Journal of Human Movement Science (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jhms.2024.050112
ISSN 2523-5850 Vol. 5 Num. 1

Research on Athlete Momentum Based on GA-BP Neural Network

Mengyang Li[1,a,*,#], Siyuan Peng[2,b,#], Peijia Li[1,c]

[1]*College of Arts and Sciences, Northeast Agricultural University, Harbin, China*
[2]*College of Economics and Management, Northeast Agricultural University, Harbin, China*
[a]*18846190672@163.com*, [b]*pengsiyuan_2022@163.com*, [c]*L18846455785@163.com*
[*]*Corresponding author*
[#]*Co-first author*

Keywords: Athlete Momentum, GA-BP, Neural Network, Non-linear regression score model, tactical strategies

Abstract: In the 2023 Wimbledon men's singles final, frequent changes in momentum have become one of the focus of attention. This paper aims to explore the performance, momentum advantage and momentum changes in athletes. First, the data were characterized by dimension reduction and feature extraction. Then, nonlinear regression was used to construct a time point-player Performance score evaluation model, in which the performance score of each player at each time point in each game was calculated according to the characteristics. Then the GA-BP neural network prediction model was constructed based on the athlete scores, and we found that the model had good results for predicting the inflection point in the game, with an accuracy of 88.64%. According to the contribution rate of the characteristic indicators to the fluctuation, give the suggestions for different players in the new game, covering the technical level, physical factors and tactical strategies.

1. Introduction

In the men's singles final of Wimbledon in 2023, young Spanish player Carlos Alcaraz had a fierce confrontation with experienced Novak Djokovic. This game not only ended Djokovic's rule at Wimbledon, but also witnessed the rise of new stars. What is impressive in the game is the fluidity of the score and the frequent change of momentum. The audience was shocked by the superb skills and excellent tactical performance in this game. It's hard to predict the winner of the game when Alcatraz and Djokovic score alternately. This dynamic game has always been the dream of tennis fans. Alcatraz's calm and outstanding performance showed the championship potential, not only played a subtle winning goal in the attack, but also remained calm and calm at the crucial moment. As a legendary player, Djokovic also showed indomitable spirit, but Djokovic's indomitable spirit eventually lost to young vitality. These questions have aroused the interest in the study of players' performance, momentum advantage and momentum change in the competition.

2. Related Works

In recent years, the applications of artificial intelligence and machine learning have infiltrated the field of sports, providing new perspectives and tools for event outcome prediction and athlete performance analysis. Min et al. [1] developed a composite framework to improve the prediction accuracy of football match outcomes by combining different prediction models. This approach allows to exploit the advantages of multiple models and reduce the dependence of a single prediction approach. Next, Baulch[2] demonstrated how to use machine learning techniques to predict sports competition outcomes, providing a practical and systematic methodological framework for the analysis of sports competitions. While McCabe and Trevathan[3] specifically explore the use of AI in sports prediction, showing how AI helps improve the prediction of event outcomes. Smith et al. [4] predicted the Cy Young award winners of baseball events using data mining techniques, whose work highlights the value of statistics in sports analysis. Similarly, through visual data analysis, Luhr and Lazarescu[5] proposed a tool to evaluate, compare and detect changes in athlete performance during the competition, which has direct implications for sports science and athlete training. In basketball, Atlas and Zhang [6] use fuzzy neural network agents for online NBA scout ting, showing how intelligent systems can help team management assess player performance. While Polese et al. [7] proposed a system based on data mining, which supports tactical decision-making and confirms the role of data mining in strategy formulation. Finally, the research of Deng and Liu [8-10] focuses on the semantic analysis of sports video content and video event mining. This study provides a new method to understand and analyze the video content of sports competitions, which has potential value for improving the audience watching experience and the game analysis of coaches.

This paper focuses on analyzing player performance, momentum advantage, and shifts during the 2023 Wimbledon men's singles final. A model was established to track scores and applied to match data, which underwent preprocessing, imputation, and feature extraction, leading to a dimension reduction to three factors for analysis. A GA-BP neural network prediction model was then constructed, achieving an 88.64% accuracy in predicting turning points. Based on index contribution rates to score fluctuations, tailored advice was offered for players facing various opponents, emphasizing technical skills, physical fitness, and tactical strategies.

