Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
C		2422034

Who Can Turn the Corner? Who is the Winner?

Summary

Tennis sport has always been deeply loved by people. When people watch a tennis game, they will find that the advantage of the game is not always biased towards a single player, that is, the game is volatile. And the swings that happen in the dominant players are often boiled down to momentum. We analyze the relationship between momentum and game swings by building reasonable models, and predict swings in future games.

Several models are established: Model I: **Define Momentum Based on the Elo System**; Model II: **LSTM Model**; Model III: **Grey Relation Analysis** Model[1]cite; Model IV: **A predictive model based on Quadratic Discriminant Analysis**.

For Model I: We have improved on the basis of the traditional Elo. In addition to the cumulative scores of the two players, we also consider whether the player is the server, the mistakes and physical consumption of the players and the consecutive number of games won by the players, and define the formula of momentum through optimization modeling. And use momentum as an important indicator to predict game swings and performance of scoring players.

For model II: The performance of the two players in the competition is not easy to measure directly. We customize an indicator to measure the performance of two players for the number of sets and games won, and the total score. The LSTM model predicts the one who performs better and how much better by the two players on the next scoring point, and comparing it with the true value shows that our model predicted better.

For model III: Grey Relation Analysis can quantitatively describe and analyze the development and change trend of a system, reflecting the degree of correlation between indicators. We chose the Grey Relation Analysis Model to analyze the correlation between all indicators and competition swings, and found that the correlation degree of momentum and competition swings was **0.88**, with strong correlation. The most relevant factors are **the number of consecutive wins** and **error situations**, and the correlation degree reaches **0.89**.

For model IV: We compared the accuracy of multiple algorithms for training our predictive model and ultimately selected Quadratic Discriminant Analysis to train a model for predicting turning points in the game.

We solve the model and evaluate the following conclusions. There is a strong correlation between momentum and competition swings; the most important factors are **the number of consecutive wins** and **error situations**; players should positively reflect the predicted turning points and adjust their physical and mental state.

Finally, we wrote a **memo** for the tennis coaches with our models, results, and advice.

Keywords: Momentum; LSTM; Grey Relation Analysis; Quadratic Discriminant Analysis; Elo Rating

Contents

1 Introduction

1.1 Problem Background

Novak Djokovic is a Serbian professional tennis player who is ranked number one in the world in men's singles by the Association of Tennis Professionals (ATP). He has achieved great success and won numerous Grand Slam titles, including Wimbledon Gentlemen's singles. However, in the 2023 Wimbledon Gentlemen's final, Novak Djokovic was defeated by the 20-year-old Spanish tennis player Carlos Alcaraz, breaking his previous streak of success.

Looking back at the match, Novak Djokovic won the first set with a score of 6-1, seemingly on track to easily win the match. However, the tide turned in the second set, and Novak Djokovic lost to Alcaraz with a score of 6-7. Alcaraz then took control of the third set and won it with a score of 6-1. But the situation changed again in the fourth set, as Novak Djokovic regained control and won it with a score of 6-3. In the early stages of the final set, Novak Djokovic still had the upper hand and led by four points. However, there was a turnaround in the game, and Alcaraz ultimately won the set with a score of 6-4, securing the final victory(see in figure 1).

Novak Djokovic	6	6	1	6	4
Carlos Alcaraz	1	7	6	3	6

Figure 1: Score performance in five sets.

1.2 Restatement of the Problem

For the information content of the documents published by MCM, we analyzed and built a model to solve the following problems:

- Build a model that **predicts when each point appears**, and determine **which player performs better** at the point moment, and how much better, to predict which player is scoring the next point. And to describe the **matching process** based on the visualization of this model. Finally, the model is used in many games.
- Find the factors that determine the turning point of a match, and develop a model to predict the turning point of a match. Based on the change of momentum and the developed model, give suggestions for different players in a match.
- Find some more relevant indicators from the previous analyzed momentum characteristics, and analyze when the dominant situation shifts from one player to another player. Build a model to predict the swing of momentum in the game and find the most relevant indicators, and give suggestions for different players for new matches.
- The model is evaluated in two aspects: the **effectiveness** of predicting tennis momentum swings and the **adaptability and universality** of the model to other

games such as Women's tennis matches.

- Make a memo on the role of momentum in the game and advise players on how to deal with the fluctuating momentum in the game.

1.3 Our Work

To avoid complicated description, intuitively reflect our work process, the flow chart is shown in Figure 2:

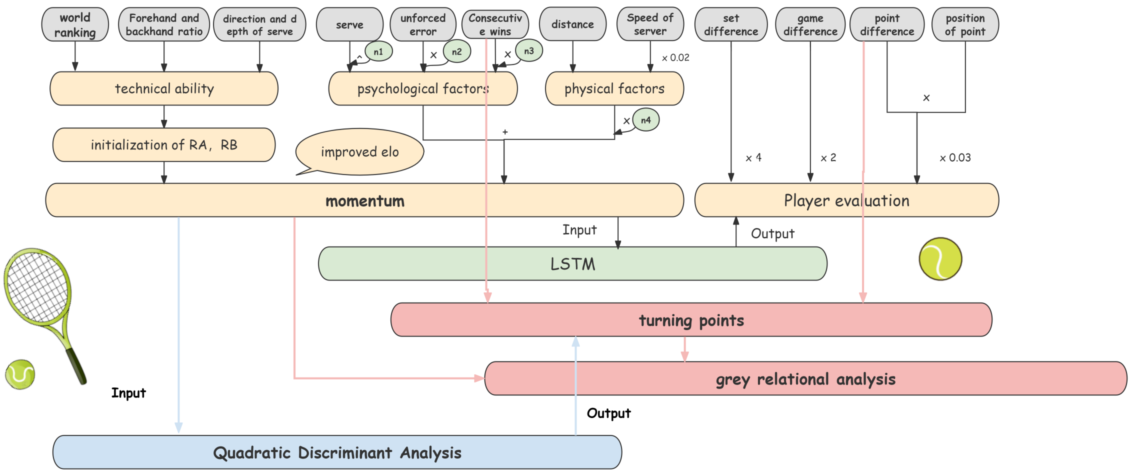


Figure 2: Flow Chart of Our Work

2 Assumptions and Explanations

- **The initial physical indicators of the players are the same, and the physical strength is the same.**
  ↪ When we define momentum, we use the physical exertion of the player as a measure. In order to exclude the interference of other factors, it is assumed that each player is in similar physical condition before the competition.
- **The player's condition is not disturbed by external factors.**
  ↪ We only take into account the physical and mental factors. It ignores the influence of external factors such as coach's guidance and audience during the match.
- **Ignore the luck of the players.**
  ↪ Luck can play a part in a game, such as a player's lucky shot. However, the luck factor is usually random and unpredictable, so it is not taken into account in this article.
- **Ignore the impact of different venues on players.**
  ↪ Different types of surfaces, such as grass, hard and clay courts, can have an impact on a player's performance and momentum. We assume that the players are highly professional and have strong adaptability to different competition venues.

