Athlete Momentum Visualization Based on Logit Regression Model and LSTM Prediction Model

Jinhao Jiang[1, *, #], Yuqian Li[2, #]

[1]School of Physics, Hefei University of Technology, Hefei, China, 230009

[2]School of Instrument Science and Opto-electronics Engineering, Hefei University of Technology, Hefei, China, 230009

*Corresponding author: 2022215021@mail.hfut.edu.cn

[#]These authors contributed equally

Abstract. With the development of information technology, big data analytics predicts the likelihood of a player's victory by capturing the game situation, thus providing the player with a timely and appropriate strategy to win the game. In order to investigate whether the player's "momentum" has an impact on the game results, this paper extracts four indicators, namely, the player difference, the number of consecutive wins, the number of breaks of serve, and the number of successful breaks of serve based on the data of the Wimbledon Men's Tennis Championships in 2023, in order to quantify the impact of the "momentum" on the game results. Logit regression models were then developed to find strong correlations between these metrics and the number of wins. Therefore, in order to further investigate the influence of these metrics on the game process, an LSTM prediction model was built to predict the outcome of the game. The training results of the final model show that the MSE is 0.316 and the MA is 0.387, with low prediction errors. Based on the prediction results of LSTM, this paper provides appropriate suggestions for players by analyzing the turning points of win/loss tilts and changes in players' status.

Keywords: Momentum, Logit regression, LSTM prediction.

1. Introduction

With the advancement of information technology, big data analytics can be used to describe the match conditions which has become the key to improve the winning percentage of the players at the Wimbledon Gentlemen's Championships in 2023. The official collection of the Wimbledon Tennis Championships, which publishes the score statistics of the first two rounds of men's tennis, can be used to predict the likelihood of a player's victory by capturing the match conditions, which can help the player to make the appropriate strategic adjustments to win the match.

For an athlete, whether in tennis or any other sport, there are many factors that influence the outcome of a match, and most of the time it is attributed to "momentum". The high momentum gained by a player through the influence of various factors is very likely to stimulate the player's potential to win the championship; low momentum may undermine the player's confidence and disrupt their mindset, but it may also stimulate their power of survival. Therefore, this paper hopes to quantify the size of "momentum" through big data analysis, so as to judge the impact of momentum on the game.

2. A statement on momentum

2.1. The structure of BP neural network [1]

We decided to extract several relevant metrics as components of momentum and demonstrated that these metrics are highly correlated with race wins through a logit regression model.The overall model construction idea is shown in the Figure 1.

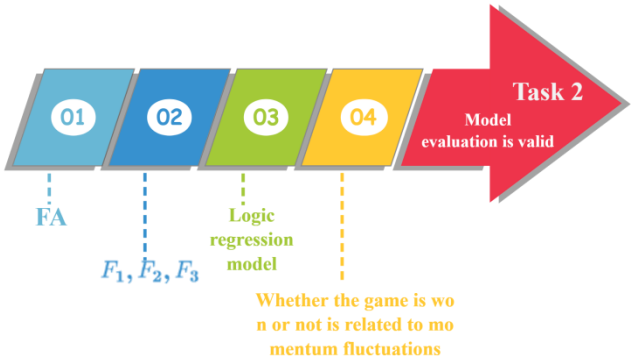


Figure 1. Flowchart

2.2. The Index was Determined by Factor Analysis [2]

2.2.1 New Indicator

In order to quantify player momentum, we have added the following metrics to describe player momentum:

1)The difference SD_{ik}^{m} between player 1 and player 2, that is, the difference in the score of the player at each scoring point, is expressed as:

$$SD_{ik}^{m} = PW_{ik1}^{m} - PW_{ik2}^{m} \tag{1}$$

2) i, j, m, k denote group number, player number, match number, and tournament number respectivelyWhere PW_{ik1}^{m} and PW_{ik2}^{m} represent the points of player 1 and Player 2 in game k of set i in match m, respectively.

3)Number of consecutive matches won by a player $PS_{ij}^{m} : PS_{ij}^{m} = k-1$ is recorded if a player wins k games in a row in each set

4)Player breaks BP_{j}^{m} : indicates the cumulative number of player breaks in each match;

5)Whether the player wins the next game WNG_{ij}^{m}: $WNG_{ij}^{m} = 1$ if he wins, $WNG_{ij}^{m} = 0$ if he doesn't.

