

Unlocking Tennis Dynamics: A Hidden Markov Model Approach to Match Scoring

Yaxian Cao ¹, Xiaoyun Zhu ², Yatai Xu ^{3,*}

¹ College of Economics and Management, Tianjin Agricultural University, Tianjin, 300384, China

² College of Food Science and Biological Engineering, Tianjin Agricultural University, Tianjin, 300384, China

³ College of Engineering and Technology, Tianjin Agricultural University, Tianjin, 300384, China

* Corresponding author: xyt19102148603@163.com

Abstract. The aim of this paper is to develop a scoring process model for tennis matches that is applicable to other match scenarios and is able to recognize a player's performance at a specific moment in order to focus on the player's status and match flow. First, we search for specific moments and optimize the model based on the known key moments as well as assigning certain weights to the serving player based on the player's match data and performance metrics. We constructed a model of the scoring process of a tennis match based on a Hidden Markov Model (HMM) to identify specific moments. By converting the match scoring event data into a time series dataset and subsequently assigning service weights, we used the Hidden Markov Model to predict match points. Ultimately, we obtain the probability that each match point belongs to a winning state.

Keywords: Tennis, Hidden Markov Model, Match Point Prediction, Momentum Scoring.

1. Introduction

Tennis being a highly competitive sport, the course of a match is influenced by many factors, including a player's skill level, mental state, and key moments in the match. Understanding and predicting these critical moments is crucial for coaches, players and spectators. Therefore, it is important to develop a model that can accurately predict the course of the match and evaluate the players' performance [1-2].

Traditional scoring models for tennis matches usually only provide information on match scores and lack an in-depth understanding of match progress and player performance. In contrast, this study aims to develop a new scoring process model using hidden Markov model, combined with data visualization and momentum scoring, so as to deeply analyze the conduct of tennis matches and players' performance.

Through this model, we can better understand the conduct of tennis matches and the players' state changes, predict the key moments in the matches, and provide more comprehensive and precise tactical guidance for the players, as well as provide a richer viewing experience for the spectators. This study will deeply explore the application of Hidden Markov Model in tennis matches and analyze the accuracy and reliability of the model to provide new perspectives and methods for the analysis and evaluation of tennis matches [3].

2. The scoring process model for a tennis match

In developing the scoring process model for tennis matches, it should be applicable to other match scenarios and capable of identifying players' performance at specific times, in order to focus on player status and match flow. To achieve this, we first seek out specific moments. Based on known crucial moments and assigning certain weights to serving players, we optimize the model according to players' match data and performance indicators [4-5].

2.1. The match flow based on the Hidden Markov Model

To identify specific moments, we read some literature. Therefore, based on historical match data and tennis match rules, predicting the occurrence of match points, a scoring process model for tennis matches was established [6]. Taking the winning of previous sets as a feature, if a player has already won two sets out of the first four, the occurrence of a match point in the third set is directly predicted. If a player in a match point state successfully wins the game, they win the match. If the opponent saves the match point and wins the game, the match continues until the next match point occurs or enters a tiebreak or set-deciding game. By converting the event data of match scores into a time-series dataset, and subsequently assigning service weights, we employ the Hidden Markov Model for predicting match points. Ultimately, we obtain the probability of each match point belonging to the state of winning the match.

Hidden Markov Model (HMM) is a probabilistic model for time series data, describing the process where an unobservable state sequence is generated by a hidden Markov chain, and then each state generates an observation, resulting in an observed sequence [9]. Each position in the sequence can be seen as a moment in time. The diagram below illustrates the relationships in a Hidden Markov Model. Figure 1 shows the relationship of the hidden Markov model.

