

Research on Momentum Analysis of Tennis Match Based on Machine Learning

Yijie Hao^{*, #}, Jiachen Chang[#]

School of Science, China University of Geosciences (Beijing), Beijing, China, 100083

*Corresponding author: 548654861@qq.com

#These authors contributed equally

Abstract. This study aims to quantify momentum changes in tennis matches by building a machine learning model and investigate their impact on match outcomes. Firstly, a momentum model is established to quantify the momentum value of players by defining basic points, break points, guaranteed points and continuous points. Then, the momentum curve of players is calculated using real match data, and its high match with the match trend is verified. Then, the correlation between momentum value and match result is tested by statistical test method. The results show that there is a significant correlation between momentum value and match victory. Further, the displacement test verifies that the change of momentum value is not random, but reflects the actual trend of the game. This study proves that momentum plays an important role in tennis matches, and its change has a significant impact on match results, which provides an important reference for further research and application of momentum.

Keywords: Machine learning; Momentum model; Statistical testing; Replacement test; Tennis match.

1. Introduction

Tennis is played in both singles and doubles formats. Players use tennis rackets to hit the ball over the net and onto the other team's court. It goes back and forth like this until one player hits the ball out of bounds or doesn't receive the ball. Before an official match is played, it needs to be determined who will serve first for the match^[1-3]. Throughout the match, both players take turns serving. The serving player should stand in the area behind the end line, between the center point and the assumed extension of the sideline before serving. The ball served shall cross over the net and land in the opposing team's serving area on the opposite corner. The first point of each set is recorded as 15, the second as 30, and the next as 40. In each set, at least two more points must be scored than the opponent to end the set. Within the new rules of doubles, matches can be won with just one more point^[4-6].

In competitive sports, there is an interesting phenomenon, when a person has a certain momentum in the game, he will have a greater probability of winning the game, this momentum is not the momentum of physics, but the athlete in the sport has the power, this power consists of a lot of aspects, such as the enthusiasm of the audience, the player's state of mind and so on. This power is like a booster, which can make the player's play go to the next level, and we call this momentum^[7].

The main research content of this paper is to use machine learning model to quantify the momentum change in tennis matches and study its impact on match results.

2. Model building and solving

2.1. Construction of momentum analysis model of tennis match based on machine learning

2.1.1 Problem analysis

A mathematical model is needed to determine an indicator. This metric is enough to change the direction of the game. To solve this problem, let's define this metric as "momentum." Momentum is very important in all kinds of competitive events around the world, but especially in sports. It can be thought of as a psychological force that can have a significant impact on a player's performance in a

game. While momentum itself is very abstract, we can visualize this phenomenon through key metrics in the game. For example, scoring consecutive runs, bowling success, breakthrough success, and so on. At the same time, momentum is also related to the confidence of the players and the reaction of the crowd. The solution is shown in Figure 1.

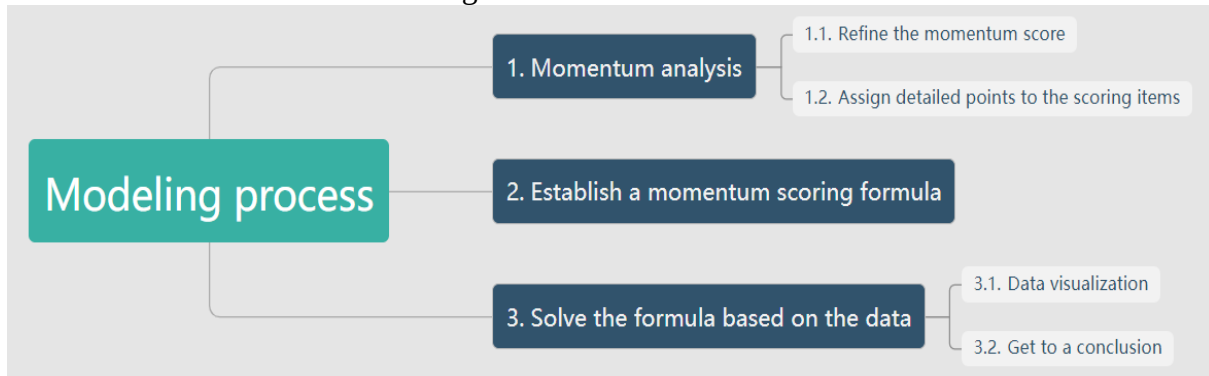


Figure 1. Mind map

2.1.2 Establishment of momentum model

(1) Momentum analysis

In order to quantify the momentum, and at the same time ensure the accuracy of the quantification, we need to define some key indicators. Each indicator corresponds to a certain momentum score:

- Base score: For each point won, the momentum score is increased by one point.
- Break score: Winning a point on the opponent's serve adds an extra four points to the momentum score, reflecting the difficulty of breaking score and the importance of the impact on the opponent's mindset.
- Guaranteed score: Gaining a point in the middle of her serve and adding an extra two points to her momentum score reflects the value of a successful hold.
- Continuous score: If three or more consecutive points are scored in a game, an additional three points are added to the Momentum Score, reflecting the important impact of consecutive points on momentum.

Based on the rules of tennis and the unique characteristics of the game, we have created the above indicators. Each indicator has been set up taking into account psychological factors as well as match strategy.

(2) Momentum score calculation formula

For a player, his total momentum score can be calculated using the following formula:

For every scoring point:

$$S_{\text{momentumi}} = S_{\text{base}} + S_{\text{break}} + S_{\text{hold}} + S_{\text{continuous}} \quad (1)$$

Momentum scores are cumulative, so a player's Momentum in a match is the sum of the Momentum scores from all scoring points:

$$S_{\text{total}} = \sum_{i=1}^n S_{\text{momentumi}} \quad (2)$$

where n is the total number of points scored during the game.

2.1.3 The solution of momentum model

Taking the numbered 1701 match (2023 Wimbledon Championships Men's Singles Semi-Final), in which Carlos Alcaraz defeated Novak Djokovic, as an example, our calculations can ultimately result in a momentum score change curve as shown in Figure 2.

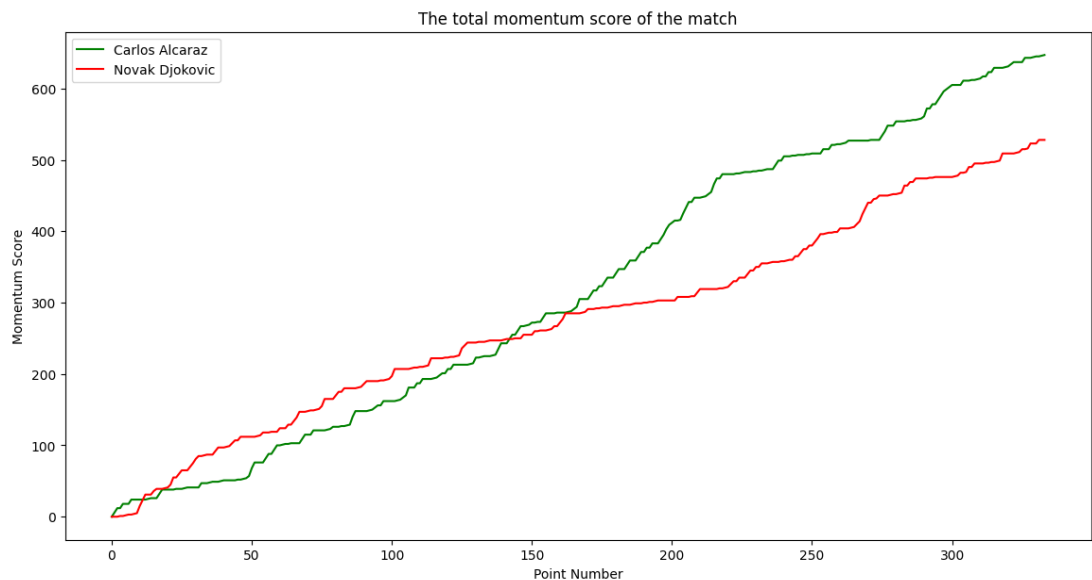


Figure 2. Momentum change curve for Carlos Alcaraz vs Novak Djokovic matches

According to Figure 2, we can see that the momentum change closely matches what happened between the two in the actual match. Novak Djokovic went from having the momentum in the beginning to being overtaken in the middle, to falling behind in the momentum and eventually losing the match.

