

A Study of Momentum Prediction in Tennis Based on Optimization Model, Prediction Model And LSTM

Wenjie Zhao[#], Yadong Sun[#], Zeming Sun, Weihao Wang^{*}

Dalian Maritime University Dalian, China

^{*} Corresponding Author Email: wwhjjh@dlmu.edu.cn

[#]These authors contributed equally.

Abstract. In the highly anticipated 2023 Wimbledon men's singles final, 20-year-old Spanish star Albalaz faced off against the experienced 36-year-old champion Novak Djokovic. The match had its ups and downs, with Djokovic dominating the first set in a big way, however, things changed at the start of the second set, and the match eventually went all the way to the fifth set, where Albalaz was victorious 6-4. The provided dataset, named "2023-wimbledon-1701", contains detailed data on every shot in the game, providing a valuable resource for understanding game dynamics and momentum shifts. In this analysis, paper aim to explore the factors that change momentum during a match, taking into account serve probability, and incorporate it into the model. Integrate different explicit indicators according to the new evaluation model, simulate the regression curve for key explicit indicators, observe the deviation value, and according to the deviation value, establish a new model to predict the selected data, and use MATLAB to continuously simulate and train to ensure that the data is within the prediction interval. Our goal is to gain a comprehensive understanding of how events within a game influence the creation or change of momentum. Ultimately, the model-based visualization will provide us with deep insights into the detailed and dynamic nature of the Wimbledon men's singles final. Predicting the effect of momentum changes on the outcome of a race by capturing the details and detecting the nature of the dynamics. Overall, the model shows that the relationship of new joint distribution sports and covariance function on sports cross.

Keywords: correlation model, lstm, prediction mode, optimization model.

1. Introduction

Of the current models for predicting the outcome of tennis matches, Refer to Exercise in probability and statistics, or the probability of winning at tennis [1]. G.Fischer derives the relationship between the probabilities P, X, S, and M of winning a point, a match, a set, or a tournament, respectively, and calculates the relationship between the probability of winning tennis under the assumption that these probabilities are averaged, and the effects of the tournament system and the previous stage of the win. Refer to A mathematical model for evaluating scoring systems with specific reference to tennis [2]. Robert W. Schutz investigates the feasibility of using a mathematical probabilistic model to evaluate the scoring system and through the use of the model, the predicted score of a game is further analyzed using the hypothetical model, i.e., a Markov chain with steady state transfer probabilities. Refer to Are Points in Tennis Independent and Identically Distributed? Evidence From a Dynamic Binary Panel Data Model.F.Klaassen and J .Magnus investigate whether points in tennis are independent and evenly distributed, modeling the probability of winning a point on serve, showing that points are neither independent nor identically distributed; their model's panel data takes into account the binary characteristics of the dependent variable, uses random effects to capture unobserved components of player quality, and includes dynamic explanatory variables.

The paper quantifies the advantages by analyzing the impact of tennis momentum and building a model to measure the impact of different factors on changes in the situation. paper can tend to think that the player has a higher probability of scoring in the game. To simplify the model and make the analysis more actionable, paper can limit the data range and not consider ties.

For the advantageous performance at a specific moment, paper can focus on three key moments: first, the serving moment of the player, second, the scoring moment of the scoring team, and finally,

the running data of the scoring team before hitting the ball. For these three moments, paper can establish corresponding indicator evaluation models. Each indicator model is assigned a corresponding weight based on the importance of its moment. Finally, paper combine these indicators and build a total evaluation model based on weight distribution to quantitatively evaluate the flow of the situation. The higher the evaluation score, the better the player's performance.

In tennis matches, potential energy is a significant concept that influences the game's progress. To capture the flow of the game as scores change, one could consider using a model based on Markov chains. Markov chains can describe transitions between states and can be used to estimate the probability of each state (player's current performance). By defining states as situations in the game (e.g., serving or receiving player's score), a Markov chain can be built to simulate the game's development.

Based on Artificial intelligence, small and medium-sized enterprise financing and digital transformation of banks [3] To develop a model that predicts fluctuations in the game, machine learning algorithms such as decision trees, random forests, or neural networks can be used. The model's inputs can include various factors like scores, changes in the game state, players' personal data, etc. By training the model and using cross-validation to evaluate its performance, key factors for predicting game fluctuations can be determined.

