

Research on the Construction and Correction of Sports Performance Prediction Model

Fan Zhang

Abstract—This work presents a prediction model of sport performance based on the error correction of the extreme learning machine. Firstly, grey model is used to predict sports performance. Secondly, the prediction error of grey model is corrected by extreme learning machine. Finally, the performance of the model is tested using actual sports results.

Index Terms—Sports results, prediction model, extreme learning machine.

I. INTRODUCTION

With the continuous development of sports, the Chinese mainland government has paid more and more attention to the sports education. The scientific prediction of sport performance can be used to understand the physical function of students or athletes, and provide the basis for scientific training and accurate quantitative evaluation of physical function. The research of sports performance prediction model has important application value in sports education, humanity science and health evaluation.

II. TIME SERIES ANALYSIS OF SPORTS PERFORMANCE

A. Sports Results Data Sequence Reconstruction

Submit your manuscript electronically for review. Sports performance is a set of nonlinear time series, which can be analyzed and predicted by nonlinear time series analysis method. Through the preliminary statistics and sampling of sports results, we can construct a single variable time series of sports performance, set as $\{x_n\}$, usually, the time series of sports performances are scalar time series, phase space reconstruction method is adopted to develop high dimensional vector of sports performance [1]. The method of reconstructing the phase space of sports performance is described as: select the minimum embedding dimension m , optimal time delay τ ; a differential equation is constructed to express the information flow model of sports performance.

Manuscript received May 24, 2017; revised September 15, 2017. This work was supported in part by the Project of Nanjing Forest Police College pre-research under Grant LGY201603, in part by the Project of the Fundamental Research Funds for the Central Universities under Grant LGZD201709, in part by the Project of China Postdoctoral Science Foundation under Grant 2017M611849, in part by Jiangsu Qing Lan Project under Grant 2017, and in part by Nanjing Forest police College Teaching Reform Project under Grant ZD17101.

Fan Zhang is with the Police Skills and Tactics Department, College of Nanjing Forest Police, Nanjing, Jiangsu, China. He is now with the Sports Science Postdoctoral Research Station, Nanjing Normal University, Nanjing, Jiangsu, China (e-mail: zhangfan@nfpcc.edu.cn).

$$x_n = x(t_0 + n\Delta t) = h[z(t_0 + n\Delta t)] + w_n \quad (1)$$

Formula (1): $h(\cdot)$ is a multiple number of sports performance values of the time series function; w_n is the observation or measurement error of the time series of sports performance. Through the construction of information flow model of sports performance, the mapping relation between sports performance information flow model and nonlinear feature extraction is established, and feature extraction and prediction model design are carried out.

B. Design the Predictive Model Flow of Sports Performance

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables. Nonlinear time series analysis of sports performance in reconstructed phase space, extracting effective principal component features of sports performance time series, and grey model is used to predict sports performance. The overall procedure of sports result prediction designed in this paper is shown in the following steps. 1) Prior sports performance data; 2) Phase space reconstruction extract information flow model; 3) Extracting effective principal component features; 4) Grey model learning; 5) Extreme learning machine correct error; 6) Output test results.

The nonlinear time series analysis and feature extraction are carried out according to the prediction process of sports performance shown in Fig. 1. It is assumed that sports performance are generated by linear correlation of nonlinear time series and represented by the ARMA model.

$$x_n = a_0 + \sum_{i=1}^{M_{AR}} a_i x_{n-i} + \sum_{j=0}^{M_{MA}} b_j h_{n-j} \quad (2)$$

Formula (2): a_0 is the sampling amplitude of initial sports performance; x_{n-i} is the scalar time series of sports performance with the same mean and variance; b_j is the oscillatory amplitude of sports performance.

Carries on the Fourier transformation to the sports performance data information sequence, get $x(k)$, under the training of grey model, the degeneration tendency of sports performance is obtained:

$$A(t) = mt + \sqrt{am} B^H(t) \quad (3)$$

Formula (3): a is the inter-domain coefficient of variance,

and $B^H(t)$ represents correlation function.

