WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Prediction of winning and losing in tennis match based on entropy weight -TOPSIS and machine learning model

Quankun Zang [1, *], Zetong Li [1], Jiren Hu [2], Tingxu Zhang [3], Weiqi Zhang [1]

[1] School of Computer and Information Engineering, Fuyang Normal University, Fuyang, China, 236037

[2] College of Nuclear Science and Technology, Lanzhou University, Lanzhou, China, 730106

[3] School of Information and Artificial Intelligence, Anhui Agricultural University, Hefei, China, 230036

* Corresponding author: 18156988929@163.com

Abstract. Tennis is a beloved sport worldwide, and spectators often eagerly predict the winners of both sides based on various playing conditions. Research has shown that there is a significant correlation between players' performance in the game and their momentum. To explore the impact of players' momentum on match results, this paper establishes a TOPSIS evaluation model based on entropy weight method to measure players' momentum value at every moment. Additionally, this paper establishes a decision tree model to predict real-time trends in match results according to match data. Ultimately, this paper findings suggest that players with higher momentum perform better at any given time, and predictions about game outcomes can be made based on differences in momentum. This study sheds light on various indicators and psychological states of athletes during competition and training while providing scientific recommendations for developing strategies, adjusting mentality as needed, and achieving optimal results.

Keywords: Entropy weight -TOPSIS; Machine learning; Decision tree.

1. Introduction

Tennis originated in France in the 12th century. The tennis match has a high viewing value, because it is a strategic sport that requires high skills, and the result of the match often depends on the players' physical quality game and psychological quality game at key moments. Tennis is not only a competitive sport, but also a wide range of social value, it can help people to maintain health, improve physical and psychological quality, is a relatively popular game sport.

Celestine Iwendi studied how to detect and measure pressure through existing data and established various machine learning models for pressure detection. By using machine learning technology, the accuracy of three class classification problems and binary classification problems was increased to 81.65% and 93.20% [1]. Gao Z and M Sipko set up the most used random forest model and Logistic Regression model respectively for the prediction of tennis match results, and compared them with SVM and other models, and obtained the conclusion that the random forest model has universal applicability to the prediction of tennis match results [2-3]. Y Zhu described the rationality and importance of the entropy weight method for the objective division of weights, and S Chakraborty explained the subjectivity and reliability of TOPSIS in the evaluation grading model in comparing TOPSIS and improving TOPSIS models [4-5]. In his paper, Walid Briki explored whether momentum is related to sports by making cyclists confront each other, and concluded that momentum of players is related to performance [6].

Therefore, this paper find that the above scholars have proved that player pressure in tennis competitive matches is real and can be quantitatively measured and qualitatively analyzed by establishing evaluation models and machine learning models. At the same time, this paper also found in the papers of scholars that entropy weight method is too objective and simple, the process of TOPSIS model setting power is subjective and lacks objective analysis and calculation, and the machine learning model can be further optimized and improved to make the prediction more universal. Therefore, the problem to be solved in this paper is to combine entropy weight method and TOPSIS model to conduct subjective and objective quantitative scoring of momentum, and then determine that the machine learning model algorithm after comparison optimization can accurately predict the momentum level of the two players in the game and predict the result of the win or loss of the game, and extend the model to all kinds of competitions, so that the prediction result is still relatively accurate Rate.

2. TOPSIS evaluation model and machine learning prediction model based on entropy weight method

2.1. TOPSIS evaluation model based on entropy weight method

The entropy weight method can determine the weight of data index under objective conditions according to the variability of data index. The idea of entropy weight method is to determine the weight of an index according to its variation degree. When the information entropy of a certain index is relatively small, it indicates that the degree of variation is large and the information provided is relatively rich, so the proportion in the comprehensive evaluation is larger. On the contrary, if the information entropy of a certain index is large, its variation degree is relatively small, and the information provided is also less, so the influence on the score in the comprehensive evaluation will be weakened accordingly. In the evaluation process, more attention is given to indicators with greater variability, as they can provide more and more representative information. TOPSIS can rank evaluation objects according to how close they are to the idealized goal. The basic idea is to set up positive and negative ideal solutions, and then measure the distance between each sample and these two solutions, so as to measure their closeness to the ideal solution. The closer the sample is to the positive ideal solution and the farther the sample is to the negative ideal solution, the higher the degree of its quality. The TOPSIS model based on entropy weight method can get the corresponding score more scientifically by combining subjective and objective methods.

