Research on trend prediction and factor analysis of tennis potential energy

*Zhaoqing Zhu, Chenxuan Song, Yuantong Kong, Dongping Sheng**
Changzhou Institute of Technology, Changzhou, China

ABSTRACT

This paper proposed a model to capture score changes and potential energy flows in tennis matches, which could be used to explore the indicators that can predict changes in the direction of matches. The study verified the feasibility of potential energy and provide data-based strategy recommendations for coaches and players to help them better understand and exploit potential energy changes in matches. By model establishing and analysis, there are several conclusions could be obtained. First, points advantage and games won have a significant impact on player 2's performance, but compared with player 1, winning points have a slightly greater impact on player 2, which may reflect that Player 2 takes a more active offensive strategy in the game. Second, increasing the weight of serve advantage can enhance the potential energy of players to a certain extent, although the enhancement effect is relatively small. Sensitivity analysis further revealed the sensitivity of potential energy to the change of service advantage weight. Finally, by calculating the change rate of potential energy, it could be observed that, with the increase of weight, the sensitivity showed a slight fluctuation trend, which was amplified to the range of 0 to 1, and the potential energy was less sensitive to the change of the serve advantage weight.

Keywords: Potential Energy; Score Model; Critic Method, Neural Network Regression

1 INTRODUCTION

This study aims to capture score changes and potential energy flows in tennis matches by developing a model, as well as to explore whether there are indicators that can predict changes in the direction of matches [1]. Through the analysis of match-specific data from Wimbledon 2023, this study attempts to verify the existence of potential energy and provide data-based strategy recommendations for coaches and players to help them better understand and exploit potential energy changes in matches.

2 MATHEMATICAL MODEL

2.1 Model of Tennis Potential Energy Evaluation

The core concept of potential energy is to quantitatively analyze the scoring advantage of a player in a specific time window, and then evaluate the degree of performance advantage over the opponent. To this end, we provide a potential energy calculation model that comprehensively considers multiple match statistical indicators, together to more accurately capture and quantify player performance [2]. Several potential energy indexes are defined as follows.

(1) Score advantage: Scoring advantage is measured by comparing the difference in scoring between two players in a specific time window. It directly reflects the intensity of a player's performance at a particular stage of the game. Which could be defined as:

$$Momentum_{Advantage}^{Points} = Points_{Player1}^{Win} - Points_{Player2}^{Win} \tag{1}$$

$$Momentum_{Advantage}^{Serving} = Serve_{Wins} \times Serve_{Weight}^{Advantage} \tag{2}$$

(2) Service Advantage: In tennis, the server usually has a big advantage. Therefore, considering the fraction of points won by the server and its break ability, that is, winning its own service game, is crucial for the evaluation of potential energy [3].

(3) Break point: The number of break points won is an important measure of a player's ability to win a match on the opponent's serve, and defined as:

$$Momentum_{Win}^{BreakPoints} = BreakPoints_{Player1}^{Win} - BreakPoints_{Player2}^{Win} \tag{3}$$

(4) Unforced errors: Fewer unforced closes are generally seen as better performance and higher potential, and defined as：

$$Momentum_{Errors}^{Unforced} = -Unforced_{Player}^{Errors} \tag{4}$$

(5) The number of games and sets won: The number of games and sets each player wins in the time window is one of the important measures of potential energy, and A defined as

$$Momentum_{Win}^{Games} = Games_{Player1}^{Win} - Games_{Player2}^{Win} \tag{5}$$

$$Momentum_{Win}^{Sets} = Sets_{Player1}^{Win} - Sets_{Player2}^{Win} \tag{6}$$

(6) The winning point: The number of winning points is an important measure of a player's aggression and initiative in the game. A high winning score not only shows a player's skill, but also his ability to take the initiative and take control of the game [4].

$$Momentum_{Winners} = Winners_{Player1} - Winners_{Play2} \tag{7}$$

The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between the two variables:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{8}$$

The above formula defines the overall correlation coefficient, which is often represented by the Greek lowercase letter symbol ρ . Estimate the covariance and standard deviation of the sample to obtain the Pearson correlation coefficient, which is commonly represented by the English lowercase letter γ .

$$r = \frac{\sum_i (x_i - x)(y_i - y)}{\sqrt{\sum_i (x_i - x)^2 \sum_i (y_i - y)^2}} \tag{9}$$

γ can also be estimated from the mean of the standard scores of the letter (X_i, Y_i) sample points, resulting in an expression equivalent to the above:

$$\gamma = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma_X}\right)\left(\frac{Y_i - \bar{Y}}{\sigma_Y}\right) \tag{10}$$

Where $\frac{X_i - \bar{X}}{\sigma_X}$, $\bar{X}$ and σ_X are the standard score, sample mean and sample standard deviation of sample X_i respectively.