In the time domain of grey information exchange of data characteristics, sports results sequences have strong nonlinear characteristics. In this paper, nonlinear time series analysis is used to predict and analyze sports performance sequences and scalar time series, it is necessary to construct time series signal model of sports performance. In the presence of multiple known disturbances, in order to make the discrete data of sports performance discrete and analytic, an alternative data method is used to handle the oscillatory amplitude randomization of sports performance [2]. The analytic model of the grey model in sports result prediction is as below.

$$z(t) = x(t) + i y(t) = a(t) e^{i\theta(t)} + n(t) \quad (4)$$

Formula (4): $x(t)$ is the real part of the time series of sports performances; $y(t)$ is the imaginary part of the time series of sports performances; $a(t)$ is the phase randomization amplitude; and $n(t)$ is the disturbance vector.

Through the nonlinear time series analysis of sports performance, the principal component feature extraction of sports performance time series is carried out, and the step is described as follow.

- 1) Carries on the Fourier transformation to the sports performance time sequence, get $x(k)$;
- 2) The D dimensional compact manifold vector field of sports performance was measured, and the sport scores were randomly processed by means of surrogate data, get $x'(k)$;
- 3) The nonlinear features of sports performance are tested by using the autocorrelation function feature matching method to generate surrogate data. Inverse transformation of Fourier for Substitution of data for sequences of sports performances, get $x'(n)$. The generated sports performance substitution data preserves the linear autocorrelation function of the original sports results data, and extracts the principal component features of the data stream of sports results data as below:

$$C_{or3} = \langle (x_n - \bar{x})(x_{n-d} - \bar{x})(x_{n-D} - \bar{x}) \rangle / \langle (x_n - \bar{x})^3 \rangle \quad (5)$$

Formula (5): x_n refers to the nonlinear time series of athletic performance; d refers to the time interval in which students take their sports performance, $D = 2d$; \bar{x} refers to the mean; $\langle x(n) \rangle$ refers to the mean for $x(n)$:

$$\langle x(n) \rangle = 1/N \sum_{n=1}^N x(n) \quad (6)$$

III. MATH MODELING AND PREDICTION OF SPORTS PERFORMANCE

A. Implementation of Grey Model Training Prediction

Through the analysis of nonlinear time series of sports performance, the prior knowledge and principal component features of sports performances are extracted. Vector time series of sports performance track in high dimension phase space is $\{x(t_0 + i\Delta t)\}$, $i = 0, 1, 2, \dots, N-1$, grey model is used to predict sports performance, the basic idea is to measure the time series time-varying of sports performance, and to measure the stochastic nature of linear correlations between x and $x_{n+\tau}$, the grey model training function $C(\tau)$ is defined as:

$$C(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)x(t+\tau) dt \quad (7)$$

Formula (7): τ is the time delay of sports performance in reconstructed phase space, and the correlation between sports performance changes that represent t and $t+\tau$ time.

According to the correlation degree, the sports performance trend prediction is carried out, and the sampling interval of the fixed sports performance is i , and the curve of time τ ($\tau = 1, 2, \dots$) of the training model of grey model is constructed. As shown in Fig. 1, the optimal sampling interval is obtained according to the intersection of the time delay curve, and the rationality of the sports result prediction is determined.

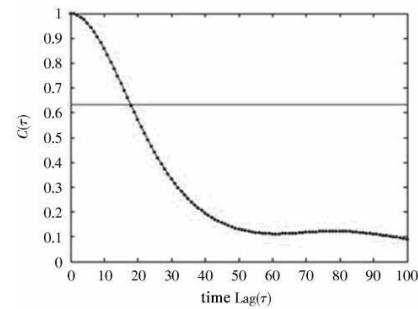


Fig. 1. The curve of time τ of grey model training function.

