Problem Chosen	2024	Team Control Number
C	MCM/ICM Summary Sheet	2421624

The Flow of Match Hidden in "Momentum"

Summary

As the oldest tennis tournament in the world, Wimbledon Championships is widely considered by many as the world's premier tennis tournament. When everyone concentrates on who wins the match, we want to figure out how can a player win the game. Therefore, we used **AHP** (the Analytic Hierarchy Process), **EWM** (the Entropy Weight Method), **XGBoost** (eXtreme Gradient Boosting), and **DE** (Differential Evolution) to build a model, exploring the role of "momentum" in tennis matches.

For **Task 1,** we mainly used the geometric mean of the results from **AHP** and **EWM** to acquire the combination weight of evaluating indicators, and then calculate the momentum. We selected six indicators to comprehensively evaluate the performance of each player at each point, which is the player's PM (Point Momentum). Considering that the current momentum will act on a period, we use the Gaussian function as the coefficient to sum the score momentum in multiple days to get the momentum at the current moment.

For **Task 2,** we proved the difference between the distributions of two sub-data sets which should be the same random distribution according to the coach's statement. We selected these two samples by the distance from the previous break point obtained by a certain player, and proved their non-normality by **Shapiro-Wilk** Test. With the p-values of $\mathbf{2.42 \times 10^{-9}}$ and $\mathbf{6.38 \times 10^{-4}}$ in **Kolmogorov-Smirnov Test** and the p-values of $\mathbf{6.09 \times 10^{-6}}$ and $\mathbf{6.31 \times 10^{-3}}$ in **Mann-Whitney U Test**, we confirmed that the distributions of point momentum in the two samples are different.

For **Task 3,** we obtained the prediction model and offered some suggestions to players for preparing a new match based on our analysis. We first calculated the **Pearson Correlation Coefficient** between the momentum to be predicted and all the variables and found that the momentum is affected differently in different matches. Using **XGBoost** and **DE Algorithm** to build and optimize our prediction model, the MSE (Mean Square Error) of player 1 and player 2 improved from 0.0828 to 0.0619 and 0.0861 to 0.0743, meanwhile, the MAE (Mean Absolute Error) of from 0.219 to 0.196 and 0.221 to 0.238. According to the feature importance of the optimized XGBoost in **Figure 9**, we obtained three indexes that have the greatest influence on momentum: p2_games, p1_points_won, and p1_unforced_error.

For **Task 4,** we used the trained model to predict the match flow of 2023-wimbledon-1305, in which the prediction is shown in **Figure 10**. According to our calculation, the MSE of player 1 and player 2 is 0.183 and 0.163, proving that our model has a certain degree of fitting in terms of parameters for the performance of men's competitions. For other types of competitions, the framework of our model is still appropriate. Therefore, after feature adjustment and parameter fitting, our model can be transplanted to other similar competitive competitions.

In conclusion, our model has advantages in predicting the momentum of players and the flow of play. To predict the momentum in the game as accurately as possible, we need more data and included factors, such as the player's status of home or away, past game performance, weather conditions, players' physical condition, etc.

Keywords: Momentum; Analytic Hierarchy Process; Entropy Weight Method; XGBoost; Differential Evolution

Contents

1 Introduction

1.1 Problem Background

The Wimbledon Championships (hereinafter, “the Wimbledon”), arranged in Wimbledon, London, is prestigious for its long history, unique tradition, and worldwide popularity. Since its founding in 1877, Wimbledon has become one of the four Grand Slam tournaments. During the fortnight, there typically are hundreds of thousands of guests sitting around the grass courts and millions of impressions received by its official social media.

2023 witnessed a magnificent occurrence in the Wimbledon final where the 20-year-old player, young Carlos Alcaraz, defeated the experienced defending champion, Novak Djokovic. With several incredible turning points, we can feel that there are regular patterns within success. Within each small winning, the “momentum” in players acts a critical role.

1.2 Restatement of the Problem

Combining background information and related conditions identified in the problem statement, we need to accomplish the following tasks:

- **Task 1: Prove that "Momentum" Exists by Evaluating the Performance of Players.**
  Develop a model to illustrate the flow of play at the occurrence of points, which can quantify and rank players’ performance. Then describe the match flow with the model and visualized tools.
- **Task 2: Rethink the Coach’s Skeptical Opinion of “Momentum”.**
  Offer comments to the view that the swings in a match are determined by randomness instead of "momentum".
- **Task 3: Predict the Flow of Play according to "Momentum".**
  Develop a model to predict the flow of play and clarify the influencing factors of “momentum” and, if any, the most influential one. Then according to the model, provide players with suggestions to play in other matches against new opponents.
- **Task 4: Examine the Generalizability of the Model.**
  Apply data from other matches to test the accuracy of the model when predicting the swings in the match. If occasional deviation happens, improve the model with factors that may be taken into consideration. Then detect the generalizability of the model under different genders of players, types of matches, surfaces of courts, and events of matches.

1.3 Our Work

Our work flow of this paper is showned in **Figure 1** on **Page 3**.

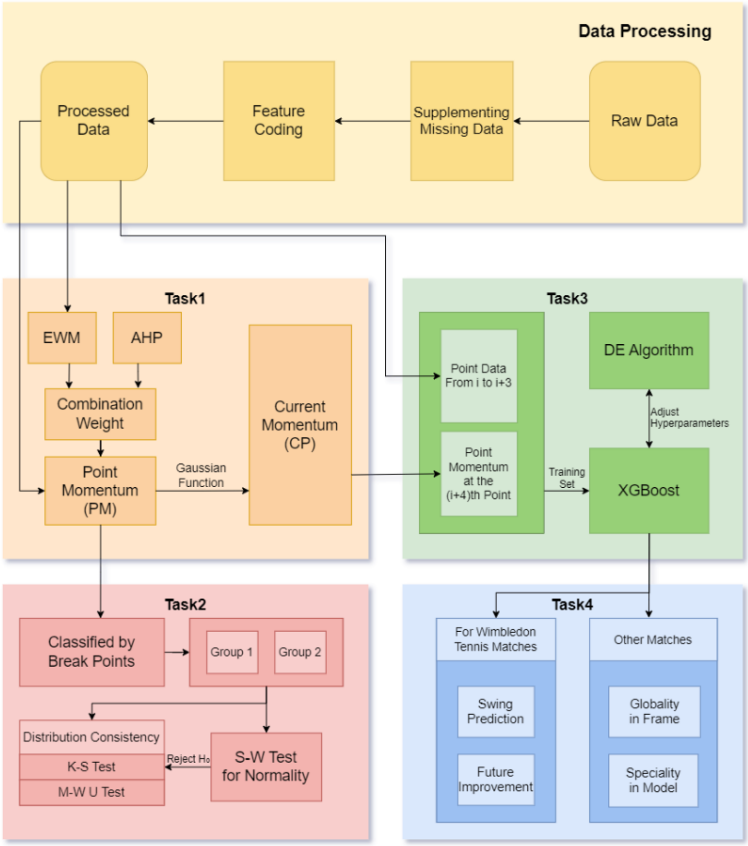


Figure 1: Work Flow

2 Assumptions and Justifications

- **Assumption 1:** Momentum describes the current status of a player and has no relation to personal ability, which enables us to compare and analyze the momentum of both sides even if the abilities of the two players are different.
- **Assumption 2:** The influence of momentum (including current momentum and point momentum) on players' performance will last for a while and decline with time.
- **Assumption 3:** The value of momentum is a comprehensive score considering who gets the point, how the player gets or loses the point, the number of rallies before the occurrence of the point, etc. Such evaluating method can exhibit the overall status of a player, contributing to the following analysis.
- **Assumption 4:** Due to the limitation of data information, we assume that momentum changes only when a point occurs and keeps stable before the next point comes. This assumption can effectively simplify the model without increasing errors.