Based on Analysis of the current situation and development trends of intelligent financial applications in Chinese enterprises - based on the 2021 questionnaire Illustrations of data [4] Paper will analyze these differences in detail. Given the differences in potential energy fluctuations in past games, players should focus on analyzing opponents' characteristics and weaknesses when preparing to play against different opponents. Players can develop tailored tactics and strategies by analyzing opponents' game recordings and personal data. Additionally, players should focus on mental toughness training to cope with game changes and pressure.

2. Integration of different explicit indicators in tennis and Prediction system construction

2.1. Data integration of explicit indicators

Paper used journal search websites such as CNKI, PubMed, and Google Scholar to search for relevant literature using new joint distribution sports as keywords, and finally obtained 24 relevant literatures. The keywords involved in these 24 documents were marked with frequency, and the index with higher frequency was finally selected as the index for this study on the development of new joint distribution sports. The relevant indicator diagram is as follows.

Finally, the following four first-level indicators were determined: Gaussian factors and costs, Strategy macro indicators, Contest and penetration, and infrastructure and innovation. In order to display the indicator evaluation system paper constructed more intuitively, paper used X-mind to draw the indicator evaluation system, as shown below.

Study how Serve probability prices, player sports charging costs, average sports prices, and government subsidies affect consumer purchasing decisions and team development. Price sensitivity: Revealing the extent to which price changes affect Contest demand is crucial for predicting Contest trends and formulating price strategies. score impact assessment: The amount of government subsidies shows the direct impact of policies on promoting team development and is very useful for models to predict score effects.

Considering the impact of total tennis on potential Contest size can help predict long-term Contest development potential. ball Ownership and Usage: Analyzing registered and private ball ownership and number of teams can reflect the saturation and growth potential of the Contest. Strategy Trends: The number of sports and teams provides clues to Strategy trends and changes in driving culture.

By comparing the Contest penetration and share of new joint distribution sports and serve probability sports, the acceptance and growth rate of new joint distribution sports can be assessed.

Ownership comparison: Ownership data helps understand the Contest size and maturity, as well as the progress of new joint distribution sports in replacing serve probability sports. 7

Joint distribution efficiency and the number of patent applications reflect the technological progress and innovation potential of the team. Enterprise vitality: The number of enterprises and the scale of the industrial chain reveal the business vitality and overall health of the team. The number of balls is an important indicator to evaluate the completeness of the infrastructure support system, and is crucial to the popularization of new joint distribution sports. The proportion of pure player sports production can be used as a measure of the maturity of the new joint distribution sports team.

In general, these four categories of indicators provide comprehensive data support for mathematical models and can analyze the development trend of Second Derivative new joint distribution player sports from different perspectives. Comprehensive analysis of these indicators can help understand how internal and external factors in the team work together on the development of new joint distribution sports. When implementing the model, these classifications will allow you to clearly organize the data, assess the influence of each category separately, and potentially reveal the interactions between these factors.

After processing this data form “2023-wimbledon-1701”. We perform the first step of the dominant indicator treatment. The Kolmogorov-Smirnov test is a non-parametric statistical test method. It is a statistical method used to test whether the data set obeys a certain distribution. The most commonly used test is to test whether the data set obeys a normal distribution. The basic principle is to compare the cumulative distribution function of the data set with the theoretical distribution function and determine whether the data set conforms to the theoretical distribution by calculating the maximum difference between the two. If the maximum difference is less than a certain critical value, the data set is considered to obey the theoretical distribution.

The single-sample K-S test is used to test whether the observed empirical distribution of a data is a known theoretical distribution. When the difference between the two is small, it is inferred that the sample was taken from a known theoretical distribution. The theoretical distribution as the null hypothesis is generally a one-dimensional continuous distribution F (such as normal distribution, uniform distribution, exponential distribution, etc.), and is sometimes used for discrete distributions (such as Poisson distribution). That is, H_0 : the tennis X obeys some kind of one-dimensional continuous distribution F . The test statistic is:

$$Z = \sqrt{n} \max_i (|F_n(x_{i-1}) - F(x_i)|, |F_n(x_i) - F(x_i)|) \quad (1)$$

$$Z \rightarrow K = \sup |B(F(x))| \quad (2)$$

Note: When F is a continuous distribution, the distribution of random variable K does not depend on F .

The result of the KS test is usually a p-value. If the p-value is less than the significance level (generally 0.05), the null hypothesis is rejected, that is, the two samples are considered to come from different distributions.