By continuing to compare the rest of the matches, we can conclude that the model is indeed able to recognize which player performs better in a match and which player wins the final match. And the momentum change curve has a high match with the actual game situation, which can show the game flow well.

2.2. Research on the randomness of momentum change in tennis matches

2.2.1 Problem analysis

For this question, we want to assess whether momentum, as we define it, is a random variable. We target our analysis to the momentum model presented in the previous question. The goal of our analysis is to determine whether shifts in momentum over the course of a game are statistically significant and whether they indicate that momentum is not an indicator of randomness. To address this question, we decided to use two approaches:

(a)The extent to which momentum scores correlate with the final outcome of the match: we analyze the relationship between momentum and the outcome of the match. If the relationship between momentum scores and match wins is very strong, this may indicate that momentum has some degree of influence on the outcome of the match.

(b)Statistical Tests: In order to more rigorously assess the randomness of momentum, we compare the difference between actual game data and randomly simulated game data by applying statistical tests, which have the advantage of not relying on the specific distribution of the data. For example, we can use a permutation test to assess whether momentum scores vary more than the level of variation expected in a randomized situation (Figure 3).

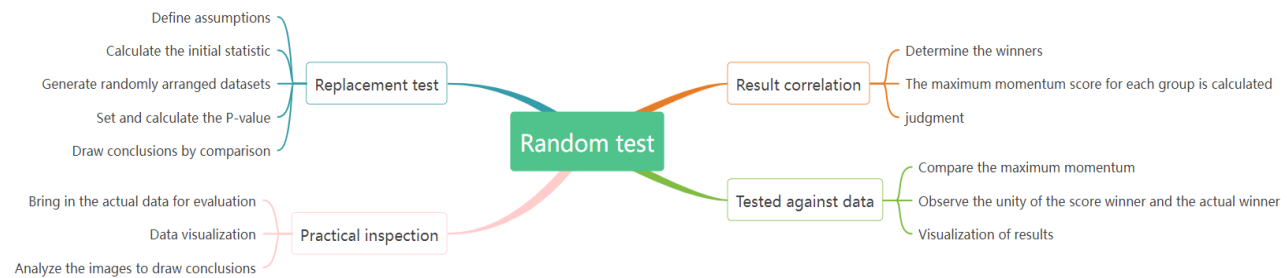


Figure 3. Second question mind map

2.2.2 Random test—Based on the correlation between momentum scores and game results

(1) Modeling Steps and Data Handling

(a)First of all, we have to determine the final winner of the game based on the data given to the person who scored the last point. Also we assume that the side with the higher momentum at the last scoring point is the final winner of the game.

(b)Second, for each game, we calculate the maximum momentum score achieved by both players in the game and use the group by and agg functions to group and calculate the maximum value for each group.

(c)Finally, we will integrate the Maximum Momentum Score with the data on the winner of each match, compare whether the person who reached maximum momentum in each match was the same person as the eventual match winner.

(2) Reality check

We assume that in the actual game, the momentum scores of the two players at the ath scoring point are $S_{Player1}^{(a)}$ as well as $S_{Player2}^{(a)}$, and the maximum momentum scores of the two players in the game are:

$$S_{Player1}^{max} = \max_a (S_{Player1}^{(a)}) \quad (3)$$

$$S_{Player2}^{max} = \max_a (S_{Player2}^{(a)}) \quad (4)$$

And the eventual winner of the game can be determined by comparing the momentum points scored by both teams on their last scoring point:

$$Winner = \begin{cases} Player1, & \text{if } S_{Player1}^{(last)} > S_{Player2}^{(last)} \\ Player2, & \text{if } S_{Player1}^{(last)} < S_{Player2}^{(last)} \end{cases} \quad (5)$$

where $S_{Player1}^{(last)}$ and $S_{Player2}^{(last)}$ are the momentum scores of the last scoring point of both players in the middle of the game, respectively.