2.2.2 Factor Analysis

This problem requires us to establish an assessment of the relationship between players' momentum and whether they can win the match. Therefore, we conduct factor analysis on four indicators: the number of players' breaks, the number of successful breaks, the difference of opponents, and the number of consecutive wins, and comprehensively deal with these four indicators to extract common factors that have a greater impact on the outcome of the match. The core principles are as follows:

Given four indicators: $SD_{ik}^{m}, PS_{ij}^{m}, BP_{j}^{m}, BPP_{j}^{m}$ We decided to reduce dimensionality to three factors using principal component analysis equation:

$$\begin{cases} F_1 = 0.495BPP + 0.15SD + 0.295PS + 0.357BP \\ F_2 = 0.412BPP + 0.982SD - 0.116PS + 0.01BP \\ F_3 = 0.32BPP + 0.226SD + 1.255PS - 1.434BP \end{cases} \tag{2}$$

2.3. Player Win Probability Evaluation Based on Logic Regression Model

2.3.1 Logit Regression Model was Established

Three main indexes are obtained according to the principal component analysis F_1, F_2, F_3, the Logit regression method [3] is used to analyze the player's momentum and verify its correlation with the player's win probability in each match. Take the winning probability of each player in the i plate

of the m match as WP_{ij}^{m}, and fit the coefficient of the index in the regression equation through the index. The model is as follows:

$$WP_{ij}^{m} = \frac{1}{1+e^{-\sum_{k=0}^{3}\beta_k x_{kj}}} \tag{3}$$

Where $x_{kj}, k=1,2,3$, respectively, represents the three indicators corresponding to each player, β_k is the regression coefficient of the three indicators, then the maximum likelihood function of player j 's victory is:

$$L(\beta) = \prod_{i=1}^{s^m}\left(\frac{1}{1+e^{-\sum_{k=0}^{3}\beta_k x_{kj}}}\right)^{y_{ij}^{m}}\left(1-\frac{1}{1+e^{-\sum_{k=0}^{3}\beta_k x_{kj}}}\right)^{1-y_{ij}^{m}} \tag{4}$$

Where y_{ij}^{m} is the known probability of winning. Then the cost function is:

$$J(\beta) = -\frac{1}{s^m}\sum_{i=1}^{s^m}\left[y_{ij}^{m}\ln\left(\frac{1}{1+e^{-\sum_{k=0}^{3}\beta_k x_{kj}}}\right)+\left(1-y_{ij}^{m}\right)\ln\left(1-\frac{1}{1+e^{-\sum_{k=0}^{3}\beta_k x_{kj}}}\right)\right] \tag{5}$$

By calculating the cost function $\min J(\beta)$, the regression coefficient is finally obtained as shown in the Table 1:

Table 1: The Value of Regression Coefficient β

Regression Coefficient	β_1	β_2	β_3
Corresponding Value	0.113	0.088	-0.024

2.3.2 Forecast Result

According to the Logic regression established above, we can conclude that the closer the value of WP_{ij}^{m} is to 1, the greater the probability of the player winning, and the closer the value of WP_{ij}^{m} is to 0, the smaller the probability of the player winning. The probability of players winning in each match is predicted as figure 2:

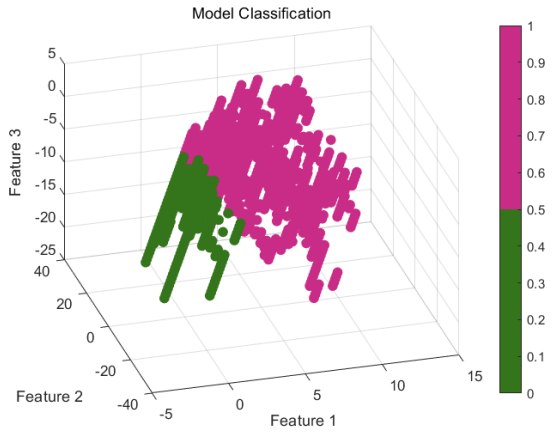


Figure 2: Player Win Probability Prediction

The 3D indicator image shows player's F_1, F_2, F_3 prediction of the probability of winning the next match under the influence of these three indicators, and believes that: if $WP_{ij}^{m} > 0.5$, we judge the player to have won, $WNG_{ij}^{m} = 1$ can be recorded; if $WP_{ij}^{m} \le 0.5$, we think the player has lost in the next match, $WNG_{ij}^{m} = 0$ can be recorded. This binary image verifies that the three different indicators

each have different effects on the players' win probability, which also shows that the fluctuations in the game are not random, so we believe that the coach's view is wrong.

