A Novel prediction model for champions' scores of men's 110 - meter hurdle in Olympic Games

Xiaoxin Li

School of Physical Education, China University of Mining and Technology, Xuzhou, Jiangsu Province, 221116

Corresponding author's e-mail: jsxzkd62@163.com

Received 01 October 2013, www.cmnt.lv

Abstract

In order to improve the prediction accuracy of the grey model for champions' scores of men's 110-meter hurdle in Olympic Games, a nonlinear grey Bernoulli model (NGBM) has been built on the base of the GM(1, 1) model, and a genetic algorithm (GA) has been adopted to optimize the parameters of the model. Based on the statistics of champions' scores of men's 110-meter hurdle in Olympic Games during the 1948 - 2012 period, the NGBMGA model is employed to predict the performance of the 2016 and 2020 Olympic Games which is contrasted against the prediction result of the GM (1, 1) model. The results show that the NGBMGA model has higher prediction accuracy, with its feasibility and veracity verified.

Keywords: GM(1,1) model, nonlinear grey Bernoulli model, predictive analysis, 110-meter hurdle

1 Introduction

The Olympic event of men's 110-meter hurdle has received much focus from the countrymen for many years, especially the impressive performance Liu Xiang has achieved in men's 110-meter hurdle, which not only demonstrated his extraordinary skill but manifested the countrymen's dramatic breakthrough in men's short-distance track events. Actually, many prediction techniques and methods have been applied in competitive sports [1]. Since in Olympic races, the champion must be created under the optimal condition to which competitors adjust their own status, there are fewer factors influencing champions' scores of men's 110-meter hurdle in Olympic Games. Thus, it can be seen the key to prediction is building a scientific and rational prediction model.

The predictive analysis of champions' scores of men's 110-meter hurdle event in Olympic Games belongs to an issue of time sequence. Currently, in dealing with the prediction problem of time sequence, the grey prediction model GM (1, 1) is known as a well applied model [2]. However, the traditional grey model GM (1, 1) is constructed on exponential function, while the development trend of champions' scores does not necessarily follow the rule of the grey model's monotonous change [3-5] but tend to fluctuate, therefore it is required to conduct pointed modification over the grey model to improve its prediction accuracy. An effective modified model is nonlinear grey Bernoulli model (NGBM). The principle of which is to introduce the Bernoulli equation among differential equations on the base of the GM(1, 1) model and to combine both to produce a new prediction model [6]. Which enjoys the advantages of simple inferring process of the GM(1, 1) model and smaller sample data size; moreover, based on an appropriately selected power exponent, the solution of equation after conversion can make the generated sequence via one-time accumulation become optimized by fitting.

In this paper, the NGBMGA(1, 1) prediction model is built. A prediction analysis is conducted with champions' scores of men's 110-meter hurdle in the Olympic Games from 1948 to 2012 with the object. Then the predictive abilities of NGBMGA(1, 1) model and GM(1, 1) model are contrasted with percentage error and mean absolute percentage error as evaluation criteria to screen out the prediction model with higher accuracy.

2 Sample statistics

With statistics of champions' scores of men's short-distance track event in previous Olympic Games, the data are selected since 1948 (See Table 1), all originating from the official website of International Olympic Committee (http://www.olympic.org/) and reflecting the development level and tendency of performance of 110-meter hurdle in Olympic Games since World War II.

TABLE 1 Champions' Scores of Men's 110-meter Hurdle in Olympic Games (s)

Year	**1948**	**1952**	**1956**	**1960**	**1964**	**1968**	**1972**	**1976**	**1980**
Score	13.9	13.91	13.7	13.98	13.67	13.33	13.24	13.3	13.39
Year	**1984**	**1988**	**1992**	**1996**	**2000**	**2004**	**2008**	**2012**	
Score	13.2	12.98	13.12	12.95	13	12.91	12.93	12.92	

3 Nonlinear Grey Bernoulli Model

NGBM follows the following steps:

Step 1. Let $X^{(0)}$ denote the non-negative data of original time series:

$$X^{(0)} = \left\{ x^{(0)}(t_i) \right\}, \quad i = 1, 2, \cdots, n. \tag{1}$$

where, $x^{(0)}(t_i)$ denotes the system's numeric value at time t_i; n denotes the size of data sample.

