,254.5,323.5,405,484\right)\end{aligned} \tag{3}$$

So,

$$B=\begin{bmatrix} -z_2^{(1)} & \left(z_2^{(1)}\right)^2 \\ -z_3^{(1)} & \left(z_3^{(1)}\right)^2 \\ \vdots & \vdots \\ -z_9^{(1)} & \left(z_9^{(1)}\right)^2 \end{bmatrix}, Y=\begin{bmatrix} x_2^{(0)} \\ x_3^{(0)} \\ \vdots \\ x_9^{(0)} \end{bmatrix} \tag{4}$$

Least-squares estimation of the parameter columns $\hat{\alpha}=\left(a,b\right)^T$ is performed

$$\hat{\alpha}=(B^TB)^{-1}B^TY=\begin{bmatrix} -0.448219317 \\ -0.000623125 \end{bmatrix} \tag{5}$$

The Verhulst model is

$$\frac{dx^{(1)}}{dt}-0.448219317x^{(1)}=-0.000623125 \tag{6}$$

Its time response is

$$\hat{x}_{k+1}^{(1)} = \frac{\hat{a}}{\hat{b}x_1^{(0)} + (\hat{a} - \hat{b}x_1^{(0)})e^{\hat{a}k}} = \frac{-0.448219317}{-0.000623125x_1^{(0)} + (-0.448219317 - 0.000623125x_1^{(0)})e^{-0.448219317k}} \tag{7}$$

2.3 Model Accuracy Test

Let k = 0, 1, ..., 8, and the predictive value $\hat{x}^{(1)} = \left(\hat{x}_1^{(1)}, \hat{x}_2^{(1)}, \cdots, \hat{x}_9^{(1)}\right)$ of $x^{(1)}$ is obtained. The last obtained predictive value and error analysis data are shown in Table 1.

Table 1 Raw data, predicted values and Verhulst model errors

Year	Initial data $x^{(0)}$	Predicted value $\hat{x}^{(0)}$	Residual $x^{(0)} - \hat{x}^{(0)}$	Relative error Δ_k
1984	32	32	0	0
1988	28	16.87	11.13	0.40
1992	54	24.81	29.19	0.54
1996	50	35.35	14.65	0.29
2000	59	48.18	10.82	0.18
2004	63	61.83	1.17	0.02
2008	75	73.49	1.51	0.02
2012	88	79.80	8.20	0.09
2016	70	78.57	-8.57	0.12

The average relative error is 18.54%, and model accuracy is level four. We also abtain the absolute correlation degree which is 0.3610.

According to Figure 2,we find that the fluctuation is obvious around 1988 and 1992. It is worth noting that the fluctuation is caused by the political struggle between the former Soviet Union and the United States and other countries according to relevant historical facts. Then, the unsatisfactory accuracy of our model can be understood.

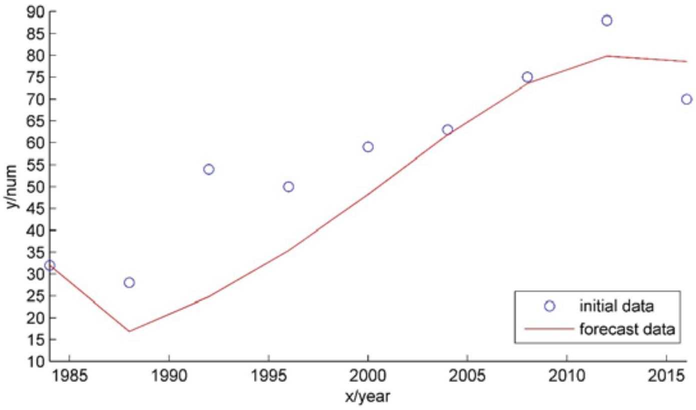

Figure 2 forecast result

3. Analysis on the Influence of National Comprehensive Strength

In this section, we will propose a grey correlation calculation model[5]. We select some the important components of national comprehensive strength ,which are affirmed by international community generally , as indicators. And take the representative of countrys in the 2016 Olympic medals table as our research target. By studying the degree of similarity between the indicators and the number of medals to measure the impact of national comprehensive strength on the forecast results.

3.1 Data Transformation and Processing

We select the top 15 countries in the 2016 Olympic medal table as the object of study, and select 7 factors as GDP, active labour force (15-64 years old) and patents. In order to ensure the quality of modeling and the correct results of system analysis, the collected original data are transformed and processed to eliminate the dimension and be comparability.

- **Nondimensionalization**

$x_i^{(k)}$ is the original data sequence, where the sequence of gold medals is $x_0^{(k)}$ 。 The process of the nondimensionalization of eight sets of data is as follows ：

$$y_i^{(k)} = \frac{x_i^{(k)}}{S_i}, i = 0,1,\cdots,7 \tag{8}$$

where i denotes the i-th index and k denotes the k-th country. $y_i^{(k)}$ stands for the transformed value of the i-th index of the k-th country and S_i stands for the global value of the i-th index. Data processing shown in Table 2.

