Visualisation of Athlete Performance Based on Markov Chains and EWMs

Senlong Yu*, Jianzhe Sun, Yusen Hu

College of Information Engineering, Zhejiang Ocean University, Zhoushan, China, 316000

* Corresponding Author Email: 15967619201@163.com

Abstract. The aim of this study is to synthesize the effects of several factors on the performance of sports players, such as skill level, score differences, and psychological factors, and to establish a corresponding evaluation index system. To address this goal, this paper propose a performance evaluation framework to quantitatively assess the state of sports players. In this paper, this paper take the performance scores and winning rate in the game flow as the evaluation indexes, and firstly, this paper measure the performance scores of sports players by calculating nine indexes such as the winning rate of the first serve, the winning rate of the second serve, etc., and use the entropy weighting method to determine the weights of each index. Then, this paper used Markov model to calculate the winning rate of each point, combined with the performance score and the winning rate to establish the regression equation, the results show that the coefficient of determination are close to 1, indicating that the model has a good fitting effect. In this paper, this paper continue to calculate the winning percentage of a player at a certain point of winning this game through Markov chain. This study provides a comprehensive quantitative method for evaluating the performance of sports players, which helps to more accurately determine the players' status and level, and provides a scientific basis for training and competition.

Keywords: Markov chain, Motion Performance Visualization, EWM, Regression.

1. Introduction

With the birth of the four Grand Slam tournaments, tennis has become a popular sport all over the world. A player's performance not only depends on his/her skill level, but also affected by other factors, such as point difference, psychological factors and so on.[1][2] Therefore, we will establish evaluation metrics to determine the player's status, so as to derive the probability of winning and establish a mathematical model. In order to visualize the performance of the players. Some researchers quantitatively assess the stability of the athletic performance of the players by summarizing and analyzing the effective climbing performance and the number of successes of the athletes in the competition, applying mathematical statistical methods to statistically determine the success rate, the performance rate, and the stability of the effective performance in the preliminaries and finals, and analyzing the frequency of the best performance by using the chi-square test, thus quantitatively assessing the stability of the athletic performance of the athletes[3]. Some researchers have quantified the performance of professional tennis men's singles players by analyzing the relationship of five physical performance indicators (work efficiency, impact load, high-intensity transition, sprint running, and total athletic distance) with stroke errors and match results in tennis.[4] Some other researchers have quantified the performance of soccer players through standardized score conversion methods, using radar charts to display the indicators and establishing standardized profiles to provide intuitive and timely feedback on athlete performance and to assess the strengths and weaknesses of opponents.[5] Combining these research results, this paper defines two evaluation indexes of athletes' performance scores and win rates, the performance scores are calculated by the EWM model[6][7] from the primary serve score rate, secondary serve score rate, serve-receive score rate, break score rate, ACE rate, total score rate, number of unforced errors, and the number of double faults, and the win rates are calculated by the Markov chain model[8][9][10] from the POINT win rate, GAME win rate, and set wins were calculated by Markov chain model. The data for this paper was obtained from https://www.comap.com/contests/mcm-icm.

2. Visual model of athlete performance

As the game progresses, players' performances and scoring fluctuate, and players' performances can be reflected by various statistics of the game. Usually, momentum is a factor reflecting the real-time performance of the player, and the ups and downs of the momentum will affect the trend of winning the game. Based on the game data, this paper establishes a model to reflect the performance of players through their performance scores and win rate prediction.

2.1. Entropy weighting method - performance scoring

Match statistics include first serve percentage, second serve percentage, ACE rate, double serve error rate, break points, etc. The contribution value of each index to the player's performance degree is different, and this paper determines the weight for each index according to the entropy weighting method.

2.2. Data standardization

Based on the data of 2023 Wimbledon men's singles final, the steps of this paper to derive the weights according to the entropy weighting method are as follows:

In order to eliminate the effects of magnitude and direction and better explain the effects of weights on the model, this paper standardizes the above indicators.

For these positive indicators such as first-shot scoring rate, ACE rate and net scoring rate, the standardized values are obtained according to the formula and direction and better explain the effects of weights on the model, this paper standardizes the above indicators.