3 Notations

In this paper, some important notations are listed in **Table 1**.

Symbol	Description
PM (Point Momentum)	the momentum of a certain player when a point occurs
CM (Current Momentum)	the momentum that has an impact on the future occurrence of several points
SN (Serve Number)	the number of serves
AP (Ace Point)	the point which the server obtains on serve
CP (Common Point)	the point which the server obtains after serving, or the break point that the receiver gets
DF (Double Fault)	the point lost by the server because of missing both of the serves
CF (CommonFault)	the point which the server or receiver loses when facing an untouchable winning shot
RC (RallyCount)	the number of shots during the point
SAM (Server's Average Momentum)	a player's average momentum score as a server
RAM (Receiver's Average Momentum)	a player's average momentum score as a receiver

Table 1: Notations

4 Task 1: Evaluation Model Based on AHD-EWM

4.1 The Definition of "Momentum"

According to our analysis of the problem, we render that:

Momentum is divided into Point Momentum (PM) and Current Momentum (CM). PM represents the momentum of a certain player when a point occurs, and CM is the symbol of the momentum that has an impact on the future occurrence of several points. Momentum reflects on and can be evaluated by future performance. Also, momentum is influenced by abundant factors.

4.2 The Identification of Evaluation Indicators

To construct a consistent system for apprising the performance of players in different games, we select several indicators as follows:

1. **SN (the Number of Serves)** [negative]

   SN=1 represents a normal performance, while SN=2 means an abnormal performance.

2. **Untouchable Winning Shot** [positive]

   - **AP (Ace Point):** the point which the server obtains on serve
   - **CP (Common Point):** the point that the server obtains after serving, or the break point that the receiver gets

3. **Unforced Error** [negative]

   - **DF (Double Fault):** The server misses both the serves and loses the point.
   - **CF (Common Fault):** The server or the receiver loses the point for facing an untouchable winning shot.

4. **RC (Rally Count)** [negative]: the number of shots during the point

4.3 The Establishment of Evaluation Model

After selecting the indicators, we need to determine the weight of each element. According to our requirements, we choose AHP (the Analytic Hierarchy Process) first, giving weight through human experience. Although reasonable, it is relatively poor in objectivity. To increase the credibility of our evaluating system, we combine it with EWM (the Entropy Weight Method) and determine the objective weight by the size of the index variability. As for the final weight of each indicator, the Lagrange Multiplier method gives the combination weight.

4.3.1 AHP Subjective Evaluation Model

Based on decision-making theory, the Analytic Hierarchy Process is a structured technique for complex systematic decisions. It provides a general and rational framework to construct a decision goal, quantify symbolic elements, connect them with the final goal, and assess the alternatives. Involving more than one indicator in the course of evaluating the performance of players, we regard the flow of play during each point as an alternative solution, quantifying the weights of decision criteria with AHP and a combination of expert experience and indicated data.

Step 1: Decompose the decision problem "Performance Evaluation" and build a hierarchy of more easily comprehended sub-problems with SN, CP, AP, DF, CF, and RC.

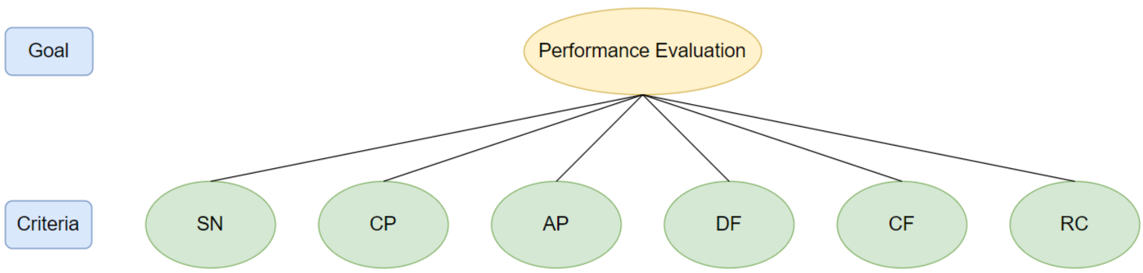


Figure 2: AHP Hierarchy

Step 2: Construct a judgment matrix to compare each pair of indicators at a time and establish their priorities concerning their relative meaning and importance. Thereinto, $\boldsymbol{A} = \left(a_{ij}\right)_{6\times 6}$, $a_{ij} > 0$, $a_{ij} = \frac{1}{a_{ji}}$.

	SN	CP	AP	DF	CF	RC
SN	1	1/4	1/6	1/2	2	1
CP	4	1	1/3	1	5	4
AP	6	3	1	3	7	7
DF	2	1	1/3	1	3	2
CF	1/2	1/5	1/7	1/3	1	1/2
RC	1	1/4	1/7	1/4	2	1

Table 2: The comparison between indicators in AHP

Step 3: Check the consistency of the judgmental elements. Recognizing that Matrix A is a reciprocal matrix of n = 6, we calculate the maximum eigenvalue λ, and substitute it into Equation (1) to obtain CI (the Consistency Index). In the meanwhile, we introduce RI (the

Randomness Index), and find out that RI is 1.26 when n=6. Then we use Equation (2) and find that CR (the Consistency Ratio) is 0.023176, smaller than 0.1. This result proves that the inconsistency degree of Matrix A is within the allowable range, which means that the model passes the consistency test.

$$CI = \frac{\lambda - n}{n - 1} \tag{1}$$

$$CR = \frac{CI}{RI} \tag{2}$$

Step 4: Compute the weight vector. The model having passed the consistency test, we normalize the eigenvector as a subjective weight vector (w_1) which is (0.070, 0.207, 0.446, 0.170, 0.044, 0.061).

4.3.2 EWM Objective Evaluation Model

The Entropy Weight Method (EWM) is a significant information weight model that gives weight by measuring the value dispersion in decision-making. Information is a measure of the degree of order of the system, while entropy is a measure of the degree of disorder of the system. The entropy theory in information theory reflects the degree of disorder of information and can be applied to evaluate the amount of information. The greater the difference value among a certain indicator is, the smaller the entropy is, indicating that this indicator contains more information and should be given a greater weight.