On the basis of the optimal sampling interval τ , the minimum embedding dimension of grey model training m is calculated, by averaging mutual information method, the grey model training embedding dimension of sport achievement sequence is calculated. The mutual information of the sports performance time series in vector space is:

$$I(\tau) = -\sum_{ij} P_{ij}(\tau) \ln \left[\frac{P_{ij}(\tau)}{P_i P_j} \right] \quad (8)$$

Arbitrarily point of reconstructing the time series phase space of sports performance is X_n , its nearest neighbor point in the high dimensional grey model of sports performance is expressed as $X_{\eta(n)}$, the training function of m gray model is set up as below.

$$X(n) = \{x(n), x(n+\tau), \dots, x(n+(m-1)\tau)\} \quad (9)$$

$n = 0, 1, 2, \dots, N$, in the grey model, the covariance matrix C of the historical data of sports performance is calculated as below:

$$C = \frac{1}{N} [X - \bar{X}_l] [X - \bar{X}_l]^T \quad (10)$$

Among formula (10): $l = [1, 1, \dots, 1]_{1 \times N}$, $\bar{X}_l = \frac{1}{N} \sum_{k=1}^N X_k$, $X = [X_1, X_2, \dots, X_m]$

Constructing characteristic equation of sports result prediction:

$$(\lambda I - S)U = 0 \quad (11)$$

The characteristic value λ of S is calculated, and then the prediction of sports performance based on grey model is realized.

IV. EXTREME LEARNING MACHINE CORRECT PREDICTION ERROR

K sample data set of sports performance of input extreme learning machine [3] $\{x_i, y_i\}$, $i = 1, 2, \dots, k$, k represents the number of samples of time series for sports performance, the data is normalized, load the historical data $\{x_i\}_{i=1}^N$ of the sports performance test through the extreme learning machine into the high dimension phase space S . In high dimension phase space, construct autoregressive process of sports performance of nonlinear time series [4], expression as below:

$$f(x) = \omega^T (\phi)_x + b \quad (12)$$

Formula (12): ω is the error vector matrix, and b is the deviation vector.

Historical data of sports performance were chosen as predictive models to process features dimensionality reduction [5]. Adaptive weighting is used; giving extreme learning machine correction error coefficient ξ , the discriminant function obtained by calculation is below:

$$L_\xi = \begin{cases} f(x) - y, & |f(x) - y| \geq \xi \\ 0, & |f(x) - y| < \xi \end{cases} \quad (13)$$

According to the contribution degree of cumulative variance, the error correction and feature dimension reduction are adopted to reduce the complexity of the model, and the minimum error model of sports result prediction is obtained as below:

$$\min_{\omega, b, \zeta_i, \zeta_i^*} = \frac{1}{2} \omega^T \omega + c \sum_{i=1}^l (\zeta_i + \zeta_i^*) \quad (14)$$

Formula (14): ζ_i and ζ_i^* represent redundancy and principal component features respectively. c represents the cost factor of error correction for extreme learning machines. The greater the c value, the better the prediction accuracy [6]. Finally, the output time series estimation function is obtained as below:

$$f(x) = \sum_{i=1}^l (a_i + a_i^*) k(x - x_i) + b \quad (15)$$

The convergence of the prediction error is:

$$d_m(0) = \|X_m - X_k\| \quad (16)$$

According to the principle of Lyapunov convergence [7], [8], the prediction error will converge to 0, and the time series prediction of sports performance growth index is obtained as below:

$$x'(t_{n+1}) = X_{m+1}(m) \quad (17)$$

Analysis shows that the above process can achieve zero error prediction of sports performance, and the accuracy is better.

V. EXPERIMENT AND RESULT ANALYSIS

The test sample of sports performance comes from 6 sports, 1.5 km, 100 m, 4 x 10 m run back, standing jump, push-ups, sit-ups and so on, from the 2016 grade student of Nanjing Forest Police College. The sampling time sequence of prior data information is shown in Fig. 2.

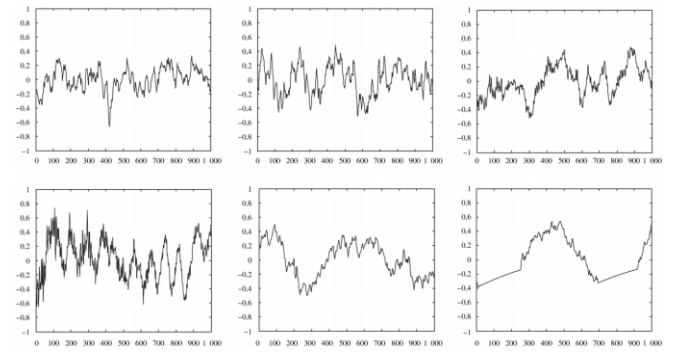


Fig. 2. Prior information sampling time series of sports achievement.