For these positive indicators such as first-shot scoring rate, ACE rate and net scoring rate, the standardized values are obtained according to the formula R_{tj}:

$$R_{tj} = \frac{x_{tj} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{1}$$

R_{tj} is the standardized value of the first indicator at the t point, x_{tj} is the original value on the jth indicator at point t.

For negative metrics like the number of unforced errors and double faults, use the formula:

$$R_{tj} = \frac{\max(x_j) - x_{tj}}{\max(x_j) - \min(x_j)} \tag{2}$$

Substituting the normalized data into the entropy weight method of calculation.

2.3. Calculating relative entropy

First, the relative entropy of the indicator is calculated:

$$E_j = -\frac{1}{\ln(n)} \sum_{j=1}^{n} R_{tj} \ln(R_{tj}) \tag{3}$$

where E_j is the relative entropy of the jth indicator and n is the number of indicators. The information entropy reflects the expected value of the amount of information, when the smaller the information entropy, the greater the amount of information obtained.

2.4. Calculation of weights

Based on the relative entropy data, the weights of each indicator were calculated according to the formula:

$$w_j = \frac{1 - E_j}{n - \sum_{j=1}^{n} E_j} \tag{4}$$

w_j is the weight of the jth indicator.

The resulting weights were normalized to obtain weights for each indicator.

Table 1. weighting table

Parametic	Novak Djokovic	Carlos Alcaraz
First Serve Win Rate	0.0174	0.0311
Second Serve Win Rate	0.0125	0.0867
Total Points Win Rate	0.0238	0.0057
Net Points Win Rate	0.0148	0.0781
Ace Rate	0.1315	0.2412
Serve Receive Win Rate	0.0378	0.0603
Break Point Conversion Rate	0.0380	0.1396
Unforced Errors Count	0.2355	0.1801
Double Faults Count	0.4889	0.1772

The weights assigned to each indicator are shown in Table 1. For Novak Djokovic's metric weights, Double Faults Count and Unforced Errors Count had a greater impact on his performance; for Carlos Alcaraz's metric weights, Ace Rate and Unforced Errors Count had a greater impact on his performance. Among them, Double Faults Count and Unforced Errors Count need to be normalized, and the smaller their values are, the better. From the fact that ace rate and double faults count account for the largest proportion, it is not difficult to see that in the top games, whether or not the serve can be scored largely determines whether or not the player can win, and unforced errors count is also very important, whoever has less errors is more likely to win in a top match.

3. Markov Chain Model Predicts Winning Percentage

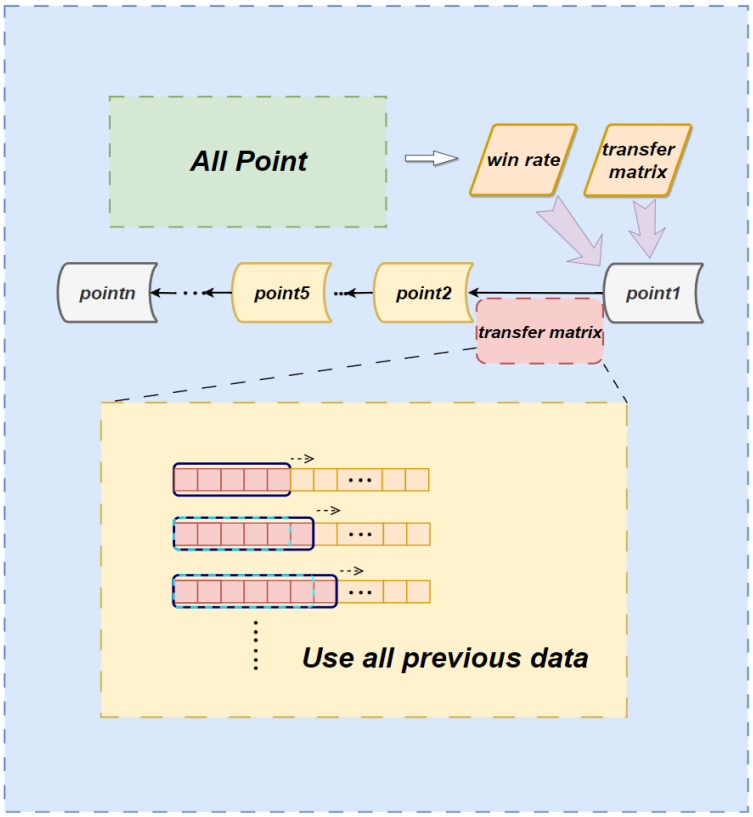


Figure 1. Markov Chain flow chart

3.1. Markov Chain modeling

Obviously, the state space of a tennis match is win or lose. As for the initial probability distribution, in order to ensure the accuracy of the prediction at the beginning of the match, this paper chooses the