Step 1: Standardize the measured values. As a consequence of the different measures among those indicators, we need to eliminate their dimensions. Record our six indicators as $X_1, X_2, \ldots, X_6$, in which $X_i = \{x_1, x_2, \ldots, x_n\}$ (n refers to the number of data in the selected training set). Since all negative indicators are positive by adding a negative sign, we record the standardized indicators as $Y_1, Y_2, \ldots, Y_6$, and the calculation method is as follows:

$$Y_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)} \times 0.99 + 0.01 \tag{3}$$

Step 2: Calculate the ratio of each index in each case. To obtain the variation of the index size, we calculate the proportion of the jth index in the ith case as follows:

$$p_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{n} Y_{ij}} \quad (i = 1, 2, ..., n; \; j = 1, 2, ..., 6) \tag{4}$$

Step 3: According to the definition, the information entropy of a data set can be calculated as follows:

$$E_j = -\frac{\sum_{i=1}^{n} p_{ij} \cdot \ln p_{ij}}{\ln n} \tag{5}$$

Apparently, $E_J \geq 0$. In the actual evaluation using the EWM, $p_{ij} \cdot \ln p_{ij} = 0$ is generally set when $p_{ij} = 0$ for the convenience of calculation.

Step 4: Determine the weight of each indicator. We record the information redundancy and the of the jth indicator as D_j and the weight of the jth indicator under EWM as w_{2j}. The calculation method is as follows:

$$D_j = 1 - E_j \tag{6}$$

$$w_{2j} = \frac{D_j}{\sum_{j=1}^{n} D_j} \tag{7}$$

After calculation, the calculated weight vector of EWM is (0.059, 0.229, 0.666, 0.005, 0.036, 0.003).

4.3.3 The Calculation of the Combination Weight

Combining the subjective weight w_{1j} and the objective weight w_{2j}, we can find a combination weight w_j (j=1, 2, . . . , 6). Premising that w_j should be as close as possible to w_{1j} and w_{2j}, we can calculate it according to the principle of minimum relative information entropy and Lagrange Multiplier method:

$$w_i = \frac{\left(w_{1j} w_{2j}\right)^{0.5}}{\sum_{j=1}^{6} \left(w_{1j} w_{2j}\right)^{0.5}} \quad (i = 1, 2, ..., 6) \tag{8}$$

After calculation, the combination weight vector is (0.071, 0.238, 0.597, 0.032, 0.043, 0.016).

4.4 The Evaluation of "Momentum" and the Visualization of the Model

Step 1: Evaluate PM (point momentum). We use the combination weight to calculate the score of each PM which shows the players' performance when obtaining or losing a point.

Step 2: Eliminate the influence of advantages. For each player, we calculate his average momentum score as a server (SAM) or a receiver (RAM) in advance. When calculating the comprehensive momentum score, we divide SAM into the server's PM score and RAM into the receiver's PM score so that the results will not contain the advantage for server and personal capability. After such calculation, the change of PM in the first set of the match between Carlos Alcaraz and Nicolas Jarry can be exhibited as follows:

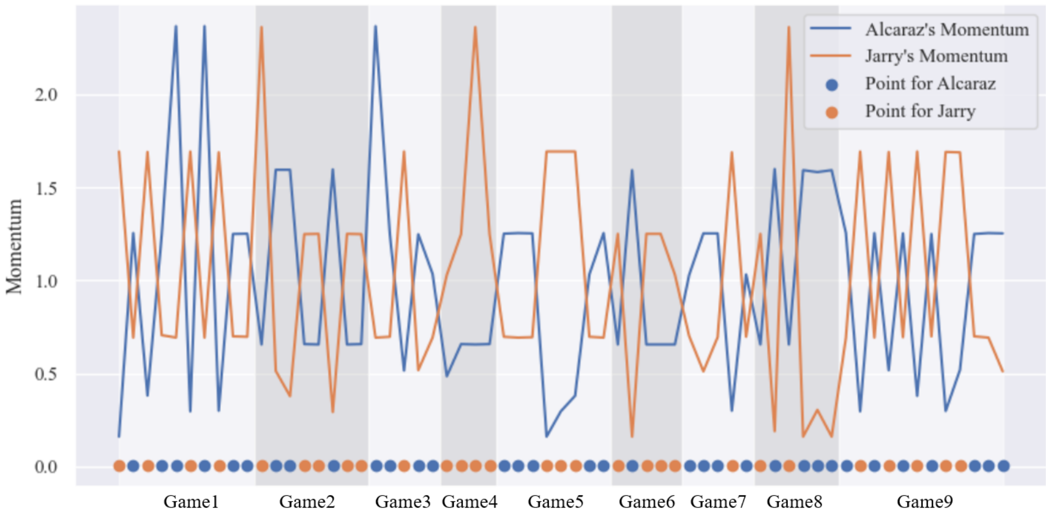


Figure 3: The Change of PM in the First Set of the Match between Alcaraz and Jarry

Step 3: Consider the temporal trend. According to **Assumption 3**, the influence of personal CM will decline with time. We therefore process the afterward PM scores of a certain CM by multiplying the Gaussian function, the decaying function, as listed:

$$f(x) = kae^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{9}$$

Taking $a = 1$, $\mu = 0$, $\sigma = 1$, and k as the normalized coefficient we regard x as the serial number of the afterward PM (x=0, 1,. . . ,9). Defining that personal CM reflects the afterward performance, we add current and future PM scores to get the CM score. According to the CM scores in different periods, we draw a figure to depict the flow of momentum in a set:

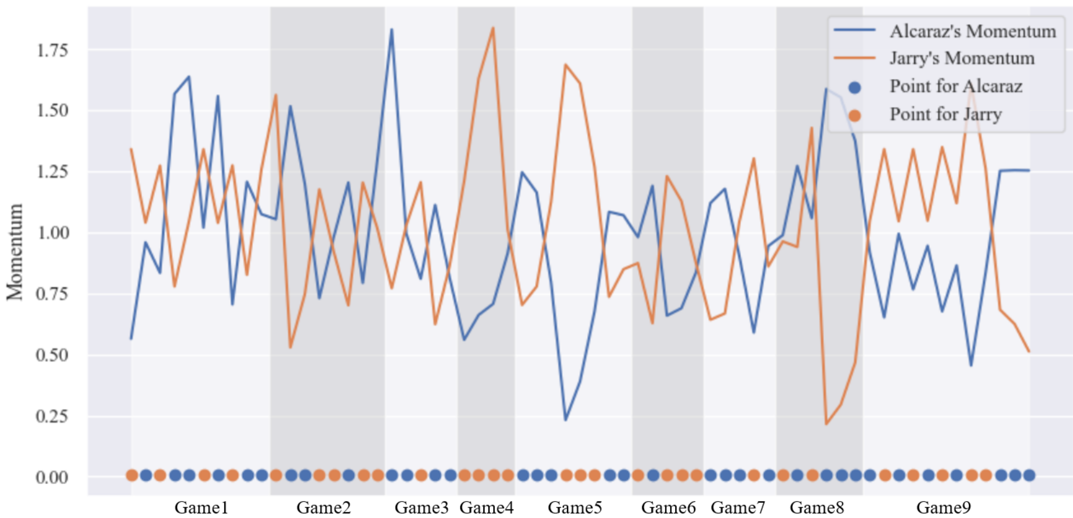


Figure 4: The Change of CM in the First Set of the Match between Alcaraz and Jarry

It is easy to figure out that momentums are commonly in a high degree when a player gets a point and vice versa.