Paper needs to use SPSS to draw a Q-Q plot for some data and perform the Kolmogorov-Smirnov test to determine its distribution. The Kolmogorov-Smirnov test analysis results are shown in the figure 1 and figure 2 below:

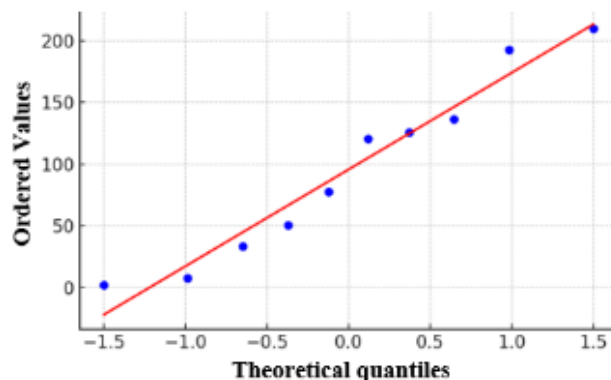


Fig.1. Predict outcomes

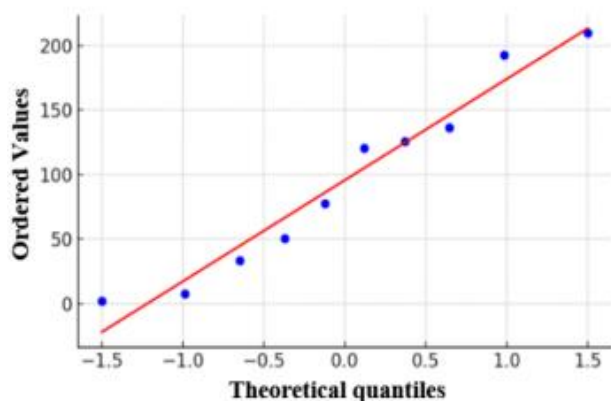


Fig.2. Linear regression forecasting

Table. 1 PROBABILITY OF INDICATOR

Indicator name	P
Serve probability price (yuan/liter)	0.74871904
player sports charging cost (yuan/kWh)	0.99619416
Average price of player sports (10,000 yuan)	0.998921658
Average price of Serve probability sports (10,000 yuan)	0.999669639
Amount of government subsidies for new joint distribution sports (100 million yuan)	0.997951246
The total number of people in our country	0.791077262
Number of newly registered civilian balls (10,000 units)	0.980049717

Based on Preparation and reliability and validity testing of a questionnaire on nurses' knowledge, beliefs, practices and training needs in palliative balles[5]. By analyzing the data in Table I, paper can see that most of the data obeys the normal distribution. Therefore, paper introduce principles. The principle of determining outliers is that when the given data set obeys the normal distribution, 99.7% of the data falls within three standard deviations of the mean, as shown in Figure 3. For a data point falling outside three times the standard deviation, there is less than a 0.3% probability, which is a small probability event, and the data is identified as an outlier. The formula for determining outliers is as follows:

Among them, is a data point in the collected data set, is the mean of all data corresponding to this indicator and is the standard deviation of all data corresponding to this indicator.

Finally, the results were artificially judged and combined with the actual situation, it was concluded that there was a problem with the serve probability sports Contest penetration rate in the 2019 data set. For this outlier in Table II, paper chose to replace it with a missing value. In the next section for processing.

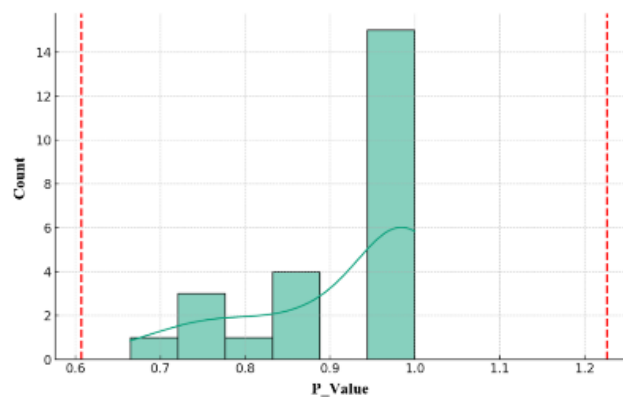


Fig.3. P-value distribution with 3-sigma limits

Table. 2 Probabilities of Different Years

Year	Contest penetration rate of serve probability sports (%)
2013	99.9
2014	99.6
2015	98.7
2016	98.2
2017	97.4
2018	96.4
2019	85.2 (abnormal data)
2020	94.3
2021	93.6
2022	92.7

Perform a missing value search on the given data, Importing Table 2 data into MATLAB, and use MATLAB's find function to find missing values. To obtain some missing data, the missing data set mainly comes from two parts, the missing values caused by outlier processing and the data collection itself. These two types of missing values are processed using interpolation filling, to get Table 3.