Step 2. Establishing a successively accumulated AGO sequence $X^{(1)}$, namely:

$$X^{(1)} = \left\{x^{(1)}(t_i)\right\}, \quad i = 1, 2, ..., n. \tag{2}$$

where:

$$x^{(1)}(t_k) = \sum_{i=1}^{k} x^{(0)}(t_i), \quad i = 1, 2, ..., n. \tag{3}$$

Step 3. $X^{(1)}$ is a monotonously increasing sequence, so the Bernoulli differential equation can be introduced to solve:

$$dX^{(1)}/dt + aX^{(1)} = b[X^{(1)}]^r, \tag{4}$$

where: a is development coefficient, b is grey coefficient, and r is a random numeric value; when $r = 1$ the model becomes a grey model GM(1, 1).

Step 4. In order to determine parameters a and b, Equation (4) can be approximately expressed as

$$\Delta X^{(1)}(t_k)/\Delta t_k + aX^{(1)}(t_k) = b[X^{(1)}(t_k)]^r, \tag{5}$$

where:

$$\Delta X^{(1)}(t_k) = x^{(1)}(t_k) - x^{(1)}(t_{k-1}) = x^{(0)}(t_k), \tag{6}$$

$$\Delta t_k = t_k - t_{k-1} \tag{7}$$

If the interval of the sample is selected at 1, namely.

Adopt:

$$z^{(1)}(t_k) = px^{(1)}(t_k) + (1-p)x^{(1)}(t_{k-1}), \; k = 2, 3, ..., n, \tag{8}$$

to replace the term $X^{(1)}(t_k)$ in Equation (5), and we get:

$$x^{(0)}(t_k) + az^{(1)}(t_k) = b[z^{(1)}(t_k)]^r, \quad k = 2, 3, ..., n, \tag{9}$$

In Equation (8), p is generating coefficient of background value, and $p \in (0,1)$; when $p = 0.5$ the model becomes NGBM(1, 1).

Step 5. In Equation (9), the values of a and b can be determined using the least square method, namely:

$$[a,b]^T = \left(B^T B\right)^{-1} B^T Y_N, \tag{10}$$

where:

$$B = \begin{bmatrix} -z^{(1)}(t_2) & z^{(1)}(t_2)^r \\ -z^{(1)}(t_3) & z^{(1)}(t_3)^r \\ \vdots & \vdots \\ -z^{(1)}(t_n) & z^{(1)}(t_n)^r \end{bmatrix}, \tag{11}$$

and

$$Y_N = \left[x^{(0)}(t_2), x^{(0)}(t_3), \cdots, x^{(0)}(t_n)\right]^T. \tag{12}$$

Step 6. Equation (4) can be rewritten into:

$$\hat{x}^{(1)}(t_k) = \left[\left(x^{(0)}(t_1)^{(1-r)} - b/a\right)\mathrm{e}^{-a(1-r)(t_k - t_1)} + b/a\right]^{1/(1-r)} \\ r \neq 1, \; k = 1, 2, \cdots, n. \tag{13}$$

Step 7. Using a reversely accumulated sequence to define $\hat{x}^{(1)}(t_k)$, the prediction sequence $x^{(0)}(t_k)$ can be expressed as:

$$\hat{x}^{(0)}(t_1) = x^{(0)}(t_1), \tag{14}$$

$$\hat{x}^{(0)}(t_k) = \hat{x}^{(1)}(t_k) - \hat{x}^{(1)}(t_{k-1}), \quad k = 2, 3, ..., n. \tag{15}$$

4. Parameter optimization of NGBM

4.1 PARAMETER OPTIMIZATION MODEL.

A reference to Equation (13) informs that when NGBM is adopted for prediction, the decisive condition affecting the model's prediction accuracy is the value range of parameters a, b and r. Among them, when $r = 0$, the NGBM model becomes the simple GM(1, 1) model; when $r = 2$, it becomes the Grey-Verhulst model. Adjustment of parameter r grants the NGBM model with better flexibility than the GM(1, 1) model and the Grey-Verhulst model so as to make a better imitative effect between the predicted value and the actual value.

Through Equation (10), we can discover that parameters a and b are related with the data of original time sequence $X^{(0)} = \left\{x^{(0)}(t_1), x^{(0)}(t_2), \cdots, x^{(0)}(t_n)\right\}$ and generating coefficient p of background value. $X^{(0)}$ is a set of historical data, therefore p is a control parameter. Zhuang [7] has substantiated the following relation existing between parameter p and the development coefficient a in GM(1,1):

$$p = \frac{1}{a} - \frac{1}{e^a - 1} \tag{16}$$

Induction of Equation (16) reveals that as $a \to 0$, the value of p approximates 0.5. It is not difficult to find that it is wrong to set the generating coefficient p of background value at 0.5 for a very large development coefficient a, therefore Chang [8] points out the prediction accuracy can be improved by using modified and optimized parameter p.