The above selected feature data is divided into very large data and very small data, and the very small data is positively transformed to meet the calculation method of very large data. Then the data are normalized to form a new normalized numerical matrix. And calculate the proportion of sample i among j indicators:

$$P_{ij} = \frac{x_{ij}}{\sum_{i}^{m} x_{ij}} \tag{1}$$

Where P_ij is the proportion of sample i among j indicators.

Then use the information entropy formula to calculate the entropy value:

$$e_j = -\frac{1}{\ln m}\sum_{i}^{m} P_{ij} \ln P_{ij} \tag{2}$$

Where e_j is Information entropy.

Calculate the degree of difference of each indicator j, and define the weight formula.

$$d_j = 1 - e_j \tag{3}$$

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

Where d_j is optimal solution distance and w_j indicates worst solution distance.

The information entropy of A and B is calculated for the processed numerical matrix, and the coefficient of anisotropy and weight W_1 and W_2 are calculated respectively.

In order to unify the calculation method of weights, this paper determine the final weights.

$$W = 0.5(W_1 + W_2) \tag{5}$$

Where W is comprehensive weight.

The data is normalized again, and the weighted matrix is calculated, and the optimal solutions D_+ and D_- are obtained according to the matrix.

$$D_+ = \sqrt{\sum_{j=1}^{m}\left(z_j^{+} - z_{ij}^{+}\right)^2} \tag{6}$$

$$D_- = \sqrt{\sum_{j=1}^{m}\left(z_j^{-} - z_{ij}^{-}\right)^2} \tag{7}$$

Where D_+ is optimal solution distance and D_- indicates worst solution distance.

Through the above processing, the TOPSIS comprehensive evaluation model based on entropy weight method can be established. The specific process is shown in Figure 1:

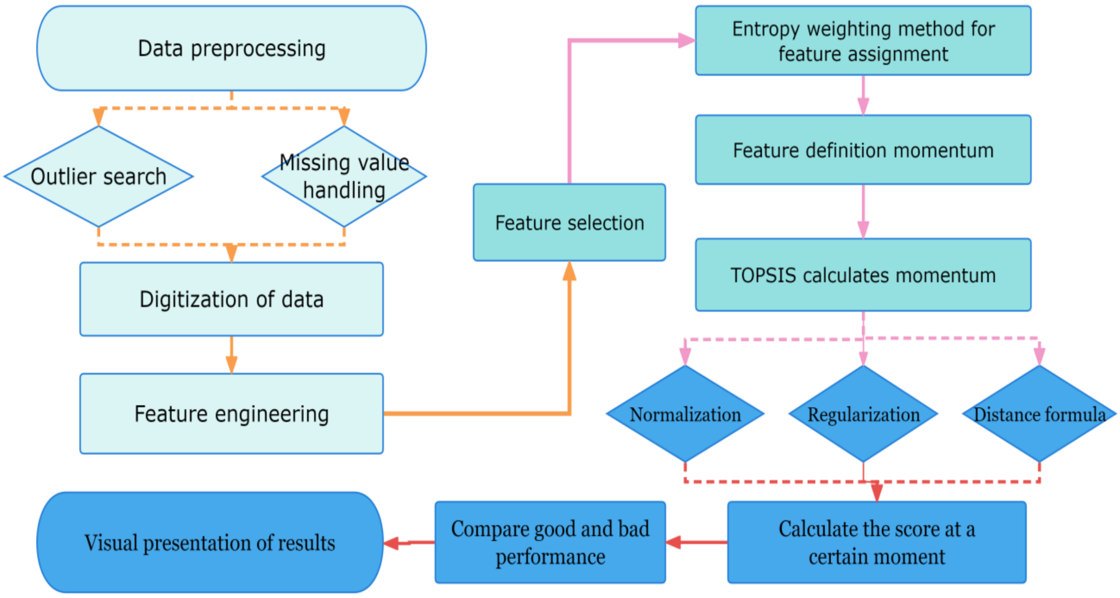

Figure 1. Model flow diagram description

2.2. Comparative evaluation of machine learning models

Machine learning is the process by which computers learn from patterns and patterns in data to make predictions and decisions. In combination with the papers of Solanki S and Shankar K, this paper found that the machine learning model can better predict the results of tennis matches given the serving parameters of tennis players and the way of hitting the ball and scoring the goal [7-8]. In this paper, a decision tree classifier is established, which is a model that displays decision rules and classification results with a tree data structure. As an inductive learning algorithm, its focus is to transform seemingly disordered and chaotic known data into a tree model that can predict unknown data by some technical means. Each path from the root node (the attribute that contributes the most to the final classification result) to the leaf node (the final classification result) represents a decision rule; The basic idea of SVM model is to solve the separation hyperplane that can correctly partition the training data set and has the largest geometric spacing [9]. Logistic regression model, which often uses the relu function to estimate the probability of an event according to a given data set of independent variables. Since the result is a probability, the range of the dependent variable is between 0 and 1; Random forest classifier is an integrated algorithm based on decision trees [10], that is, a collection of several different decision trees. In this paper, the above model is used to train each model and predict the momentum fluctuations in the model. Moreover, the model with the best effect after 100 training times is selected from these machine learning prediction models, and the selected model is treated with grid tuning optimization.

3. Results

3.1. The establishment of simulation model

In this summary, TOPSIS comprehensive evaluation model based on entropy weight method and multiple machine learning prediction models are used to successfully solve the problem of how to judge the momentum of tennis players and predict the momentum volatility of both sides of the match.

3.2. Analysis of experimental results

After the initial data of players are processed by entropy weight method and sorted in descending order of weight, the final result is shown in Table 1:

Table 1. Average scores of players with different momentum

Feature	Entropy	Coefficient of variability	Weight
p_break_pt_won	0.502947029	0.497052971	0.192758526
p_ace	0.652996238	0.347003762	0.135111733
p_net_pt_won	0.717245632	0.282754368	0.111085023
p_net_pt	0.757577541	0.242422459	0.095575964
game_victor	0.789629999	0.210370001	0.082124524
p_winner	0.798718773	0.201281227	0.078328156
p_double_fault	0.773567569	0.226432431	0.070395906
p_break_pt_missed	0.780702582	0.219297418	0.068250708
p_unf_err	0.880206386	0.119793614	0.038379313
p_sets	0.923438239	0.076561761	0.029953548
server	0.924352825	0.075647175	0.029531246
p_score	0.941753667	0.058246333	0.022568223
rally_count	0.951104054	0.048895946	0.019088065
p_games	0.962224604	0.037775396	0.014752371
p_points_won	0.974820515	0.025179485	0.009890797
speed_mph	0.994349386	0.005650614	0.002205894

Therefore, this paper obtains the weight ratio of each feature to momentum through subjective and objective comprehensive analysis. In known tennis matches, the server is much more likely to win points, so considering that server and momentum are related and cannot be ignored, this paper select the top 12 features with W>0.02 as momentum quantification indicators.

The weight of each data feature obtained by entropy weight method combined with TOPSIS comprehensive evaluation model to get the momentum score of both sides in tennis match.

In order to reflect the good performance of players A and B, according to the momentum scores of both players at each moment, this paper use the difference between the momentum of player A and the momentum of player B to reflect the performance gap between the player and the other player at this moment, judge the performance of player A and player B through the momentum gap, and mark the moment at the match point or match point, result is shown in Figure 2:

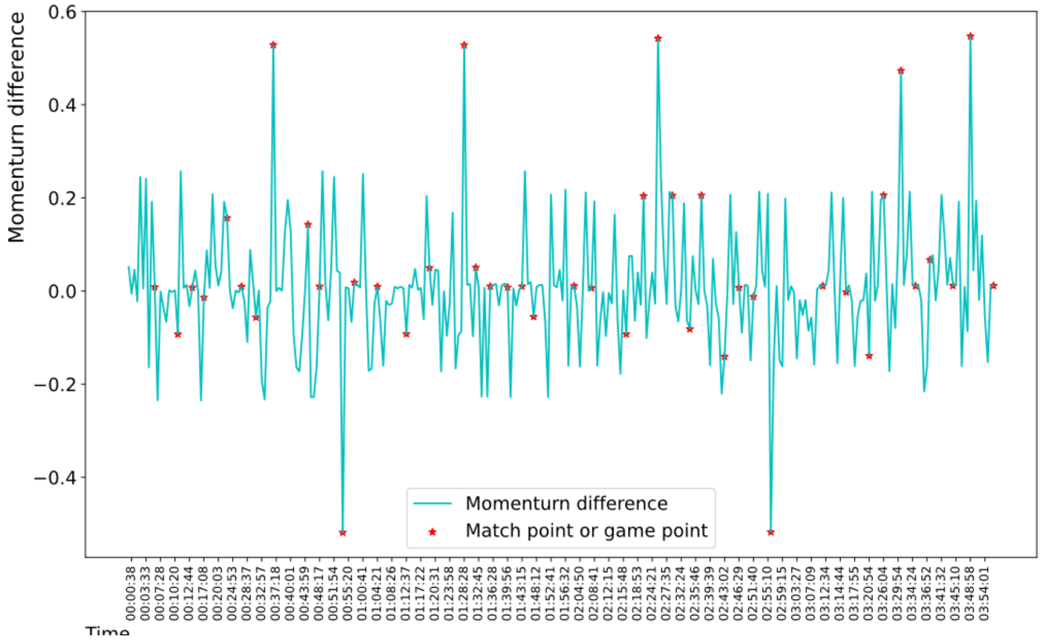

Figure 2. Momentum score difference between player A and Player B

By observing the chart, it can be found that the momentum difference between A and B at the match point or game point in this match is usually between 0, indicating that the momentum difference between A and B is not much at most moments, and the performance is similar. However, according to the whole match, when A is at the match point or game point, the momentum is much higher than B (that is, the image appears the peak), the proportion is higher, about 71.43%, while the momentum of B is much higher than A (that is, the image appears the trough), the proportion is low, about 28.57%. At the same time, this paper know that the final winner of the match is A, so this paper preliminarily conclude that the more times a player has higher momentum than his opponent at match point and game point in the whole match, the easier it is to win the match.

After the score of tennis players is obtained, the decision tree classifier, SVM, logistic regression and random forest classifier constructed in this paper are trained using the data of the same tennis match respectively. The accuracy results of 100 training sessions of each model are visualized, as shown in Figure 3:

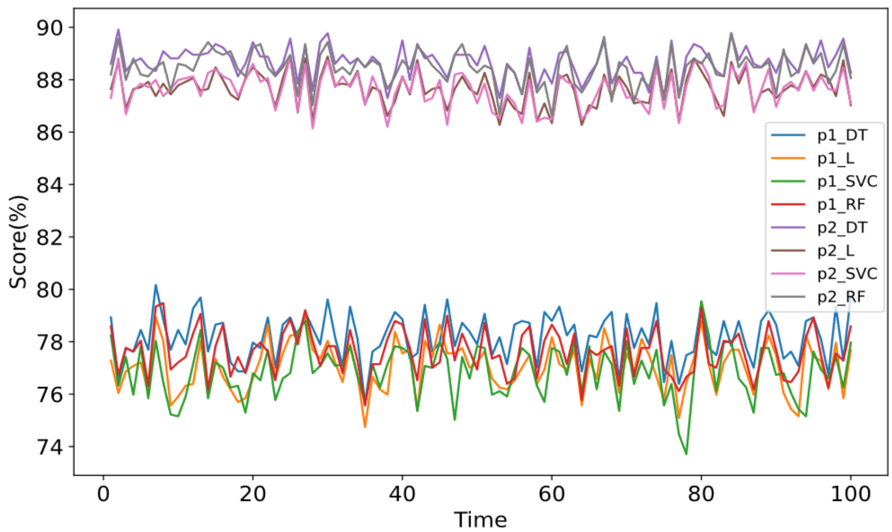

Figure 3. Accuracy of each model trained 100 times

According to Figure 3, this paper come to the conclusion that the accuracy rate of training p1 test set by using machine learning algorithm is generally about 74%-80%; The average accuracy of p2 test set training is concentrated in 866-90%, and the accuracy curve of decision tree algorithm is generally higher than other algorithms in both p1 and p2 test set. In the actual training process, using the decision tree model to predict the efficiency is also high.