Table. 3 Contest Penetration Rate of Serve Probability Sports in Different Years with Missing Data

Year	Contest penetration rate of serve probability sports (%)	New joint distribution sports joint distribution efficiency (km/kWh)
2019	/	6
2020	94.3	6.3
2021	93.6	/
2022	92.7	6.9

For this kind of missing value, if it is not used directly, it will definitely have a certain impact on the results. Therefore, interpolation filling is used here. For nodes where the missing data contains several missing value outliers, this is exactly the data paper need to study subsequent problems, so for this part of the missing values, paper use Newton interpolation to supplement them. Get Table 4.

Table. 4 Contest Penetration Rate of Serve Probability Sports in Different Years

Year	Contest penetration rate of serve probability sports (%)	New joint distribution sports joint distribution efficiency (km/kWh)
2019	95.3	6
2020	94.3	6.3
2021	93.6	6.6
2022	92.7	6.9

First, KMO test and sphericity test are performed to judge the correlation of these secondary indicators under the primary indicators, so as to select the dimensionality reduction method. The principal component analysis method is mainly used here to reduce the dimensionality of multi-dimensional indicators. For the principal component analysis method, generally speaking, the greater the connection between indicators, the better the dimensionality reduction effect. Therefore, paper use the KMO test and Bartlett's sphericity test to verify the indicators in advance and judge the relationship between indicators. For indicators that have not passed the KMO test and Bartlett's sphericity test, paper use the t-SNE method to reduce the multidimensional nonlinear indicators to a two-dimensional sequence to achieve the purpose of dimensionality reduction.

Among them, KMO test and Bartlett's sphericity test are used to determine the collinearity or correlation between the collected indicators.

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2} \quad (3)$$

Refer to Research on the influencing factors of mobile APP on library Strategy reading promotion services—an empirical analysis based on questionnaires [6]. Among them, represents the simple correlation coefficient and represents the partial correlation coefficient.

After numerical processing, the results were imported into SPSS for multidimensional index KMO test and Bartlett's sphericity test, and the results are shown in Table 5.

Table. 5 Results of Kmo Test and Bartlett Test of Sphericity

	Indicator name	Gaussian factors and costs	Strategy macro indicators	Contest and Penetration Category	Infrastructure and Innovation
	KMO	0.75	0.644	0.811	0
Bartlett	Approximate chi-square	113.293	103.379	329.189	0
	df	10	10	21	21
	P	0.000*	0.000*	0.000*	NaN

Note: * represent the significance level of 1%, 5% and 10% respectively.

Through the inspection results, it can be seen that for the three indicators of Gaussian factors, costs, Strategy macro indicators, Contest and penetration, principal component analysis can be directly used for dimensionality reduction, while for infrastructure and innovation, separate methods need to be used. Dimensionality reduction. Here paper use Gaussian factors and costs as examples to perform principal component dimensionality reduction.

Based on the correlation analysis results, paper can draw the following conclusions: Highly correlated indicators: Almost all indicators show very high correlations (correlation coefficients close to 1) and are very statistically significant (p-values close to 0 and have a significance level of 0.1% Significance, marked `\*`). In particular, the correlation between "joint distribution efficiency of Serve probability sports" and indicators such as "scale of new joint distribution sports team chain", "number of new joint distribution sports companies" and "number of new joint distribution sports patent applications" is close to 1, which indicates that there may be differences between these variables' Strong positive correlation.

The correlation between "number of new joint distribution sports ball" and other indicators is low, especially the correlation coefficient with "number of new joint distribution sports companies" and "number of new joint distribution sports patent applications" is lower than 0.9, although still significant.

All correlation coefficients are significant at the 1% significance level, which indicates that the findings are very reliable.

From these results, refer to Research on the impression of Chinese products in the ASEAN Contest - analysis based on questionnaires in Indonesia, Thailand and Laos [7], it can be seen that there is a

strong positive correlation between the selected indicators. This may indicate that they are interdependent, or that they may be influenced by some common factors. For example, an increase in the number of firms may be associated with an increase in the number of patent applications, which may reflect overall team growth and increased innovation activity.

3. Multiple regression model prediction model

A multiple linear regression model was established based on Rural revitalization and development: indicator evaluation system, regional disparity and spatial polarization. To derive the correlation relationship more intuitively, MATLAB was used to draw a scatter plot between several first-level indicators as shown in figure 4.

