Research on Prediction of Key Points in Tennis Matches Based on Neural Network and Momentum Analysis

Yiwen Lu *

School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China, 100044

* Corresponding Author Email: luyiwen@bjtu.edu.cn

Abstract. In modern sports, advanced data analysis sheds light on match dynamics. Our study focuses on momentum and leverage in Wimbledon 2023 matches using ATennision Model, a neural network model with Transformer architecture. We analyze winning transitions and key turning points by tracking real-time changes in winning percentages, our model obtains an average accuracy of 94.1%. Leverage, the impact of specific points on overall match outcomes, is a key metric. We quantify leverage effects and assess momentum's influence on match flow. Statistical tests challenge the belief that momentum is insignificant. We develop a Turning Point Prediction model for strategic guidance and achieve high predictive accuracy. Our study enhances understanding of tennis dynamics and equips players with strategies to capitalize on or counteract momentum shifts, aiming to decode winning transitions in this dynamic sport.

Keywords: ATennision Model, Momentum analysis, Turning point prediction.

1. Introduction

As data analytics mature, sports prediction not only enables athletes to improve their training methods, but also enables them to change their game strategies during competitions [1]. A typical example includes a scenario where a tennis player appears to be dominating a match, poised for an easy win, only for the dynamics to shift dramatically due to an unexpected game or point won by the opponent. Such dramatic turns, affecting numerous points or even entire games, are frequently ascribed to the concept of "momentum."

To solve the problem, we propose the model ATennision: Win Probabilities Model Based on Transformer. This is a prediction model of key points in tennis matches and can explore the impact of "momentum". This paper defined the momentum which indicates the flow of play as the exponential decay moving average of the derivation of the win probability. Meanwhile, this paper obtained and built a mask to predict the turning point according to the momentum graph. We evaluate ATennision model on the 2023 Wimbledon men's match data. The experimental results prove the superiority and versatility of our model.

2. Preliminaries

2.1. Data Preprocessing

After analyzing the problem, we collected 2023-wimbledon-points.csv from https://github.com/JeffSackmann, which includes data from all matches in men's singles, women's singles, men's doubles, and mixed doubles. It consists of 65 dimensions and 48,677 data entries.

To enhance the data quality, we use data preprocessing to make it suitable for modeling [2].

Data Cleaning: Before we build the model, we first process and remove the missing data. There are several columns of data that are completely missing, so we delete them. Moreover, we perform mean interpolation for quantitative data and mode interpolation for qualitative data.

Data Normalizing: To use the data as input for the Transformer and accelerate model convergence, we normalize continuous data.

Data Encoding: To use the data as input for the Transformer and accelerate model convergence, we perform one-hot encoding on discrete data.

2.2. Positional Encoding

Positional encoding is a crucial component of the Transformer architecture introduced in the seminal paper "Attention is All You Need" by Vaswani et al. (2017) [3]. This mechanism addresses the inherent limitation of Transformers, which lack sequential information, by injecting positional information into the input embeddings.

In this work, the encodings are sine and cosine functions of different frequencies and are assigned to alternating columns of the pe matrix.

$$PE_{(pos,\,2i)} = \sin\left(pos/10000^{2i/d_{model}}\right), \tag{1}$$

$$PE_{(pos,\,2i+1)} = \cos\left(pos/10000^{2i/d_{model}}\right), \tag{2}$$

Finally, the positional encoding matrix is registered as a buffer tensor to ensure its synchronization with the model parameters during training.

2.3. Transformer

The self-attention mechanism can be represented by the following mathematical formulas:

Given an input sequence $X = (x_1, x_2, ..., x_n)$, where x_i is the i-th element of the input sequence and n is the sequence length, the self-attention mechanism assigns a weighted representation y_i for each input position x_i, which is a weighted combination of representations from all input positions. The computation of this weighted representation can be broken down into three steps:

Calculation of Query, Key, and Value: For each position xi in the input sequence, we compute its Query, Key, and Value:

$$\begin{gathered} q_i = W_q \cdot x_i,\ k_i = W_k \cdot x_i,\ v_i = W_v \cdot x_i, \\ where\ W_q, W_k\ and\ W_v\ are\ weight\ matrices. \end{gathered} \tag{3}$$