5 Task 2: "Momentum" or Randomness?

From a tennis coach's perspective, there is little relation between "momentum" and performance during a match. Instead, he assumes that "swings in play and runs of success by one player are random". Based on the result of our model, we believe that "momentum" absolutely exists and makes a significant difference to the flow of a match.

To assess this opinion, we use SPSS to conduct an autoregressive test, which cannot confirm the autocorrelation of the momentum. In other words, we have no evidence to determine the randomness of the results. As a consequence, we take the data in a match as the total and separate it into two samples according to a series of break points from one player. One, recorded as A for player 1 and B for player 2, is half of the afterward momentum data near the break point, and the other is the half farther recorded as C for player 1 and D for player 2. Sample A and C are in Group 1, and B and C, Group 2.

After the selection, we apply the following methods to conduct further tests:

1. **Shapiro-Wilk (S-W) Test**

   To filtrate the appropriate test methods, we first use the S-W Test to see whether the selected samples conform to normal distribution.

   The basic principle of the S-W Test is to test the null hypothesis that a sample X_1, $\ldots, X_n$ came from a normally distributed population based on the covariance between the observed values in the sample and the sample mean. It concentrates on the shape of the data distribution within the sample, but does not involve a comparison with a specific theoretical distribution. It examines whether the data points are symmetrically distributed around the sample mean.

   The test statistic is

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{10}$$

   where

   - $x_{(i)}$ is the ith order statistic.
   - $\bar{x} = \dfrac{(x_1 + \ + x_n)}{n}$ is the sample mean.

   The coefficients a_i are given by:

$$(a_1, \ldots, a_n) = \frac{m^T V^{-1}}{C} \tag{11}$$

   where C is a vector norm: $C = \|V^{-1}m\| = \left(m^T V^{-1} V^{-1} m\right)^{1/2}$, is made of the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution; finally, V is the covariance matrix of those normal order statistics.

The result of S-W Test are as follows:

Sample	P-value	Comparison
A	$\mathbf{1.903 \times 10^{-23}}$	0.05
B	$\mathbf{4.499 \times 10^{-24}}$	0.05
C	$\mathbf{7.348 \times 10^{-20}}$	0.05
D	$\mathbf{1.026 \times 10^{-19}}$	0.05

Table 3: The result of S-W Test

For each sample, the p-value is less than the significance level of 0.05. Therefore, we can reject the assumption that the sample comes from the normal distribution, which means that samples A, B, C, and D do not obey the normal distribution.

2. **Kolmogorov-Smirnov (K-S) Test**

To determine whether the distribution of Sample Y and Sample N is the same, we performed the KS test. The basic principle of the KS test is to compare the maximum difference between the observed data distribution (empirical distribution function) and the cumulative distribution function (CDF) of the theoretical distribution. In the case of two samples, the KS test can be used to test whether the one-dimensional probability distribution of the two samples is different.

The test statistic is

$$D_{n,m} = \sup_x \left| F_{1,n}(x) - F_{2,m}(x) \right| \tag{12}$$

where

- $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of the first and the second samples respectively.
- sup is the supremum function.
- n and m are the sizes of the first and second sample respectively.

For large samples like Y and N, the null hypothesis is rejected at a level α if

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{n \cdot m}} \tag{13}$$

The value of of $c(\alpha)$ is given in generally by $c(\alpha) = \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1}{2}}$, so that the condition reads

$$D_{n,m} > \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1 + \frac{m}{n}}{2m}} \tag{14}$$

We conducted two tests and the p-values of the two groups are $\mathbf{2.423 \times 10{-}9}$ and $\mathbf{6.383 \times 10^{-4}}$, both less than the significance level of 0.05. Therefore, we reject the null hypothesis that two samples from one group have the same distribution, which means there are differences between A and C from Group 1 as well as B and D from Group 2.

3. **Mann-Whitney U (M-W U) Test**

   Since the compared variables between A and C or B and D are continuous and the data is non-normal distribution, we choose M-W U Test to detect whether the shapes of the two distributions are similar by ranking the observed results.

   Mann-Whitney U test, also known as Wilcoxon rank sum test, is a non-parametric statistical test. In essence, the Mann-Whitney U test is to test whether the two samples have the same shape in terms of data. Therefore, the hypothesis of the Mann-Whitney U test is :

   - $\boldsymbol{H_0}$: The distribution shapes of the two sample data are similar.
   - $\boldsymbol{H_1}$: The distribution shapes of the two sample data are not similar.

   Before calculating the test, we selected the significance level $\boldsymbol{\alpha = 0.05}$. The values of the whole sample are sorted in order from small to large, and a test statistic $\boldsymbol{U_i}$ is generated according to the ranking and sample size by

$$\boldsymbol{U_i = n_1 n_2 + \frac{n_i\,(n_i + 1)}{2} - R_i \quad (i = 1, 2)} \tag{15}$$

   where

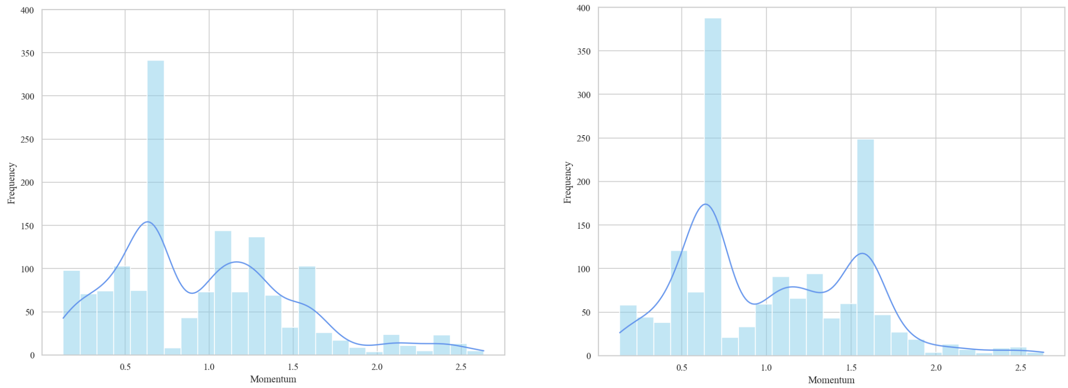
   - $\boldsymbol{n_1, n_2}$ are the sample size of each of the two samples.
   - $\boldsymbol{R_1, R_2}$ are the sum of the rankings of the two samples.