3. Theory and Method

3.1 Data Preprocessing

Table 1: Transformatiosn Rules

Processing index	Original symbol	Transformation result
p1_score	AD	50
p2_score	AD	50
winner_shot_type	B	1
winner_shot_type	F	2
speed_mph	NA	0
serve_width	B	1
serve_width	BC	2
serve_width	NA	5

In order to ensure the integrity of the game data, we need to process the acquired data first. Use interpolation to fill in missing values, including game time, score and serve. Unify the time format to ensure that the same score representation is used at the beginning of each ball. And the transformation rules are shown in Table 1.

3.2 Feature Extraction

The following extracts a series of performance characteristics of both players during the competition, which are then used to evaluate the performance of the election. And the Characteristic Indicators are shown in Table 2.

Table 2: Table of Characteristic Indicators

Characteristic	Explain
Leading innings	The number of wins minus the number of defeats
Strong ball	Statistical values
Error rate	Double number / (single number + double number)
Hit the ball and score	Statistical values
Distance_run	The distance a player moves in one round.
Rally_count	Player's return times
Speed_mph	Measured value
Whether to serve	Statistical values

3.3 Data dimension reduction model based on factor analysis

In order to make more representative classification, it is necessary to reduce the dimension of features. In this paper, the interdependence of continuous attributes is digitized by factor analysis, and a large number of relationships are classified into a few comprehensive factors. By analyzing the variance contribution effect between variables, the original system is represented as a new factor combination to reproduce the internal relationship. Modeling is as follows:

$$\begin{cases} x_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + e_1 \\ x_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + e_2 \\ \quad \cdots \cdots \\ x_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + e_p \end{cases} \tag{1}$$

Its matrix form is:

$$X = AF + \varepsilon \tag{2}$$

Where $X = (x_1, x_2, \dots, x_p)^T$ is the observable random vector, which is the continuous attribute given in the question. $F = (F_1, F_2, \dots, F_m)^T$ is a common factor and an unpredictable random vector. $A = (\alpha_{ij})_{p*m}$ is called the factor load matrix.

$$X = (x_1, x_2, \dots, x_p)^T F = (F_1, F_2, \dots, F_m)^T \varepsilon = (e_1, e_2, \dots, e_p)^T A = (\alpha_{ij})_{p*m} \tag{3}$$

Therefore, the factor score model can be written as:

$$\begin{cases} F_1 = \omega_{11}x_1 + \omega_{11}x_2 + \cdots + \omega_{1p}x_p \\ F_2 = \omega_{21}x_1 + \omega_{22}x_2 + \cdots + \omega_{2p}x_p \\ \quad \cdots \cdots \\ F_m = \omega_{m1}x_1 + \omega_{m2}x_2 + \cdots + \omega_{mp}x_p \end{cases} \tag{4}$$

3.4 Nonlinear regression score model

Through the above-mentioned feature screening process, we can establish a regression analysis model about the relationship between time points and athletes' performance scores, and quantify the relationship between time points and athletes' performance scores through regression analysis. We can express the relationship as:

$$y_i = f\left(x_1^i, x_2^i, \dots, x_j^i, \theta_1, \theta_2, \dots \theta_p\right) + \sigma_i \varepsilon (i = 1,2, \dots, n) \tag{5}$$

Where y_i is the true value; i is the first set of data; $f\left(x_1^i, x_2^i, \dots, x_j^i, \theta_1, \theta_2, \dots \theta_p\right)$ is a multivariate nonlinear function, which represents the deterministic part; $\theta_1, \theta_2, \dots \theta_p$ is the unknown model parameter of multivariate nonlinear function; ε is a random part and a random variable that obeys $N \in (0,1)$ distribution.