Since the minimum mean absolute percentage error (MAPE) demonstrates the model's higher predictive ability[9], the research target of this paper is to minimize the model's MAPE value by searching for the optimal value of parameters p and r, namely to reach the highest prediction accuracy.

At first, this paper needs to optimize parameters p and r and modifies the numeric value of the prediction result $\hat{x}^{(0)}(t_k)$ further, so as to minimize the MAPE value of the prediction result and the actual result. The following objective function is created as a detection factor of the prediction model.

$$\min MAPE = \frac{1}{n-1}\sum_{k=2}^{n}\left|PE(t_k)\right| = \frac{1}{n-1}\sum_{k=2}^{n}\left|\frac{\hat{x}^{(0)}(t_k)-x^{(0)}(t_k)}{x^{(0)}(t_k)}\times 100\%\right|,\ k=1,\cdots,n \tag{17}$$

where: n is the size of data, $x^{(0)}(t_k)$ and $\hat{x}^{(0)}(t_k)$ are the actual value and the model's predicted value, respectively, at time t_k.

The parameter optimization of NGBM model is concerned with global optimization of p and r, its target being appropriate value of both parameters. Genetic algorithm [10, 11] has high ability in global optimization and fast computation speed as compared with other optimization algorithms, therefore GA is adopted for parameter optimization of the NGBM model in this paper.

4.2 GA-BASED NGBM MODEL

GA is a stochastic optimization and searching method evolving from the evolutionary law of the living nature (namely the genetic mechanism known as survival of the fittest) and first put forward by Prof Holland in 1975. This theory is characterized by the following:

1) Straightforward operation on the target without limit on derivation and continuity of function;

2) Advantage in steady internal parallel and strong global optimization capability;

3) Using randomization theory while seeking for optimization to voluntarily get and lend evidence to optimizing the selected interval and voluntarily change the direction of selection without fixed limit by rule.

NGBMGA(1, 1) is computed by:

$$\hat{x}_{GA}^{(1)}(t_k) = \left[\left(x^{(0)}(t_1)^{(1-r)} - \tilde{b}/\tilde{a}\right)e^{-a(1-r)(t_k-t_1)} + \tilde{b}/\tilde{a}\right]^{1/(1-r)}$$
$$r \neq 1,\ k=1,2,\cdots,n, \tag{18}$$

where:

$$\hat{x}_{GA}^{(0)}(t_1) = x_{GA}^{(0)}(t_1), \tag{19}$$

$$\hat{x}_{GA}^{(0)}(t_k) = \hat{x}_{GA}^{(1)}(t_k) - \hat{x}_{GA}^{(1)}(t_{k-1}),\quad k=2,3,...,n. \tag{20}$$

5. Result & Analysis

By applying the historical scores of 110-meter hurdle in Olympic Games and utilizing the NGBMGA(1, 1) model built in the above text, the score of the next session can be predicted and analyzed. In the NGBMGA(1, 1) model, the parameters p and r need optimizing. In the solving process of the model, the crossing probability of GA is 0.6, the mutation probability is 0.001, and the population quantity is 1000, according to literature [12]. The solving method is: recipe, the permitted iterations being 5000. Apply the NGBMGA (1, 1) model to get p=0.5123, r=0.2107, with the result of prediction shown in Table 2.

TABLE 2 Prediction Result of Men's 110-meter Hurdle

Time of Final	Score of Champion (s)	Binomial Fitting		GM(1,1)	
		Predicted Value	PE(%)	Predicted Value	PE(%)
1948	13.90	13.90	0	13.90	0
1952	13.91	13.81	-0.71	13.89	-0.13
1956	13.70	13.74	0.286	13.79	0.65
1960	13.98	13.67	-2.24	13.72	-1.88
1964	13.67	13.60	-0.54	13.68	0.043
1968	13.33	13.52	1.462	13.54	1.612
1972	13.24	13.45	1.618	13.46	1.694
1976	13.30	13.38	0.631	13.39	0.706
1980	13.39	13.31	-0.57	13.36	-0.19
1984	13.2	13.24	0.337	13.27	0.564
1988	12.98	13.18	1.504	13.23	1.889
1992	13.12	13.11	-0.10	13.17	0.354
1996	12.95	13.04	0.679	13.07	0.911
2000	13.00	12.97	-0.23	13.03	0.217
MAPE(1-14)			0.839		0.835
2004	12.91	12.90	-0.06	12.98	0.558
2008	12.93	12.83	-0.74	12.95	0.191
2012	12.92	12.77	-1.18	12.93	0.059
MAPE(15-17)			0.659		0.269
2016		12.70		12.88	
2020		12.64		12.84	