2.3.3 Model Evaluation

Table 2 shows four important metrics for Logic regression model evaluation, all of which are above 0.75, reflecting the model's high accuracy:

Table 2: Four important indicators for Logit regression model evaluation

Precision	Sensitivity	AUC	F1 Score
0.807	0.816	0.776	0.811

At the same time, it can also be seen from the confusion matrix in Figure 3 and the ROC curve in Figure 4 that in the binary classification model, the proportion of positive diagonal line is significantly higher than that of negative diagonal line, and the ROC curve is near the upper left corner. As can be seen from the above table, AUC=0.776, the closer the value is to 1, the better the classification effect of the model. Therefore, we can conclude that the Logic regression model established in this paper has high sensitivity, low misdiagnosis rate and good prediction performance.

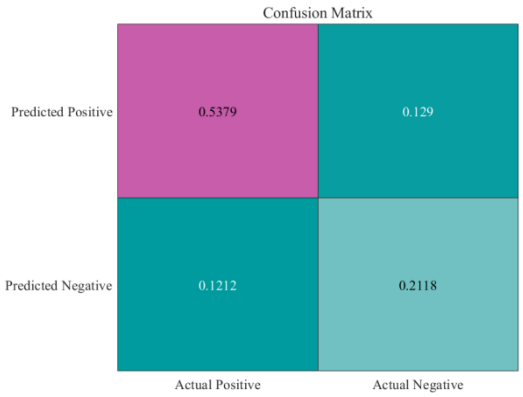


Figure 3: Confusion Matrix

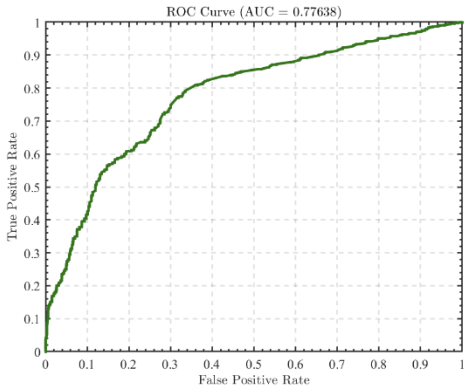


Figure 4:ROC Curve

3. Determine the metrics that affect a player's score

3.1. Spearman Correlation Analysis

We quantified the momentum of players by using the aforementioned opponent differences, the number of consecutive wins of players, the number of breaks of players and the number of successful breaks of players, and judged the correlation between the above four indicators and the indicator of whether a player won the match. The specific ideas of the correlation method used are as follows:

The Spearman correlation coefficient[4] is actually calculated using the rank correlation coefficients of the two variables. We took the momentum indicators of the three players as x and whether the players won the game as y respectively, and calculated them by Spearman correlation formula:

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i)-\overline{R(x)}\right)\cdot\left(R(y_i)-\overline{R(y)}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i)-\overline{R(x)}\right)^2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R(y_i)-\overline{R(y)}\right)^2}} \tag{6}$$

The correlation coefficient between the above four indicators and whether the next round is won (WNG) is shown in Figure 5:

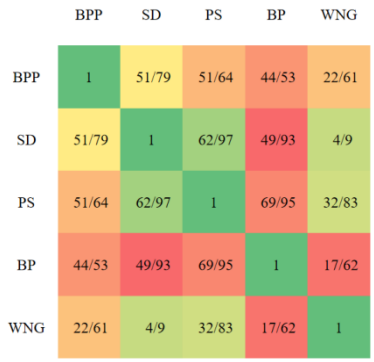


Figure 5: Spearman Correlation Coefficient Matrix

It can be seen from Figure 5 that these four characteristic indicators describing the momentum of the player are significantly correlated with whether the player wins the game, indicating that these four indicators have an impact on whether the player wins the game. Therefore, a neural network model can be used to predict the outcome of a match through four indicators with high correlation

3.2. Match result prediction based on LSTM model

3.2.1 Model Building

In tennis matches, the number of breaks, the number of successful breaks, the opponent's margin, and the number of consecutive wins are all serial data that change over time. So we define the above four features as input variables and the output variables as predictions of the player's next match outcome. Finally, for the selection of the prediction model, we decided to use the LSTM deep learning model, the flowchart of which is shown in Figure 6, and the principles of the application of the model are described in detail below.