3 Notations

Some important mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description
H_p	The scores of the two players
α_1 ,α_2	Player 1,2 serves
$\beta_1[k]$,$\beta_2[k]$	Player 1,2's error situation in the kth point
$W_1[k]$, $W_2[k]$	Player 1,2's physical exertion in the kth point
$\mathrm{Z}_1[k]$,$\mathrm{Z}_2[k]$	Player 1,2's winning streak in the kth point
$R_1[k]$,$R_2[k]$	The momentum of Player 1,2 in the kth point
$E_1[k]$,$E_2[k]$	The expected outcome of player 1,2
$S_1[k]$,$S_2[k]$	The actual winning value of player 1,2

*There are some variables that are not listed here and will be discussed in detail in each section.

4 Data Preprocessing

We first treat the vacancies of the data that need to be used.

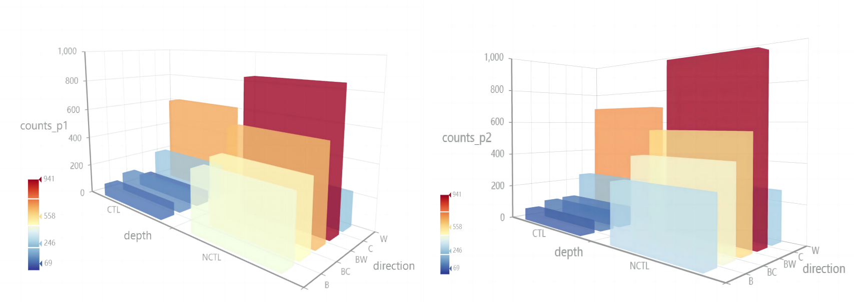
- The vacancy value in the speed mph was filled with **the average of the player's serve speed**.
- Take the mode of the player to fill the vacancy in serve_width.
- Take **the mode of the player** to fill the vacancy in serve_depth.
- Take **the mode of the player** to fill the vacancy in return_depth.

5 Factors that Alter the Course of Momentum

In a single set, there are many factors that can change the situation. We comprehensively consider the **technical ability**, **physical condition** and **psychological factors of the contestants**, and believe that these three characteristics are the most important factors affecting the change of momentum.

5.1 Technical Skill

The technical level of a player has a direct impact on the overall scoring situation of the game and serves as a significant factor that tilts the competition in favor of a particular player. It is evident that players with a high technical level are more likely to assert dominance in the game. To evaluate the technical level of a player, we consider factors such as **the number of points won on serve**, **the ratio of forehand to backhand shots**, **the distribution of serves in terms of direction**, and **the depth and speed of serves**. These factors serve as indicators of a player's technical proficiency(see in figure 3.)



(a) image 1 (b) image 2

Figure 3: Left: Serve direction and depth of player 1; Right: Serve direction and depth of play 2

Table 2: Technical Ability

Forehand-Backhand Ratio	Direction	Depth	World Ranking
5.313	3.087	4.188	2
5.216	3.003	4229	18

5.2 Physical Factors

The physical factor of the players is also an important factor that causes the momentum swing. In the course of a game, the players be affected due to physical exhaustion and other factors, thus losing their original advantage. We considered several indicators of physical factor such as serving speed, and the distance run during the point.(see in figure 4)

5.3 Psychological Factors

In the process of the competition, the score, whether the server, the number of points will interfere with the psychology of the players (see in figure 4), thus affecting the situation of the competition. For example, a player will have more anxiety as the number of missed points increases, which will affect his performance, which may lead to a momentum bias towards another player.

(a) image 1 (b) image 2

Figure 4: Left: Physical Factors; Right: Psychological Factors

6 Definition of Momentum

In order to better express the definition of momentum with a formula, we selected the following four indicators to measure momentum.

- The first indicator we selected is whether the player is the server. Given by the problem, in tennis, the server has a higher chance of winning the game. According to the data in the table, **we calculate that the probability of the server winning the point is 0.673**. So we think whether to serve is an important factor in the momentum of the player. The information of the sever is given in the database, and we treat the data as follows:

$$\alpha_1 = \begin{cases} 1 & \text{if } sever = 1 \\ 0 & \text{else} \end{cases} \quad \alpha_2 = \begin{cases} 1 & \text{if } sever = 2 \\ 0 & \text{else} \end{cases} \tag{1}$$

  When sever=1, α_1=1 indicates Player 1 serve, α_2=0 . And when sever=2, α_2=1 indicates Player 2 serve, α_1=0.

- The error status of the player. If a player makes mistakes during the course of the game, it will put greater psychological pressure on him and may affect his momentum. So we use this indicator to measure the momentum of the players.The database gives the data of two players in a match making unforced error and their serving errors . We sum these two indicators of each player and get the error situation in the kth point of the player. The specific calculation method is as follows:

$$\beta_1[k] = P_{1_{unf-err}} + P_{1_{double-fault}} \tag{2}$$

  $\beta_1[k]$ indicates Player 1's error situation in the kth point. $P_{1_{unf-err}}$ represents Player 1's unforced error, and $P_{1_{double-fault}}$ represents Player 1's double severing fault.

- Assessment of the players' physical exertion. In this article, we believe that the player's momentum is related to his physical exertion. Assessment of physical exertion was defined by the analysis of factors such as serving speed and the distance run during one point. The specific calculation methods are described as follows:

$$W_1[k] = \sum_{i=1}^{k} P1\_distance\_run[i] + \lambda \cdot S_{mph} \tag{3}$$

  $W_1[k]$ indicates Player 1's physical exertion in the kth point. $P1\_distance\_run[i]$ has been given in the sheet. The weight assigned to serving speed in measuring player's physical exertion, denoted as λ, is set as 0.02 in this study. And S_{mph} indicates the serving speed of the player 1.