2.2 Model of potential energy index weight based on CRITIC method

The factors of the game environment and the opponent also need to be taken into account to adapt to different game situations. Weight analysis allows data normalization to ensure that the contribution of each indicator is fair and appropriate [5]. The CRITIC method to determine the weight and calculate the potential energy is an objective method and the most appropriate choice for this problem, and its model theory is as follows:

For the CRITIC method, when the standard deviation is constant, the smaller the conflict between indicators, the smaller the weight. The greater the conflict, the greater the weight; In addition, when the degree of positive correlation between the two indicators is greater (the correlation coefficient is closer to 1).

In the form of standard deviation, Sj represents the standard deviation of the JTH index.

$$\bar{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{11}$$

$$S_j = \sqrt{\frac{\sum_{i=1}^{n} \left(x_{ij} - \bar{x_j}\right)^2}{n-1}} \tag{12}$$

It is expressed in the form of correlation coefficient, and *r*、*i*、*j* represents the correlation coefficient between the evaluation index *i* and *j*.

$$R_j = \sum_{i=1}^{p} \left(p - r_{ij}\right) \tag{13}$$

The correlation coefficient is used to represent the correlation among indicators. The stronger the correlation is with other indicators, the smaller the conflict between the indicator and other indicators [6]. The larger the C_j is, the greater the role of the JTH evaluation index in the whole evaluation index system, so more weight should be assigned to it.

$$C_j = S_j \sum_{i=1}^{p} \left(1 - r_{ij}\right) = S_j \times R_j \tag{14}$$

So the objective weight of the JTH index is W j.

$$W_j = \frac{C_j}{\sum_{j=1}^{p} C_j} \tag{15}$$

2.3 Potential Energy Prediction and Competition Strategy Based on PSO-bq

In order to develop new game strategies for players to prepare for different opponents, the model established in section 2.2 can be used to simulate and predict their battle process. By establishing a goal planning function, a player can be set up to maximize potential energy by adjusting variables that the player can actively adjust [7]. Then, relevant adjustment operations can be summarized to rationally allocate energy and rhythm. The specific method is as follows:

1) Model simulation: Using the established potential trend change model to simulate the course of Carlos Alcaraz's race under different strategies.

2) Goal programming function: Build a goal programming function designed to maximize Alcaraz's race potential energy. To maximize the output of race E for the model, which is a function of the following variations:

$$\begin{aligned} E = f(&p1_ace, p1_winner, p1_double_fault, \\ &p1_unf_err, p1_net_pt, p1_net_p) \end{aligned} \tag{16}$$

2.4 SHAP Model Establishment

SHAP uses an idea called "force orientation," which uses the influence of each feature in the model to interpret the results of the model. By calculating the valid value of each feature, the model's decision is inferred, and a meaningful interpretation result is obtained. This method can generate a graph according to the input data provided by the user, and show the behavior of the model through the graph [8]. The advantage of SHAP compared to other methods of interpreting models is that it can provide an effective contribution to each feature and explain the decision-making process of the model.

Calculate the average absolute SHAP value for each feature: For feature j, its global importance can be estimated by calculating the average absolute value of the SHAP value for that feature in all samples:

$$importance(j) = \frac{1}{N}\sum_{i=1}^{N} \left|SHAP_{ij}\right| \tag{17}$$

3 MODEL ANALYSIS

3.1 Tennis Potential Energy Evaluation

3.1.1 Exploratory analysis of potential energy index

The Fig.1 shows the value of a correlation coefficient in the form of a heat map, and the size of the value is mainly represented by the color depth. This heat map shows the correlation analysis between the potential energy indicators of players 1. Each cell represents the correlation coefficient between the two indicators, which ranges from -1 to 1.