Long-term memory network (LSTM) [5] is a variant of recurrent neural network (RNN) for sequence data with good sequence modeling capabilities. Because LSTM is able to retain and leverage historical information and capture long-term dependencies in these sequences, this is critical for predicting match outcomes. The model works as follows:

(1)First, the information retained and forgotten in the cell state of the network is controlled through the forget gate in the LSTM network. It determines the discarded information by reading the previous output h_{t-1} and the current input x_t, and then output the vector f_t:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{7}$$

(2)Then the acceptance of the new input information is controlled by the input gate to achieve the purpose of updating the memory unit in the network:

$$\begin{aligned} i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \\ \tilde{C}_t &= \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \end{aligned} \tag{8}$$

The value to be updated is first identified through the sigmoid layer and then added to the state through the tanh layer to add a new candidate value.

(3)Next we need to update the old cell state, which is C_{t-1} to C_t:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{9}$$

Multiply the old state f_t and to determine what information we need to discard, and add new candidate values to update my cell state. Because the cell state is like a conveyor belt, with only a few linear interactions, information flows on it unchanged.

(4)Finally, the output of the memory unit is controlled through the output gate, and which part of the information will be passed to the output of the network:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t) \tag{10}$$

The output part of the cell state is determined by the sigmoid layer first, and then the cell state is multiplied by the output of the sigmoid layer after processing in the tanh layer, and the output result is finally determined.The following diagram also shows the core principles of the LSTM model

Based on the establishment principle of the LSTM model above, we trained the neural network model 500 times by using the data given in the title, and tested and analyzed it. The error results of the model were obtained as shown in Table 3:

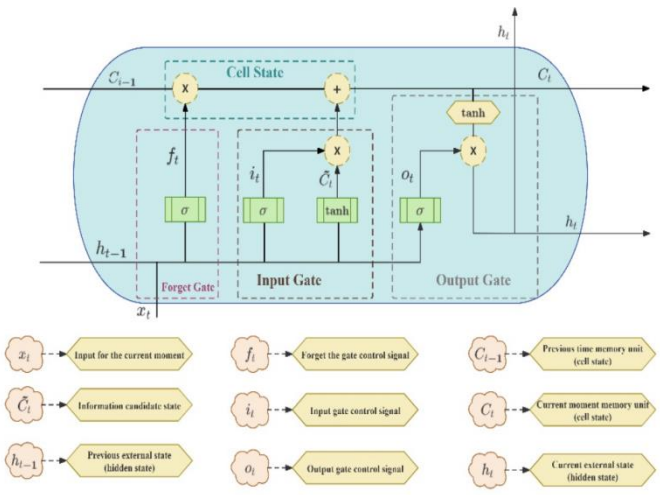


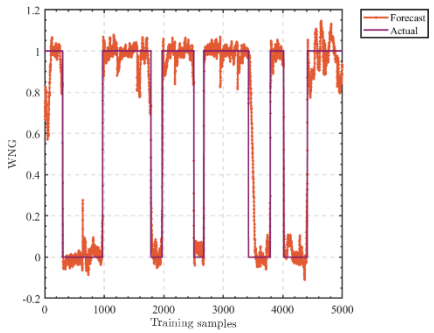
Figure 6: Establishment of LSTM

Table 3: Error Result of LSTM

RMSE	MSE	MA	MAPE
0.5619	0.31573	0.38666	0.57716

3.2.2 The Results of Prediction

The training results and prediction results obtained by the LSTM model are shown in Figure 7 and Figure 8:



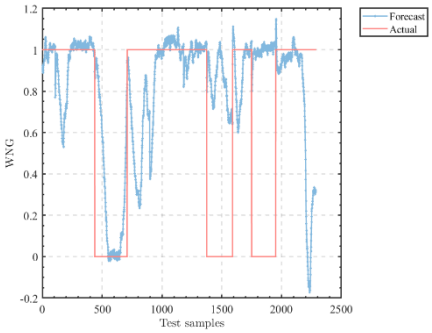


Figure 7: Training Results **Figure 8:**Test Results

Figure 7 and Figure 8 show the training results and test results of the LSTM model respectively. It can be seen from the training results that the model has a good discrete effect on discrete data, so we can use the model to predict the winning situation of players in the next game.