- Whether the player wins continuously. If a player wins in a row, it may cause some psychological pressure on the disadvantaged player, and the player may also make mistakes due to the joy of winning. So we take whether the player wins continuously as an indicator of momentum. The specific definition is given as follows:

$$\mathrm{Z}_1[\mathrm{K}] = \begin{cases} x & \text{if } player1 \quad has \quad won \quad x \quad points \quad in \quad a \quad row \\ -x & \text{if } player1 \quad has \quad lost \quad by \quad x \quad points \quad in \quad a \quad row \\ 0 & else \end{cases} \tag{4}$$

Elo is a well recognized method to predict the winning probability of two tennis players[2]. It was originally created by Hungarian-American physicist Arpad Elo to measure the level of various types of chess games. The level of players is evaluated based on the Statistical principle, which is more objective than other rating methods. We use the Elo rating system in this paper to measure the momentum of the two players. The original Elo calculation method has only considered one factor in the score of two players. The formula is as follows:

$$R_{a+1} = R_a + \gamma \cdot (S_a - E_a)$$

We have refined the traditional ELo in this paper. Considering the impact of the four selected factors on momentum, our improved formula for Elo defines momentum as follows:

$$R_1[k+1] = 0.9 \cdot R_1[k] + (n_1^{\alpha[k]} \cdot \gamma - n_2 \cdot \beta_1[k] - n_3 \cdot W_1[k] + n_4 \cdot Z_1[k]) \cdot (\mathrm{S}_1[k] - \mathrm{E}_1[k]) \tag{5}$$

The formula is explained as follows:

- The formula is defined by player 1's momentum in kth point and related characteristics.
- $S_1[k]$ indicates whether player 1 won kth point.

$$\mathrm{S}_1[k] = \begin{cases} 1 & \text{if } 1\_wins \\ 0.5 & \text{if } draw \\ 0 & \text{if } 1\_loses \end{cases} \tag{6}$$

- $\mathrm{E}_1[k]$ indicates expected outcome of the player 1.

$$E_1[k] = \frac{1}{1 + 10^{(R_2[k] - R_1[k])/400}} \quad E_2[k] = \frac{1}{1 + 10^{(R_1[k] - R_2[k])/400}} \tag{7}$$

- γ is a constant coefficient. The greater the γ, the greater the impact of winning or losing on the player. So we set γ as 48 if a set won by one player, set γ as 32 if a game won by one player, and set γ as 16 if a point won by one player.
- For player 2, the indicators are defined in the same way.

This formula has four coefficients n_1, n_2, n_3, n_4. In order to better select the coefficients, we obtain these four coefficients by solving the following optimization problem.Using this formula, we calculate the momentum for each player separately and get their momentum difference.

$$\Gamma = \min M_j$$

$$s.t. \begin{cases} M_{k+1} = M_k + (-1)^{victor[k]} \cdot C_k, \quad k = 0, 1, 2... \\ C_k = R_1[k] - R_2[k] + 0.01 \cdot (R_1[k]^2 + R_2[k]^2) \\ R_1[k+1] = 0.9 \cdot R_1[k] + (n_1^{\alpha_1[k]} \cdot \gamma - n_2 \cdot \beta_1[k] - n_3 \cdot \mathrm{W}_1[k] + n_4 \cdot \mathrm{Z}_1[k]) \cdot (\mathrm{S}_1[k] - \mathrm{E}_1[k]) \\ R_2[k+1] = 0.9 \cdot R_2[k] + (n_1^{\alpha_2[k]} \cdot \gamma - n_2 \cdot \beta_2[k] - n_3 \cdot \mathrm{W}_2[k] + n_4 \cdot \mathrm{Z}_2[k]) \cdot (\mathrm{S}_2[k] - \mathrm{E}_2[k]) \\ R_1[0] = 130, \quad R_2[0] = 100 \end{cases} \tag{8}$$

- C_k indicates the momentum difference between the two players.At the same time, a regularization term $0.01 \cdot (R_1[k]^2 + R_2[k]^2)$ is added to prevent overfitting.
- According to the players' world rankings, we define that Player 1 and Player 2 have initial momentum of 130 and 100 respectively if player1's rank is higher than player2's.
- And victor[k] indicates the outcome of the kth point between the two players. If Player 1 wins, the value is 1. If Player 2 wins, the value is 2.
- M_{k+1} is the sum of the momentum differences. The purpose of "$\Gamma = \min M_j$" is to maximize the momentum gap of the winning and losing players

For each set (31 sets in total) we find the optimal solution for coefficients, and then we take the average of these optimal solutions to represent the final n_1, n_2, n_3, n_4.

Table 3: the Values of the Four Coefficients

n_1	n_2	n_3	n_4
0.838862	**1.65989**	**0.000633403**	**0.3571146**

In the database for the value of the second set, with the momentum formula(5) we defined, we draw two players' momentum and the winner of the game (see in figure 8). It can be found that the the general direction of momentum can basically fit the winner of the game, thus proving that the formula can accurately represent the momentum of the player.

7 Model for Predicting Which Player Will Perform Better

According to the competition process of Novak Djokovic and Carlos Alcaraz, no one always occupy the dominant position in sports competition, with a change in the situation in every scoring point. We wanted to build a model to analyze the impact of various factors on the changes in the momentum of scoring points and predict which player the next scoring point will be better and better based on the momentum of that point.

7.1 Data Selection and Processing

We conducted a preliminary analysis and judgment based on the data provided in the database , and selected the following indicators to train the model.

- The momentum of each player is an important factor to consider. We can calculate the momentum of each player using the formula we defined earlier and use them as training inputs for the model.
- Given that in tennis, the server has a higher probability of winning the game, we consider whether the player is serving as one of the indicators.
- The number of sets won by each player is an important indicator as it reflects their psychological state, which can further impact their performance.
- The number of games won by each player is similar to the number of sets won and serves as an indicator of their current performance and psychological state.

- The total score of each player is another important indicator. It can influence the psychological state of the players and subsequently impact their performance. Therefore, we will include the total score of each player as an input for training the model.
- The winner of every point is also included.
- To ensure data consistency and improve analysis performance, we use **MinMaxScaler** in python to normalize the above features before using them to train the model. This normalization process will scale the values of the features to a standard range, such as between 0 and 1, without changing their relative relationships. By doing so, we can enhance the effectiveness of the analysis and training process.

7.2 LSTM Model

LSTM (Long Short-Term Memory) is a variant of recurrent neural networks (RNN) widely used for processing and predicting time series data. Compared to traditional RNNs, LSTM has a more powerful memory capacity, allowing it to capture and learn long-term dependencies. This is the primary reason for selecting this model for analysis and processing. LSTM effectively handles long-term dependencies and mitigates the issues of vanishing or exploding gradients through its gated mechanism, enabling it to better capture patterns and structures in time series data. The gates include the forget gate, input gate, and output gate. These characteristics make LSTM excel in various tasks, such as language modeling, machine translation, speech recognition, and stock prediction.

