Exploration and Prediction of Factors Influencing Athlete Momentum in Tennis Matches

Zihan An[1, *], Jiaxu Du[2, #], Dinggen Li[3, #]

[1]School of Astronautics, Northwestern Polytechnical University, Xi'an, China, 710129

[2]Wuxi School of Medicine, Jiangnan University, Wuxi, China, 214122

[3]School of Law, Beijing Institute of Technology, Beijing, China, 100081

* Corresponding Author Email: woaizhongguo@mail.nwpu.edu.cn

[#]These authors contributed equally.

Abstract. In the field of sports competition, the momentum fluctuation of athletes in a match has an important impact on the final performance, especially in a high-intensity confrontational sport such as tennis. In this paper, based on the men's singles data of the 2023 Wimbledon Tennis Championships, a prediction model based on a decision tree and a BP neural network was established to analyze the momentum and fluctuation of athletes in a match. This study contributes to an in-depth understanding of athletes' performance during matches and provides a scientific basis for their training and tactical adjustments.

Keywords: Decision tree, BP neural network, Logistic regression, Momentum.

1. Introduction

In the field of sports competition, accurate analysis of the momentum fluctuations of athletes during a match is of great significance in understanding the outcome of the match. Especially in high-intensity antagonistic sports such as tennis, momentum changes may directly affect the outcome of the match. To address this issue, this paper establishes a prediction model based on the men's singles data of the 2023 Wimbledon Tennis Championships using methods such as decision trees and BP neural networks[1], aiming to analyze the momentum and fluctuations of the athletes in the matches in depth. By quantifying and trending momentum, we will be able to better understand the performance of athletes at critical moments and provide a more scientific basis for training and competition strategies[2].

Decision trees can obtain the dynamic change of optimal solutions when multiple variables change simultaneously[3], and have been widely used in the fields of fruit classification and identification[4], reservoir flood control and scheduling[5], and civil power supply[6]; BP neural networks have strong nonlinear simulation ability, and they have a high prediction accuracy, strong self-learning and self-adaptive ability[7]. This makes them an effective tool for analyzing sports and athletic data such as momentum and fluctuations of athletes during competition.

In this work, this study firstly used decision tree classification to select different factors affecting athletes' performance each time for three training sessions to get the weights of the selected factors and then used BP neural networks to make predictions and map the predictions into percentage scores to quantify the scores of the athletes' performance and to judge the performance of the athletes. Secondly, 11 factors are considered to have an impact on the athlete's momentum, and the momentum of two people is calculated two times by the predicted characteristics of one of them. The coefficients of each factor were also determined by parameter optimization[8] and the correlation between the calculated momentum and the direction of the match was tested using the Spearman correlation coefficient method.

2. Evaluation and Prediction Analysis of Athlete Performance

A player's performance is influenced by various factors such as serving side, running distance, ball speed, landing point, and so on[9]. Therefore, for such a complex dynamic system, we adopt decision tree classification and BP neural network regression to solve it. The decision tree model generates rules and decision paths to visualize which features are most important to the athlete's performance[10]; For nonlinear problems such as situation changes, BP neural networks can provide more flexible and accurate modeling capabilities[7]. Therefore, it is more accurate to judge the performance of athletes based on these two methods

2.1. Screening of influencing factors

The decision tree parameters were set up using the various features of the data table as the basis for categorization, and the parameters are shown in Table 1.

Table 1. Decision Tree Parameters

Parameters	Data Value
Minimum number of samples for internal node splitting	10
Minimum number of samples for leaf nodes	5
Leaf node sample minimum weight	0
Maximum depth of the tree	10
Maximum number of leaf nodes	50

Taking 70% of the data as a training set and 30% of the data as a test set for three predictions, the prediction results are shown in Figure 1.