win rate of the whole match as the initial probability, and the initial transfer matrix is derived based on the player's whole performance dat.

(1) State space: The set of values of the random variable X_t is the same at each moment t, and it can be continuous or discrete.

(2) Initial probability distribution: The probability of 0 at moment X_0 is $P(X_0)$

(3) Transfer probability matrix: The probability of a random variable X_t at any moment depends only on X_{t-1}. This is Markovianity. However the probability of moving from each state to another is represented by a probability matrix, which is usually a matrix of $n \times n$.

$$P = \begin{bmatrix} P_{ii} & P_{ij} \\ P_{ji} & P_{jj} \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} p_{t+1} \\ 1 - p_{t+1} \end{bmatrix} = P \cdot \begin{bmatrix} p_t \\ 1 - p_t \end{bmatrix} \tag{6}$$

n is the number of states, P is the transfer matrix, P_{ii} is the probability of going from winning to winning, P_{ij} is the probability of going from winning to losing P_{ji} is the probability of going from a loss to a win, and P_{jj} is the probability of going from a loss to a loss. p_t is the probability of winning at the current point.

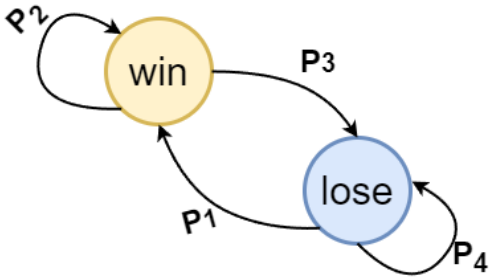


Figure 2. Markov Chain

The Markov chain process is shown in Figure 2, P_1 is the probability of shifting from losing to winning, P_2 is the probability of transferring from winning to winning, P_3 is the probability of moving from winning to losing, P_4 is the probability of transferring from losing to losing.

3.2. Solving the Markov model

We used python code to implement the above model, calculated the win rate at each point to win the point, from the win rate, the beginning of the win rate fluctuation is large, to the back of the tend to stabilize, in a value of above and below the small fluctuations in the actual game may be due to the differences in experience and ability of the athletes, so fluctuations are large, with the advancement of the game, the two gradually familiar with each other's style of play, so it tends to be stabilized, so that the model is considered reasonable.

4. Polynomial regression

A polynomial regression was used to fit the data on player performance ratings and win shares to produce a curvilinear relationship.

The general equation for polynomial regression is:

$$Y = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_n t^n + \varepsilon \tag{7}$$

Y for performance rating, $\alpha_0, \alpha_1, \ldots \alpha_n$ is the coefficient used to fit the polynomial, ε is the error factor.

In order to achieve a balance between overfitting and underfitting, this paper plots the learning curve, compares the number of training samples with the goodness of fit, calculates the error coefficients at each order, and determines the optimal order 2, which results in a second-order polynomial regression equation:

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \varepsilon \tag{8}$$

When performing second order polynomial regression, it is necessary to extend the feature matrix.

In this paper, based on python, we use the PolynomialFeatures class in the sklearn library to transform the features so as to achieve the fitting, and obtain the following visualization results.