Calculation of Attention Scores: Compute the attention scores

$$\begin{gathered} Attention(q_i, k_j) = soft\max\left(\frac{q_i \cdot k_j}{\sqrt{d_k}}\right), \\ where\ d_k\ is\ the\ dimensionality\ of\ Key. \end{gathered} \tag{4}$$

Weighted Sum to Obtain Output: Use the attention scores to perform a weighted sum of the Values to obtain the output:

$$y_i = \sum\nolimits_{j=1}^{n} Attention(q_i, k_j) \cdot v_j \tag{5}$$

3. Methods

3.1. ATennision Model: Win Probabilities Model Based on Transformer

3.1.1. Design of Transformer

Fig.1 shows the overall workflow of the Transformer framework within the ATennision model. Originally in Transformer, it used Encoder as a feature extractor and then Decoder as a parser to output the result we want. We modified the Transformer to take out the Decoder and use only the Encoder as the feature extractor [4]. The extracted “features” are then used for the task of win probabilities prediction, through which the parameters are corrected.

Initially, an input data of 56 dimensions is provided. The encoder, which consists of N = 2 identical layers, processes this input data. Each layer within the encoder comprises two sub-layers. The first sub-layer implements a multi-head self-attention mechanism, while the second sub-layer is a simple position-wise fully connected feed-forward network. A residual connection is applied around each of these two sub-layers. Additionally, layer normalization is employed after each sub-layer.

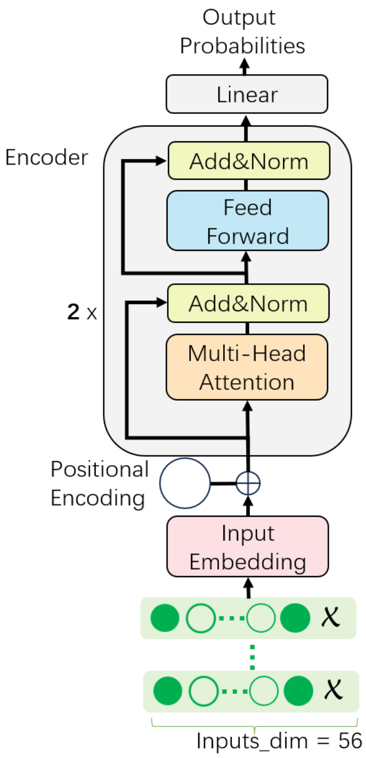

Figure 1. Transformer in ATennision.

3.1.2. Result of Transformer Predictions

We applied early stopping techniques to mitigate overfitting when working with larger models. We are confident in our model's ability to generalize, as evidenced by its consistent accuracy across our training, validation, and test sets (94.1%).

Our test results and comparisons are detailed in Table 1. "K-M Logit Elo" represents a point-based model [5] employed by Gollub [6] for outcome prediction using point-by-point data. "Logistic Regression" is a straightforward classifier [7] that utilizes the 13 Gollub models as input features. Our model achieved a 20% improvement over the best mid-match Gollub model and a 23% improvement over the basic predictor. Additionally, we provide results for our model at various data points after each set.

Table 1. Model Results.

Model	Net Accuracy	Set1	Set2	Set3	Set4
ATennision	94.1	90	90	96	95
K-M Logit Elo (Gollub)	76.5	NA	NA	NA	NA
Logistic Regression	73.4	NA	NA	NA	NA

Remarkably, our model's performance exhibits a noteworthy improvement, reaching 96% accuracy after Set 3, a substantial increase compared to the initial stages of the match or the overall accuracy. This indicates that our model has acquired crucial insights about the match dynamics beyond merely considering the score. The model's predictive capability for Set 1 of this match is quite poor. This is entirely reasonable given that a player's performance is influenced by multiple factors. However, the model's predictive accuracy greatly improves after Set 2, and it becomes increasingly accurate as more sets are played.

To see it more clearly, we choose the match: Alexandre Muller VS Arthur Rinderknech, whose final winner is Alexandre Muller. The win probabilities by points show in Fig.2.