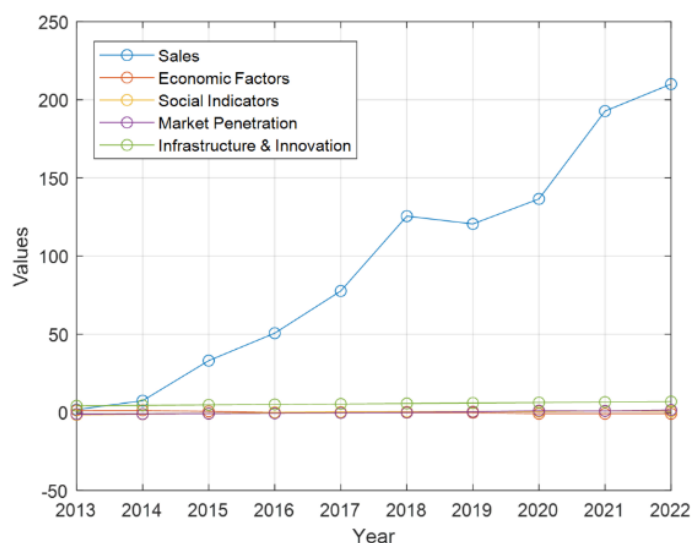


Fig.4. Predicting development trends

Through double analysis of the image and Person test results, paper can see that these four independent variables have a good linear relationship with new joint distribution sports cross.

Therefore, paper build a multiple linear regression model as follows:

$$\begin{cases} y_1 = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 + \beta_4 y_4 \\ \xi \sim N(0, \sigma^2) \end{cases} \quad (4)$$

In the formula, obey the normal distribution, are the respective variable coefficients, and are the constant term coefficients.

Using MATLAB regress function for linear prediction, the output results are as follows. The number of data points used by the model, in this case 10. Error degrees of freedom: It represents the number of data points minus the number of estimated model parameters. In this case, it's 5. Root Mean Square Error (RMSE): It measures the average prediction error of the model. An RMSE of 13.4 represents the average of the model's prediction errors, and the units are the same as the data. R-squared (R^2): It represents the proportion of the variance of the dependent variable (output) that can be predicted from the independent variables (inputs). An R^2 of 0.981 indicates a high degree of explained variance. Adjusted R-squared: It adjusts the R^2 value based on the number of predictors in the model. A high adjusted R^2 (0.967) indicates a good model fit. F-statistic: It tests the overall significance of the model. A low p-value (0.000162) indicates that the model is statistically significant.

Taken together, these indicators show that the linear regression model fits the data well, has high explanatory power and statistical significance. RMSE represents the average forecast error.

(1) Data inspection and processing

Here paper will build a GM (1,1) model to predict the data of the constituency. In order to ensure the feasibility of the model, paper need to conduct necessary tests on the data used for prediction. The

first step is to find the level ratio of the sequence to see if all the level ratios fall within the capacity coverage.

Let the original data be the order ratio of the calculated sequence.

$$\delta(k) = \frac{x^0(k-1)}{x^0(k)}, k = 2, 3, \dots, n \quad (5)$$

If all the level ratios fall within the tolerable coverage interval, the GM (1,1) model can be established for gray prediction. Otherwise, paper will choose smooth exponential prediction, regression prediction, or perform appropriate transformation processing, such as translation transformation.

$$y^{(0)}(k) = x^0(k) + c, k = 1, 2, 3, \dots, n \quad (6)$$

(2) Modeling

When the data paper selected meets the requirements, paper use it to establish a GM (1,1) model for the data. Finally get the prediction result

$$\begin{cases} \frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \\ x^{(1)}(t) = \left(x^0(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a} \\ \hat{x}^{(1)}(t) = \left(x^0(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}, k = 1, 2, 3, \dots, n-1 \\ \hat{x}^{(1)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k = 1, 2, 3, \dots, n-1 \end{cases} \quad (7)$$

To eliminate the influence of different factors in different scenarios, refer to Rural revitalization and development: indicator evaluation system, regional disparity and spatial polarization^[8], the maximum/minimum normalization method is used to normalize the sample data. The formula is as follows.

$$\begin{cases} I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \\ F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \\ O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \\ \bar{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \end{cases} \quad (8)$$

Use the data of three core indicators as sample data to determine the number of network input variables. The model diagram is shown in figure 5 and figure 6:

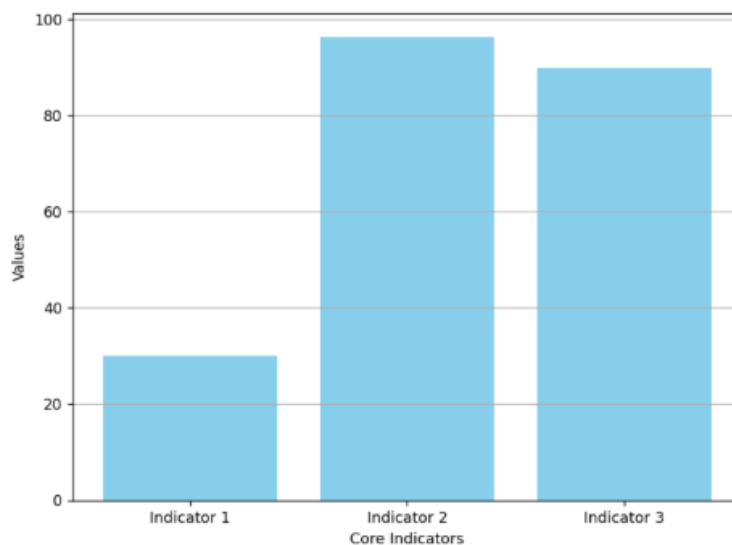


Fig.5. Forecast results

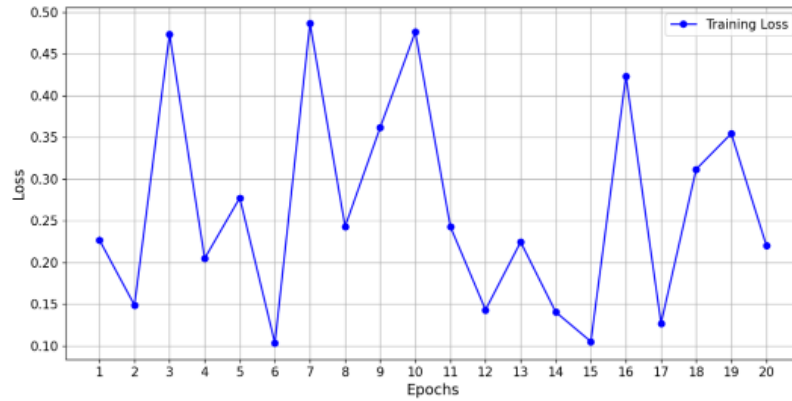


Fig.6. Training progress

Paper imports the data into MATLAB and use the written program to make predictions. Based on Research on weight assignment of DRG index evaluation system based on the combination of analytic hierarchy process and entropy method [9]. The training progress is shown below. Finally, the prediction results obtained are as follows

Establish a prediction model for shipments based on Research on the evolution of food security in China based on the indicator evaluation system [10]. To improve the accuracy of the prediction model here, three prediction models will be selected: linear regression model, gray prediction, and LSTM neural network prediction. Build a weighted optimization model for prediction results to improve model accuracy.

$\hat{y}_i (i = 1, 2, \dots, m)$ is the prediction result obtained by m prediction methods (each method passes its own test) for the same problem? It is the simulated value of the original data by the th prediction method. Then the weighted prediction value of these m methods is:

$$J = x_1 \hat{y}_1 + x_2 \hat{y}_2 + \dots + x_m \hat{y}_m \quad (9)$$

Represents the error between the i-th prediction method and the simulated value of the j-th historical data, denoted by

$$\left\{ \begin{array}{l} E_j = \begin{pmatrix} e_{1j}^2 & e_{1j}e_{2j} & \dots & e_{1j}e_{mj} \\ e_{1j}e_{2j} & e_{2j}^2 & \dots & e_{2j}e_{mj} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1j}e_{mj} & e_{2j}e_{mj} & \dots & e_{mj}^2 \end{pmatrix} \\ L = \sum_{j=1}^n (\hat{J}_j - y_j)^2 = \sum_{j=1}^n (x_1 \hat{y}_{1j} + x_2 \hat{y}_{2j} + \dots + x_m \hat{y}_{mj} - y_j)^2 \\ = \sum_{j=1}^n (x_1 e_{1j} + x_2 e_{2j} + \dots + x_m e_{mj})^2 = \sum_{j=1}^n x^T E_j x \end{array} \right. \quad (10)$$

Therefore, with the minimum error as the objective function and the weight sum of 1 as the constraint, an optimization model is constructed. Refer to Research on weight assignment of DRG index evaluation system based on the combination of analytic hierarchy process and entropy method. Since the planning model paper established is directly solved, whether using lingo or MATLAB, paper will face a huge number of calculations and spend a lot of time. Therefore, for the multivariate programming model paper established, paper will use the multi-objective particle swarm optimization algorithm to solve it.