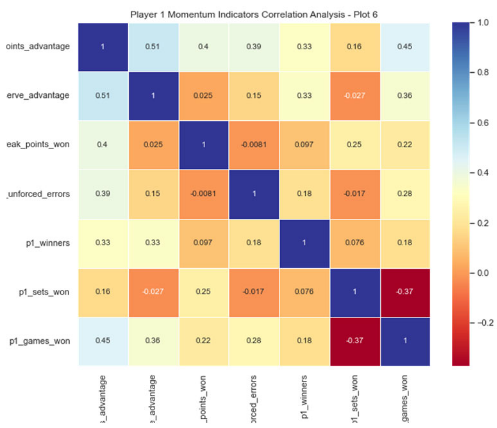

Figure 1: Momentum indicators correlation analysis of player 1

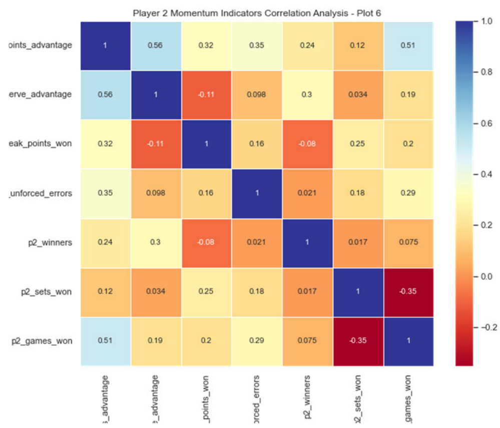

Figure 2: Momentum indicators correlation analysis of player 2

It can be found that point advantage has a positive correlation with winning points and service advantage, which indicates that the increase of point advantage is often related to the increase of winning points and service advantage. This is logical because strong serves and winning points can lead directly to an increase in points.

The relatively low correlation between the number of break points won and other metrics may be due to the fact that break success is associated with specific situations in the match that may not occur often or have a strong direct correlation with other metrics [9].

The Fig.2 shows the correlation break between the players 2 potential energy indicators. Similar to the correlation analysis for player 1, this heat map reveals correlations between different game metrics to provide insight into player performance. You can see:

The correlation of scoring advantage with other indicators, especially the positive correlation with serve advantage and winning points, reflects how Player 2's advantage in serving and active attack translates into a scoring advantage in the match.

By comparing the correlation analysis between player 1 and player 2, the analysis reveals the unique performance characteristics of each player, which provides the basis for customized training and game strategies. Understanding these correlations is critical to optimizing a player's match preparation and improving their competitiveness.

3.1.2 Potential energy index weight analysis by CRITIC method

Fig.3 shows the weight calculation results of player 1 based on CRITIC method. According to the results, the weight of each indicator is analyzed. The results show that, The weight of p1_points_advantage is 30.909%, that of p1_serve_advantage is 2.23%, that of p1_break_points_won is 5.182%, that of p1_unforced_errors is 15.078%, and that of p The weight of 1_winners is 17.03%, the weight of p1_sets_won is 3.835%, and the weight of p1_games_won is 25.737%. The maximum indicator weight is p1_points_advantage (30.909%). The minimum value is p1_serve_advantage (2.23%), as shown in Fig.3.

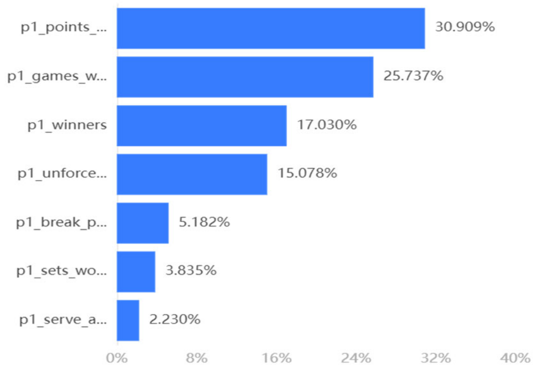

Figure 3: Potential energy weight distribution bar graph player 1

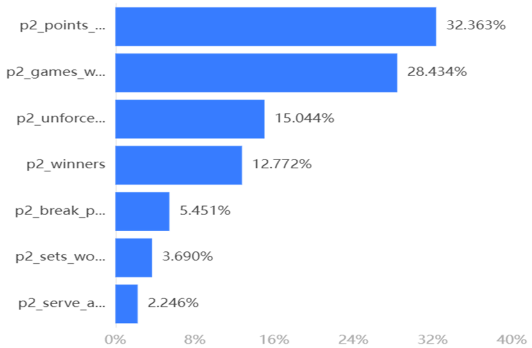

Figure 4: Potential energy weight analysis player 2

It can be seen that the index weight of player 1 is concentrated in points advantage and games won, which reflects that the number of sets successfully won in the match has the greatest impact on the overall performance of the player.