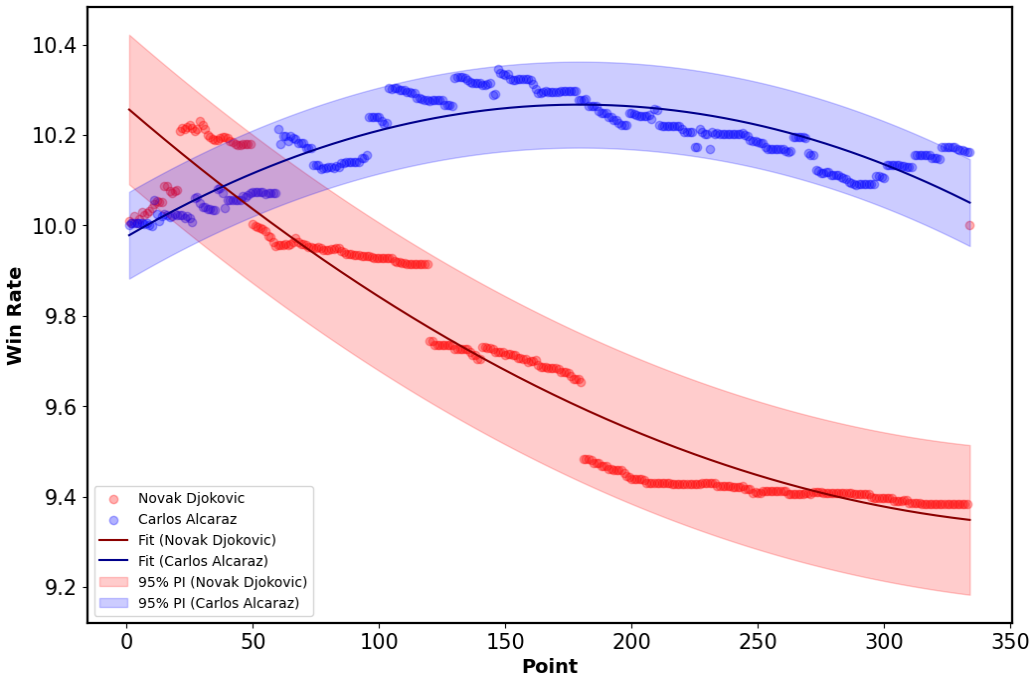


Figure 3. Player Performance Evaluation Fitting Chart

The performance of Djokovic and Alcaraz is shown in Figure 3, it can be seen that the fitted curves for the performance ratings are closer to the actual data. For the performance rating function, the goodness of fit is 0.88 for Djokovic and 0.79 for Alcaraz, both of which are close to 1. It can be seen that the predicted values are close to the actual data, so the model is considered to be more effective. Djokovic was better than Alcaraz until 45 points, but his performance kept slipping, on the contrary, Alcaraz started poorly but kept getting better and better and ended up far better than Djokovic.

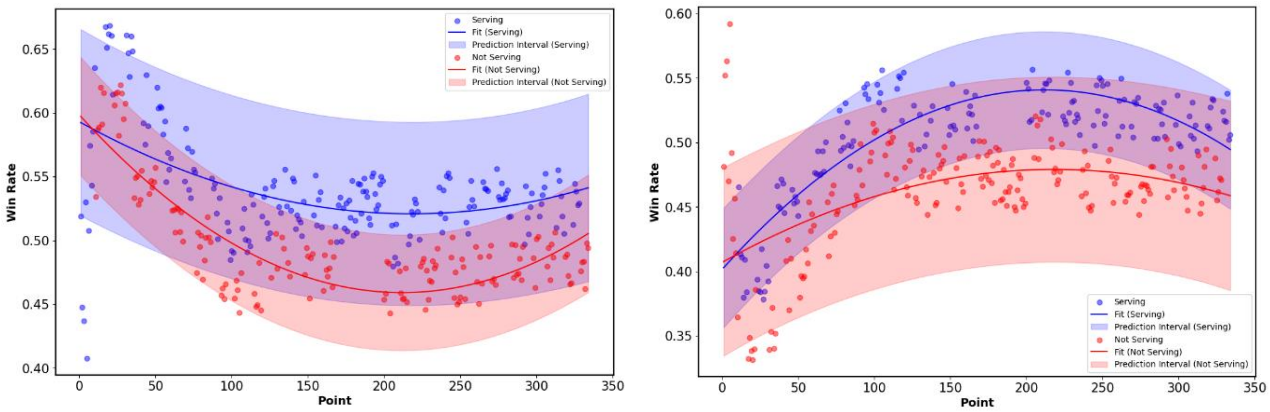


Figure 4. Djokovic VS Alcaraz Win Prediction

Djokovic and Alcaraz Win Prediction is shown in Figure 4, look at the blue curve on the serve and the red curve on the receive. Looking at the prediction graphs shows that the predicted winning percentage is a better fit to the actual. In the final, Djokovic's serve win rate function has a goodness

of fit of 0.83 and a goodness of fit of 0.87 for the reception set, and Alcaraz's serve win rate function has a goodness of fit of 0.78 and a goodness of fit of 0.89 for the reception set, so the fit is considered to be better performance.