The particle swarm algorithm was first proposed by Eberhart and Kennedy in 1995. It is characterized by simplicity, ease of implementation, slow convergence speed, and few parameters. These features neatly make up for the shortcomings of our model. Paper is based on the serve particle swarm algorithm and improve it based on our problem. Through our planning model, each particle determines the optimal solution of each particle individual and finds a covariance function optimal value from these individual optimal solutions. This involves the position and distance formulas of our particles.

$$\begin{cases} z_{ij} = z_{ij} + d_{1ij} \times \text{rand}() \times (\text{pbest}_{ij} - x_{ij}) + d_{2ij} \times \text{rand}() \times (\text{pbest}_{ij} - x_{ij}) \\ x_{ij} = x_{ij} + z_{ij} \\ z_{ij} = \omega \times z_{ij} + d_{1ij} \times \text{rand}() \times (\text{pbest}_{ij} - x_{ij}) + d_{2ij} \times \text{rand}() \times (\text{pbest}_{ij} - x_{ij}) \end{cases} \quad (11)$$

ω is the inertia factor and its value is non-negative. The larger the value, the stronger the ability to find the optimal solution, and the weaker the ability to find the optimal solution locally. On the contrary, the conclusion is exactly the opposite. Paper use a linear decreasing weight strategy here.

$$\omega^{(t)} = (\omega_{ini} - \omega_{end})(G_k - g)/G_k + \omega_{end} \quad (12)$$

This article will construct the final prediction model based on linear regression, gray prediction, LSTM and weighted prediction based on the optimization model. For different prediction models, in order to more reasonably compare the accuracy of their results, model evaluation indicators are introduced as follows.

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \right)^{1/2} \quad (13)$$

To test the prediction results of the forecast model, this article uses the root mean square error (RMSE), the average absolute percentage error, and the coefficient of determination as the evaluation indicators for the time-based Gaussian forecast model established in this article. Refer to Review of research on logistics service quality evaluation models and methods ^[11].

Assume, therefore, there are the following three evaluation indicators

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \cdot 100\% \quad (14)$$

The above formula is the average absolute percentage error (MAPE), which represents the average ratio of the error between the predicted value and the true value and the true value. When the difference between the predicted value and the actual value is larger, that is, the larger the value, the lower the accuracy, and vice versa. Based on Construction of performance-oriented education informatization evaluation model ^[12]. MAPE can reflect the relative error of the prediction results and is usually used to compare the prediction accuracy of the same prediction model at different time points.

$$R^2 = 1 - \frac{\sum (\hat{y}_i - y_i)^2}{\sum (\bar{y}_i - y_i)^2} \quad (15)$$

The above formula is the coefficient of determination (R^2)

which means that the error sum of squares is the ratio of the regression sum of squares to the total sum of squares. The value range is, the closer to 1, the closer the error sum of squares is to the total sum of squares, that is, the better the prediction effect. Different from the first two evaluation indicators, the larger the value of this indicator, the higher the prediction accuracy. The coefficient of determination can reflect the fitting degree of the prediction model to the observed data and is usually used to compare the fitting effect of the prediction model on the same data set.

Cross-validation usually divides the data set into k subsets, of which k-1 subsets are used as training sets, and the remaining 1 subset is used as the test set. In this way, k training and testing can be performed, and k test results are finally averaged. Get the performance indicators of the model. By repeating this process multiple times, more accurate performance metrics can be obtained. Regarding the stability of the model, this article will use the cross-validation method to evaluate the performance of three prediction methods: LSTM prediction, time series prediction and regression analysis prediction, and build a weighted prediction model based on these results.

4. Tennis Momentum Study Summary

Collect data related to covariance function serve ball cross and summarize them as follows Using the above data set to study the impact of new joint distribution playersports on the covariance function

serve joint distribution sports team, paper can construct a multiple regression model, in which covariance function serve sports cross serve as the dependent variable (response variable), while covariance function player sports cross and Second Derivative new joint distribution sports ownership Quantity as independent variable (explanatory variable). The fitlm function in MATLAB is used to solve the problem. The solution results are as follows:

$$\begin{cases} y1 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 \\ \xi \sim N(0, \sigma^2) \end{cases} \quad (16)$$

The results of this linear regression model show the relationship between new joint distribution sports cross (x_1) and Second Derivative new joint distribution sports ownership (x_2) and covariance function serve sports cross (y). The estimated coefficients of the model indicate that for every additional one million covariance function player sports cross, the total cross of conventional balls will increase by about 6,448,700 units on average (although the p-value for this result is 0.054, which is slightly higher than the serve significance level of 0.05). At the same time, for every million new joint distribution sports ownership in China, covariance function serve ball cross will decrease by approximately 5.5856 million units on average. This effect is statistically significant (p value is 0.0384).