Fig.4 shows the weight calculation results of player 2 based on CRITIC method. According to the results, the weight of each indicator is analyzed. The results show that, The weight of p2_points_advantage is 32.363%, that of p2_serve_advantage is 2.246%, that of p2_break_points_won is 5.451%, that of p2_unforced_errors is 15.044%, and that of P2_points_errors is 15.044% [10]. The weight of p2_winners was 12.772%, that of p2_sets_won was 3.69%, and that of p2_games_won was 28.434%. The maximum indicator weight was p2_points_advantage (32.363%). The minimum value is p2_serve_advantage (2.246%).

It can be seen that the index weight of player 2 also shows that points advantage and games won have a significant impact on player 2's performance, but compared with player 1, winning points have a slightly greater impact on player 2, which may reflect that Player 2 takes a more active offensive strategy in the game.

3.1.3 Potential energy composite score

Based on the weight analysis of 2.2, the comprehensive scores of players 1 and 2 can be calculated by weighted potential energy index. This comprehensive scoring method based on weights provides a comprehensive perspective for evaluating and comparing player's performance, helping coaches and players better understand which aspects of performance are most critical to their overall success, and adjust their training and competition strategies accordingly [11]. The line chart shows how the combined potential score of Player 1 and Player 2 has changed over a series of matches, as shown in Fig.5. By comparing the scoring routes of the two players, we can observe the fluctuations in their performance from game to game and the overall performance trend.

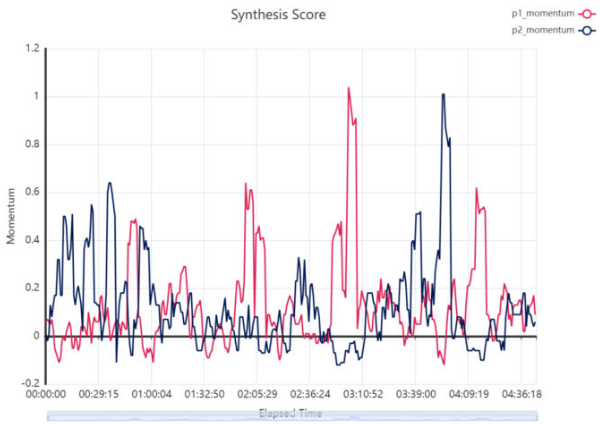

Figure 5: Potential energy composite score

3.2 Influence Analysis of Player's Potential Energy Change Factors Based on SHAP

(1) Single sample interpretation

The Fig.6 shows the feature impact analysis for the first sample, quantifying the contribution of each feature of the first sample to the model prediction relative to the baseline prediction.

Figure 6: The first sample feature impact level

The left side lists several features that affect the model's predictions, and the right side is a horizontal bar chart, as shown in Fig.7. The length of each bar represents the feature. The degree of influence on the model output, the direction of the bar indicates how the feature affects the output. To the left indicates that the feature will decrease the model output value, to the right indicates that it will increase the model output value. Color is often used to distinguish the positive and negative effects of features on the output, where red represents lower and blue represents higher.

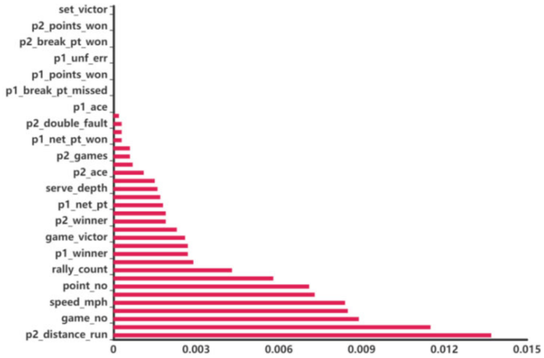

Figure 7: SHAP value feature important analysis diagram

The most important features: p2_distance_run(0.01368) and p1_distance_run(0.011499) are the two important features, indicating that the distance the player moves on the field has a significant impact on the prediction results. This may reflect the importance of luck in the game, such as endurance and mobility.

