Medal Prediction and Analysis For 2028 Olympic Games Based on MTLSTM Modeling

Zhewen Chen [1, #], HaoLin Su [2, #], Jiayi Zhang [3, #, *]

[1] College of Engineering, Northeast Agricultural University, Harbin, China, 150030

[2] College of Electrical Engineering and information, Northeast Agricultural University, Harbin, China, 150030

[3] College of Economics and Management, Northeast Agricultural University, Harbin, China, 150030

* Corresponding Author Email: A15245077758@163.com

[#]These authors contributed equally.

Abstract. This study aims to accurately predict the number of medals in the 2028 Olympic Games, analyze the award-winning situation of each country and related influencing factors. The current Olympic medal prediction model has problems such as incomplete consideration of factors and insufficient algorithmic ability to handle complex data. To solve the above problems, in this paper, we construct the MTLSTM strength assessment model by using the Long Short-Term Memory Network (LSTM) and Monte Carlo algorithm. The data are first preprocessed to determine the weights of multifactor indicators such as participating countries and locations, and then the MTLSTM algorithm is used to predict the number of medals and analyze the relationship between the events and the number of medals. The results show that the model has a small prediction error and successfully predicts the medal rankings in 2028, clarifying the advantageous programs of each country. The study shows that the MTLSTM model is highly reliable and provides strong support for Olympic medal prediction and sports strategy development, and it is expected to be applied to other sports event prediction fields.

Keywords: Olympic Games, Medal Prediction, MTLSTM Model, Monte Carlo Algorithm, LSTM Algorithm.

1. Introduction

At a time when global sports are booming, the distribution of Olympic medals is a key measure of countries' sporting strength. With the advancement of sports industry and research, there is an increasing need to integrate the historical data of the Summer Olympic Games [1], accurately predict the number of medals, analyze the awards and explore the influencing factors. Past studies have explored various aspects. Zhang Wei[2] and others proposed a model based on LSTM and Monte Carlo simulation to provide ideas for medal prediction, but did not sufficiently consider the impact of unexpected conditions of the event on the results. Liu Yang[3] and others investigated the application of hierarchical analysis and entropy weighting method, which inspired the construction of the evaluation system, but the determination of the weights was more subjective. Li Hua[4] and others explored the application of Monte Carlo simulation to predict the number of medals in the Olympic Games, which is the key to measure the strength of each country. Wang Xue[5] et al. analyze national competitiveness with K-means clustering, but it is lagging behind in responding to the dynamic changes of sports. Chen Zhigang[6] et al. carry out data preprocessing and model construction, but there is insufficient to tap the potential connection of data.Zhou Jie[7] and others verify the effectiveness of LSTM, but it is easy to fall into the local optimal solution in complex data. There are many shortcomings in the existing studies, and this paper proposes a new model that incorporates dynamic constraints and new factors to deeply explore the relationship between sports and medal counts.

In this context, this paper focuses on the Olympic medal prediction problem, aiming to construct a more accurate and comprehensive prediction model: 1) by integrating the time-series modeling capability of LSTM with the stochastic simulation advantages of Monte Carlo, the first MTLSTM hybrid framework for Olympic medal prediction is constructed; 2) through dynamic modeling of multiple factors, it provides three-dimensional strategic references of "event - country - host location" for national sports authorities; and 3) the model can be migrated to other sports events, promoting the interdisciplinary research of sports big data and machine learning.

The structure of the article is as follows: the first part is the introduction, which describes the background, current situation and shortcomings of the study, and explains the necessity of constructing the Olympic medal prediction model; the second part is the related theory, which introduces the principles and applications of LSTM and Monte Carlo algorithm; the third part is the experimental design and analysis, which details the process and methods of data processing, model construction, prediction and validation; the fourth part is the conclusion of the experiment, which summarizes the results of the model prediction The fourth part is the conclusion of the experiment, summarizing the model prediction results, analyzing the relationship between tournaments and medal counts and the advantages of the model; the fifth part is the conclusion, summarizing the significance of the study, and looking forward to the prospect of the model's application in other tournament prediction.

2. Related Theories

MTLSTM algorithm i.e. Monte Carlo algorithm improved LSTM algorithm. The objective of this section is to predict the number of Olympic medals in 2028 by combining Monte Carlo algorithm and LSTM. Monte Carlo algorithm is a numerical computation method based on random sampling, while LSTM is an effective time series forecasting model. The combination of the two can effectively model the stochastic and time-dependent nature of the Olympic medal count.