   Choose a smaller U value as the M-W U test statistics and check the reference table of the critical value to determine the critical value $\boldsymbol{U_\alpha}$. When $\boldsymbol{U \le U_\alpha}$, we can reject $\boldsymbol{H_0}$ and accept that the two sample shapes are not similar.

   During the test of our samples, we transfer the $\boldsymbol{U_\alpha}$ into p-value and compare it with 0.05. As the p-values of Group 1 and Group 2 are separately $\boldsymbol{6.092 \times 10^{-6}}$ and $\boldsymbol{0.006}$, both less than 0.05, we can reject the null hypothesis that the two samples from one group is similar.

Through testing and observation, results prove that the distribution of the two samples in each group is different, indicating that bp has an impact on momentum. Since our two groups of samples are divided by break points, it shows that bp affects momentum. Since momentum is defined according to the results of the competition and the performance of the players, swings in play and runs of success are not completely random.

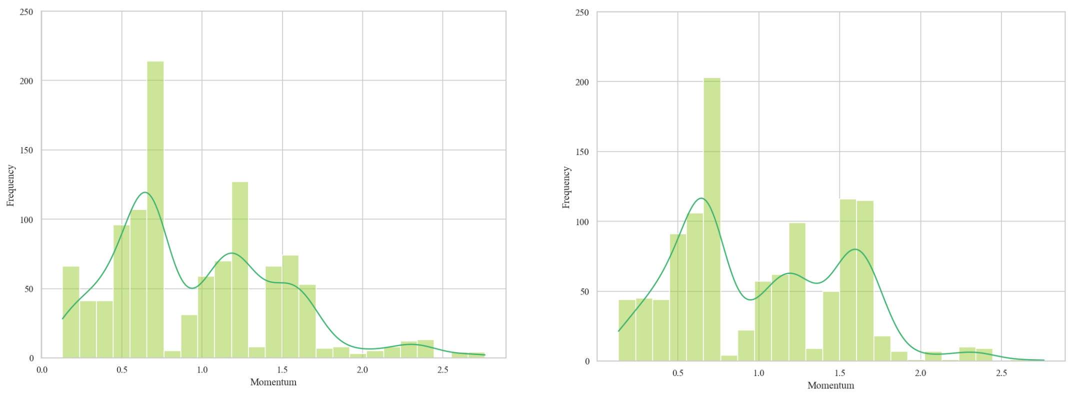
Through the graphics, we can see clearly the differences in the distribution of the samples in these two groups:



(a) The Frequency of Sample A

(b) The Frequency of Sample C

Figure 5: The Conparison of Sample A and Sample C



(a) The Frequency of Sample B

(b) The Frequency of Sample D

Figure 6: The Conparison of Sample B and Sample D

6 Task 3: Predicting the Flow of Play

To predict the fluctuation of momentum in the match as accurately as possible, we take a match as a unit and differentiate the player 1 and player 2 in it. We calculate the Person Correlation Coefficient between the momentum to be predicted and the 40 variables provided in the data of all the matches. Finding that the influential factors and the influencing degree on players' performance are changing with the chosen object, we select the 2023-wimbledon-1406 match and then make the following heat map of its Person Correlation Coefficients:

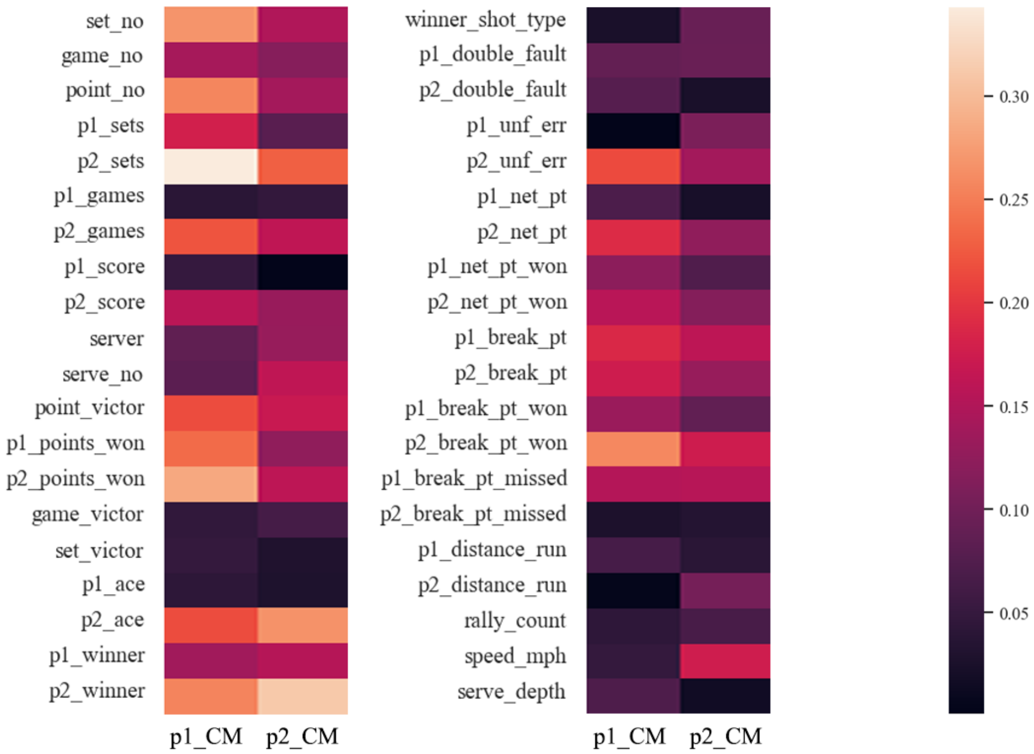


Figure 7: Thermodynamic Diagram

From the thermodynamic diagram, we can see that momentum is related to many variables. Because of the factors of great diversity and the data of small scale, we choose XGBoost as the model according to its characteristics.

6.1 The Establishment of Swings Prediction Model Based on XGBoost

XGBoost (eXtreme Gradient Boosting) is a gradient boosting algorithm, which is usually built on the basis of the decision tree. Because of the ability to select features, it can deal with high-dimensional data and a large number of features, and automatically filter out the most important characteristics. For small sample sets, XGBoost uses regularization techniques and pruning methods to reduce the risk of overfitting and improve the generalization performance of the model.

Therefore, we choose XGBoost to predict the momentum to obtain the flow of play. Considering that there are two-minute rest between each two sets, which may cause a change in the player's momentum, we begin our prediction from the 3rd point and delete the last point if each set to eliminate the influence of rest time.

Step 1: Initialize the model with a constant value. For the prediction of momentum, we use the mean of all samples as the initial model.

Step 2: Construct the objective function. We choose MSE (mean square error) as the loss function $\boldsymbol{l}$ to calculate the loss between the predicted value and the true value. Meanwhile, the complexity of the model is controlled by the regularization item in case that the model is over-fitted. Therefore, the objective function can be constructed as:

$$\mathcal{L}(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \tag{16}$$

Step 3: Repeat the training process. The model will generate a decision tree which is established with the determination of the split node to minimize the XGBoost objective function. As the objective function is very complex, XGBoost uses the Taylor series to approximate it and reduce the time of calculating the objective function by transforming the problem from calculating the objective function to calculating the variation of the objective function after the tree is generated.

