Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 5, No. 2, 2025

Research on Olympic Medal Distribution Prediction Based on Exponential Smoothing, Frequency Probability, and Logistic Regression Models

Hongyu Zhao*, Qingyang Li, Guyu Ding

[1]Southeast University, Nanjing, China
* **Corresponding author**: Hongyu Zhao (Email: 213220240@seu.edu.cn)

Abstract: This paper proposes the Exponential Smoothing Model, Frequency Probability Model, and Logistic Regression Model, focusing on the prediction methods of multiple models in scenarios with varying data characteristics. First, for scenarios with large datasets or stable trends, the Exponential Smoothing Model is introduced. It captures trend changes through weighted averages of historical data (with higher weights assigned to more recent data) and, through cross-validation, determines the parameters to predict future total and categorized quantities of the target object while analyzing their changing trends. Next, for scenarios with smaller datasets or larger fluctuations, the Frequency Probability Model is used, estimating future probabilities based on the frequency of historical events. By comparing data at different stages of typical cases, the model quantifies the impact of specific factors on increasing the probability of an event occurring. Finally, for zero-sample scenarios, the Logistic Regression Model is constructed, incorporating variables such as participant scale and historical participation frequency. Adaptive optimization algorithms are used to solve the parameters and predict the probability of first-time events, identifying high-potential objects and fields, and providing recommendations for resource allocation. The Exponential Smoothing Model is effective for predicting stable data, the Frequency Probability Model quantifies the probability characteristics of repeatable events, and the Logistic Regression Model provides a multi-factor probability prediction framework for zero-sample scenarios. Together, the three models offer diverse predictive approaches for different data characteristic scenarios.

Keywords: Data Feature Adaptation, Exponential Smoothing Model, Frequency Probability Model, Logistic Regression Model.

1. Introduction

This paper focuses on the data prediction problem in multiple scenarios, aiming to explore prediction methods and influencing factors under different data characteristics by constructing differentiated models [1]. First, for scenarios with large data scales or stable trends, the Exponential Smoothing Model is introduced. By processing historical data with weighted averages, this model captures trend changes in data sequences and enables dynamic prediction of future totals and categorized indicators of the target object. The evolution of these indicators is then analyzed and compared [2]. Next, for scenarios with small data scales or large fluctuations, the Frequency Probability Model is used. Based on the historical event occurrence frequency, it estimates future probabilities [3]. Finally, for zero-sample scenarios, a Logistic Regression Model is constructed. It predicts the probability of a first-time event occurring for the target object, identifies potential objects and fields, and provides targeted resource allocation strategies based on the model results [4-6]. Experimental results show that the Exponential Smoothing Model demonstrates excellent predictive performance in stable data scenarios, the Frequency Probability Model effectively characterizes the probability features of repeatable events, and the Logistic Regression Model provides a multi-factor analysis-based probability prediction framework for zero-sample scenarios. Together, the three models provide multi-level methodological support for prediction problems and resource allocation decisions in different data characteristic scenarios.

2. Exponential smoothing model: medal prediction method for stable data scenarios

In order to solve the Olympic medal distribution prediction problem, a variety of different statistical models were constructed for different scenarios to predict the medal distribution of the 2028 Los Angeles Olympic Games and to analyze the main factors affecting the medal distribution. In feature engineering, key information is extracted or computed from historical data, such as the number of historical medals, the number of participants, and the dominance of the host country, among other characteristic variables, which are essential for understanding the medal distribution patterns and constructing accurate prediction models.

For countries dealing with a sufficiently large or stable number of medals, an exponential smoothing model is introduced to ensure effective medal forecasting under stable data conditions. The exponential smoothing model is a time series forecasting method that captures trend changes in the data by performing a weighted average of the historical data, giving higher weights to the most recent observations, which is used to make smoothed forecasts of future values.

2.1. Exponential smoothing model implementation steps

Step 1: Select the smoothing coefficient α : The smoothing coefficient is between 0 and 1, which determines the influence range of the current observed value on the prediction results linear measure. A larger α value makes the model more sensitive to the most recent observations, while a smaller A value weights the model to the historical data More uniform.

Step 2: Initialization: Select the initial level.

Step 3: Iterative update: Calculate the smoothing value at each time point step by step.