By analyzing the relationship between the maximum momentum score and the final winner of the game, we can assess whether momentum has a relatively large impact on the final outcome. After integrating the calculations, we can obtain a complete table of the relationship between the momentum score and the final winner of the game, some of which are shown in the figure 4.

set_no	game_no	winner	max_moment um_p1	max_moment um_p2	max_moment um_winner	momentum_mat ches_winner
1	1	player1	24	5	player1	TRUE
1	2	player2	26	38	player2	TRUE
1	3	player2	38	47	player2	TRUE
1	4	player2	41	82	player2	TRUE
1	5	player2	47	93	player2	TRUE
1	6	player2	51	105	player2	TRUE
1	7	player2	51	113	player2	TRUE
2	1	player2	57	119	player2	TRUE
2	2	player2	100	125	player2	TRUE
2	3	player2	103	157	player2	TRUE

Figure 4. A partial table of complete momentum scores in relation to the eventual winner of the match

After comparing the data in the complete table, we can see that the winner of the maximum momentum score is indeed the winner of the competition. We have labeled the maximum momentum

score with a blue dot when it coincides with the winner, and with a red dot when it does not. We can visualize this in Figure 5.

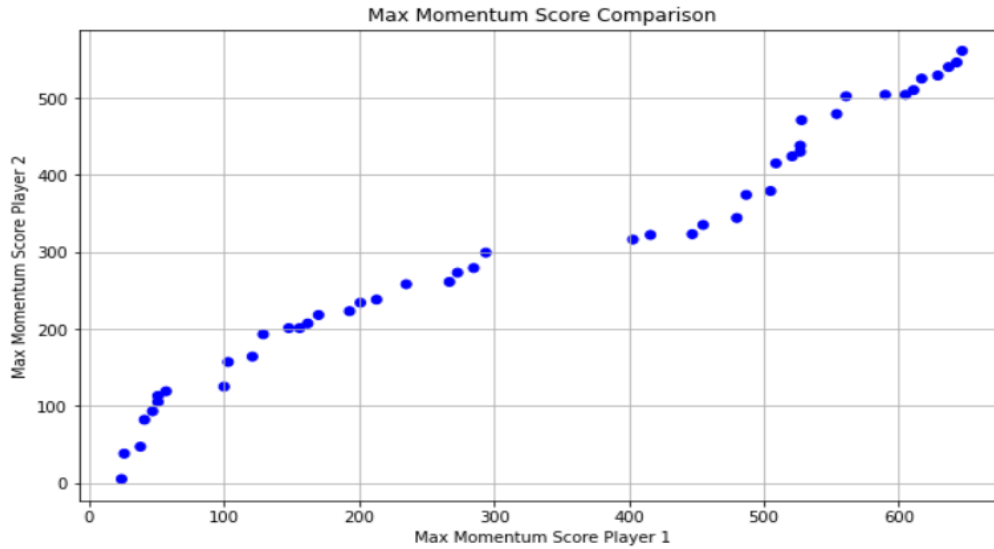


Figure 5. Correlation table between maximum momentum scorers and final winners of the game

We can see by looking at the chart that all the points are blue and there are no red dot, side by side verifying that momentum scoring is very closely linked to the final match victory.

2.2.3 Random test——based on the permutation test

(1) Modeling Steps and Data Handling

(a) First, we need to define the original hypothesis (H0) as well as the alternative hypothesis (H1):

▼ The original hypothesis(H0): All observations come from the same distribution.

▼ Alternative hypothesis(H1): Some or all observations come from different distributions.

(b)Next, we need to compute the most initial statistic. We choose an appropriate statistic T to measure the difference between the sample groups. This difference can be the difference in means or some other measure appropriate to the problem. For example, if we assume that there are two distinct sample groups A and B, then our original statistic can be the difference in the means of the two samples, which can be expressed as:

$$T_{\text{average}} = \bar{A} - \bar{B} \quad (6)$$

(c)After that, generate multiple randomly-arranged datasets by randomly swapping group labels of sample points, These randomized permutations represent the permutations of data that might be observed if the original hypothesis were true. and compute the same statistic as the original statistic for all data sets obtained T_{primvel} .