Moreover, LSTM is relatively flexible in terms of data requirements, making it suitable for handling the data in this particular task.

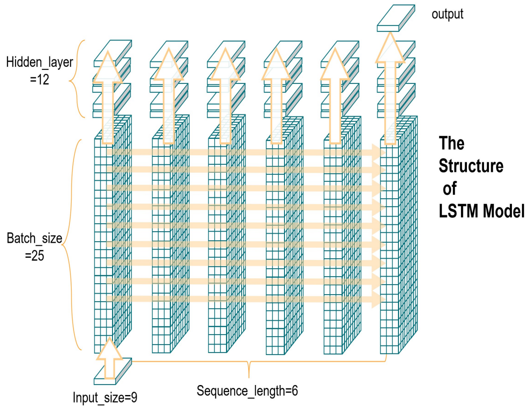
7.3 The Training Process of LSTM

We choose to use an LSTM model to predict which player will gain the advantage at the next scoring point based on the current situation of the players. In our model, we will input the normalized results obtained from the data processing stage to fit the output defined by ourselves that evaluates the performance of players.

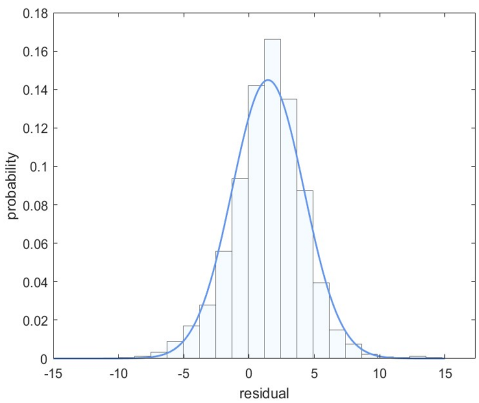
The output we defined is as follow:

$$\begin{aligned}\mathrm{H}_p = 4(P1\_sets[k] - P2\_sets[k]) + 2(P1\_games[k] - P2\_games[k]) + \\ 0.03(P1\_pwg[k] - P2\_pwg[k]) \cdot \delta_{pois}\end{aligned} \tag{9}$$

- H_p indicates the index evaluating the performance of the two players.
- $P1\_sets[k] - P2\_sets[k]$ indicates the difference of set scores between the two players in the kth point.
- $P1\_games[k] - P2\_games[k]$ indicates the difference of game scores between the two players in the kth point.
- $P1\_pwg[k] - P2\_pwg[k]$ indicates the difference of points between the two players in the kth point.
- δ_{pois} represents the position of the point in the game.



(a) image 1



(b) image 2

Figure 5: Left: Workflow Diagram for LSTM; Right: The residual of the model fits a normal distribution

We use the model we trained to predict the H_p. And we plot the predicted H_p and the real H_p. (see in figure 6). **The MAE(Root Mean Square Error)** of the model is **2.473592. The RMSE (Root Mean Square Error)** of the model is **3.1105382.** Both MAE and RMSE are indicators that measure the predictive power of a model and are used to evaluate the accuracy of the model. Moreover, seeing the probability distribution of residual in (b) image 2, we can concluded that our model fits the real H_p well.

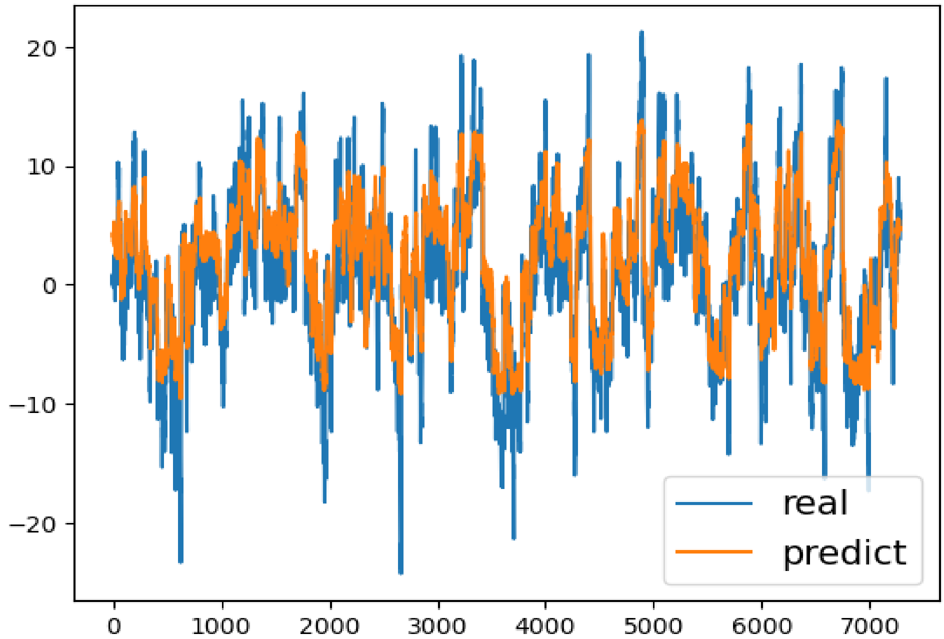


Figure 6

7.4 Visualization of the Match Process

We plotted the graph showing the predict H_p our model produced and the real H_p we calculated.The abscissa represents the total point number during the game. And we use data of match 2023-wimbledon-1301 to observe the accuracy of our model.

The ordinate is the value of the index we defined (H_p). If this value is greater than 0 means that player 1 performs better than player 2, and less than 0 means that player 2 performs better than player 1. The greater the absolute value is, the better one is performing. By comparing the two lines in charts, we can see that our predicted curve is a good fit to the real curve. And the winner of each set corresponds with H_p trend, as is observed that the H_p changes with the swing of the points(see in figure 7).