Step 4: Prediction: Use the last calculated level value to predict the future value.

$$Y_t = \alpha Y_{t-1} + (1-\alpha)Y_{t-2} + \cdots + (1-\alpha)^n Y_{t-n}, \quad \alpha \in (0,1) \tag{1}$$

Crossing validation: the data set was divided into training and test sets, and the training set was used to fit the model to calculate prediction errors at different a values Difference and then select the one with the lowest prediction error.

As a result, α was chosen as 0.8 after cross-validation. A bar chart has been plotted comparing the total number of medals and gold medals won by the top ten major countries in 2024 with the predicted total number of medals and gold medals for 2028.

2.2. Result and analysis

Considering that the first few countries are almost stable, predictions are made and visualized for these countries.

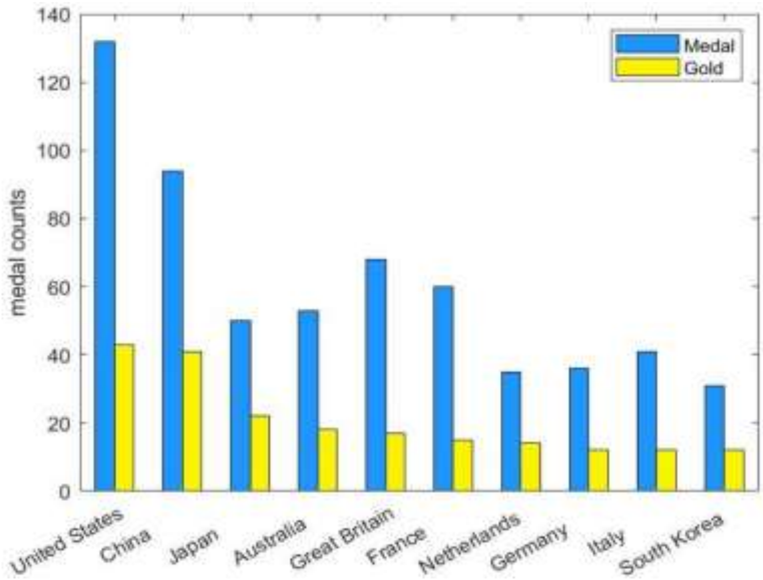

Figure 1. Predicted top 10 countries in 2028 with total and gold counts

In Figure 1, the predicted top 10 countries for 2028 are analyzed, and their performance in the prediction is compared with the actual performance in 2024. The comparisons are shown in Figure 2 and Figure 3.

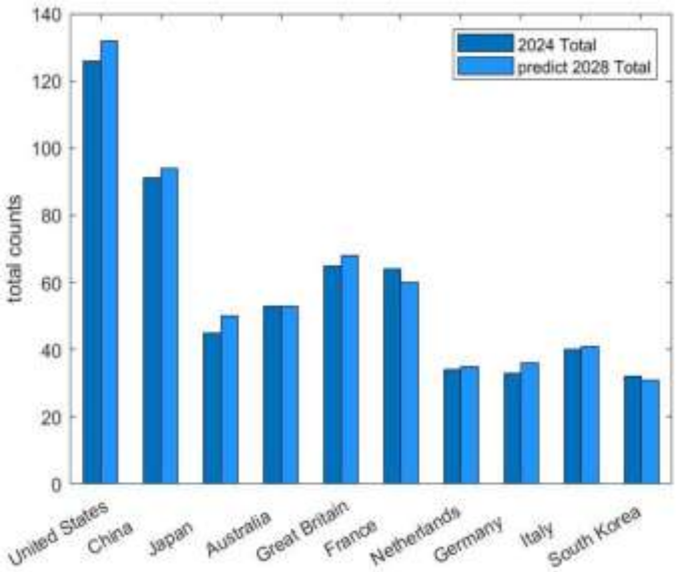

Figure 2. Comparison of predicted top 10 countries in 2024 and 2028(total)