(d)Then, we set the P-value, which is the position of the observed statistic T_{average} among all the statistics generated by the randomized permutations, which itself means the probability of the observed statistic or a more extreme statistic appearing if the original hypothesis is true, and we assume that the number of all the permutations is E, which is expressed by the formula:

$$P = \frac{\text{Number} (T_{\text{primvel}} > T_{\text{average}})}{E} \quad (7)$$

(e) Ultimately, a comparison was made based on the resulting p-value and our assumed significance level of $\alpha = 0.05$. If $p \leq \alpha$, the original hypothesis is rejected, indicating that the difference between the sample groups is statistically significant. Conversely, if $p \geq \alpha$, the original hypothesis is accepted, indicating that the difference between the sample groups is not statistically significant^[8-10].

(2) Reality check

We use python to solve the final result of the replacement test obtained is:

★Raw statistic:1.0, Among the raw data, this represents a 100% probability that the maximum momentum scorer is the winner of the match, as shown in Figure 6.

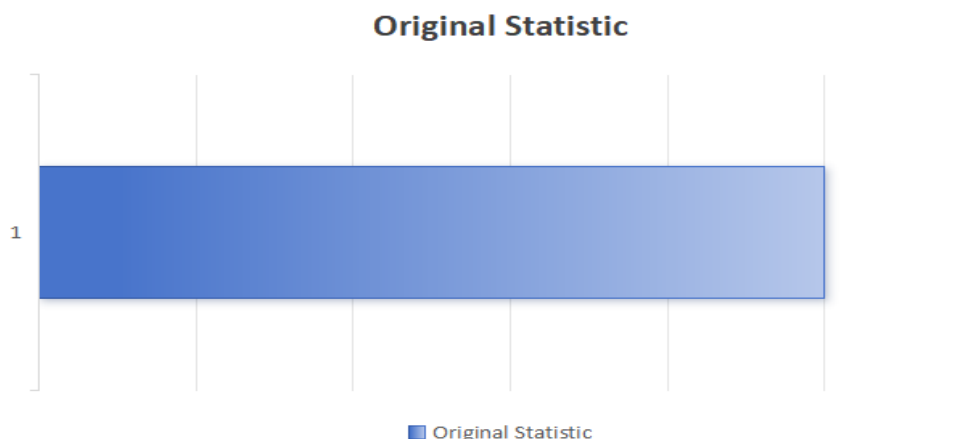


Figure 6. Raw statistics

★P-value:0.0, This demonstrates that in all of the permutation tests, for any of the randomly permuted datasets, the proportion of its maximum momentum scorers who agreed with the winners did not reach the proportion observed in the original data. In our case, the P-value is close to zero, suggesting that it is almost impossible to observe a ratio of maximum momentum scores consistent with winners that is as high as the raw data in the case of randomized permutations. This suggests that the raw observations are unlikely to have occurred by chance and that the consistency between momentum scores and match winners is statistically significant.

We present the histograms and initial statistics of the permutation test results combined, as in Figure 7.

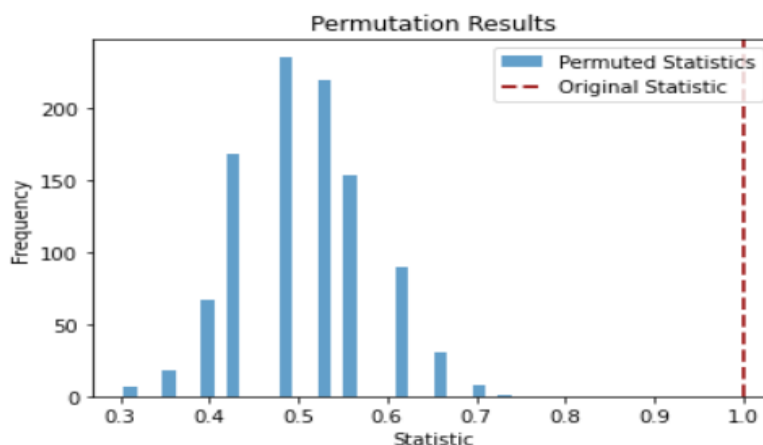


Figure 7. Results of the replacement test and the initial statistics combined

By looking at the histogram, we can analyze that this histogram represents the distribution of the statistics calculated under all the random permutations generated during the permutation test. Each blue bar represents the number of times the statistic falls within a specific range across all permutations. This distribution reflects the distribution of the statistic that we would expect to see under conditions of complete randomization.