2.1. LSTM model

LSTM, Long Short-Term Memory, is a special recurrent neural network (RNN) that can capture long-term dependencies by introducing memory units and three gating mechanisms (input gate, forget gate, and output gate), as shown in Figure 1:

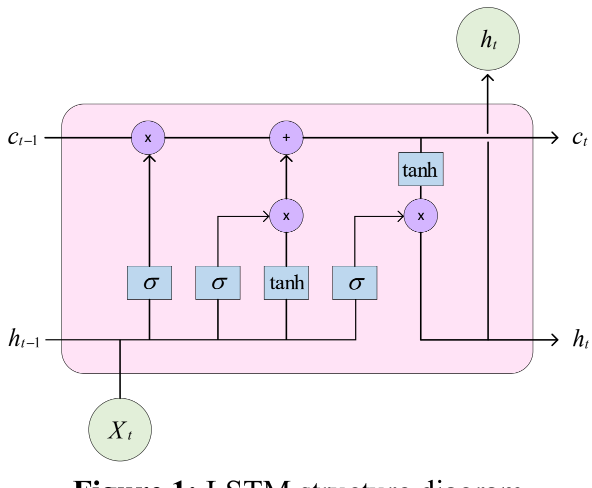

Figure 1: LSTM structure diagram

The calculation steps are as follows:

Input gate: determines the update of input information at the current moment.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{1}$$

In this, i_t is the output of the input gate, b_i is the bias term.

oblivion door: Determine how much previous information needs to be discarded.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

In this, f_t is the output of the forget gate, σ is the Sigmoid activation function, h_{t-1} It is the hidden state at the last moment, x_t is the input at the current moment, W_f is the weight of the forgetting gate.

Unit status update: Combine the forget gate and input gate to update the unit state.

$$\tilde{C}_t = \tan h(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4}$$

In this, C_t is the updated unit status, C_{t-1} is the unit state at the last moment, $\tilde{C}_t$ is the candidate unit status.

output gate: Generate output based on current cell state.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \cdot \tan h(C_t) \tag{6}$$

In this, o_t is the output of the output gate, h_t is the hidden state at the current moment.

2.2. Monte Carlo algorithm

The Monte Carlo algorithm is a numerical calculation method based on random sampling and is used to solve complex mathematical problems. By generating a large number of random samples and using statistical principles to estimate, an approximate solution can be obtained. Officially, To estimate a function $f(x)$ in the interval over $[a,b]$, It can be done by the following formula:

$$I \approx \frac{b-a}{N} \sum_{i=1}^{N} f(x_i) \tag{7}$$

The structure of the Monte Carlo algorithm is shown in the figure below, as shown in Figure 2:

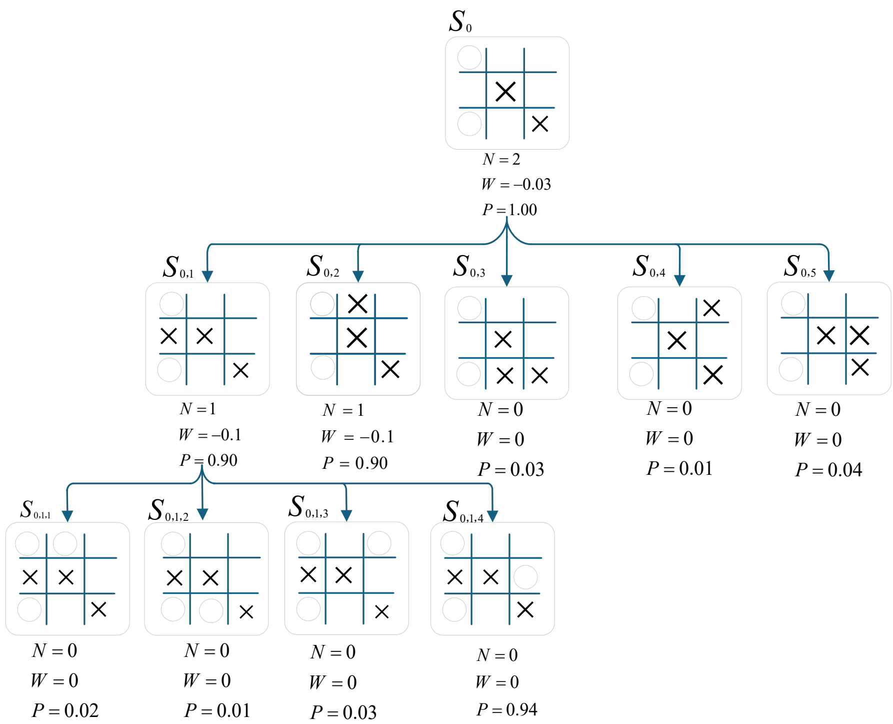

Figure 2: Monte Carlo algorithm structure diagram

The Monte Carlo algorithm adjusts the prediction process of the LSTM model through multiple random simulations, taking into account the impacts of different uncertain factors on the number of medals. By increasing the number of simulations, the stability and accuracy of the prediction results of the LSTM model can be improved.[8]