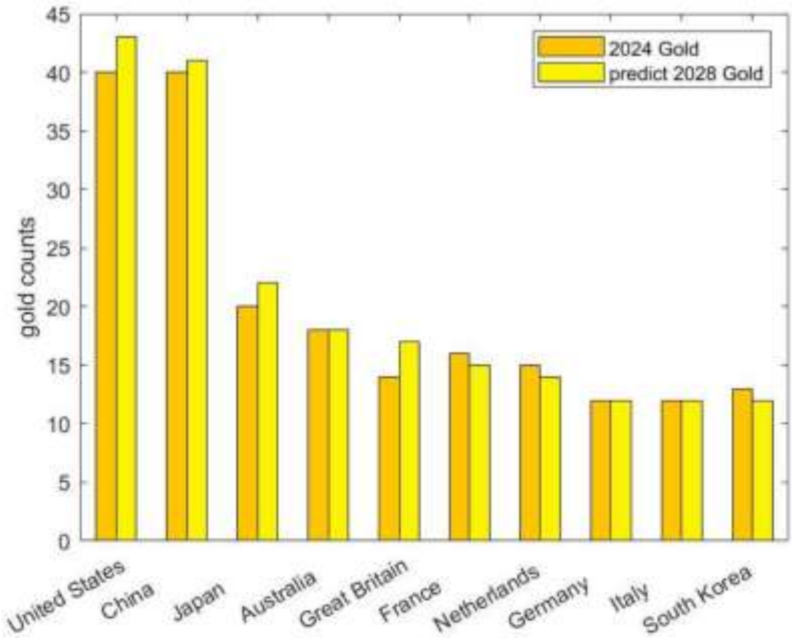

Figure 3. Comparison of predicted top 10 countries in 2024 and 2028(gold)

Among these countries, the medal counts of some have shown significant progress or regression. Therefore, predictions were made and presented in the Figure 4.

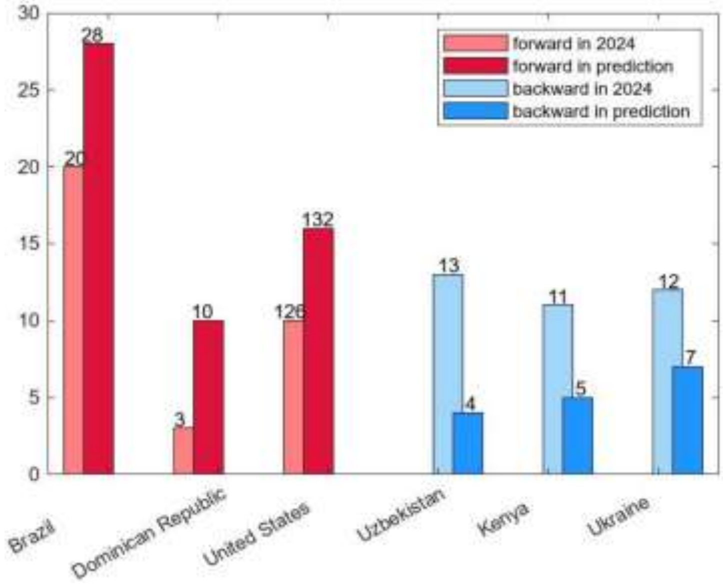

Figure 4. Predicted top 3 progressive and regressing countries

3. Frequency probability model: award probability estimation for small sample or fluctuating data

Frequency probability modeling is a method of estimating the probability of an event occurring in the future based on how often it occurs in history, and is suitable for situations where the sample size is large and the event is highly repetitive.

The frequency probability model is introduced for countries dealing with insufficient number of medals or erratic medal winning. Athletes are analyzed to assess the likelihood of winning medals in Olympic competitions. The method portrays the medal propensity of an athlete or team in terms of frequency-approximate probability by counting the frequency of medal wins in historical participation data.

The basic formula for the probability model is:

$$\mathrm{P} \approx \mathrm{f} = \mathrm{n} / \mathrm{N} \tag{2}$$

Where n denotes the number of times the event occurred (e.g., the number of awards) and N denotes the total number of attempts at the event (e.g., the number of entries). The central idea of the model is that if a certain type of event has occurred at a stable frequency over a large number of past trials, that frequency can be used as an approximation of its future probability of occurrence.

In this way, it is possible to estimate the effect of the "Great Coach" effect on the medal winning rate. Three representative international coaches are selected: Lang Ping (China), Omatsu Hirofumi (Japan) and Tom Maher (Australia), corresponding to volleyball, gymnastics and basketball respectively. The f_c and f_{nc} were calculated by counting the medals won before and after their coaching. Using a frequency estimation probability model, it was estimated that the presence of great coaches could improve the medal-winning rate by 17.5%.