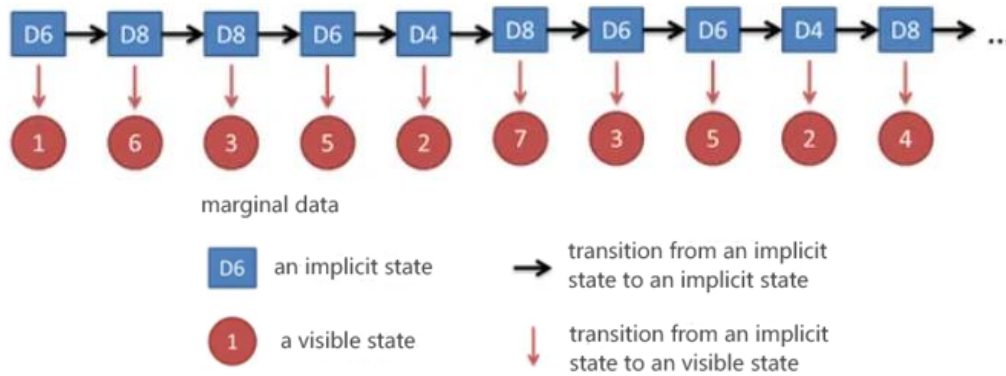


Figure 1. Relationship of the hidden Markov model

Initial State Probability Vector (s): The initial state probability vector specifies the initial state of the player in terms of gaining match points. It sets the probability distribution of the initial state based on the player's match data and the current competitive situation [7-8].

$$s = (s_i)$$

$$s_i = p(i_1 = q_i)$$
(1)

Where s_i is the probability of being in state q_i at time $t = 1$.

Transition Probability Matrix (P): This matrix describes the probability of transitioning from one hidden state to another at each time step. Since there are only two players in a singles tennis match, there are two hidden states. Therefore, the transition probability matrix will be a 2x2 matrix.

$$P = [a_{ij}]_{N \times N}$$

$$a_{ij} = p(i_{t+1} = q_i | i_t = q_j)$$
(2)

Where a_{ij} represents the probability of moving to state q_i at time t if you are in state q_j at time $t + 1$.

1. Observation Probability Matrix (B): This matrix represents the probability of observing each possible observation value in each hidden state. It estimates the player's state based on the observed score situation and the probability of observing match points in each state. The two defined hidden states are "Alcaraz" and "Djokovic", while the observations correspond to the score situation, with match points being considered as special events.

$$Q = [b_j(k)]_{N \times M}$$

$$b_j(k) = p(h_t = v_k | i_t = q_j)$$
(3)

Where $b_j(k)$ represents the probability of generating observation v_k at time t in state q_j .

The hidden Markov model is expressed as:

$$\eta = (P, Q, s)$$
(4)

2.2. Result analyze of model

Match Point Prediction Results: The model outputs the prediction results for each match point, indicating which player has a higher probability of winning the match in each game situation. The results are represented as 0 and 1, where 0 indicates that the model believes there is a higher probability of Alcaraz winning, and 1 indicates that the model believes there is a higher probability of Djokovic winning. The time-series evolution state diagrams shown in Figures 2 and 3 are predicted by incorporating the data from two matches into the Hidden Markov Model.