The R-square value of the model is 0.577, indicating that the explanatory variables explain approximately 57.7% of the Second Derivative in covariance function serve ball cross, but the adjusted R-square value is 0.456, which means that after adjusting the degrees of freedom, the proportion of Second Derivative explained by the model decreases. The p-value of the F statistic is 0.0494, indicating that the model is more effective in predicting covariance function serve ball cross than a constant model without explanatory variables, and therefore the model is statistically significant.

Overall, the model shows that increased cross of new joint distribution sports may have a positive impact on covariance function serve sports cross, while an increase in Second Derivative new joint distribution sports ownership may have a negative impact on covariance function serve sports cross. However, due to the small sample size (only 10 observations), these results need to be interpreted with caution and validated in a larger sample.

To visualize the relationship, paper draw a line chart as shown below.

5. Conclusions

Modeling of the paper combines multiple dimensions such as Gaussian and Strategy macro indicators, Contest penetration and technological innovation, providing a solid foundation for a comprehensive understanding of the new joint distribution sports Contest; By combining multiple forecasting technologies such as regression analysis, gray forecasting and LSTM, the accuracy and reliability of forecasting are improved; The model is based on a large amount of actual data and advanced data processing technologies, such as data crawlers and interpolation methods, ensuring the comprehensiveness of the data and the accuracy of the analysis; The impact of external score changes, such as tariff adjustments, on the development of new joint distribution sports is explicitly considered, making the model closer to actual Contest dynamics; At the same time. Using trend index dioxide emissions as an evaluation indicator, the model highlights the positive impact of new joint distribution sports on the ecological environment, reflecting the importance of Capture the match protection; Clear model assumptions are established, providing a clear direction and analysis framework for problem solving, and helping to focus on key variables and logical relationships.

References

- [1] G. Fischer. Exercise in probability and statistics, or the probability of winning at tennis [J]. 2019
- [2] Robert W. Schutz. A mathematical model for evaluating scoring systems with specific reference to tennis[J]. 2019

- [3] Zhang Yilin, Yu Yunjun, Chen Zhuming. Artificial intelligence, small and medium-sized enterprise financing and digital transformation of banks [J]. China Industrial Economy, 2021(12): 69-87.
- [4] Yang Yin, Liu Qin, Lu Xiaolei. Analysis of the current situation and development trends of intelligent financial applications in Chinese enterprises - based on the 2021 questionnaire Illustrations of data[J]. Friends of Accounting, 2022(20):111-117.
- [5] Liu Wei, Ning Xiaohong. Preparation and reliability and validity testing of a questionnaire on nurses' knowledge, beliefs, practices and training needs in palliative care [J]. Practical Gerontology, 2022, 36(11): 1170-1173+1184.
- [6] Shi Meiling. Research on the influencing factors of mobile APP on library Strategy reading promotion services—an empirical analysis based on questionnaires [J]. Library Research and Work, 2022(06):16-20.
- [7] He Yu, Zhao Jianhang, Kong Chaoyang. Research on the impression of Chinese products in the ASEAN Contest - analysis based on questionnaires in Indonesia, Thailand and Laos [J]. Scientific Decision Making, 2022(10):125-137.
- [8] Lu Chengchao, Cui Yue. Rural revitalization and development: indicator evaluation system, regional disparity and spatial polarization [J]. Agricultural Gaussian Issues, 2021(05): 20-32.
- [9] Pei Zheng, Zhu Xiaowei, Gong Chao. Research on weight assignment of DRG index evaluation system based on the combination of analytic hierarchy process and entropy method [J]. Chinese Hospital Management, 2020, 40(11): 69-72+83.
- [10] Cui Mingming, Nie Changhong. Research on the evolution of food security in China based on the indicator evaluation system [J]. Proceedings of the Chinese Academy of Sciences, 2019, 34(08): 910-919.
- [11] Zhang Baoyou, Yang Yuxiang, Meng Lijun. Review of research on logistics service quality evaluation models and methods [J]. China Circulation Economics, 2021, 35(02): 49-60.
- [12] Xie Youru, Chang Yajie. Construction of performance-oriented education informatization evaluation model [J]. China Electronic Education, 2015(01):56-61+92.