Figure 5 illustrates the medal winning rates of the three coaches at different stages of the corresponding national teams:

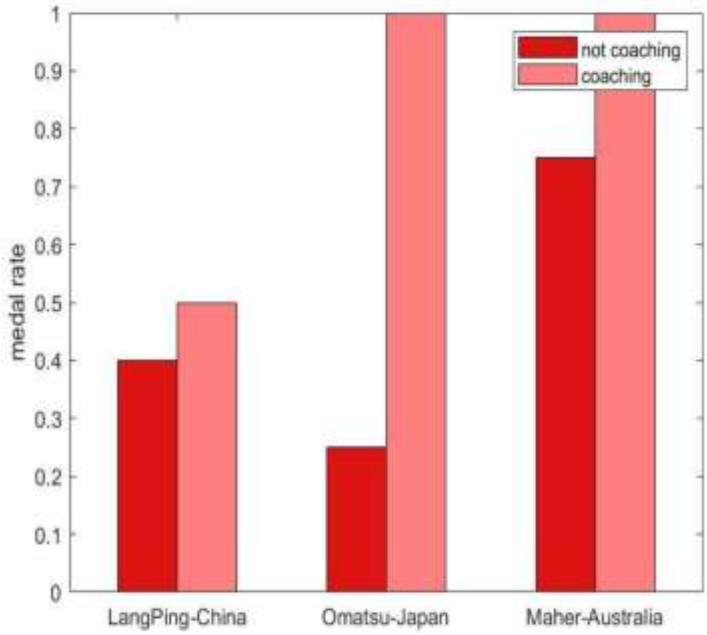

Figure 5. Medal rates under coaching or not with "Great Coach"

From Figure 5, it can be observed that during the coaching period of Lang Ping, Omatsu Hirofumi and Tom Maher, the medal winning rate of their countries is significantly higher than that of the non-coaching period, especially in the two groups of Omatsu-Japan and Maher-Australia, the medal winning rate increases significantly. This result confirms the positive role of the "Great Coach" effect in enhancing national Olympic competitiveness, indicating that the introduction of high-level coaches can significantly improve the overall performance of athletes and increase the probability of winning medals to a certain extent.

4. Logistic regression model: first-time medal probability prediction framework for countries with no medal records

For countries that have not yet won a medal, it is difficult to predict their likelihood of winning in future Olympic Games based on historical medal records alone. Therefore, a Logistic Regression Model (LRM) is introduced to construct a probabilistic prediction model to estimate the likelihood of a country winning an Olympic medal for the first time. The method quantifies the influencing factors of a country's first-time award through historical participation behavior and related characteristic variables, and provides decision support for resource allocation and strategic guidance.

4.1. Model structure and variable definitions

In the model for predicting first-time award-winning countries, the likelihood of receiving a first prize is hypothesized to be positively correlated with the number of project participants. Additionally, historical participation experience (cumulative number of sessions) is considered to increase the probability of winning a first prize. Therefore, logistic regression modeling was employed, with the formula as follows:

$$P_{\text{medal}} = \frac{1}{1+e^{-\left(\eta_0+\eta_1\cdot E+\eta_2\cdot T_{\text{participatior}}\right)}} \tag{3}$$

4.2. Parameter estimation and optimization methods

To solve for the optimal parameter β, the maximum likelihood estimation (MLE) method is used to construct the objective function:

$$\log L(\eta) = \sum_{i=1}^{n}\left[y_i \log P_{\text{medal}} + \left(1-y_i\right)\log\left(1-P_{\text{medal}}\right)\right] \tag{4}$$

y_i indicate whether the country has won the award for the first time (1 is the first award and 0 means no award). For the solution of parameters, the Adam optimization algorithm is introduced to replace the traditional gradient descent method, which combines the first-order and second-order momentum information, dynamically adjusts the learning rate of each parameter, and improves the stability and convergence speed of the algorithm under small batch of data.