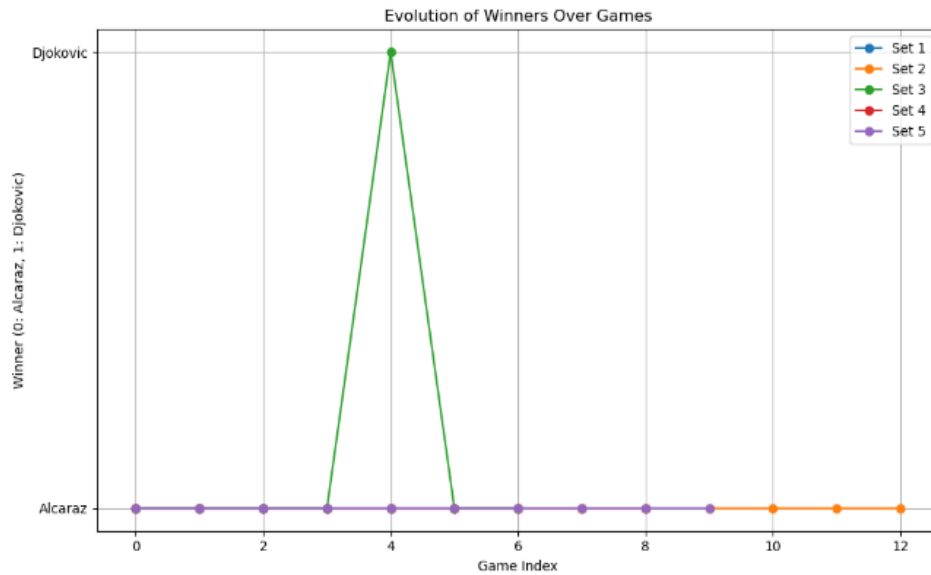


Figure 2. Result1

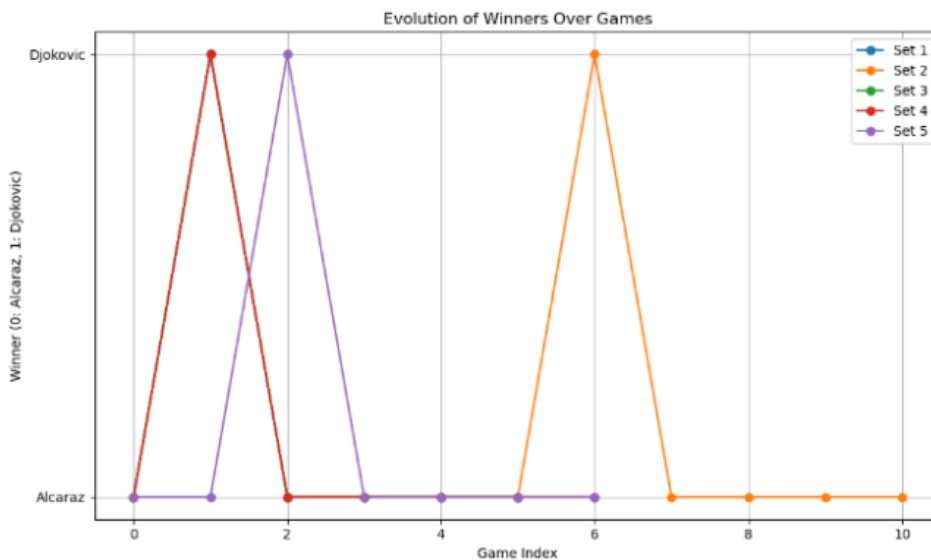


Figure 3. Result2

Probability of Winning Match Point: Outputs the probability of winning the match for each match point, indicating the likelihood that the corresponding player will win when each match point occurs. A higher probability value suggests a greater likelihood of the player winning, thereby indicating better performance by the player at that specific moment.

Model Evaluation: By comparing the match point prediction results with the probability of winning the match, the accuracy of the model's match point predictions can be assessed. If the model's predicted match point results align with the actual winning player and the probability of winning the match at those match points is high, it indicates that the model's predictions are reliable. The model was evaluated using Python language, yielding an accuracy of 84.0% and an AUC of 0.92 so that these metrics indicate that the model performs well.

2.3. data visualization

Figure 4 shows the visualization based on the model.