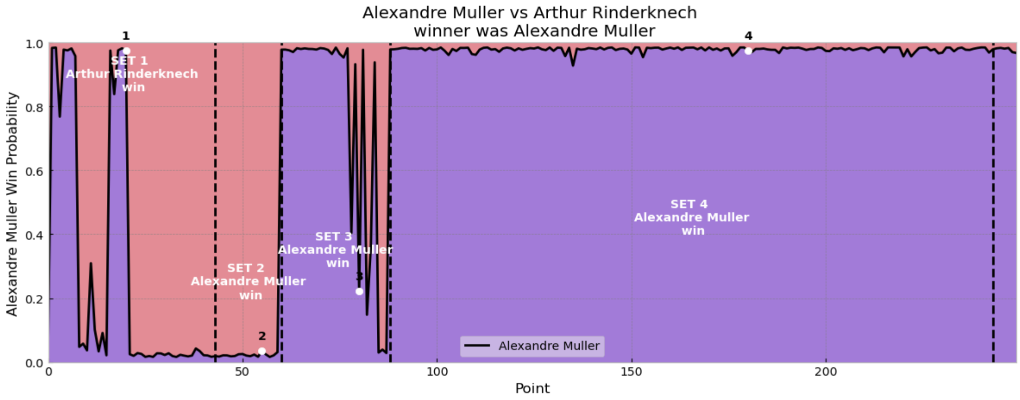

Figure 2. Win Probability Prediction.

The win probability curve experiences significant fluctuations in the early stages of Set 1 and towards the end of Set 3. This is attributed to these moments being critical phases in the match, characterized by continuous shifts in momentum between player1 and player2.

In order to provide a more intuitive representation of the match dynamics, we employ heatmaps to visualize the relationship between match phases and features. Show as Figure 3.

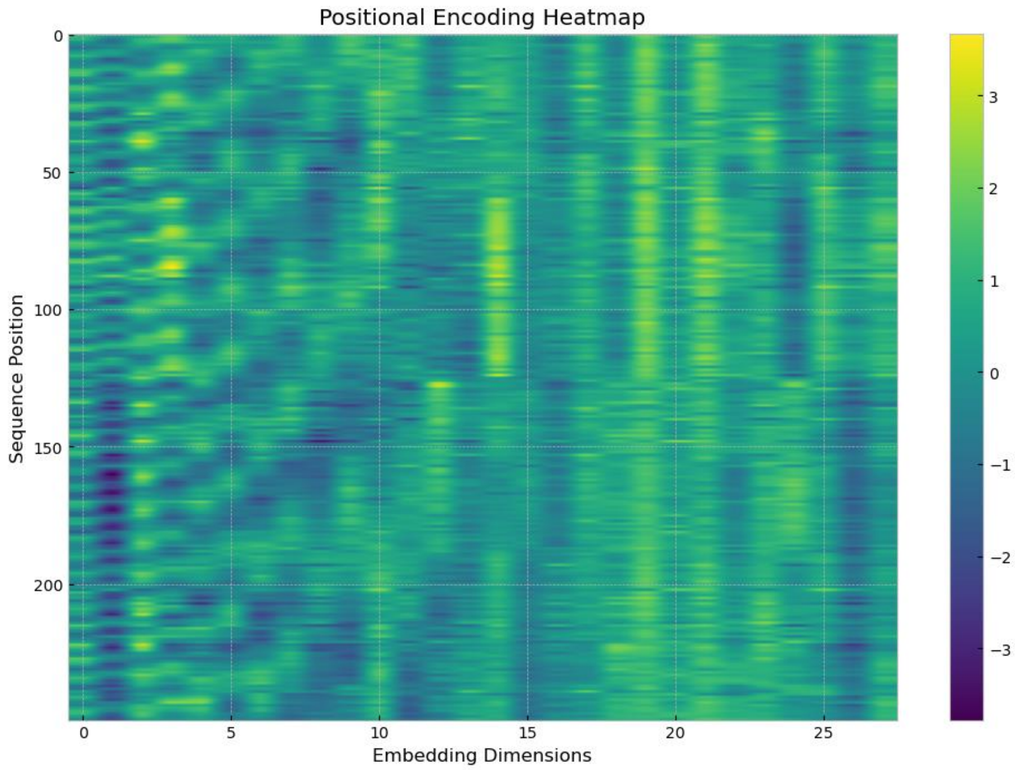

Figure 3. Heat Map.

Within a single match, certain consecutive rallies may exhibit similar statistical characteristics or occur during comparable stages of the game, which manifests as visual patterns in the heatmap with position encoding. The color variations in the heatmap convey the importance or impact of events at specific time points. Here, changes in color intensity can indicate shifts in a player’s momentum or leads and deficits in the score. The vertical axis (sequence position) of the heatmap represents time points within a tennis match, such as different stages of the game (e.g., beginning, mid-match, and end). The horizontal axis (embedding dimensions) represents various features related to the tennis match, such as player positions, ball speed, shot types (forehand, backhand), and scoring situations.