In this model, the score is the athlete's performance and x is the time point. In this paper, the performance of players in the 2023-wimbledon-1301 competition is selected for analysis.yx

At present, the construction of most nonlinear regression models depends on empirical or experimental methods. However, the empirical method may introduce large errors, while the experimental method takes a long time, but the correct model can be better selected through the experimental results. Therefore, this paper determines the regression model suitable for this relationship through experiments as follows:

Player Carlos Alcaraz:

$$y = p_1 + p_2 x^{\frac{1}{2}} + p_3 x + p_4 x^{\frac{3}{2}} + p_5 x^2 + p_6 x^{\frac{5}{2}} + p_7 x^3 \tag{6}$$

Player Nicolas Jarry:

$$y = p_1 + p_2 x^{\frac{1}{2}} + p_3 x + p_4 x^{\frac{3}{2}} + p_5 x^2 + p_6 x^{\frac{5}{2}} + p_7 x^3 \tag{7}$$

3.5 A Prediction Model Based on GA-BP Neural Network

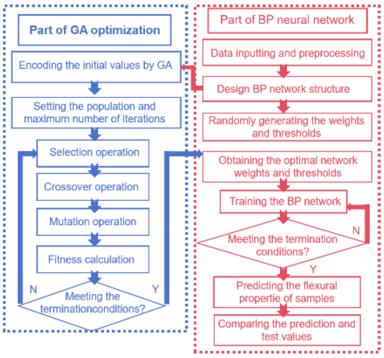

Figure 1: Flow chart of prediction model of GA-BP neural network

To address the computational complexity in deep neural network training, this paper presents the Generalized Approximate Back Propagation (GA-BP) algorithm as an enhancement to the traditional Backpropagation algorithm. The GA-BP algorithm reduces the computational complexity by approximating gradient information, leading to improved efficiency and convergence speed in large-scale deep neural network training. By integrating genetic algorithm with the BP neural network, this paper optimizes neural network parameters to enhance training efficiency and reduce computational burden. Genetic algorithm, an artificial intelligence technique simulating biological evolution, aids in achieving global optimization by evolving towards the optimal solution. This approach helps prevent the BP neural network from getting stuck in local optima, optimizing network parameters to enhance prediction accuracy and effectiveness. The flow chart of the GA-BP neural network prediction model is shown in Figure 1.

BP neural network is a multi-layer feedforward neural network, and it is also called error Back Propagation Neural Networks because its error adjustment process proceeds from the output layer layer by layer. BP neural network consists of three neuron layers, namely, inputlayer, Hidderlayer and 0utput layer, and its topological structure is shown in Figure 2. Nodes in the same layer do not interfere with each other and are independent of each other, and nodes in different layers are connected nonlinearly through neural network parameters such as threshold and weight.

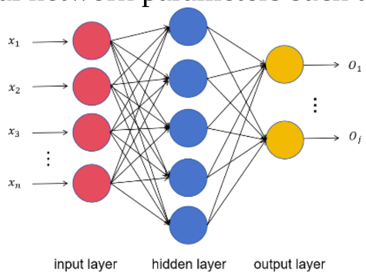

Figure 2: Topological structure diagram of BP neural network

The general process of optimizing BP neural network by genetic algorithm is as follows:

Input variables are input into BP neural network, the initial weights and thresholds of BP neural network are coded, and the individual length is calculated.

$$T = H_{num} \times I_{num} + H_{num} + O_{num} \times H_{num} \tag{8}$$

Where is the individual length $\mathrm{T H_{num} I_{num} O_{num}}$, are the number of input variables, the number of hidden layer neurons and the number of output variables, respectively. Compute the fitness value by calculating the absolute error between the predicted value from the BP and the actual value. The fitness value F is determined using the formula:

$$F = \sum_{i=1}^{n} |Y_i - y_i| \tag{9}$$

Where n is the number of output nodes; is the expected output and predicted output of the first node of BP neural network.$\mathrm{Y_i y_i}$i. The optimal fitness value is found through crossover and mutation. The greater the fitness, the greater the corresponding probability and the easier it is to be selected.

4. Results and Discussions

4.1 Solution of Factor Analysis Model

The regression estimation method is used to calculate the factor score coefficient matrix, and the factor score coefficient matrix is obtained, as shown in Table 3.