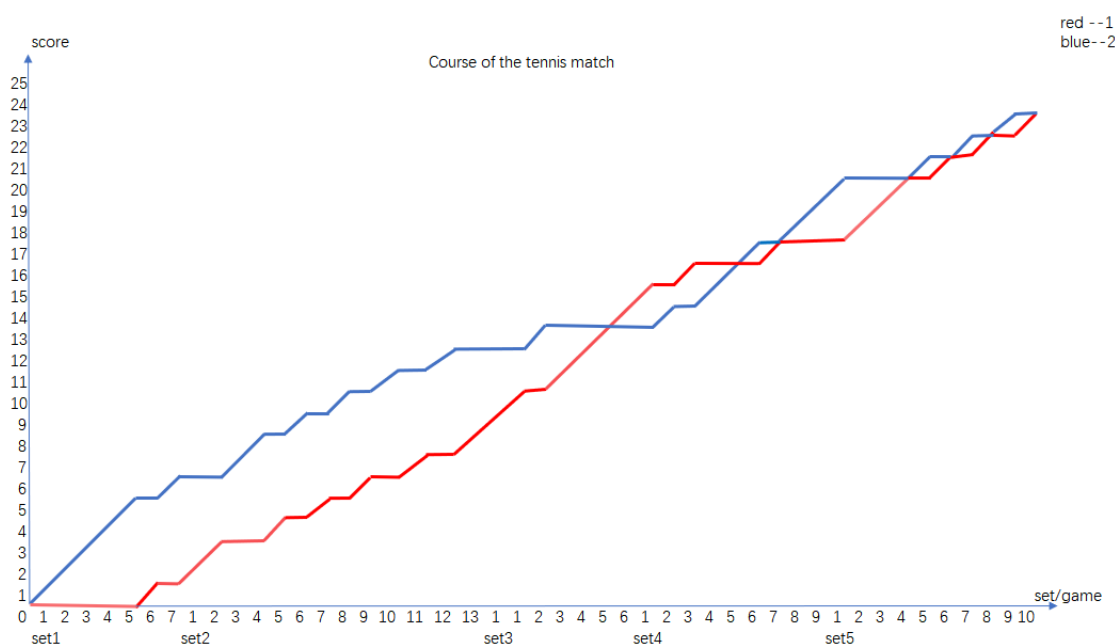


Figure 4. Course of the tennis match

3. Evaluate whether momentum plays a random role in the game

The outcome of competition depends not only on physical skills but to a large extent on psychological factors as well. This contention frequently is illustrated in sports. Players who are evenly matched in physical abilities often rely upon psychological skills to gain an edge over their opponents [9]. Coaches or players who single out particular games or stages during which they feel they can gain a psychological edge maybe indulging in an unwise practice [10].

Iso-Ahola and Mobily (1980) defined psychological momentum as “an added or gained psychological power which changes interpersonal perceptions and influences an individual's mental and physical performance” [7].

3.1. Calculating momentum score

According to the correlation between match point victory and indicators, five indicators are selected as the criteria for evaluating Momentum Scoring, and the Momentum Scoring System is constructed, which takes into account the following five factors:

- Score Points: For every point earned, momentum points are added.
- Serving Advantage: The server has a higher probability of winning on the exchange points, so the points earned in the service game are given a higher weight.

- Consecutive Points: Winning consecutive points adds additional momentum points, reflecting an overwhelming advantage in the game.

- Untouchable winning serve: A perfect serve that can increase momentum points.

Missed both serves and lost the point: Two missed serves affect the server's momentum.

Let S_t time point t be the momentum score; F_t is the score of the point, $F_t=1$ for win-ning, $F_t=0$ for losing; V_t is the weight of the serve, the server $V_t=1.2$, the receiver $V_t=1.0$; C_t is the bonus of consecutive points. The first consecutive win $C_t=1$, and the subsequent consecutive win C_t increases by 0.2. U_t is the bonus for the untouchable winner, $U_t=1.2$ when the ball is sent, $U_t=1$ when the ball is not sent; M_t is the effect of sending two errant balls, $M_t=0.8$ when this occurs and $M_t=1$ when it does not. You can get:

$$S_t = S_{(t-1)} \times (F_t \times V_t \times C_t \times U_t \times M_t) \quad (5)$$

According to formula 15, the trend chart of momentum score over time is obtained, as shown in Figure 8. According to the formula 15 Get a graph of momentum score over time, as shown in Figure 8.