The simplified objective function is:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{17}$$

Step 4: Update and output the model. The model will repeat the establishment and judgment process of decision tree updates the model. When reaching the preset number of iterations, which is 20, or failing to improve the model in a new tree, the calculation stops and the latest model is output.

6.2 Parameter Optimization based on DE Algorithm

After the initial establishment of the model, we use the **DE** (Differential Evolution) algorithm to optimize the hyperparameters, in order to improve the accuracy of the model.

The idea of differential evolution is derived from the early proposed genetic algorithm (GA), which simulates the crossover, mutation, and reproduction in genetics to design genetic operators. In DE algorithm, the population of DE is driven by the mutation and selection process. The mutation process, including mutation and crossover operations, is designed to exploit or explore the search space, while the selection process is used to ensure that the information of promising individuals can be further utilized. The mutation vector of DE algorithm is generated by the parent difference vector, crossed with the parent individual vector to generate a new individual vector, and directly selected with its parent individual. Based on the ideas above, we optimize the hyperparameters of the model.

Step 1: Initialize. Taking (learning rate, number of iterations, proportion of samples selected when training each tree, regularization parameter 1, regularization parameter 2, stopping condition) as XGBoost's hyperparameters, the model will generate 30 individuals randomly and evenly in all solution spaces and each individual is a 6-dimensional vector as follows:

$$X_i(0) = \left(x_{i,1}(0),\ x_{i,2}(0),\ \ldots,\ x_{i,6}(0)\right) \quad (i = 1, 2, \ldots, 30) \tag{18}$$

The jth-dimensional value of the ith individual is:

$$X_{i,j}(0) = L_{j\_min} + rand(0,1)\left(L_{j\_max} - L_{j\_min}\right) \tag{19}$$

where $i = 1, 2, \ldots, 30$ and $j = 1, 2, \ldots, 6$.

Step 2: Mutate. In the gth iteration, the model will select the optimal solution $X_{best,G}$ from the population in the way of 'selecting the best', and two different individuals $X_{r1,G}$ and $X_{r2,G}$ in random from the remaining samples. The mutation vector generated by these three individuals is:

$$V_{i,G} = X_{best,G} + F \times \left(X_{r1,G} - X_{r2,G}\right) \tag{20}$$

where $\left(X_{r1,G} - X_{r2,G}\right)$ is a difference vector and F is a scaling factor which is a random value of $[0.5, 1]$ in our algorithm.

Step 3: Cross. For the variation individuals obtained by the target individual, only part of the obtained variation parameters is retained, and the others are restored to the original values. Here we use the binomial crossover method, which is a random selection according to the preset probability CR, but at least one mutation result is retained. The binomial cross formula we used is as follows:

$$u_{i,G}^{j} = \begin{cases} v_{i,G}^{j}, & \text{if } rand_j[0,1] < CR \text{ or } j = j_{rand}; \\ x_{i,G}^{j}, & \text{else} \end{cases} \tag{21}$$

where

- $CR = 0.7$,
- j_{rand} is a generated random number that is used to ensure that at least one post-mutation parameter is retained.

Step 4: Choose. The model compares the target individual and the mutant individual and keeps the better one as follows:

$$X_i(g+1) = \begin{cases} v_i(g), & f(V_i(g)) < f(X_i(g)) \\ x_i(g), & \text{else} \end{cases} \tag{22}$$

For each individual, the obtained solution is better than or equal to the individual through mutation, crossover, and selection to achieve the global optimal.

6.3 The Result of Prediction

According to the research above, we develop a model to predict players' momentum. The result of the test set predicted by the model and the true value of the current momentum are compared as follows:

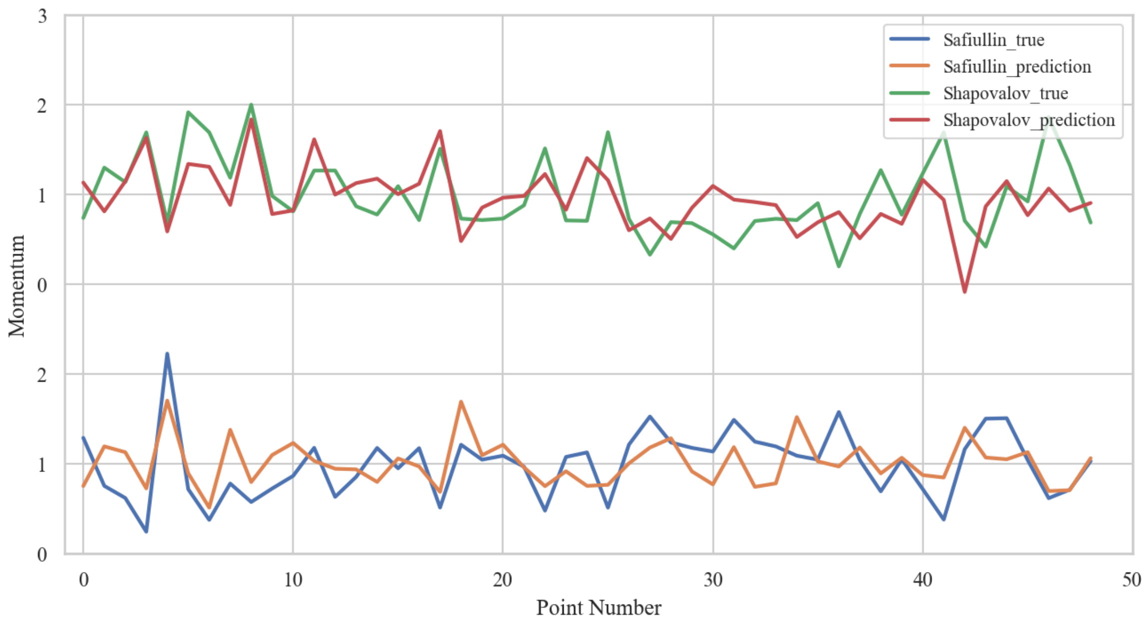
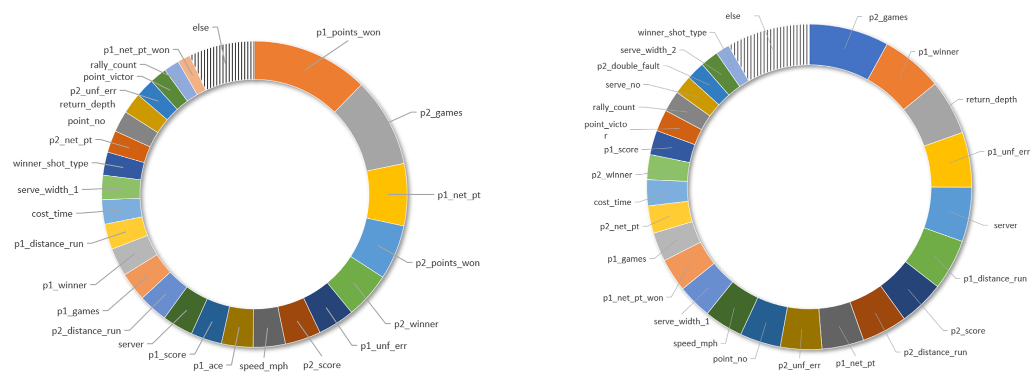


Figure 8: The Prediction of Test Set

Using XGBoost (eXtreme Gradient Boosting) and DE (Differential Evolution) algorithm to build and optimize our prediction model, the MSE (Mean Square Error) of player 1 and player 2 improved from 0.0828 to 0.0619 and 0.0861 to 0.0743, meanwhile, the MAE (Mean Absolute Error) of from 0.219 to 0.196 and 0.221 to 0.238.