Table 3: Factor Score Coefficient Matrix

Characteristic	F1	F2	F3
Leading innings	-0.001	-0.037	-0.006
Strong ball	-0.074	-0.046	0.639
Single-shot error rate	0.354	-0.003	-0.067
Hit the ball and score	-0.099	0.035	0.646
Distance_run	-0.012	0.520	-0.048
Rally_count	0.024	0.520	-0.033
Speed_mph	0.376	0.029	-0.048
Whether to serve or not	0.389	0.007	-0.070

At this point, the factor analysis model can be written as:

$$\begin{cases} F_1 = -0.001x_1 - 0.074x_2 + \cdots + 0.389x_8 \\ F_2 = -0.037x_1 + 0.046x_2 + \cdots + 0.007x_8 \\ F_3 = 0.006x_1 + 0.639x_2 + \cdots - 0.070x_p \end{cases} \tag{10}$$

4.2 Solution of Nonlinear regression score model

We solve the parameters by using the differential evolution algorithm, and the results are shown in Table 4.

Table 4: Parameter Optimization Results

Parameter	y_1	y_2
p_1	0.5123	0.1283
p_2	-0.0255	0.01540
p_3	0.0014	-0.0005
p_4	3.8731	7.2635E-6
p_5	5.6972E-7	-4.2070E-8
p_6	-4.1852E-9	7.1927E-11
p_7	1.2026E-11	

Then draw the curves of both y and x as shown in Figure 3.

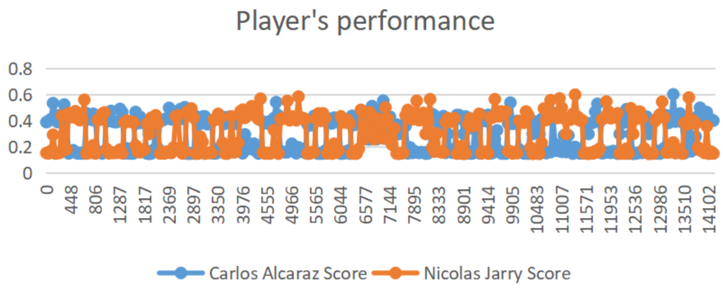

Figure 3: Players' Score Curve

In order to further see the performance of our two players at the time point, we selected the performance of the players at some time points to draw the curves of y and x, as shown in Figure 4.

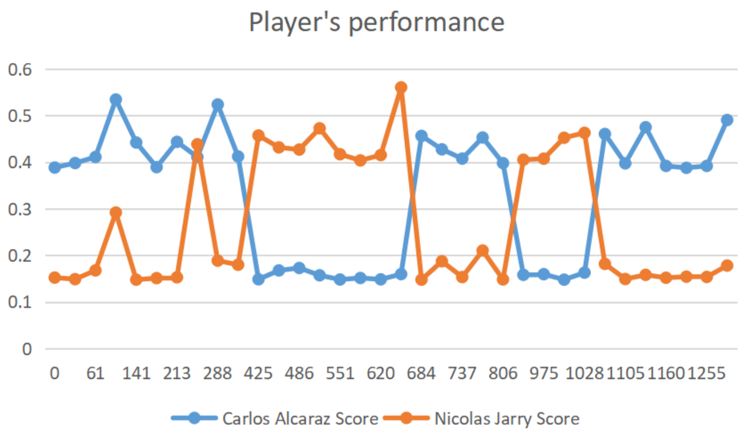

Figure 4: Score curve of local players

It can be clearly seen from the figure that Carlos Alcaraz and Nicolas Jarry performed better at a specific time. Carlos Alcaraz performed better than Nicolas Jarry in the whole game, even at a specific time, Carlos Alcaraz was far ahead of Nicolas Jarry.

4.3 Model Solution Results of GA-BP Neural Network

In order to show the effect of prediction, we show the prediction results of the top 100 inflection points as shown in Figure 5.