With the above sampled sports performance as the test set, the predicted output is shown in Fig. 3.

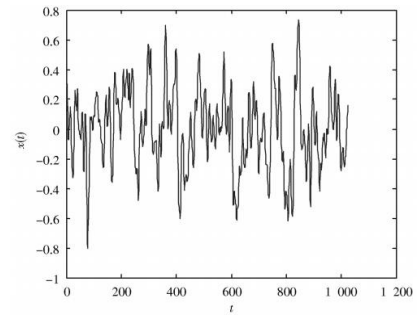


Fig. 3. Sports result prediction.

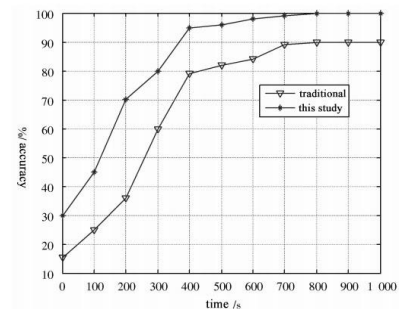


Fig. 4. Sports result prediction.

As can be seen from Fig. 3, using this model to predict sports performance can effectively track the characteristics of sports performance, and have better performance in dynamic tracking of sports performance changes, and show better prediction ability. In order to compare the model performance, this model and the traditional model are used to predict the sports performance, and the results of the accuracy contrast are shown in fig. 4. The model of this study has higher prediction accuracy and better performance.

VI. CONCLUSION

This paper studies the accurate prediction model of sports performance, the data flow model of sports performance prediction and nonlinear time series analysis of sports performance were carried out, extracting effective principal component features of sports achievement time series. Through the grey model learning, the prediction of the sports performance is carried out, and the prediction error of the grey model is modified by the extreme learning machine, so that the prediction error converges to zero quickly.

The algorithm analysis and experimental results show that the model has higher accuracy in sports result prediction, and the error is lower than the traditional method, and has good application value.

REFERENCES

- [1] Y. Z. Ma and X. F. Meng, "Research on indexing for cloud data management," *Journal of Software*, vol. 26, pp. 145-166, January 2015.
- [2] L. Yang, G. P. Li, and P. Zhang, "Improved wolf step prediction of network anomaly flow detection method," *Bulletin of Science and Technology*, vol. 30, pp. 47-49, February 2014.
- [3] Y. C. Chang *et al.*, "Cluster based self-organization management protocols for wireless sensor networks," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 1, pp. 75-80, 2006.
- [4] J. Liu, Y. Liu, Y. He *et al.*, "Joint probabilistic data association algorithm based on all-neighbor fuzzy clustering in clutter," *Journal of Electronics & Information Technology*, vol. 38, pp. 1438-1445, June 2016.
- [5] X. H. Lu and P. H. Chen, "Traffic prediction algorithm in buffer based on recurrence quantification union entropy feature reconstruction," *Computer Science*, vol. 42, pp. 68-71, April 2015.
- [6] J. Choi *et al.*, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 295-309, 2011.
- [7] J. P. Kolhe *et al.*, "Robust control of robot manipulators based on uncertainty and disturbance estimation," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 1, pp. 104-122, 2013.
- [8] Y. T. Liu, "K-pruning algorithm for semantic relevancy calculating model of natural language," *Journal of Theoretical and Applied Information Technology*, vol. 48, on. 3, pp. 231-235, 2013.



China.

Fan Zhang was born in Jiangsu province, China. He received the Ph.D. degree in sports science. In July 2016, he entered the post-doctoral research station of sports science in China Nanjing Normal University. He is currently a lecturer of College of Nanjing Forest Police, China. His research interests include sports training and health promotion. HE has published several articles and books. He has hosted or attended several fund projects of