Step 1: Initial parameters: initialized to zero

Step 2: Calculate gradient:

$$\nabla \log L(\eta) = \sum_{i=1}^{n}\left(y_i - P_{\text{medaf}}\right)X_i \tag{5}$$

Step 3: Parameter update:

$$\eta \leftarrow \eta + \alpha \nabla \log L(\eta) \tag{6}$$

Where α is the learning rate. Choose 0.01.

In this study, the parameter optimization process was enhanced by implementing the Adam optimization algorithm in lieu of standard gradient descent. Adam, an acronym for Adaptive Moment Estimation, is a widely adopted machine learning optimization technique that combines the advantages of RMSProp and momentum-based gradient methods. It dynamically adjusts the learning rate for individual parameters by leveraging first-order and second-order moment estimates of the gradient, resulting in faster convergence and greater stability.

The Adam algorithm starts by initializing the parameters, along with the first-order and second-order moment estimates, to zero. During each iteration, these values are updated using the gradient and exponential decay rates β_1 and β_2, which are set to 0.9 and 0.999, respectively. To counteract initialization bias in the early stages, bias correction is applied. The parameters are then updated using a learning rate α, which is selected as 0.01.

Through the adoption of Adam, the logistic regression model designed to predict the probability of countries winning for the first time was refined. In comparison to traditional gradient descent, Adam's adaptive updates facilitated faster convergence and enhanced robustness against noisy gradients. This refinement enabled more effective capture of the interplay between the number of project participants, historical participation data, and the probability of a country securing its first award.

4.3. Empirical analysis and prediction results

After training the model on all the "historical non-winning countries" samples, the model outputs the predicted probability of each country winning the award for the first time, and extracts the top ten countries in terms of probability, as shown in Figure 6.

As can be seen in Figure 6, the logistic regression model predicts countries that have never won an Olympic medal and produces a ranking of their probability of winning for the first time in 2028. Among them, Honduras (HON) is ranked first with a predicted probability of more than 30%, showing that it has a strong potential for a first breakthrough. It is followed by countries such as South Sudan (SSD), Iraq (IRQ) and Tuvalu (TUV), which did not reach the podium in the historical Olympics but have high program participation and stable team base, which lays a positive condition for their future medal contention. The results validate the effectiveness of the logistic regression model in dealing with the problem of predicting countries with "zero medals".

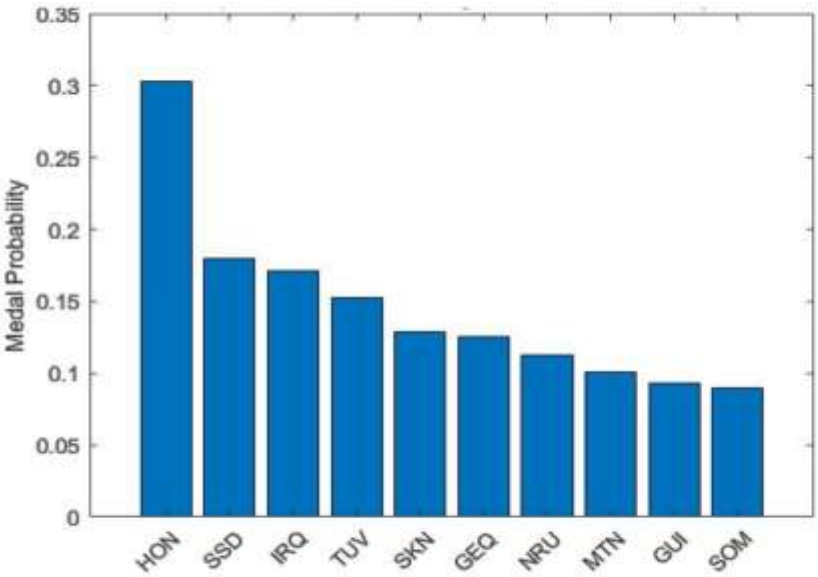

Figure 6. Top 10 countries with probability of winning the first medal

For countries that need to introduce excellent coaches, this study prioritizes those countries that have a high frequency of participation and a large number of competitors in specific sports, even though they have not yet won Olympic medals. These countries usually have a broad public base and enthusiastic participation, however, due to the lack of systematic guidance from high-level coaches, the competitive level has not yet been effectively transformed. Combined with the results of the previous analysis of the "Great Coach" effect, it can be inferred that focusing on investing coaching resources in such countries, especially introducing experienced coaching teams in potential sports, is expected to significantly improve their medal competitiveness. In order to further identify the key intervention targets, the relevant data were processed and analyzed to filter out the countries that have not won any prizes in certain key events but ranked in the top ten in terms of the number of participants. The results show that the top three sports have the most coaching efforts and potential benefits, and it is recommended to prioritize the deployment of high-level coaching resources for intervention.