By visualizing feature importance in the Transformer model, shown as Figure 4, we found that the indicators set victor 1, set victor 2, p2 games, p2 points won, serve width, point no, p1 points won, p1 games, p2 score, set victor are the top 10 important features.

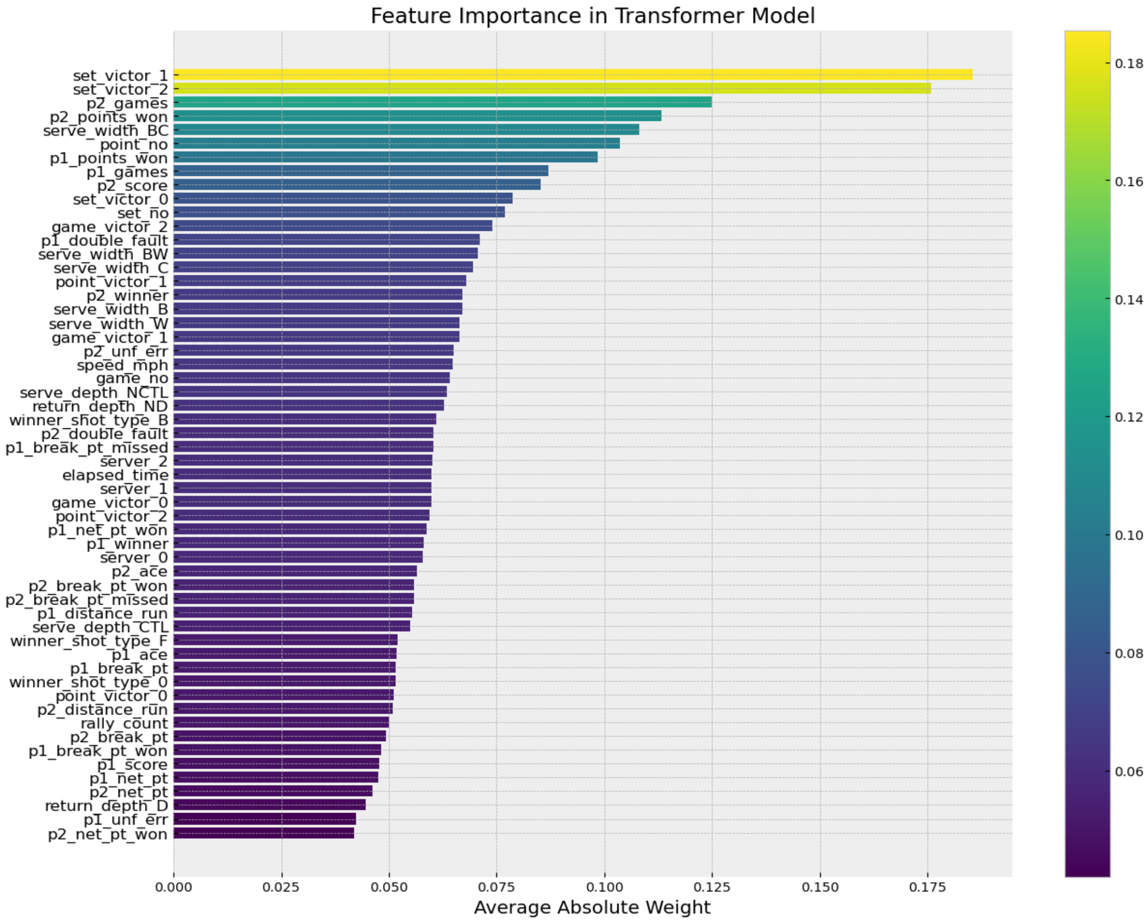

Figure 4. Feature Importance in the Transformer Model.

3.2. ATennision Model: Momentum Dynamics Model Based on Win Probabilities

3.2.1. Define the Momentum

After obtaining the win probability curve for each match, we calculate the derivative of win probability with respect to point number to obtain leverage:

$$X_t = \frac{d(Winprob)}{d(pointnumber)} \tag{6}$$

Leverage of the match: Alexandre Muller VS Arthur Rinderknech shows in Figure 5.