Additionally, after the optimization of the differential algorithm, the importance of p1 and p2 features output by our XGBoost model is shown in the following figure:



(a) The Influential Factors of Player 1 (b) The Influential Factors of Player 2

Figure 9: The Influential Factors of Two Players in a Match

where "else" contains the 15 least important features.

According to these two circular graphs, we can conclude that the 3 elements that have the greatest influence on momentum are: the number of games won by the opponent (p2_games), whether the previous point is scored by the player(p1_point_won), and their own unforced errors (p1_unforced_error).

6.4 Suggestions for Players

Aiming at players themselves, they can

- set psychological cues during the game to reduce the impact of mentality on momentum. Players can set positive goals for themselves during the game, such as "I try to play my part technically and tactically", rather than "I want to avoid being the worst of the two players".
- adjust the potential momentum and simulate the momentum change to find the optimal rhythm in training through the rational allocation of energy and rhythm

Aiming at their opponents, they can

- collect opponent player's recent game information before the game and formulate a targeted confrontation plan to perform targeted exercises. Because different players are affected by different momentum, through the model can be drawn on the other side of the opponent's most influential factors, such as the other side for the break point and the sensitivity of the ace point.
- take measures to weaken the momentum of the opponent during the game. For instance, the momentum can be weakened through a victory roar or victory gesture after winning a break point.

7 Task 4: The Analysis of the Model

7.1 The Accuracy of the Model

We use the model to predict Match 2023-wimbledon-1305, where the prediction result of the second set is shown in the following figure:

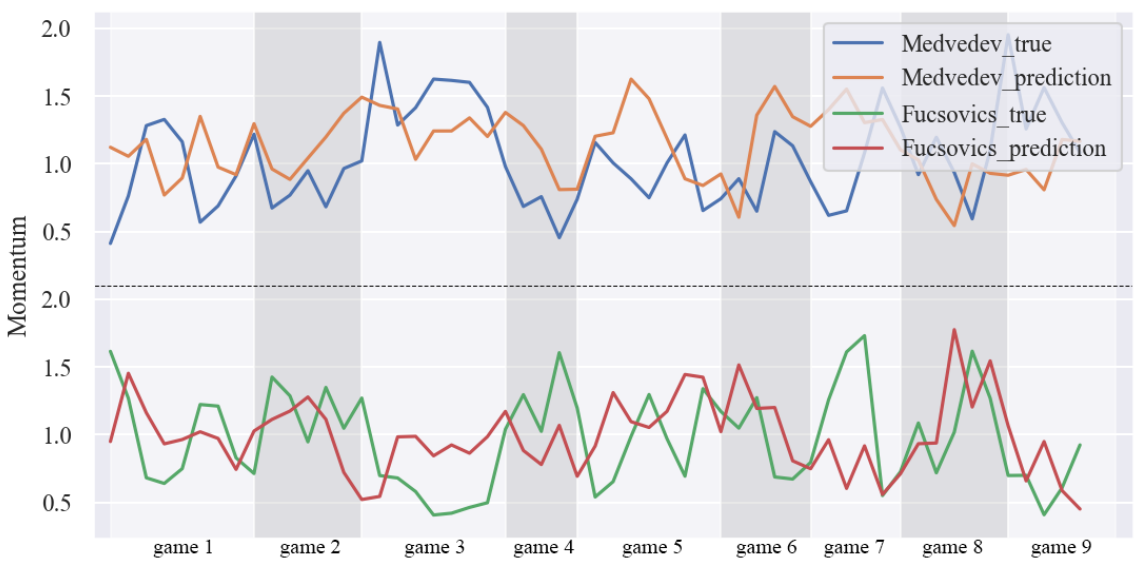


Figure 10: The Prediction of Other Tennis matches

In this prediction, $MSE_{player1}$ **= 0.183** and $MSE_{player2}$ **= 0.163**, which means the result is relatively accurate but still has less accuracy than the test set. The reason why the different predicted effects between data sets may be that different players are imposed different degrees of influence by momentum, which is in line with our conclusion that the influence of momentum is individualized.

7.2 The Improvement of the Model

1. The model needs to include the factor of whether the player is at home or away, because the audience's encouragement or boos are likely to have a greater impact on the generation of players' momentum.
2. The model needs to consider the outcome of previous matches between players and opponents, the results of which may have an impact on the initial psychological momentum of the players.
3. The model also needs to include some sudden factors, such as the weather factors of the day, the player 's injury and so on.

7.3 The Globality of the Model

The model not only has a certain degree of fitting for the performance of the men 's game in terms of parameters, but its framework is also universal for other games.

1. **For the women 's game,** we can retrain the parameters to achieve a more accurate prediction effect.
2. **For other game systems,** we need to take into account the impact of objective conditions, such as competition duration, site environmental conditions, etc.
3. **For other courts surfaces,** we need to consider the influence of site type (red soil, grass, or hard ground), ground hardness, and ground flatness on running distance.
4. **For other ball games,** taking table tennis as an example, we can increase the characteristics that affect table tennis, such as the rotation speed of the ball, and delete tennis-specific indicators, such as "serve_depth", "receice_depth", and "net_pt_won" when calculating momentum.

8 Model Evaluation and Further Discussion

8.1 Strengths and Weaknesses

8.1.1 Strengths

1. The combination of subjective and objective setting weights solves the problem of different information quantities and importance of different indicators.
2. The automatic method is used to adjust hyperparameters, which enhances the accuracy of the model and improves training speed.
3. The objective function of our model contains a regularization item to its for reducing the risk of overfitting caused by insufficient data volume.

8.1.2 Weaknesses

1. Experts' opinions are needed to give higher quality evaluation criteria when setting weights.

2. For different players and different types of games, the accuracy of prediction is insufficient or even unpredictable, and more data support is needed.

9 Conclusions

In this paper, we use the data of the Wimbledon game to quantitatively study the new concept of ’momentum’.