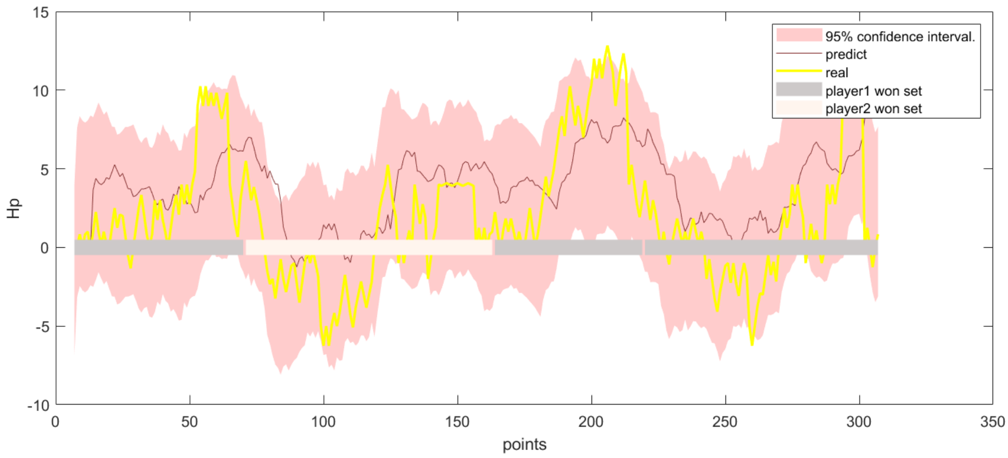


Figure 7: Visualization of 2023-wimbledon-1301

8 Correlation of Momentum and Turning Points in the Match

8.1 Turning Point in the Match

During a game, there are many factors that shift the situation to one side. We summarize five aspects to define the turning point of the match.

- In the case of one player lose two sets in a match, it is very hard for the disadvantaged player to win the match. If the disadvantaged player wins at last, every point he won after losing two sets is defined as turning point.
- If a dominant player in the last set wins by more than four games (such as 6-1), each scoring point is the turning point for the player who has lost the last set but won this set.
- If the serve player who loses the first game in a set, each point the reciever won is turning point. According to the information known from the title, the server is more likely to win the game in a tennis match, so the server is initially at advantage, but if the server loses game , the favor of servering goes to the other side.
- If one player has got AD but he lost the game, then the point the other player won was turning point. Since if one got AD, it was easy for him to win one more point but hard for the other to win continuously.

We define the above turning points as turning point in the set: obviously, as the game proceeds, the later turning point is more important than the previous one. Therefore, we assigned each turning point from the back in a game, 6,5,5,4,3,2,1,0, to get the crucial moments.

8.2 The Relevance of the Momentum and the Crucial Moments

We used **grey relational analysis** to observe the correlation between momentum and turning points in the matches of two players. Firstly, we processed the momentum on all scores (calculated using formula (5)) and turning points for both players. The results are shown in Table 4.

Table 4: Processing results of all the data

Total Score	R_A	R_B	turnning_point
1	130	100	0
2	110.6152226	98.68636707	0
3	106.0167509	81.1127898	0
...	...	...	...
62	-3.413860682	9.698973769	4
63	2.90864662	1.563970115	5
...	...	...	...
7284	15.79775108	-18.98816487	0

Subsequently, we conducted grey relational degree analysis on the processed data(see in table 5).

Table 5: Results of the correlation analysis between momentum and turning points

R_A	R_B	turnning_point
1	0.858624506	0.885772973
0.854558862	1	0.880462335
0.882454526	0.880457227	1

Through visualization, we can intuitively observe a strong correlation between momentum and turning points(see in figure 8), leading to the conclusion that the swings in the game are not random but rather have a certain relationship with the players' momentum.

9 Predicting swings in the Game

9.1 Find the Most Suitable Predictive Model

According to the previous analysis, it can be obtained that the swings in the game process are not random. We have used multiple algorithms to train a predictive model

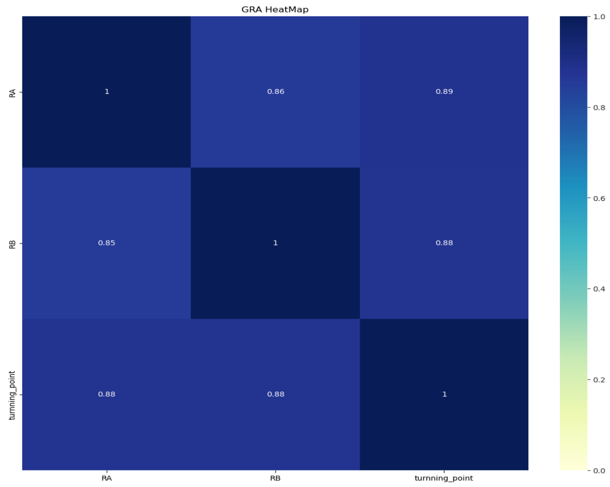


Figure 8: The Relevance of the Momentum and the Crucial Moments

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
qda	Quadratic Discriminant Analysis	0.6851	0.7018	0.2938	0.5262	0.3705	0.1859	0.2017	0.0040
knn	K Neighbors Classifier	0.6801	0.6618	0.4094	0.5079	0.4510	0.2297	0.2336	0.0080
dummy	Dummy Classifier	0.6771	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0040
nb	Naive Bayes	0.6761	0.5546	0.0000	0.0000	0.0000	-0.0020	-0.0070	0.0040
gbc	Gradient Boosting Classifier	0.6711	0.6692	0.2750	0.4895	0.3460	0.1514	0.1646	0.0210
lda	Linear Discriminant Analysis	0.6690	0.7046	0.0156	0.2810	0.0284	-0.0043	0.0085	0.0040
ridge	Ridge Classifier	0.6660	0.0000	0.0062	0.1143	0.0112	-0.0167	-0.0259	0.0040
lightgbm	Light Gradient Boosting Machine	0.6650	0.6522	0.3781	0.4861	0.4208	0.1910	0.1967	0.1510
rf	Random Forest Classifier	0.6640	0.6449	0.3625	0.4781	0.4109	0.1821	0.1865	0.0340
lr	Logistic Regression	0.6559	0.7078	0.0438	0.3081	0.0732	-0.0098	-0.0104	0.0050
ada	Ada Boost Classifier	0.6448	0.6329	0.0875	0.3319	0.1350	-0.0014	0.0013	0.0150
et	Extra Trees Classifier	0.6448	0.6355	0.3594	0.4384	0.3922	0.1462	0.1487	0.0300
dt	Decision Tree Classifier	0.6297	0.5819	0.4469	0.4338	0.4381	0.1633	0.1639	0.0050
svm	SVM - Linear Kernel	0.5943	0.0000	0.3656	0.3801	0.3674	0.0720	0.0732	0.0050

Figure 9: Evaluation of Different Models.

for predicting the swings occurring during the game process. **The accuracy results are as follows**:

From Figure 9, it is evident that the **Quadratic Discriminant Analysis algorithm** has the highest accuracy, **with an accuracy of 0.6851**. Hence, we will opt for this algorithm for our predictive model.

According to Tom Fawcett's introduction,ROC (Receiver Operating Characteristic) is a graphical tool used to evaluate the performance of a classifier[6]. It is based on binary classification problems and measures the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at different thresholds by plotting a curve. The closer the curve is to the top-left corner, the better the performance of the model. The ROC curve for Quadratic Discriminant Analysis is located in the upper left part of the graph, indicating a higher accuracy of the model(see in figure 10).