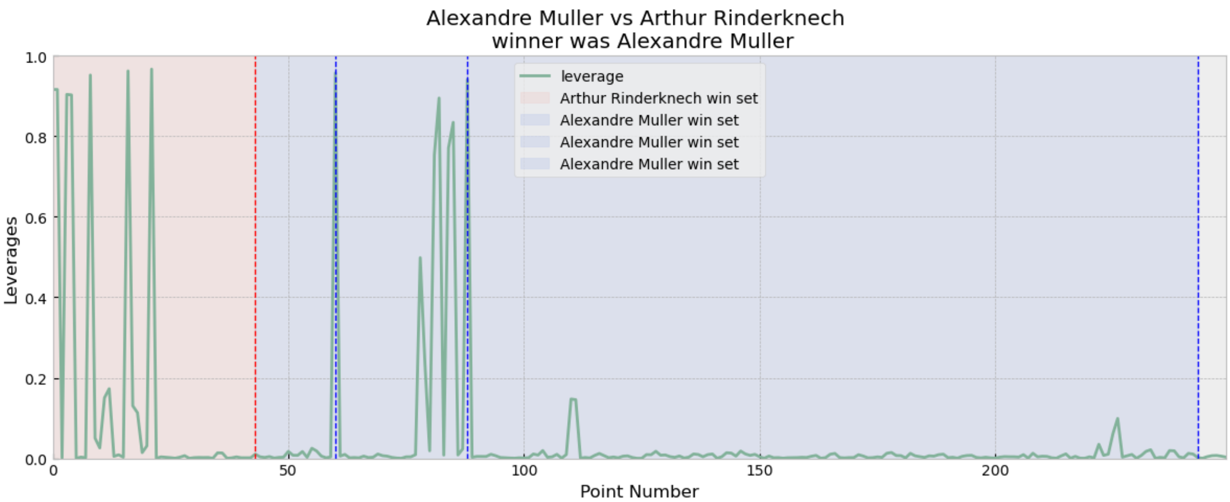

Figure 5. Leverage.

Next, we use an exponentially decaying moving average (EMA) on the derivative to define momentum [8].

$$EMA_t = \alpha \cdot X_t + (1-\alpha) \cdot EMA_{t-1}, \tag{7}$$

As we defined, momentum of the match: Alexandre Muller VS Arthur Rinderknech shows in Figure 6.

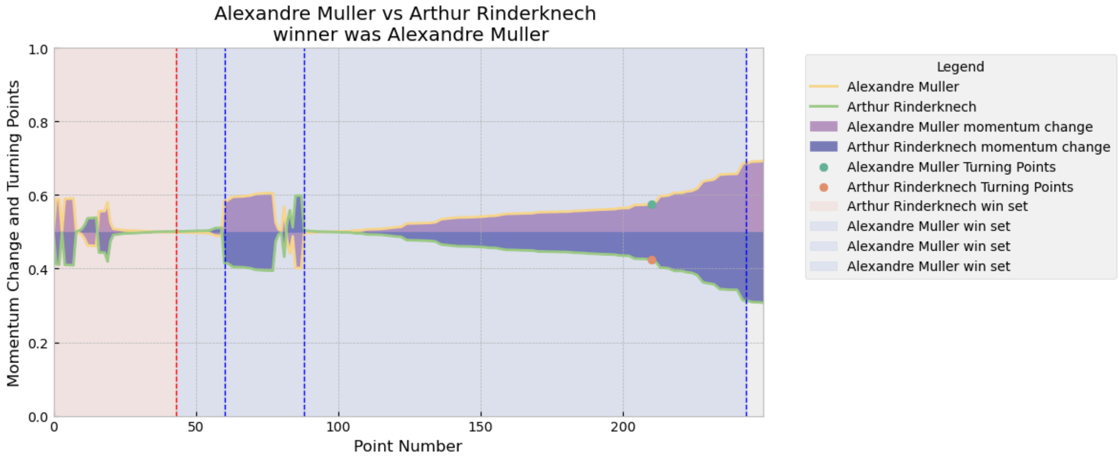

Figure 6. Momentum.

3.2.2. Results and Analysis of the Momentum

By simply subtracting the momentum of both sides, we visually obtained a more pronounced representation of the data known as the 'difference in momentum.' Show as Figure 7.