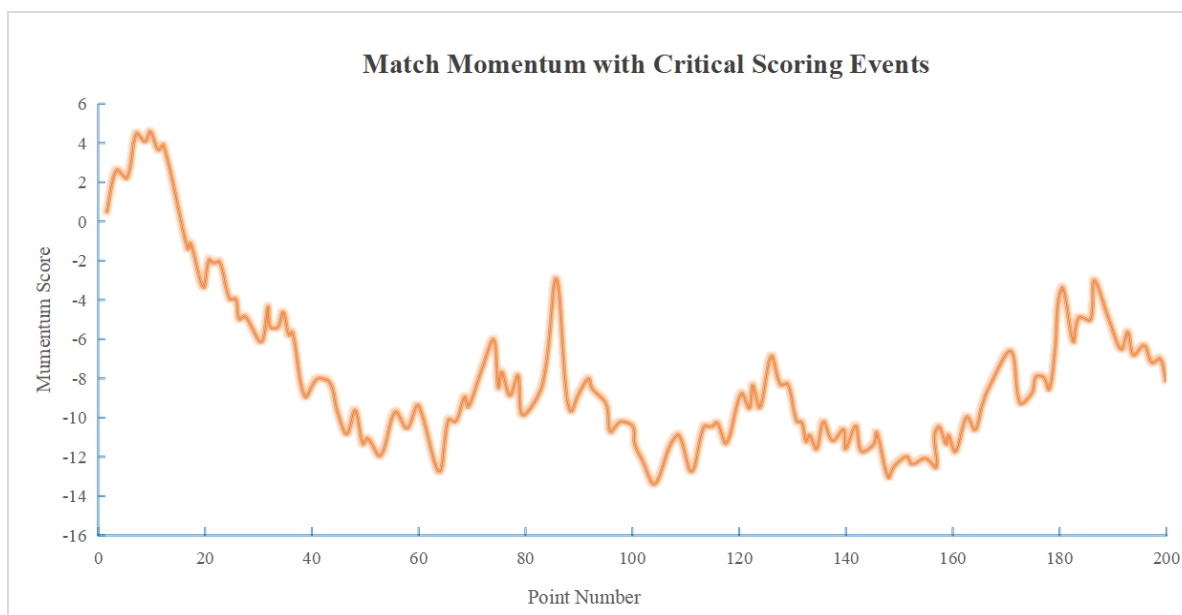


Figure 5. Match Momentum with Critical Scoring Events

In the figure 5, we can see the trend of momentum score over time. Rising curve indicates increasing trend, and the server gradually controls the game. Falling curve indicates decreasing trend, and the server needs to seize control.

Taking the third-round matching as an example, the hidden Markov model in question 1 is used to simulate the momentum score of each match point, and the momentum transition in the actual match is analyzed to observe whether there is a significant difference between them.

3.2. Hypothesis Testing

Formulate null hypothesis and alternative hypothesis

Null hypothesis (H_0): The momentum shift in a game is random and does not differ significantly from the result of a random simulation.

Alternative hypothesis (H_1): Momentum shifts in games are non-random and significantly different from the results of random simulations.

Determine the test statistics

The data in this paper is larger than 30, which is a large sample. It is considered that the sample mean follows a normal distribution and the sample standard deviation replaces the standard deviation. Therefore, the statistics are as follows:

$$Z = \frac{|\bar{x} - \mu|}{s / \sqrt{n}} \quad (6)$$

Select the significance level and determine the acceptance domain and rejection domain:

Acceptance field: $-1.96 < z < 1.96$; Reject field: $z < -1.96$ or $z > 1.96$.

In this paper, the null hypothesis is rejected because $z = 3.184 > 1.96$ falls within the acceptance domain. It can be concluded that at the significance level of $\alpha = 0.05$, the momentum transition in the match is non-random, which is significantly different from the result of random simulation.

4. Conclusion

By analyzing and evaluating the model, we found that the Hidden Markov Model is effective in predicting match points and evaluating player performance. Our model performed well in terms of accuracy and AUC, proving its reliability and validity. Therefore, we suggest applying the model to real tennis matches to help assess match progress and player performance.

References

- [1] demo. Deep Learning: A powerful tool for classifiers [EB/OL]. <https://developer.baidu.com/article/details/2563434>, 2023-12-11/2024 - 02 - 03.
- [2] SPSS School. ROC Curve analysis [EB/OL]. <https://zhuanlan.zhihu.com/p/246444894#>, 2020 - 09-15/2024 - 02 - 03.
- [3] Li Junwei, Tang Yafang, Hao Zhenghang et al. A review of cluster analysis and its applications in power systems [J]. Modern electricity, 2019, 36 (03): 1-10. DOI: 10.19725/j.cnki.1007-2322. 20181130. 001.
- [4] Li Tianhua, Yuan Yongbo, Zhang Mingyuan. Vulnerability analysis of power network nodes under earthquake action based on variable fuzzy clustering [J]. Science, Technology and Engineering, 2018, 18 (18): 126 - 130.
- [5] Guo Pengni. Research and combined application of ridge regression and quantile regression [D]. Harbin Institute of Technology, 2014.
- [6] Richardson P A, Adler W, Hanks D. Game, set, match: Psychological momentum in tennis [J]. The Sport Psychologist, 1988, 2 (1): 69 - 76.
- [7] Adler, Peter. "Momentum: A Theory of Social Action." (1981).
- [8] Dietl, Helmut; Nessler, Cornel (2017). Momentum in tennis: controlling the match. UZH Business Working Paper Series 365, University of Zurich.
- [9] Trentin E, Gori M. A survey of hybrid ANN/HMM models for automatic speech recognition [J]. Neurocomputing, 2001, 37 (1-4): 91 - 126.
- [10] Shih C, Padhy LC, Murray M, Weinberg RA. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. Nature. 1981 Mar 19; 290 (5803): 261 - 4. doi: 10.1038/290261a0. PMID: 7207618.