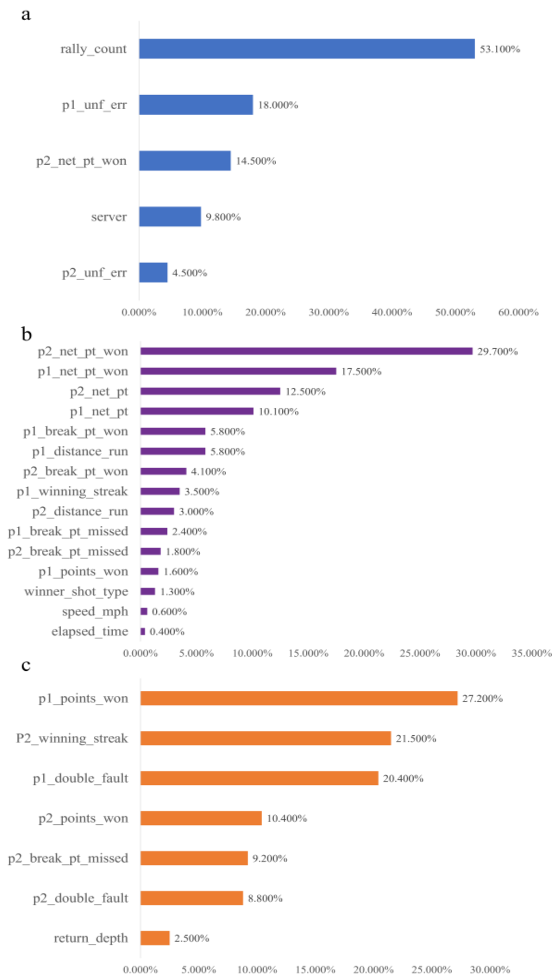

Figure 1. Decision tree prediction results

The prediction results were compared with the actual results to get the Accuracy rate、Recall rate、Precision rate and Harmonized average of precision and recall rates for each prediction and the data are shown in Table 2.

Table 2. Decision Tree Prediction Accuracy, Recall, Precision and F_1

	Accuracy rate	**Recall rate**	**Precision rate**	F_1
Train set	0.749±0.201	0.749±0.201	0.757±0.199	0.745±0.204
Test set	0.740±0.194	0.740±0.194	0.746±0.192	0.736±0.197

In the first prediction, the server was found to have 9.8% importance, suggesting that which side serves has some influence on the winner, and p2_net_pt_won was found to have an uncharacteristic 0% importance while its counterpart, p1_net_pt_won, had only 0%. So, we removed the remaining four features with higher importance (i.e., server, rally_count, p1_unf_error, p2_unf_error) and continued the decision tree prediction for the remaining terms.

In the second prediction, it can be found that there are many more influential feature terms than in the first one, in which only p2_net_pt_won was valid in the first one, whereas this time p1_net_pt_won was equally important and accounted for 17.5% of the total, second only to p2_net_pt_won. Thus, p1_winning_streak, p2_winning_streak, p1_points_won, p2_points_won, return_depth, serve_depth, serve_width, p1_break_pt, p2_break_pt, p1_double_fault, p2_double_ fault, p1_ace, p2_ace feature terms for the third prediction.

In the third prediction, p1_points_won to return_depth was found to have a higher influence.

Combining the above three predictions, the selected influential feature factors are shown in Table 3.

Table 3. Factors Affecting Characteristics

Player 1 Parameters	**Shared Parameters**	**Player 2 Parameters**
P1_net_pt	Speed_mph	P2_net_pt
P1_break_pt	Server	P2_break_pt
P1_distance_run	Rally_count	P2_distance_run
P1_double_fault	Winner_shot_type	P2_double_fault
P1_points_won	Elapsed_time	P2_points_won
P1_unf_error	-	P2_unf_error
P1_net_pt_won	-	P2_net_pt_won
P1_break_pt_won	-	P2_break_pt_won
P1_winning_streak	-	P2_winning_streak
P1_break_pt_missed	-	P2_break_pt_missed

Finally, the selected characteristic factors were predicted and the assessment results were obtained as shown in Table 4.

Table 4. Predictive Accuracy, Recall, Precision, and F1 of 25 Factors

	Accuracy rate	**Recall rate**	**Precision rate**	F_1
Train set	0.981	0.981	0.981	0.981
Test set	0.97	0.97	0.97	0.97

The precision and recall of the test data for predicting the evaluation results are both 0.97, which is an improvement compared with the 0.94 and 0.95 of the first time.

2.2. Solve using a regression algorithm

We set the BP neural network parameters with the selected 25 influencing factors as inputs, as shown in Table 5.

Table 5. BP neural network parameters

activation function	*L2 regular terms*	*Number of iterations*	*learning rate*	*Number of hidden layers*	*Number of neurons*
$f(x)=\max(0,x)$	1	1000	0.1	2	100

The resulting result out1 is transformed by the equation (1) to get out2, which is the player's performance score, and the player's performance can be compared. The higher the score, the better the performance of the player. When the out2 value of athlete 1 is greater than that of athlete 2, athlete 1 performs better, and vice versa for athlete 2.

$$out2=\frac{1}{1+e^{-3(out1-1.5)}} \tag{1}$$

The first game of 2023-wimbledon-1301 was selected to plot an image of the player's game scores against the model scores, as shown in Figure 2.