Table 2 Results after nondimensionalization

Country	Transformed Value							
	GDP	Active Labour Force	Patent Counts	Fiscal Expenditure	Military Expenditure	Generating Capacity	Import and Export	Gold Medals
America	0.2095	0.0432	0.1771	0.3964	0.3567	0.1955	0.1062	0.150
England	0.0328	0.0086	0.0092	0.1325	0.0309	0.0179	0.0313	0.088
China	0.1344	0.2097	0.4338	0.2930	0.1100	0.2635	0.1134	0.085
Russia	0.0273	0.0214	0.0177	0.0271	0.0278	0.0486	0.0212	0.062
Germany	0.0436	0.0113	0.0291	0.1600	0.0205	0.0291	0.0719	0.055
Japan	0.0623	0.0166	0.1672	0.1838	0.0239	0.0537	0.0397	0.039
France	0.0332	0.0087	0.0090	0.1456	0.0254	0.0269	0.0333	0.033
Korea	0.0160	0.0761	0.0984	0.0317	0.0192	0.0233	0.0289	0.029
Italy	0.0253	0.0081	0.0051	0.1003	0.0185	0.0140	0.0264	0.026
Australia	0.0207	0.0031	0.0019	0.0487	0.0165	0.0115	0.0126	0.026
Netherlands	0.0099	0.0023	0.0014	0.0379	0.0076	0.0054	0.0332	0.026
Hungary	0.0016	0.0015	0.0010	0.0064	0.0026	0.0018	0.0057	0.026
Brazil	0.0032	0.0283	0.0031	0.0691	0.0034	0.0227	0.0122	0.023
Spain	0.0017	0.0065	0.0019	0.0569	0.0026	0.0141	0.0179	0.023
Kenya	0.0008	0.0057	0.0017	0.0054	0.0013	0.0003	0.0006	0.020

- **Normalization**

The mapminmax function of Matlab is used to transform given sequences so that all sequences have a common intersection.

3.2 Grey Correlation Calculation Model

We select the number of medals $y_0^{(k)}$ as the reference sequence, and there are seven comparative sequences

$$y_i^{(k)} = \left(y_i^{(1)}, y_i^{(2)}, \cdots, y_i^{(15)}\right), i = 1,2,\cdots,7 .$$

The correlation coefficient of the sequence $y_i^{(k)}$ to the series $y_0^{(k)}$ is

$$\xi_i(k) = \frac{\min_s \min_t |y_0^t - y_s^t| + \rho \max_s \max_t |y_0^t - y_s^t|}{|y_0^k - y_i^k| + \rho \max_s \max_t |y_0^t - y_s^t|} \tag{9}$$

where $\rho \in [0,1]$ is the resolution coefficient. Also, $\min_s \min_t |y_0^t - y_s^t|$ and $\max_s \max_t |y_0^t - y_s^t|$ represent respectively.

3.3Grey Correlation Analysis

The results of the correlation calculation are shown inTable 3.

Table 3 The result of the correlation calculation

r_1	r_2	r_3	r_4	r_5	r_6	r_7
0.6866	0.4322	0.1034	0.2871	0.7869	0.6352	0.4059

The results show that the military strength, GDP and power generation of the national comprehensive strength are related to the Olympic Games results. Simultaneously, ,that is, The country's overall strength affects the forecasts mainly through military spending, GDP and power generation, and is dominated by military spending. This corresponds to the association shown in Figure 3.

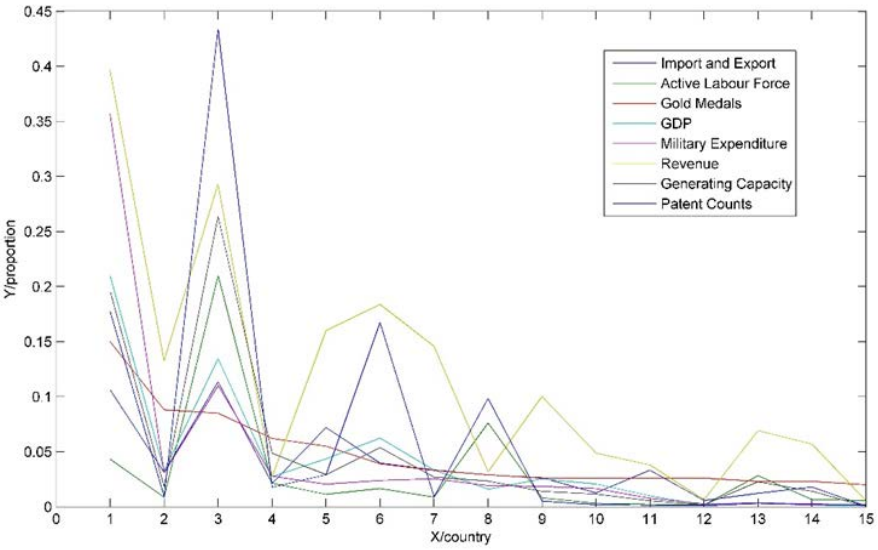

Figure 3 Data of comprehensive national strength and medals

4. Summary

This model is used to predict the results of the Olympic Games, an event with much randomicity. By comparing predictions and actual results of the 30th Olympic Games,we found that the credibility of the model is not high, but for such events it can be referenced.

According to the historical results which have no obvious regularity, the short-term Gray prediction is applicable to it. However, the error of once prediction is not small, so in the second and third questions, the differences of national conditions are taken into account, also the results of once prediction are taken as a factor influencing the final prediction results. The prediction model is given by multiple linear regression, By considering the host country's influence,we give the final prediction at the end.

References

[1] Johnson D K N, Ali A. A Tale of Two Seasons: Participation and Medal Counts at the Summer and Winter Olympic Games *[J]. Social Science Quarterly, 2002, 85(4):974–993.
[2] Bernard A B, Busse M R. Who Wins the Olympic Games: Economic Resources and Medal Totals[J]. Review of Economics & Statistics, 2004, 86(1):413-417.
[3] Matros A, Namoro S D. Economic Incentives of the Olympic Games[J]. Ssrn Electronic Journal, 2004.
[4] Wang Z X, Dang Y G, Liu S F. Unbiased Grey Verhulst Model and Its Application[J]. Systems Engineering-Theory & Practice, 2009, 29(10):138-144.
[5] Zhang F H. Grey Correlation Degree Analysis of Main Agronomic Character of Soybean Varieties in Henan Province[J]. Journal of Anhui Agricultural Sciences, 2006, 34(19):4842-4843.