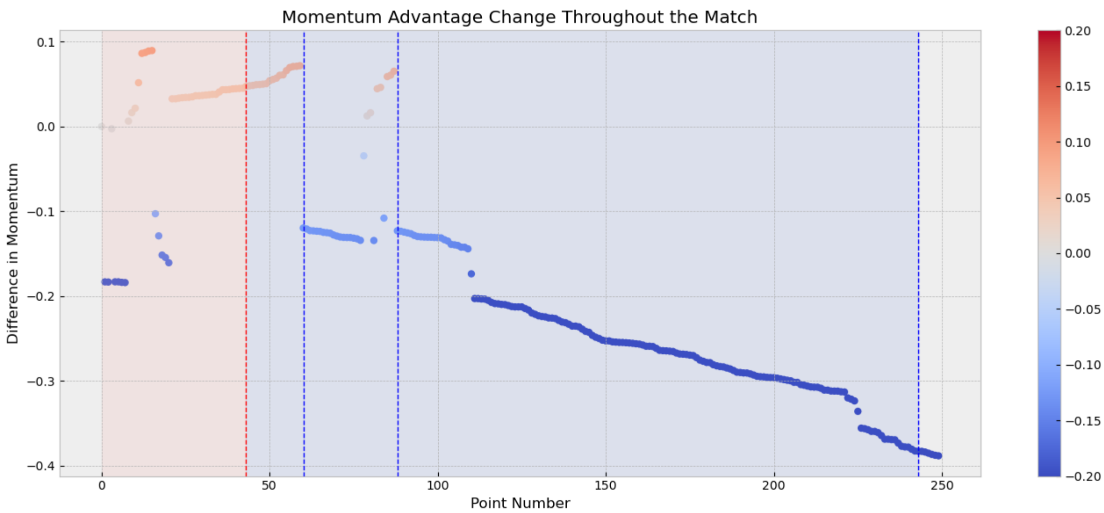

Figure 7. Difference in Momentum.

The transformations in the graph include: Continuous Decline in Momentum: The chart depicts a sustained decline in momentum over time, which may indicate that one player gradually establishes an advantage in the match, while the performance of the other player deteriorates. This could be attributed to various factors such as technical skills, physical condition, and psychological factors. Sudden Shifts in Momentum: Sharp changes in momentum at certain points may correspond to crucial moments in the match, such as a significant break point or a score after a tense, lengthy rally.

3.3. ATennision Model: Turning Point Prediction Model Based on PELT Algorithm

3.3.1. PELT Algorithm

To identify the swings of momentum in the match, we used PELT algorithm [9]. PELT (Pruned Exact Linear Time) algorithm is a statistical method used to detect turning points or mutation points in time series. It is primarily used in areas such as time series analysis and statistical modeling to discover locations of sudden changes or structures in data. We selected the match: Alexandre Muller VS Arthur Rinderknech to use the PELT algorithm. The results show in Figure 8.

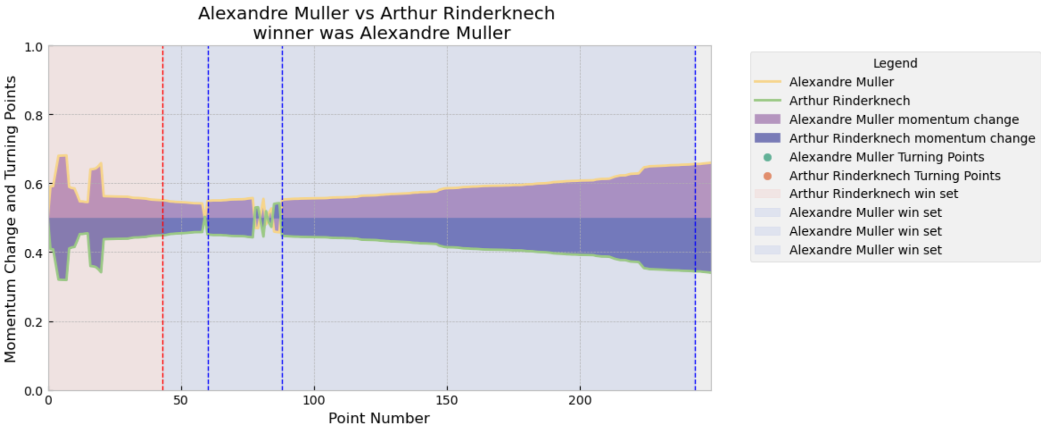

Figure 8. PELT Algorithm.