Firstly, we defined and quantified the momentum of the player according to the performance of the player after the current point. Then, through the momentum data given by the model, we prove that the fluctuation in the game is affected by momentum, instead of randomness. Then we developed a model to predict the current momentum (CP) according to the data before that certain point, and figured out several main influential factors of momentum. In the meantime, we tested the portability of the model and proposed future improvement ideas for the model. Finally, we applied the research results to practice and make suggestions to coaches and players according to the rules and models studied.

References

[1] Bergeron, Michael F., et al. "Tennis: a physiological profile during match play." *International journal of sports medicine,* 1991, pp. 474-479.

[2] Den Hartigh, Ruud JR, and Christophe Gernigon. "Time-out! How psychological momentum builds up and breaks down in table tennis," *Journal of sports sciences,*2018, pp. 2732-2737.

[3] Fu, Qiang, Changxia Ke, and Fangfang Tan. "“Success breeds success” or “Pride goes before a fall”?: Teams and individuals in multi-contest tournaments," *Games and Economic Behavior,* 2015, pp. 57-79.

[4] Gernigon, Christophe, Walid Briki, and Katie Eykens. "The dynamics of psychological momentum in sport: The role of ongoing history of performance patterns," *Journal of Sport and Exercise Psychology,* 2010, pp. 377-400.

[5] Gilovich, Thomas, Robert Vallone, and Amos Tversky. "The hot hand in basketball: On the misperception of random sequences," *Cognitive psychology,* 1985, pp. 295-314.

[6] Iso-Ahola, Seppo E., and Charles O. Dotson. "Psychological momentum: Why success breeds success," *Review of general psychology,* 2014, pp. 19-33.

[7] Meier, Philippe, et al. "Separating psychological momentum from strategic momentum: Evidence from men’s professional tennis," *Journal of economic psychology,* 2020, pp. 1-10.

[8] Reid, Machar, Stuart Morgan, and David Whiteside. "Matchplay characteristics of Grand Slam tennis: implications for training and conditioning," *Journal of sports sciences,* 2016, pp. 1791-1798.

[9] Richardson, Peggy A., William Adler, and Douglas Hankes. "Game, Set, Match: Psychological Momentum in Tennis," *The Sport Psychologist,* 1988, pp. 69-76.

[10] Weinberg, Robert, and Allen Jackson. "The effects of psychological momentum on male and female tennis players revisited," *Journal of Sport Behavior,* 1989, pp. 167-179.

Memorandum

To: coaches

From: Team #2421624

Subject: The Critical Role of "momentum"

Date: Monday, February 5, 2024

Dear trainer,

We are a group of college students with enthusiasm for ball games. Last summer, we gathered together and watched the live television relay of the "Championships Wimbledon 2023" just as hundreds of thousands of tennis fans. When it comes to this tournament, the greatest attraction must be the final of men's singles between the experienced Serbian defending champion, Novak Djokovic, and the most potential young Spanish player, Carlos Alcaraz who ranked first at that time.

Alcaraz showed with his excellently-placed forehand to win his first set of the match, but the resilience for which the 36-year-old Djokovic has become so well-known saw him through the danger in the opening game. Then he heaped the pressure back onto the Spaniard in the very next game in gusty conditions. Nonetheless, the Spaniard prevailed after nearly five hours on Centre Court, eventually winning 1-6 7-6 (8-6) 6-1 3-6 6-4, and jubilantly fell onto the grass in celebration before volleying a tennis ball into the crowd, through which he became the third-youngest Wimbledon champion in the Open Era and added a second major title to his resume after winning the US Open last year.

A series of turning points aroused our curiosity. We have heard that you have always attributed the swings of play to randomness, but we still postulate that "momentum" exists in every second of a match. To confirm our guess, we found several evidence through our models.

Firstly, we defined "momentum" with the division into Point Momentum (PM) and Current Momentum (CM). PM represents the momentum of a certain player when a point occurs, and CM is the symbol of the momentum that has an impact on the future occurrence of several points. Momentum reflects on and can be evaluated by future performance. Also, it can be influenced by abundant factors.

Secondly, we tried to depict "momentum" by giving scores to it. According to the definition, future performance is the reflection of momentum so momentum can be measured by combining different performing indicators with a weight set. To determine the combination weight, we applied the Analytic Hierarchy Process and Entropy Weight Method to balance the subjective and objective evaluation. There are 3 steps to get the final momentum score.

1. We obtained all the PM scores of each player through the method above.
2. To eliminate the influence of personal capability and advantage for server, we calculated the average point momentum of a player as a server and a receiver respectively, and divided the results into each original PM accordingly.
3. Finding momentum's property of time-lasting influence, we use future 10 PM scores with damped parameters from the Gaussian function to form the CM score at each point.

Thirdly, we tried to predict "momentum" and in turn track the flow of play. The data set we use contains diverse indicators, each of which seems relative to momentum. In order not to miss any influential factor, we examined the Pearson correlation coefficient between the CM scores of two payers in a match and 40 variables and exhibited them with the thermodynamic chart. The result shows that all the variables have influences more or less on the value of CM. The large quantities of features and the insufficiency of data reminded us of the XGBoost (eXtreme Gradient Boosting) algorithm which we used to predict the CM of momentum. To optimize our prediction, we applied DE (Differential Evaluation) algorithm and achieved our goal with MSE (mean square error) of 0.0828 and 0.0861 and MAE (mean absolute error) of o.219 and 0.221 respectively in player 1's and player 2's prediction.

Fourthly, we tried to extend our models to all types of ball games. However, the model failed to live up to our expectations. There are 3 reasons for our disappointment:

1. The indicator data we own is inadequate, which means we cannot take into consideration all the subjective and objective factors, such as the mentality of players and the weather conditions at that time.
2. The model we developed is strongly related to the owner of the data in the training set. This situation is difficult to solve because of the close connection between data and players' performance.
3. When it comes to other ball games except tennis, there are different standards for each evaluating indicator. The complexity of terminology accounts for the condition.

Though there is a necessity for improvements on our model, we still can conclude with some inspiring suggestions for you coaches to train tennis players:

1. Pay close attention to the significant role of momentum in a match. Using the right to call the games, not only can you adjust your player's low momentum, but you can also grab the opportunity to weaken the opponent's momentum.
2. As the swings in play are not determined by randomness, you can analyze your players' previous match data to avoid similar errors in the next match. Also, you can generate strategies aimed at a certain opponent by finding drawbacks from that players' momentum data.

Above are all of our findings that we want to share with you. Hope you a successful career life as a tennis coach and numerous champions from your players!

Yours respectfully,
MCM Team #2421624

Appendices

The judging table of factors' importance in AHP is as follows:

Symbol	Description
1	Two elements have the same importance compared
3	Compared with the two elements, the former is slightly important than the latter.
5	Compared with the two elements, the former is obviously important than the latter.
7	Compared with the two elements, the former is extremely important than the latter.
9	Compared with the two elements, the former is far more important than the latter.
2,4,6,8	the intermediate value of the above adjacent judgment
the inverse of 1~9	The importance of the corresponding two factors when exchanging the order of comparison

Table 4: The judging table of factors' importance in AHP