5. Conclusions

Since the coach needs to adjust the strategy according to the real-time performance of the tennis player, so as to increase the probability of the player's winning, this paper establishes a model for visualizing the player's performance, which is divided into two aspects, namely, the performance score and the win rate. For the performance score, based on the player's first-shot scoring rate, second-shot scoring rate and other indexes, we use entropy weighting to derive the weight, and calculate the score of each point, and then use binomial regression to do the visualization. visualization process, for the win rate, first use Markov to calculate the win rate of each point, and then use the Markov chain to calculate the probability of the player winning at each point, and then use binomial regression to do the visualization process. Finally, in the Wimbledon final, in the first set, Djokovic's performance points and win rate is much higher than Alcaraz, in the second and third sets, Djokovic's performance and win rate declined by Alcaraz to overtake, in the fourth set, Alcaraz's performance and win rate is slightly higher than that of Djokovic, and in the fifth set, the two are evenly matched.

References

[1] WANG Xu, TAI Xingze. Research status and application of sports performance analysis[C]//China Intelligent Engineering Research Association, China Bandi Association, Guangdong Physical Fitness Association. Proceedings of the 10th Chinese Conference on Physical Training Science (lower). Jilin Institute of Physical Education;, 2023:5. 2023.064645.

[2] M B P, T G M J, Samuel W, et al. Physical Demands of Tennis Across the Different Court Surfaces, Performance Levels and Sexes: A Systematic Review with Meta-analysis. [J]. Sports medicine (Auckland, N.Z.), 2023, 53(4): 807-836.

[3] Zhu Mingquan. Stability of competitive performance of the world's best men's speed climbing[C]// Hubei Sports Science Society. Proceedings of the First Hubei Sports Science Conference (Volume I). College of Physical Education, China University of Geosciences (Wuhan); 2023:3. 2023.078066.

[4] LI Jie, ZHANG Kebao, LI Yu. Analysis of physical fitness-related indicators and match performance of men's singles players in professional tennis tournaments[C]//Chinese Society of Sports Science. Abstracts of the thirteenth national sports science conference--Special Report (Physical Training Division). [Publisher unknown], 2023:4. DOI: 10.26914/c.cnkihy.2023.067291.

[5] LIU Hong-Yu, YI Qing, KANG Hui. Creation of a standardized profile of high-level soccer players' game performance [J]. Journal of Nanjing Institute of Physical Education (Social Science Edition), 2015, 29(01): 120-123+128. 2015.01.021.

[6] QIN Chunyan, LI Daobo, WANG Yu et al. Research on Credit Rating of Micro, Small and Medium Enterprises Based on Hierarchical Analysis and Entropy Weight Method--Taking New Third Board Enterprises in Sichuan Province as an Example [J]. China Business Journal, 2023, (24):160-163. 2023.24.160.

[7] YE Yu-Xin, TANG Wen-Xi, LI Shuai-Long et al. Research on the construction of assessment system for the development level of smart health care based on Delphi-entropy power method [J]. China Hospital Management, 2024, 44(02): 1-5.

[8] GUO Yangli, MA Feng. Research on volatility prediction of gold futures market based on Markov and mixed-frequency data model [J]. China Management Science, 2024, 32(01): 13-22. 2021.0213.

[9] Wang Ying. Analysis of rainfall prediction in Jinan based on weighted Markov chain combination model [D]. Shandong University, 2022. 2022.003062.

[10] Wu Min. Research on passenger flow prediction of urban rail transit based on Markov chain [J]. Journal of Chengdu Institute of Technology, 2023, 26(01): 62-65. 2023.01.014.