3.3.2. Run Test

In order to determine whether the detected turning points of race momentum are random or not, we propose the Run Test [10]. The Run Test is a statistical method employed to assess the presence of significant departures from randomness within a random sequence. It finds its primary application in the analysis of binary sequences, such as those consisting of 0s and 1s, with the objective of ascertaining the existence of non-random patterns or underlying structures.

In statistical hypothesis testing, the hypotheses are defined as follows:

Null Hypothesis (H0): The transitions in momentum during the match are random and do not significantly differ from the results of random simulation.

Alternative Hypothesis (H1): The transitions in momentum during the match are non-random and significantly differ from the results of random simulation. The Z statistic is employed as a quantitative measure to assess the randomness of the sequence. A larger absolute value of the Z-statistic implies a greater deviation of the sequence from randomness.

The p-value is utilized to determine whether there is sufficient evidence to reject the hypothesis of randomness. It informs us about the likelihood of observing the run distribution under random conditions. A low p-value suggests the presence of statistically significant non-random patterns within the sequence.

The Run Test results are showing in Table 2.

Table 2. Model Results.

match_id	Z1_m	Z2_m	Z1_tp	Z2_tp	P1_m	P2_m	P1_tp	P2_tp
1101	-12.73	-12.73	0.11	0.11	1	1	0	0
1102	-15.21	-15.21	0.09	0.09	1	1	0	0
1103	-16.47	-16.47	0.08	0.08	1	1	0	0
1104	-14.77	-14.77	0.00	0.00	1	1	0	0
1105	-13.72	-13.72	0.10	0.10	1	1	0	0

Therefore, we reject the null hypothesis, indicating that the postulation that swings in play and runs of success by one player are random is not supported by the data.

4. Conclusions

Massive point-by-point data in tennis matches provide the basis for predictive modeling and momentum exploration. In order to solve the inaccurate prediction caused by the large data dimension, we borrow the multi-head attention mechanism in the Transformer module to solve this problem. By developed the ATennision Model, we arrived an accuracy of 94.1%, which means our model is one of the best mid-match tennis prediction models in the world. By developed the Momentum Dynamics Model, we uncovered the mystery of momentum, which indicates that the unforced error is the most influential negative indicators and the won the point while at the net is the most influential positive indicators. Moreover, the momentum has continuity over small time windows, which indicates that momentum is greatly affected by its previous moment. By developed the Swings Detection Model, we found that the swing often happens in an important break point or a score after a long, tense round. When one player gradually establishes an advantage in the match, it often coincides with a decline in the performance of the other player. This can be attributed to various factors, including technical skills, physical condition, and psychological factors. However, when a set ends, there is a slight probability that the player who needed a mental reset may have a chance to make a comeback. The experimental results show that the ATennision model has good predictability and robustness, and has a certain practical application value.

References

[1] Daraojimba, Andrew Ifesinachi et al. "Data science in sports analytics: A review of performance optimization and fan engagement." *World Journal of Advanced Research and Reviews* (2024): n. pag.

[2] Kiran Maharana, Surajit Mondal, Bhushankumar Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings* (2022).

[3] Vaswani, Ashish et al. "Attention is All you Need." *Neural Information Processing Systems* (2017).

[4] Devlin, Jacob et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *North American Chapter of the Association for Computational Linguistics* (2019).

[5] Klaassen, Franc JGM, and Jan R. Magnus. "Forecasting the winner of a tennis match." *European Journal of Operational Research* 148.2 (2003): 257-267.

[6] Gollub, Jacob. *Producing Win Probabilities for Professional Tennis Matches from any Score.* Diss. (2019).

[7] X. Zou, Y. Hu, Z. Tian and K. Shen, "Logistic Regression Model Optimization and Case Analysis," 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2019, pp. 135-139.

[8] Klinker, Frank. "Exponential moving average versus moving exponential average." *Mathematische Semesterberichte* 58 (2011): 97-107.

[9] Gachomo, Dorcas. "The Power of the Pruned Exact Linear Time (PELT) Test in Multiple Changepoint Detection." *American Journal of Theoretical and Applied Statistics* 4 (2015): 581.

[10] Aslam, Muhammad. "The Run Test for Two Samples in the Presence of Uncertainty." *Journal of Big Data* (2023): 166.