At the same time, it can be seen from the test results that the predicted results of the model are basically consistent with the actual situation, and the fluctuation trend shown in the figure also explains the players' match results well. Accordingly, we will give the players relevant advice based on the analysis of the test results graph.

3.2.3 Momentum versus victory[6]

By analyzing the cause of the difference in "momentum"[7], that is, the turning point of the test results, it is possible to capture the influence of various characteristics on the score of the player during the game, and then make recommendations.The "momentum" turning point is shown in Figure 9.

Reason 1: we can analyze that the reason leading to the failure of players is that when players have multiple break opportunities, the number of successful breaks remains the same because of continuous failure, and the gap between opponents is also increasing, which indicates that these unfavorable factors make the players' mentality unstable. This resulted in the player losing the match.

Reason 2: We can analyze three characteristics, which are the number of successful breaks of the players, the difference between the opponents and the number of breaks of the players. During this period, the player has multiple break opportunities. However, the player repeatedly fails to break, and the score difference with Player One is gradually increasing. This causes the player to lose momentum and eventually lose the match.

Reason 3: we can analyze the four characteristics of the player's successful break times, the difference between opponents, the player's break times and the player's consecutive wins, and we can see that the four indicators of the player are constantly rising. And the four indicators are positive indicators, which makes the player's momentum continues to increase, and finally the player won the match.

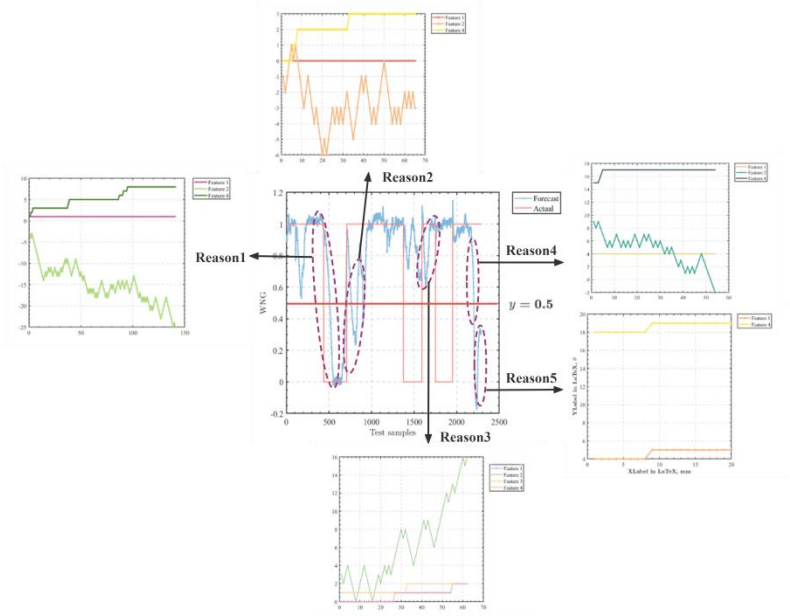


Figure 9: Turning-point Chart

Reason 4: we can analyze the three characteristics of the player's successful break times, opponent differences and player breaks times respectively. Similarly to reason 1, the player's successful break times are 0 under the condition of continuous high break times, and the increasing difference between opponents also leads to the reduction of the player's momentum, and finally the player loses in this match.

Reason 5: we can analyze the two characteristics of the player's successful break times and the player's break times respectively. Since the two positive indicators of the player's break times and successful break times are constantly on the rise, it indicates that the player's momentum is also growing, leading to the player's victory in the match.