Important race characteristics: game_no (0.008869), p2_score(0.008511) and speed_mph (0.008425) are also important features. This shows that the score of the two sides and the speed of the serve have a great influence on the prediction of the model. Serve and score: serve_width (0.007324) and point_no (0.007063) show the width of the serve and the points in the match

4 CONCLUSION

This paper proposed a model to capture score changes and potential energy flows in tennis matches, which could be used to explore the indicators that can predict changes in the direction of matches, and there are several conclusions could be obtained as:

(1) Points advantage and games won have a significant impact on player 2's performance, but compared with player 1, winning points have a slightly greater impact on player 2, which may reflect that Player 2 takes a more active offensive strategy in the game.

(2) Increasing the weight of serve advantage can enhance the potential energy of players to a certain extent, although the enhancement effect is relatively small. Sensitivity analysis further revealed the sensitivity of potential energy to the change of service advantage weight.

(3) By calculating the change rate of potential energy, it could be observed that, with the increase of weight, the sensitivity showed a slight fluctuation trend, which was amplified to the range of 0-1, and the potential energy was less sensitive to the change of the serve advantage weight.

5 ACKNOWLEDGEMENTS

This work was Supported by Special Research Project on Teaching Reform (Grant No. 30120300100-23-yb-jgkt03) and Changzhou Science and Technology Planning Project

(CE20225067).

REFERENCES

[1] Peng, R., & Li, Z. (2024). MLFEF: Machine Learning Fusion Model with Empirical Formula to Explore the Momentum in Competitive Sports. *arXiv preprint arXiv:2402.12149*.

[2] Li, Y., Kim, K., & Ding, Y. (2021). Early warning system of tennis sports injury risk based on mobile computing. *Mobile Information Systems*, *2021*, 1-10.

[3] Bozděch, M., & Zháněl, J. (2023). Analyzing game statistics and career trajectories of female elite junior tennis players: A machine learning approach. *Plos one*, *18*(11), e0295075.

[4] AlShami, A., Boult, T., & Kalita, J. (2023). Pose2Trajectory: using transformers on body pose to predict tennis player's trajectory. *Journal of Visual Communication and Image Representation*, *97*, 103954.

[5] Lennartz, J., Groll, A., & van der Wurp, H. (2021). Predicting Table Tennis Tournaments: A comparison of statistical modelling techniques. *International Journal of Racket Sports Science*, *3*(2), 39-48.

[6] Xie, Y., Bai, B., & Zhao, Y. (2022). Variation Factors and Dynamic Modeling Analysis of Tennis Players' Competitive Ability Based on Big Data Mining Algorithm. *Journal of Sensors*, *2022*.

[7] Baiget, E., Corbi, F., & López, J. (2023). Influence of anthropometric, ball impact and landing location parameters on serve velocity in elite tennis competition. *Biology of Sport*, *40*(1), 273-281.

[8] Blagus, R., Hadzic, V., Fernandez Garcia, A. I., Leskosek, B., Narang, B. J., & Filipcic, A. (2023). COVID-19 and Changes in the Model of Physical Fitness and Body Composition of Young Tennis Players. *Applied Sciences*, *13*(18), 10015.

[9] Buhamra, N., Groll, A., & Brunner, S. (2024). Modeling and prediction of tennis matches at Grand Slam tournaments. *Journal of Sports Analytics*, *10*(1), 17-33.

[10] Zhou, H., Zhang, Y., Agarwal, A., Arnold, G., & Wang, W. (2024). A preliminary study on analysis of lower limb energy during walking in the patients with knee replacement. *Heliyon*, *10*(6).

[11] Sun, Y., Wang, Y., Qu, C., Liu, Y., Nie, Y., Sun, Z., & Song, B. (2024). A novel ball-racket rebound model for table tennis robot based on continuous contact force. *IEEE Transactions on Instrumentation and Measurement*.