Table 1. Top 10 non award countries

NOC	Sport	Athlete Count
ANG	Handball	58
HON	Football	54
ERI	Athletics	39
IRQ	Football	35
ISL	Swimming	28
BAR	Athletics	20
LBR	Athletics	20
SKN	Athletics	20

Table. 1 show that among the countries that have not yet won Olympic medals, Angola (ANG), Honduras (HON) and Eritrea (ERI) have a large number of participants in handball, soccer and track and field, with 58, 54 and 39 participants respectively. This shows that although these countries have not yet achieved a breakthrough in medals at the Olympic

Games, they have a broad selection base and enthusiasm for participation in some sports. Combined with the empirical analysis of the “Great Coach” effect, it can be inferred that these sports may be more limited by the lack of high-level coaches than by the lack of talent reserves at this stage. Therefore, it is recommended to prioritize the introduction of excellent coaching resources in key sports with high participation but low output, such as Angolan handball, Honduran soccer and Eritrean athletics, in order to improve the overall training system, so as to effectively stimulate the potential of medals, and help achieve the breakthrough of zero medals.

5. Conclusions

The This paper proposes the Exponential Smoothing Model, Frequency Probability Model, and Logistic Regression Model, focusing on the applicability and performance advantages of different models in the Olympic medal prediction scenario. First, the Exponential Smoothing Model, through dynamic weighting of historical medal data (with higher weights given to more recent data), effectively captures the trend characteristics of countries with stable medal counts. It enables dynamic prediction of the total medal count and gold medals for target countries when dealing with large-scale or stable trend time series data, providing an efficient trend analysis tool for similar data scenarios. Next, the Frequency Probability Model estimates future probabilities based on historical winning frequencies, particularly suitable for countries with fewer or more fluctuating medals. By quantifying the impact of external factors, such as "great coaches," on improving the probability of winning (e.g., a 17.5% increase), it offers a probabilistic modeling approach for analyzing the influencing factors of repeat events, confirming the positive effect of high-level coaching resources on enhancing competitiveness. Finally, the Logistic Regression Model addresses the "zero-sample" scenario for countries that have not won medals, integrating multi-dimensional variables such as the number of participants and historical participation frequency. Using the Adam optimization algorithm, it constructs a framework for predicting the probability of winning medals for the first time. This provides a multi-factor potential assessment method for countries lacking medal records, which can assist in Olympic resource allocation and strategic planning.

References

[1] Li Xiang. Forecasting the number of Olympic medals based on time series analysis[J]. Computer and Digital Engineering, 2018, 46(03): 533-536+545.

[2] Wang Tian, Yan Jun, Gu Xiaoyang, et al. Research on delay prediction of congested networks based on improved exponential smoothing model[J]. Industrial Control Computer, 2023, 36(09): 101-103+114.

[3] Ran Zhihong, Dong Guohua, Yang Zifu, et al. Research on the reliability calculation method of skewed probability model based on empirical frequency distribution[J]. Sichuan Research on Building Science,2024,50(01):52-59. DOI:10.19794/j.cnki.1008-1933.2024.0007.

[4] Chen Siyu. Research on the prediction of grassland fire occurrence probability based on geographically weighted logistic regression model[D]. Inner Mongolia University of Technology, 2022. DOI:10.27225/d.cnki.gnmgu.2022.000057.

[5] Zhao Xiang, Ji Zuqian, Wang Dangwei, et al. Predicting the outflow probability of the tailrace channel of the Yellow River estuary based on logistic regression[J]. Sedimentation Research,2021,46(03): 36-42. DOI:10.16239/j.cnki.0468-155x.2021.03.006.

[6] Du Mei. Research on disease prediction method based on deep learning optimization algorithm[D]. Harbin Institute of Technology, 2022. DOI:10.27063/d.cnki.ghlgu.2022.000406.