In by looking at the dotted line, we can analyze that this dotted line represents the value of the raw statistic calculated from the actual data. In the context of momentum scoring, this could be the proportion of the raw data where the person with the largest momentum score is the same person as the eventual winner.

Also, the raw statistic is located at the far end of the histogram. This means that it is very rare to observe a statistic like the one in the raw data under randomized conditions. We can assume that the raw statistic represents a phenomenon that is unlikely to have occurred by chance, but rather a statistically significant phenomenon.

In summary, we can argue that momentum does play a role in more than tennis matches and is not completely random. This suggests that momentum shifts in matches and the winning streaks of one player are unlikely to be merely random events, but are more related to the actual circumstances of the match.

3. Conclusions

By constructing a machine learning model, this paper successfully quantifies the momentum change in tennis matches and confirms its important impact on the match outcome. The study found that the change in momentum was highly consistent with the actual trend of the game, and there was a significant correlation between it and the victory of the game. At the same time, the results of both statistical and displacement tests show that the change in momentum is not a random event, but reflects the actual trend of the game. Therefore, this study provides strong evidence for the important role of momentum in tennis matches, and lays a foundation for future research and application of momentum. Further research could be extended to other sports to test the universality of momentum changes and explore the specific mechanisms by which they affect game outcomes.

References

- [1] Chen Dongyu, Chen Hua, Fan Limin, etc. Random test strategy based on deep learning [J]. Journal of Communications, 2023,44 (06): 23-33.
- [2] Huang Jia, Liu Xiangyu. Research on the methods of classification and prediction of medical devices based on machine learning [J]. Scientific and technological Innovation and application, 2024, 14(05): 28-31.
- [3] Zhou Yifu, Tan Chunhui, Jiang Ting, etc. High-quality patent achievement screening study based on machine learning classification algorithm [J]. Modern Intelligence, 2024,44 (02): 81-91.
- [4] Zhang Guijie, Wang Xiaocan, Xing Weikang, etc. Psychological evaluation model study based on decision tree classification algorithm [J]. Journal of Jilin Normal University (Natural Science Edition), 2023, 44(04): 123-130.
- [5] Zhang Qian, Geng Xiaozhong, Yue Mengzhe, etc. Classification of motor imagination EEG signals based on cospatial patterns and decision tree support vector machine [J]. Software Engineering, 2023, 26(11): 16-19.
- [6] Wang Wei, Zhou Shu. Analysis of the game characteristics of Professional Tennis Competition [C] / Hubei Sports Science Association. Proceedings of the first Hubei Provincial Sports Science Conference (Volume 1). Hubei Normal University, 2023: 3.
- [7] Yang Jing. Research on tennis special physical training [C] / China Intelligent Engineering Society, China Bdy Association, Guangdong Physical Fitness Association. Proceedings of the 10th China Physical Fitness Training Science Conference (Part 1). Guizhou Medical University, 2023: 5.
- [8] Connor W C. Understanding New Machine Learning Architectures: Practical Generative Artificial Intelligence for Anesthesiologists. [J]. Anesthesiology, 2024, 140(3): 599-609.
- [9] Zhang S, Zhou X, Chen Y, et al. Soil organic carbon fractions in China: Spatial distribution, drivers, and future changes [J]. Science of the Total Environment, 2024, 919170890.
- [10] Wei F, Kouro T, Nakamura Y, et al. Enhancing Mass spectrometry-based tumor immunopeptide identification: machine learning filter leveraging HLA binding affinity, aliphatic index and retention time deviation. [J]. Computational and structural biotechnology journal, 2024, 23859-869.