We use the trained model to predict the swings in the match. From the figure 11, we can see that the changes in momentum for both players and the match outcomes are in good agreement. This confirms that our model has a good fitting effect and can be used to predict the swings in the match.

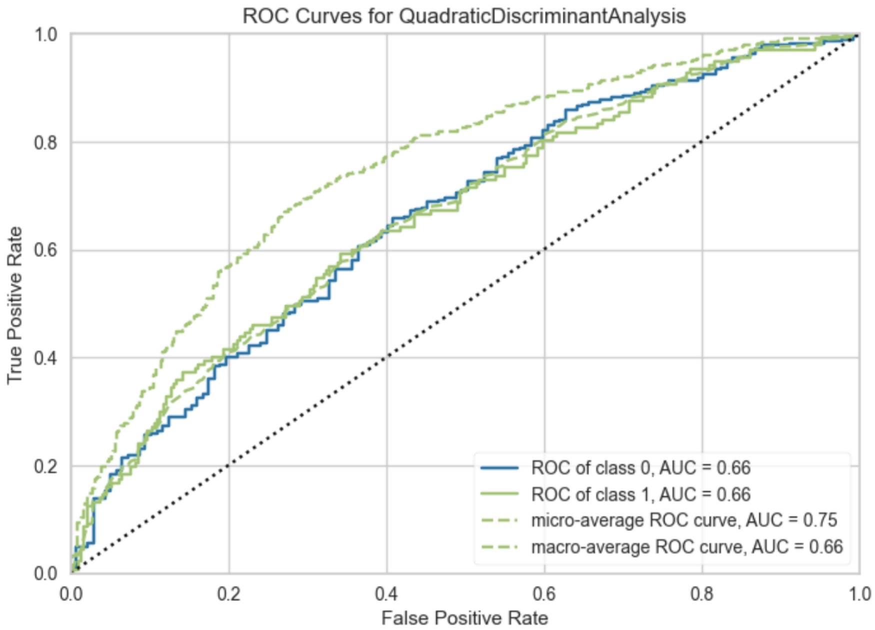


Figure 10: ROC Curve of the Quadratic Discriminant Analysis

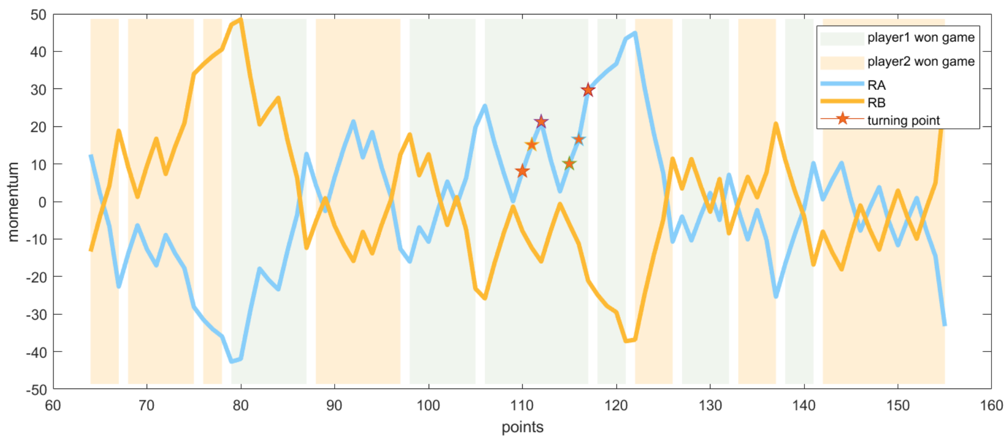


Figure 11: Predict the swings in the Match

9.2 Find the Factors with the Strongest Correlation

We first selected several indicators related to the game situation analyzed in the second question, and processed the data in the database to obtain the results shown in the following figure.

We used the method of **Grey Relation Analysis** to analyze the correlation of various factors and turning points and draw the heat map to visually present our results intuitively. The darker the color, the higher the correlation(see in figure 13.) **The correlation between the number of consecutive wins was 0.89**, which is the most correlated factor to the game situation.

9.3 Recommendations for the Players

Based on the previous analysis, we have drawn the conclusion that in a tennis game, the situation of the game does not always favor a certain player, but constantly fluc-

	K	(Sa-Ea)	(Sb-Eb)	beta_pA	beta_pB	alpha_PA	alpha_pB	z_pA	z_pB	ω_pA	ω_pB	RA	RB	turnning_point
0	16	-0.543066	0.543066	1.0	0.0	1.0	0.0	0.0	0.0	7.900	7.840	130.000000	100.000000	0.0
1	16	0.481844	-0.481844	0.0	0.0	1.0	0.0	0.0	0.0	13.613	14.934	110.615223	98.686367	0.0
2	16	-0.540267	0.540267	1.0	0.0	1.0	0.0	0.0	0.0	27.453	34.742	106.016751	81.112790	0.0
3	16	0.484531	-0.484531	0.0	0.0	1.0	0.0	0.0	0.0	78.761	110.373	89.069907	81.633888	0.0
4	16	0.462284	-0.462284	0.0	0.0	1.0	0.0	1.0	-1.0	79.050	111.186	86.642020	65.751876	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7278	16	-0.479987	0.479987	1.0	0.0	1.0	0.0	-1.0	1.0	101902.632	100890.568	26.492413	-32.893606	0.0
7279	16	0.542028	-0.542028	0.0	0.0	1.0	0.0	0.0	0.0	101941.172	100941.539	49.349994	-52.426322	0.0
7280	16	0.517160	-0.517160	0.0	0.0	1.0	0.0	1.0	-1.0	101958.525	100969.305	16.691296	-21.200650	0.0
7281	16	-0.506649	0.506649	0.0	0.0	1.0	0.0	0.0	0.0	101969.607	100987.633	-11.250548	5.904115	0.0
7282	16	0.516676	-0.516676	0.0	0.0	1.0	0.0	0.0	0.0	101977.620	100998.170	15.797751	-18.988165	0.0

Figure 12: Processing of Indicators Related to the Game Situation.