Since percentage error (PE) and mean absolute percentage error (MAPE) demonstrate good applicability in terms of prediction effect, this paper uses PE and MAPE to describe the result of prediction. PE is calculated by the Equation:

$$PE(t_k) = \frac{\hat{x}^{(0)}(t_k) - x^{(0)}(t_k)}{x^{(0)}(t_k)}\times 100\%, k=1,...,n \tag{21}$$

where n is the size of data, $x^{(0)}(t_k)$ and $\hat{x}^{(0)}(t_k)$ are the actual value and the model's predicted value, respectively, at time t_k.

Using the statistics of historical data of champions' scores of men's short-distance events in Olympic Games, this paper has built the NGBMGA(1, 1) model, and compares the predictive abilities of NGBMGA(1, 1) model and GM(1, 1) model with percentage error and mean absolute percentage error as the evaluation criteria. It has been concluded that: The MAPE values of GM(1, 1) model and NGBMGA(1, 1) model are 0.839 and 0.835, respectively in terms of prediction before 2000. The prediction accuracy of NGBM model has improved slightly, but the MAPE values are 0.659 and 0.269, respectively, in terms of prediction over the recent 3 sessions of Olympic Games, namely from 2004 to 2012. And the prediction accuracy of NGBM model has improved by 0.39, which informs the accuracy of NGBMGA(1, 1) model has improved significantly and indicates that the NGBMGA(1, 1) model has outstanding adaptability in predicting champion's score of men's 110-meter hurdle event in Olympic Game.

In the predicting champions' scores of men's 110-meter hurdle in 2016 and 2020, the prediction results of NGBM model are 12.88s and 12.84, respectively. As NGBM has higher prediction accuracy, the author thinks this score is more convincing, with which to lay certain theoretical goals and basis for coaches and athletes.

Acknowledgements

I would like to thank Su Ji for her kindly help in the data analysis.

References

[1] Derevenco P, Albu M, Duma E 2002-2003 Forecasting of top athletic performance *Romanian journal of physiology* 39-40 57-62
[2] Zhang Z 2011 Score of Men's 100-meter Running in Olympic Games Predicted by grey GM(1, 1) Model *Journal of Physical Education* **18**(4) 111-4
[3] Fan Z 2006 Grey System Model Analysis in the Field of Athletic Training Control *Journal of Physical Education* **13**(6) 104-6
[4] Fan W, Zhou A, Liu J 2002 Grey Prediction on Score of Track and Field Sports in the 28th Olympic Games *Journal of Beijing Sport University* **25**(6) 847-9
[5] Jin C, Tong L 2002 Equal-Dimension Grey Number Fill GM(1, 1) Model and Prediction of National Records of Men's Long Jump *Journal of Physical Education* **9**(5) 134-5
[6] Chen C I, Chen H L, Chen S P 2008 Forecasting of foreign exchange rates of Taiwan's major trading partners by novel nonlinear Grey Bernoulli model NGBM(1,1) *Communications in Nonlinear Science and Numerical Simulation* **13**(16) 1194-204
[7] Zhuang H Y 1993 Mechanism and application of GM(1,1) and improvement methods *Syst Eng Theor Methodol Appl* **2**(3) 56–62
[8] Chang S C, Lai H C, Yu H C 2005 A variable P value rolling grey forecasting model for Taiwan semiconductor industry production *Technol. Forecast Social Change* **72**(5) 623-40
[9] Liu S, Dang Y 2004 Grey Theory and Application *Beijing: Science Press*
[10] Boeringer DW, Werner DH 20004 Particle swarm optimization versus genetic algorithms for phased array synthesis *IEEE Transactions on Antennas and Propagation* 52 771-9
[11] Deb K 2000 An efficient constraint handling method for genetic algorithms *Computer Methods in Applied Mechanics and Engineering* 186 311-38
[12] Back T 1992 The interaction of mutation rate, selection, and self-adaptation within a genetic algorithm in: *Proceedings of the Second Conference on Parallel Problem Solving for Nature*

Author

Xiaoxin Li, 03.12.1962, Jangsu province of China.

Current position, grades: associate professor.
University studies: graduated from Shanghai University of Sport in 1985.
Scientific interest: physical education and training.
Publications: more than 20 papers, 2 textbooks.