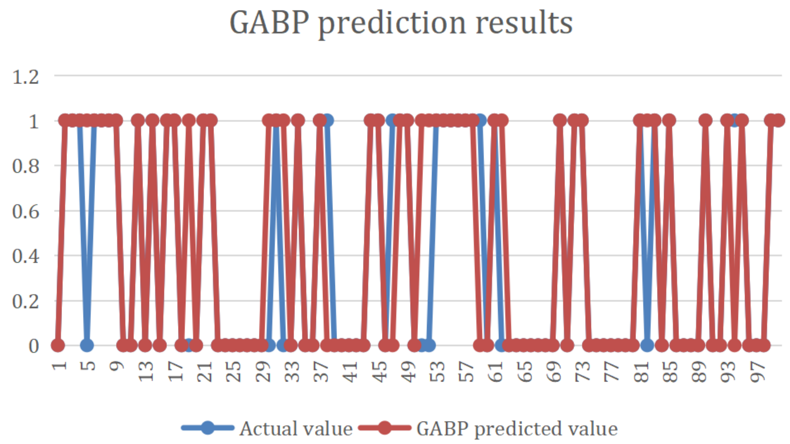

Figure 5: GA-BP prediction results

It can be seen from the figure that the effect of GA-BP prediction model in predicting inflection point is very good. Among the 14,568 inflection points predicted, 12,913 results were successfully predicted, and the accuracy rate of the model was 88.64%, indicating that the GA-BP neural network prediction model has a high accuracy rate and the performance of the model is very good.

5. Conclusion

Tennis players should improve the technical level, especially focus on the training of untouchable shot. High-level skills can help athletes hit the ball more accurately and powerfully, thus gaining a scoring advantage. The accuracy and speed of technical actions, including serving, hitting and volleying, have an important influence on the results of the game, and different technical levels may lead to different scores. And they should pay attention to physical training and improve physical fitness. In the fierce tennis match, good physical fitness and endurance are very important for athletes, because long-term running, fast moving and frequent hitting will consume a lot of physical fitness. Lack of physical fitness may lead to an increase in score fluctuation. Finally, they should study opponents and formulate corresponding tactics and strategies. Knowing your opponent and choosing the right tactics is very important for the result of the game. Different opponents may need different tactical choices, such as adopting conservative tactics or offensive tactics. Pay attention to the opponent's service ability, improve their own hitting quality, and tactical choice directly affects the score fluctuation in the game, because they determine the athletes' actions and decisions in the game.

Acknowledgment

Mengyang Li and Siyuan Peng are co-first authors.

References

[1] Min, B., Kim, J., Choe, C., Eon, H., & McKay, R. I. (2008). A compound framework for sports prediction: The case

study of football. Knowledge-Based Systems, 21(7), 551-562. https://doi.org/10.1016/j.knosys.2008.03.003
[2] Baulch, M. Using machine learning to predict the results of sporting matches. Department of Computer Science and Electrical Engineering, University of Queensland. Retrieved from http://scholar.google.com/
[3] McCabe, A., & Trevathan, J. (2008). Artificial intelligence in sports prediction. In Proceedings of the Fifth International Conference on Information Technology: New Generations (pp. 1194-1197). IEEE Computer Society. https://doi.org/10.1109/ITNG.2008.193
[4] Smith, L., Lipscomb, B., & Simkins, A. (2007). Data mining in sports: Predicting Cy Young Award winners. Journal of Computing Sciences in Colleges, 22(4), 115-121. Retrieved from http://scholar.google.com/
[5] Baird, R. Probabilistic sports prediction using machine learning honors thesis. School of Computer Science and Software Engineering, Monash University, Clayton, Vic. Retrieved from http://scholar.google.com/
[6] Atlas, M., & Zhang, Y.-Q. (2004). Fuzzy neural agents for online NBA scouting. In Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence (pp. 58-63). IEEE Computer Society. https://doi.org/10.1109/WI.2004.10016
[7] Polese, G., Troiano, M., & Tortora, G. (2002). A data mining based system supporting tactical decisions. In SEKE '02: Proceedings of the 14th international conference on Software engineering and knowledge engineering (pp. 681-684). ACM. https://doi.org/10.1145/568760.568889
[8] Lühr, S., & Lazarescu, M. (2007). A visual data analysis tool for sport player performance benchmarking, comparison and change detection. In Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (Vol. 1, pp. 289-296). IEEE Computer Society. https://doi.org/10.1109/ICTAI.2007.125
[9] Deng, L. Y., & Liu, Y.-J. (2008). Semantic analysis and video event mining in sports video. In Proceedings of the 22nd International Conference on Advanced Information Networking and Applications-Workshops (pp.1517-1522). IEEE Computer Society.
[10] Milosavljevic, B., & Vidakovic, M. (2007). Java i internet programiranje. Novi Sad: FTN Izdavaštvo.