Among them, the analysis results of the four indicators of the difference of the opponent, the number of consecutive wins of the player, the number of breaks of the player and the number of successful breaks of the player when the state is 0-1 are reason 5 and reason 3. The analysis results

of the four indicators of the opponent difference, the number of consecutive wins of players, the number of breaks of players and the number of successful breaks of players at 1-0 respectively are cause 1, cause 2 and cause 4.

From this, we can see that successive losses will frustrate the player's spirit, but as the difference between the opponent and the player gradually becomes smaller, the number of consecutive successful breaks increases, the player will recover the spirit, "momentum" gradually rises. Therefore, when the "momentum" of players is low, they should actively adjust their mentality, find the opportunity to break, improve the success rate, and gradually narrow the scoring gap and reverse the situation.

3.2.4 Proposals

In response, we have made some recommendations for coaches and athletes based on the findings above.

To Coaches:

1. Comprehensive comparative analysis of the opponent's past game data combined with the player's previous game data, and detailed analysis of "momentum" changes[8], recent game results, opponent characteristics, and player performance at critical moments, so that the coach can accurately capture situation changes and timely Provide corresponding guidance to players.

2. In response to changes in momentum, it is recommended to adjust players' techniques and tactics. Aspects such as serving, receiving, and tactical strategies may need to be modified to better suit the opponent and game situation.

To Players:

1. Maintain a high level of concentration during the game. Focus on each point and prevent distractions[9], which will help you better deal with uncertainties in the game.

2. Adjust personal tactics according to the coach's suggestions, flexibly change strategies during the game, and make adjustments according to the opponent's performance and the game situation.

3. Pay attention to psychological preparation and stay calm under various factors affecting the game through breathing control, positive attitude cultivation, etc.

4. Conclusions

In this paper, we process the data of 2023 Wimbledon Men's Tennis Championships based on big data analysis technology in today's information age. We extracted four indicators, namely, players' point difference, number of consecutive wins, number of breaks of serve, and number of successful breaks of serve, and concluded that these four indicators can form "momentum"and influence the results of the matches, and verified this conjecture by logit regression model.

In order to explore more deeply the process of momentum[10] affecting the winning and losing of matches, we built an LSTM prediction model and compared the prediction results with the actual results, and found that the rise and fall of these four indicators have a great correlation with the probability of winning the match. Accordingly, we provide coaches and athletes with strategies to cope with sudden changes in momentum, thus helping them to win better.

References

[1] Huang Y, Xiang Y, Zhao R, et al. Air quality prediction using improved PSO-BP neural network[J]. Ieee Access, 2020, 8: 99346-99353.

[2] Gorsuch R L. Factor analysis: Classic edition[M]. Routledge, 2014.

[3] Das A. Logistic regression[M]//Encyclopedia of Quality of Life and Well-Being Research. Cham: Springer International Publishing, 2021: 1-2.

[4] Yu H, Hutson A D. A robust Spearman correlation coefficient permutation test[J]. Communications in Statistics-Theory and Methods, 2024, 53(6): 2141-2153.

[5] Abbasimehr H, Shabani M, Yousefi M. An optimized model using LSTM network for demand forecasting[J]. Computers & industrial engineering, 2020, 143: 106435.

[6] Meier P, Flepp R, Ruedisser M, et al. Separating psychological momentum from strategic momentum: Evidence from men's professional tennis[J]. Journal of economic psychology, 2020, 78: 102269.

[7] Bejar M P, Shigeno T C, Larsen L K, et al. The state of diversity in the Association for Applied Sport Psychology: Gaining momentum or still swimming upstream?[J]. Journal of Applied Sport Psychology, 2022, 34(6): 1372-1386.

[8] Freitas T T, Pereira L A, Alcaraz P E, et al. Change-of-direction ability, linear sprint speed, and sprint momentum in elite female athletes: differences between three different team sports[J]. Journal of strength and conditioning research, 2022, 36(1): 262-267.

[9] Henriksen K, Hansen J, Larsen C H. Mindfulness and acceptance in sport: How to help athletes perform and thrive under pressure[M]. Routledge, 2019.

[10] Fernandes R, Bishop C, Turner A N, et al. Train the engine or the brakes? Influence of momentum on the change of direction deficit[J]. International Journal of Sports Physiology and Performance, 2020, 16(1): 90-96.