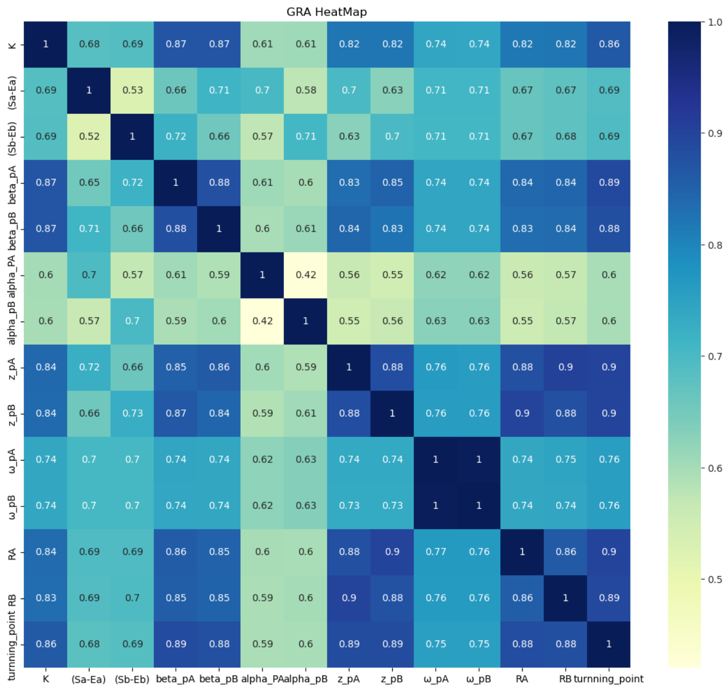


Figure 13: Correlation Between Indicators and Game Situation.

tuates, and this swing is related to the momentum and other factors. Accordingly, we give some advice when the two players play a new round of competition.

- According to the correlation analysis, it can be seen that there is a high correlation between the number of consecutive wins and the appearance of the turning point of the game, indicating that the winning streak or losing streak has a great impact on the momentum level, but according to the Cornell university psychology professor Thomas Gilovich and others on the NBA real data analysis[3] found that after three consecutive shots, the next shot and normally there is no significant difference. Therefore, we suggest that players who have won in a row should consider changing their serve or hitting mode, and those who lose should not produce too much psychological pressure and keep their normal play.

- "Hot-hand effect" comes from basketball[4]It refers to the fact that when a player makes consecutive shots, the other players generally believe that he has a "good touch" and will choose him to shoot on the next attack, but he is not always able to make a shot. The lack of necessary analysis and judgment may cause mistakes. Therefore, it is recommended that the players who have won in a row should avoid the hot-hand effect psychologically and make an accurate judgment before each serve or catch.

- "The Domino Effect of Errors[5]" is a series of chain reactions or errors triggered by an error. In some cases, each error is a separate event, but they affect each other, creating a ripple reaction that aggravates the problem. For example, in cybersecurity, a small misstep can lead to all data leaks; In the industrial field, the failure of one piece of equipment may bring down the entire production line. In healthcare, a misdiagnosis may lead to incorrect treatment that worsens the condition or leads to other consequences. Therefore, inferior players need to adjust their mentality in time to avoid more points lost by the "error chain" effect.
- In addition, the running distance of the players also has a great impact on the appearance of the turning point, so it is suggested that the players take the way of quick decision as far as possible to avoid too much running, and at the same time, they should grasp the direction and depth to increase the running of the other side.

10 Adaptability and Universality of the Model

10.1 Assessment of the Adaptation of the Model

We based on the match model of Carlos Alcaraz and Novak Djokovic to predict the swing of the tennis match, and conclude that factors such as the swing and momentum of the match have a strong correlation. Next, we put the resulting model in other events to observe whether the model can fit the swings of other games, thus assessing the universality and adaptability of the model.

- We selected the match data of the WTA women's tennis final between Diana Schneider and Julincz, which just concluded on February 4, 2024. The match result was Schneider's victory in the first set, Julincz's victory in the second set, and Schneider's victory in the final set. The analysis and predictions made by our model regarding the swings in the game were consistent with the actual results.
- First, we processed the match data to obtain the momentum of the two players and turning point in the game, as shown in the figure 14.
  And our prediction has an accuracy of 0.8974358974358975
- We utilized the model from the first question to plot the graph for predicting match swings(see in figure 17), which demonstrates a good fit of the model. This indicates that our model can be generalized to other matches and possesses a certain level of universality.

10.2 Possible Future Model Improvements and Optimizations

The prediction effect of our model did not reach 100%, with a bias in a certain range. Due to the complexity of the analysis, we did not take into account all the factors that might affect the swings of the competition, so that the model predictions and the real situation would deviate.In the future model optimization and improvement, we can also consider the following factors to make our model prediction more accurate.

- **Factors that appear in the game**
  $\mapsto$ The speed of the hit and serve.

	RA	RB	prediction_label	prediction_score_0	prediction_score_1
0	100.000000	130.000000	0	1.0	0.0
1	108.689064	121.310936	0	1.0	0.0
2	100.979568	129.020432	0	1.0	0.0
3	109.623833	120.376167	0	1.0	0.0
4	126.118835	103.881165	0	1.0	0.0
...	...	...	...	...	...
112	142.016800	87.983200	0	1.0	0.0
113	155.548325	74.451668	0	1.0	0.0
114	161.714203	68.285797	0	1.0	0.0
115	167.613327	62.386673	0	1.0	0.0
116	184.558914	45.441082	0	1.0	0.0

117 rows × 5 columns

Figure 14: Predictions of turning points in WTA matches

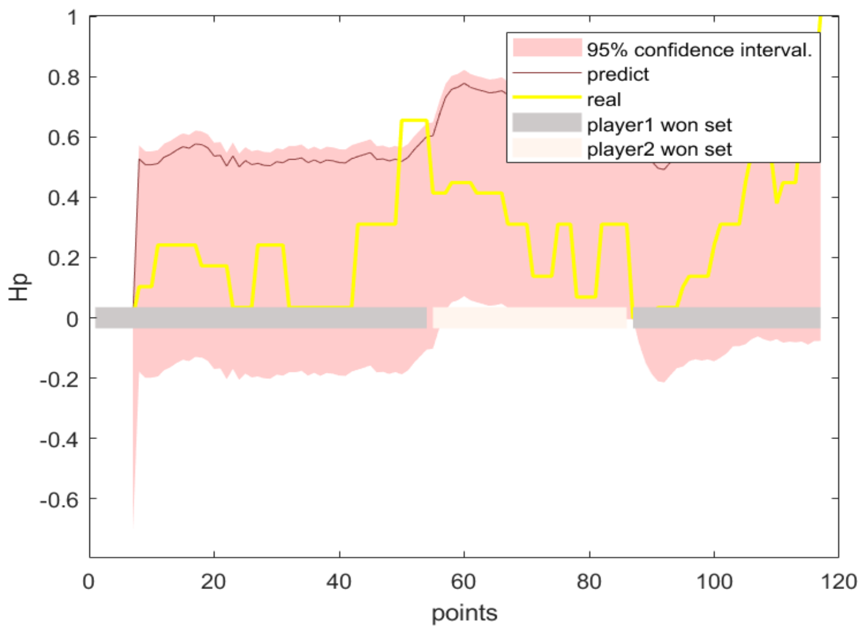


Figure 15: Prediction about the Match Process

↦The number of unforced errors per game , and the number of two consecutive serve errors.