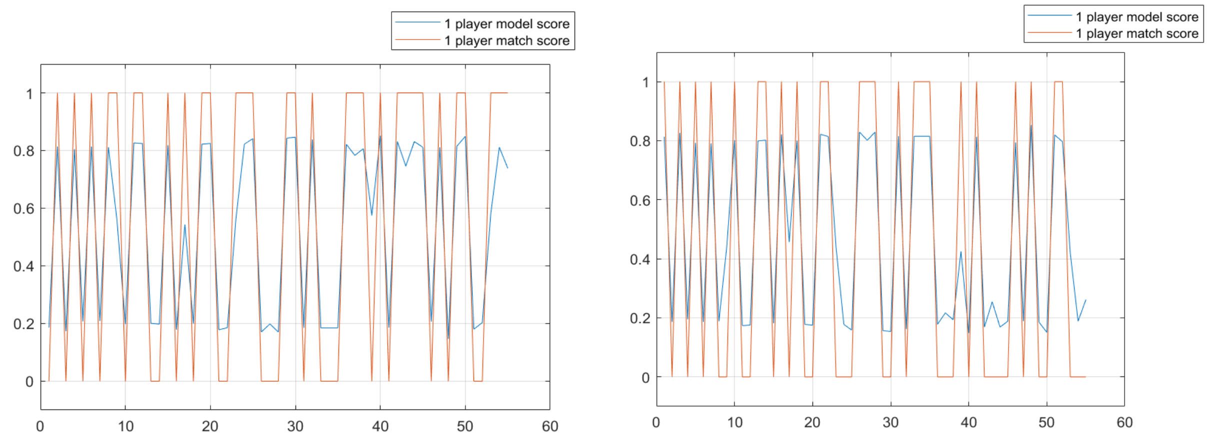

Figure 2. Scoring of the two players

Then magnify out2 by a factor of 100 to obtain out3, and draw the two player out3 scores, as shown in Figure 3.

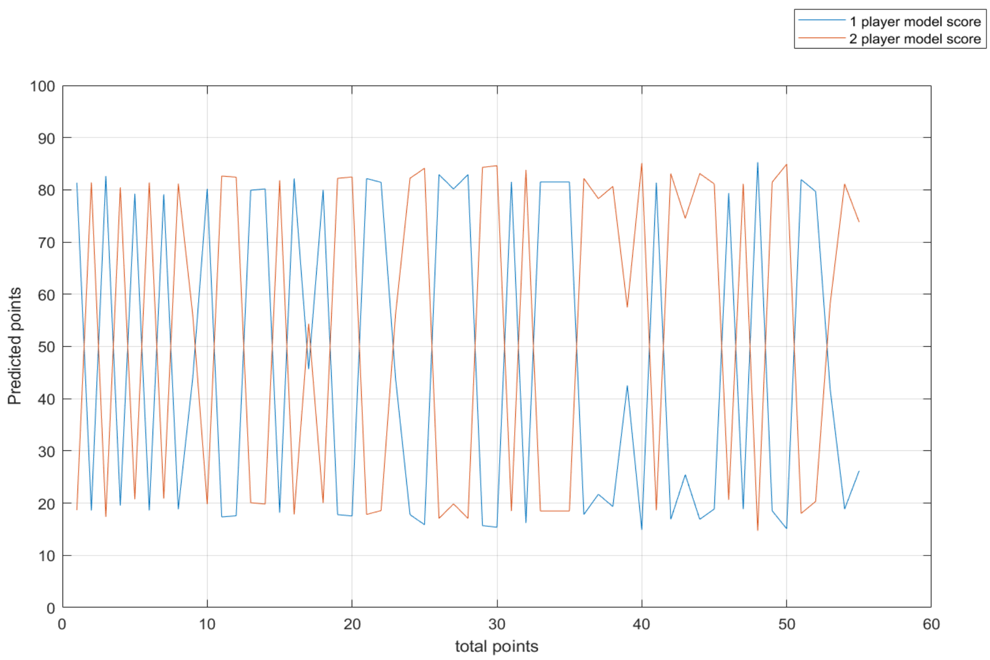

Figure 3. Comparison of scores between the two players

From the figure, it can be seen that the same player's match score is positively correlated with the model score, the model score of two players is negatively correlated, and the sum of the out2 of the two players is 1, and the sum of the out3 is 100, and the model predicts good results.