The accuracy of the above machine learning prediction models after training is summarized in Table 2:

Table 2. Accuracy of the above models

Model	P1 Accuracy	P2 Accuracy	Average accuracy
Decision Tree	78.19%	88.51%	83.35%
SVM	76.81%	87.43%	82.12%
Logistic	77.06%	87.52%	82.29%
Random Forest	77.67%	88.35%	83.01%

As can be seen from the above table, the accuracy of decision tree classifier is almost the same as that of random forest classifier. In addition, this paper only selects the data of one match, so the amount of data used is small. In order to ensure the efficiency of the model in processing large amounts of data, a faster decision tree classifier is selected to predict the momentum fluctuations. After conducting grid search on the parameters of the model to adjust the optimal parameters, it is obtained that 'gini' is adopted as the evaluation standard and its accuracy is further improved when the maximum tree depth is set to 9. The results are shown in Table 3:

Table 3. Accuracy before and after decision tree model

Model	Average accuracy (before)	Average accuracy (after)
Decision Tree	88.61%	89.73%

To sum up, this paper choose the decision tree model with the highest accuracy and the fastest computing speed among the four types of models to predict the fluctuation of the game. Using the optimized decision tree model, the momentum fluctuation of tennis players can be predicted relatively accurately at high speed. The correlation analysis of each variable of the player in the game shows that the player's momentum will increase greatly when he scores the unreacted ball, hits the net first and misses the other player. In addition, the model can also be extended to other sports to measure players' momentum and predict competition results [11].

4. Conclusion

The objective of this article is to address the issue of measuring a tennis player's momentum and predicting match outcomes based on real-time match data. Among numerous movement data features, this paper selects motion indicators that have a significant impact on players' momentum. To measure and compare athletes' momentum changes during competition, as well as assess their performance levels, this paper establishes a TOPSIS evaluation model using the entropy weight method. This method effectively screens and empowers features, enabling us to accurately calculate the value of momentum according to the TOPSIS evaluation model. By combining subjective and objective methods, this paper obtains more precise measurements of player momentum. Furthermore, multiple machine learning prediction models are constructed to predict match outcomes. After careful evaluation through grid search and cross-validation, this paper selects the decision tree model with the highest efficiency and best effect to ensure accurate and efficient prediction results. This research successfully addresses how to judge a tennis player's momentum, predict their momentum changes, and identify factors that significantly influence their momentum. Based on our findings, players should exert maximum effort when approaching the net while also being proactive in suspending play after unforced errors or unexpected balls to restore their state and maintain good momentum. The significance of this study lies in its ability to help tennis players accurately calculate their momentums, understand their current motion states better, and predict future performances based on momentous changes at any given time.

References

[1] Bobade P, Vani M. Stress detection with machine learning and deep learning using multimodal physiological data[C]//2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2020: 51 - 57. - 58.

[2] Gao Z, Kowalczyk A. Random Forest model identifies serve strength as a key predictor of tennis match outcome [J]. Journal of Sports Analytics, 2021, 7 (4): 255 - 262.

[3] Sipko M, Knottenbelt W. Machine learning for the prediction of professional tennis matches [J]. MEng computing-final year project, Imperial College London, 2015, 2.

[4] Zhu Y, Tian D, Yan F. Effectiveness of entropy weight method in decision-making [J]. Mathematical Problems in Engineering, 2020, 2020: 1 - 5.

[5] Chakraborty S. TOPSIS and Modified TOPSIS: A comparative analysis [J]. Decision Analytics Journal, 2022, 2: 100021.

[6] Briki W, Den Hartigh R J R, Markman K D, et al. How psychological momentum changes in athletes during a sport competition [J]. Psychology of Sport and Exercise, 2013, 14 (3): 389 - 396.

[7] Solanki S, Jakir V, Jatav A, et al. Prediction of tennis match using machine learning [J]. Int. J. Progress. Res. Eng. Manag. Sci, 2022, 2: 5 - 7.

[8] Colomar J, Corbi F, Brich Q, et al. Determinant physical factors of tennis serve velocity: a brief review [J]. International Journal of Sports Physiology and Performance, 2022, 17 (8): 1159 - 1169.

[9] Shankar K, Lakshmana Prabu S K, Gupta D, et al. Optimal feature-based multi-kernel SVM approach for thyroid disease classification [J]. The journal of supercomputing, 2020, 76: 1128 - 1143.

[10] Ogunseye E O, Adenusi C A, Nwanakwaugwu A C, et al. Predictive analysis of mental health conditions using AdaBoost algorithm [J]. Paradigm Plus, 2022, 3 (2): 11 - 26.

[11] Horvat T, Job J. The use of machine learning in sport outcome prediction: A review [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2020, 10 (5): e1380.