- **Skill of Player**
  ↦ Tactical changes in hitting and serving.
  ↦ The explosive power of a player.
  ↦ Physical factors such as height and weight.

11 Strengths and Weaknesses

11.1 Strengths

- **Comprehensive consideration of factors:** We considered multiple factors that may affect the swings of the competition for analysis and processing. When defining the momentum, we considered whether the players was the server, their mistakes in the competition, physical exertion and the number of consecutive wins. In addition, we considered factors such as the score of the players in each set and game. In general, this model has a more comprehensive analysis of the various factors of the players.
- **High generalizability:** We applied the obtained model to the final match of the WTA Schneider versus Zhu Lin on February 4,2024, and found that the overall trend was basically consistent. Theoretically, our model can predict the scoring and situation of all tennis matches.
- **Predict the Turning Point of the Game:** We selected five factors related to the turning point for analysis and processing. The model can predict the occurrence of the turning point in the game, which has an important reference significance for the tennis players in the training and competition process.

11.2 Weaknesses

- **The Calculation is Complex:** We adopted the Elo rating algorithm when defining the formula of momentum, and comprehensively considered the four indicators to obtain the correlation coefficient, and the calculation process is complicated.
- **Possible inaccurate use of models:** Since the measures of each sport are not completely uniform, the momentum we define in the model is selected according to the rules of tennis, and inaccuracies may occur in the prediction of other competitive events.

12 Conclusions

In this paper, we define momentum in tennis matches by building a model, analyze the correlation between momentum and fluctuations in matches, predict fluctuations in matches, and promote the model. Give clear answers to the questions given:

- For problem 1, the LSTM model can be trained to predict a player's performance at the next score point according to their performance at the current score point. The visualization of the model is presented to show the flow of the competition.
- For problem 2, momentum is defined using the improved Elo rating system. By the method of Grey Relation Analysis, it is found that there is a strong correlation between momentum and game fluctuation, and the correlation degree is 0.88.
- For problem 3, Quadratic Discriminant Analysis algorithm is used to train a model to predict the fluctuations that occur in the game. Using the method of grey correlation degree analysis, it is found that the number of consecutive wins and errors of players have the strongest correlation with the fluctuation of games.

Some suggestions are given.

- For problem 4, the universality and adaptability of our model are verified by analyzing the data of the WTA Finals on February 4, 2024.

13 Memo

TO: Tennis Coaches

FROM: Team #2422034, MCM 2024 DATA: February 5, 2024

SUBJECT: Momentum in Tennis Match

Dear Sir or Madame,

It is our honor to describe to you the momentum in tennis.

After carefully analyzing the definition of the momentum of the game and its impact on the situation of the game, we draw the following conclusions and give our suggestions.

Results

In a tennis game, the advantage situation is not always biased to one of the players, the disadvantage is the same, there are a lot of swings during the game, and these swings are related to a lot of factors, and we integrate these factors to define the momentum of the player. Through the analysis, momentum can be considered as a key factor affecting the volatility of the game.

- ***How to define momentum?***
  $\hookrightarrow$ We selected four indicators, namely, whether the player is the server, the mistakes and physical exertion in the game, and the number of consecutive wins of the player. Bring these four indicators into our improved Elo rating system to define momentum with a formula.
- ***Is there a relationship between momentum and the course of the game?***
  $\hookrightarrow$ We used the grey correlation analysis to find a strong correlation between the momentum and the turning points in the game, proving that the swings of the game situation are not random but change to some extent with the momentum.
- ***What are the key factors that influence the swing of a game?***
  $\hookrightarrow$ We first processed the data and obtained several indicators that may be related to the swing of the competition, respectively. Again, we used the gray correlation analysis method to calculate the correlation between each indicator and the swing. We found that the two indicators with the strongest correlation were the mistakes of players and the number of consecutive victories.

Recommendations for the Coaches

- ***Make good use of momentum***
  $\hookrightarrow$ According to our above analysis, we can see that momentum and the swing of the game are closely related, so the coach should be aware of the important role of momentum in the game and make momentum as one of the items of players' daily training.
- ***How should players respond to the swings in the game?***
  $\hookrightarrow$ Based on the correlation analysis between selected indicators and turning points in the game, it is evident that the number of consecutive wins and the error

situation have a significant impact on game swings, coupled with the influence of the "hot hand effect" and the "error chain effect". We recommend that players on a winning streak consider changing their serving or hitting style, while players on a losing streak should avoid placing excessive psychological pressure on themselves and strive to maintain their normal performance.

↪ The distance covered by players during the game can also have an impact on the game's dynamics. Therefore, it is advisable for players to avoid excessive physical exertion as much as possible during the game. Additionally, players can increase the opponent's physical exertion by adjusting the direction of serves and employing other strategies.

The above is the summary of our study. We sincerely hope that it will provide you with useful information.

Yours sincerely,
#2422034

References

[1] Li C, Zhao G, Meng J, et al. MULTI-OBJECTIVE OPTIMIZATION STRATEGY BASED ON ENTROPY WEIGHT, GREY CORRELATION THEORY, AND RESPONSE SURFACE METHOD IN TURNING[J]. International Journal of Industrial Engineering, 2021, 28(5).

[2] Angelini G, Candila V, De Angelis L. Weighted Elo rating for tennis match predictions[J]. European Journal of Operational Research, 2022, 297(1): 120-132.

[3] Gilovich T, Vallone R, Tversky A. The hot hand in basketball: On the misperception of random sequences[J]. Cognitive psychology, 1985, 17(3): 295-314.

[4] Bar-Eli M, Avugos S, Raab M. Twenty years of "hot hand" research: Review and critique[J]. Psychology of Sport and Exercise, 2006, 7(6): 525-553.

[5] Necci A, Cozzani V, Spadoni G, et al. Assessment of domino effect: State of the art and research Needs[J]. Reliability Engineering & System Safety, 2015, 143: 3-18.

[6] Fawcett T. An introduction to ROC analysis[J]. Pattern recognition letters, 2006, 27(8): 861-874.