3. Exploring the Impact of Momentum on Athletic Performance

3.1. Scoring model

Taking player 1 as an example, suppose A(server), B(p1_ace), C(p1_double_fault), D(p1_break_pt_missed), E(p1_break_pt_won), F(p1_distance_run), G(p2_distance_run), H(p1_winner), I(p1_unf_err), J(p1_net_pt), K(p1_break_pt), and L(p1_net_pt_won) have an effect on the player's momentum, and the metrics are quantified and considered:

$$M = A_1 + A_2 + A_3 + A_4 + b$$
$$A_1 = \alpha A + \beta B + \chi C + \delta D$$
$$A_2 = \varepsilon E + \phi \det + \varphi H$$
$$A_3 = \gamma I + \eta J$$
$$A_4 = \iota K + \kappa L$$
$$\det = F - G$$

where M is the player's momentum magnitude, which can be calculated by

$\alpha = 0.57$ $\beta = 0.078$ $\chi = 0.37$ $\delta = 0.692$ $\varepsilon = -0.856$ $\phi = -0.004$
$\varphi = -0.728$ $\gamma = 0.978$ $\eta = 0.796$ $\iota = -0.162$ $\kappa = -1.196$ $b = -0.962$

Substituting, we get:

$$\begin{aligned} M &= 0.57A + 0.078B + 0.37C + 0.692D - 0.856E - 0.004F + 0.004G - 0.728H \\ &+ 0.978I + 0.796J - 0.162K - 1.196L - 0.962 \end{aligned} \tag{2}$$

Equation (2) can be calculated by the player's momentum in the course of the game, brought into the game data for calculation, and plotted the trend of the player's momentum change in the course of the game, as shown in Figure 4.

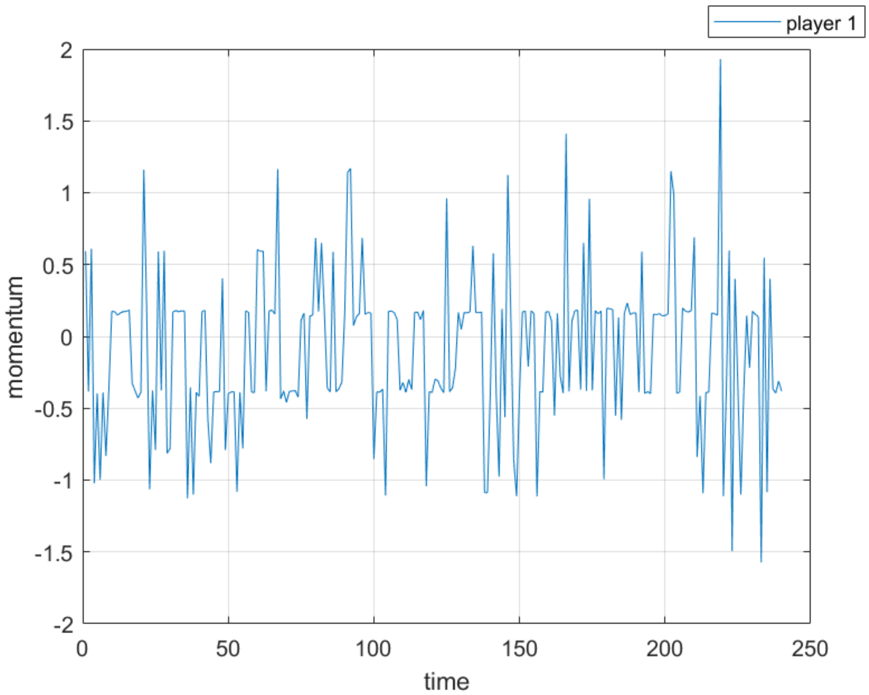

Figure 4. Momentum over time

The graph of player momentum versus player score variation is shown in Figure 5.

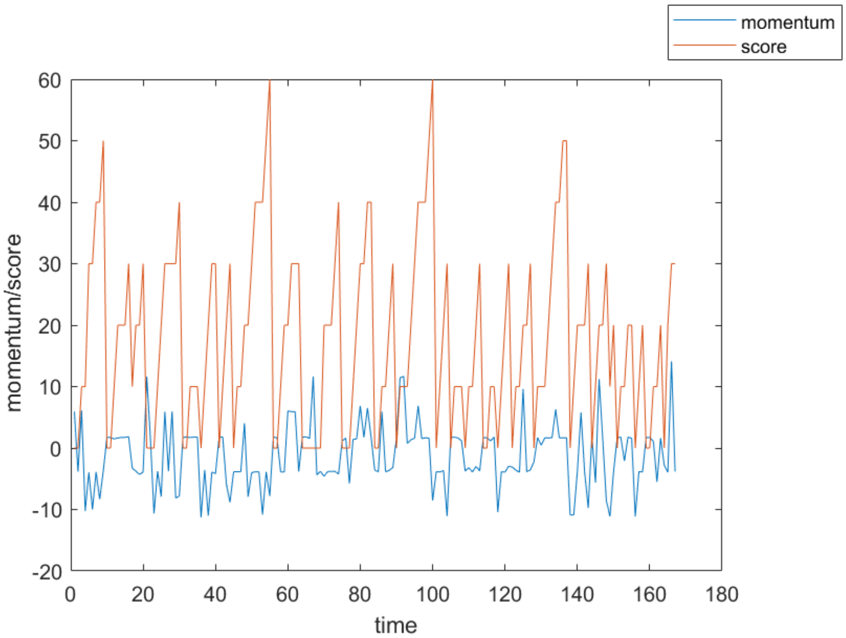

Figure 5. Momentum and Player Score Chart

In Figure 5, we zoomed in on the player's momentum five times compared it to the player's scoring, and found that the player scored when the player's momentum was positive, indicating that momentum has an effect on the athlete's scoring during the game.

3.2. Spearman correlation analysis

Spearman correlation coefficient[11]

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}. \tag{3}$$

Bringing momentum and point_victor as X and Y, respectively, into Eq. (3), a table of correlation coefficients is obtained in Table 6.

Table 6. Correlation Coefficient

significance level	ρ
1%	0.649
5%	0.927
10%	0.991

From Table 6, we can see that at a 10% significance level, the correlation coefficient of our model for assessing athletes' momentum is close to 1, while it still reaches 0.649 at a 1% significance level, which indicates that athletes' scores are related to the athletes' momentum.

4. Conclusions

In this study, 25 important factors affecting athletes' performance were screened using decision tree classification, and then the BP neural network was used to make predictions, which were mapped to percentage scores to quantify the scores of athletes' performance and judge their performance, and 11 factors were considered to have an impact on athletes' momentum. The momentum of two people was calculated two times by the predicted characteristics of one of them. The coefficients of each factor were also determined by parametric optimization and the correlation between the calculated momentum and the direction of the race was tested using the Spearman correlation coefficient method. The correlation coefficient was 0.927 at the 5% significance level.

This paper provides a research idea and framework in the field of sports science & sports psychology, as well as a reference for athletes to improve their performance and for coaching teams to better guide training and competition.

References

[1] Huang Zhiying. Research on tennis match performance prediction model based on BP neural network[D]. Fujian: Fujian Normal University,2022:9-10.

[2] Cai Yier. Research on the application of sports science and technology in competitive table tennis[J]. Sports Goods and Technology,2023,(20):154-156.

[3] Li Mingxia. Enterprise risk management method and sensitivity analysis based on decision tree[J]. Mall Modernization,2023,(23):146-148.

[4] Zhang Liang,Wang Xuanzhi. Fruit classification recognition based on decision tree algorithm[J]. Science and Technology Innovation and Application,2024,14(06):18-23.

[5] Wang Junnuo, *et al*. Research on reservoir flood control scheduling based on decision tree C5.0[J/OL]. Journal of Water Resources and Water Transportation Engineering,2024,2:1-10.

[6] Song Lili, Sun Yan. Intelligent assessment algorithm for power supply reliability of real civil buildings based on decision tree[J]. Comprehensive utilization of fly ash,2023,37(06):128-133.

[7] Wang Jiquan, *et al*. Application of BP neural network in the prediction of total power of agricultural machinery[J]. Practice and Understanding of Mathematics,2013,43(01):72-79.

[8] Li Xing, *et al*.Algorithm and topology optimization of BP neural network for predicting surface settlement parameters[J]. China Safety Production Science and Technology,2023,19(06):90-97.

[9] Zhao Mingnan, *et al*. Theoretical modeling and scale development of emotional intelligence in Chinese tennis players[J]. Psychological Research,2024,17(01):42-53.

[10] Gong Jianwei, *et al*. Research on Employee Data Visualization and Exit Prediction Based on Logistic Regression and Decision Tree[J]. Intelligent Computer and Application,2023,13 (06):162-167.

[11] Zhang Chigeng.Python program design and application of Spearman rank correlation coefficient - an empirical analysis of the correlation between book reading and tacit knowledge acquisition ability[J]. Modern Information Technology,2023,7(21):195-198.