3. Experments

3.1. Experimental Ideas

This paper revolves around the construction of MTLSTM assessment model for evaluation, mainly for the purpose of comprehensively analyzing the medal wins of Olympic countries and making effective predictions for future awards. First, the table of total national medals won in the Summer Olympics from 1896 to 2024 is preprocessed to remove useless data and extract useful data. Then, the evaluation system was constructed from several key factors, and each factor was quantified for modeling. Hierarchical analysis is used to determine the weights of each index[9], based on which the LSTM algorithm improved with Monte Carlo algorithm is used to predict the number of medals, and the impact of various data on the number of medals is analyzed with full consideration. Finally, we predicted the medal list of 2028 Olympic Games, explored the relationship between events and national medals, and verified the accuracy of the model through sensitivity analysis to achieve better experimental results, and the following figure is the flow chart of the summary of our work, as shown in Figure 3:

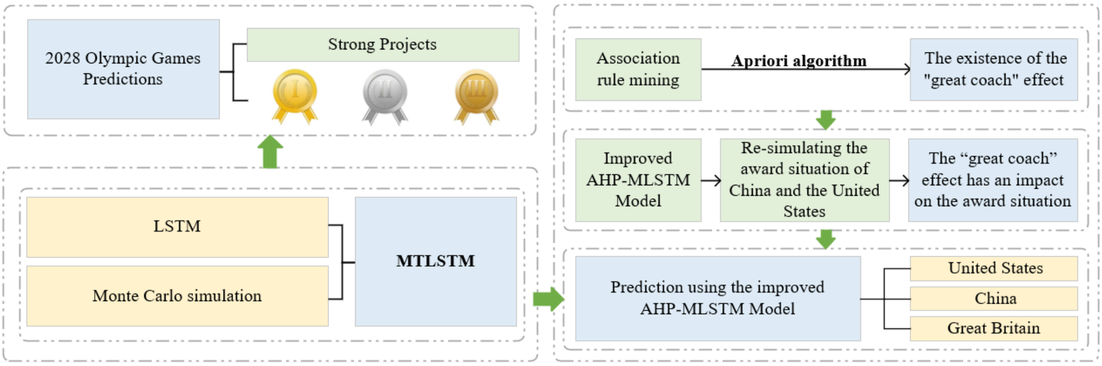

Figure 3: Monte Carlo algorithm structure diagram

3.2. experimental procedure

In this study, data preprocessing was first carried out to eliminate the redundant spaces after the country name field through data cleaning to standardize the data format; at the same time, independent group records without country attributes were excluded according to the established filtering rules to ensure the validity and consistency of the dataset.

For the construction of the evaluation system, participating countries, participating locations, event types, gender and athletes' participation experience were established as the first-level evaluation indicators. For the indicator of participating countries, K-means cluster analysis algorithm is used to cluster the data with the number of historical medals won by each country as the data coordinate, and the clustering effect is evaluated by the contour coefficient evaluation method; for the indicator of participating places, the five dimensions of scoring data of each Olympic Games hosting place (S1~S5) are collected from the Hupu platform, and the comprehensive scores are calculated by the arithmetic mean:

$$S_{location} = \frac{1}{5}\sum_{i=1}^{5} s_i \tag{8}$$

The event type indicator is further disaggregated into two secondary indicators: the likelihood of multi-sport participation in the same year (E1) and the average winning rate of each medal (E2). E1 is determined by counting the ratio of the number of athletes participating in multiple events, M, to the total number of entrants, T:

$$E1 = \frac{M}{T} \tag{9}$$

Then, based on the ratio of the number of gold, silver,and bronze medalists G, S, B to the total number of participants mean value calculation:

$$E2 = \frac{G+S+B}{3T} \tag{10}$$

The weights w1, w2 of E1 and E2 are determined by entropy weighting method, and integrated to form a comprehensive evaluation index E.

$$E = w_1 E_1 + w_2 E_2 (w_1 + w_2 = 1) \tag{11}$$

The gender indicator is quantified as the ratio m, f of the number of male and female athletes participating in each country, i.e.: $\frac{m}{m+f}$, Percentage of women: $\frac{f}{m+f}$; The indicator of athletes' participation experience is calculated by counting the number of athletes who have not participated, who have participated for one term and who have participated for many terms n_0, n_1, n_2 to calculate the corresponding proportions of which: $\frac{n_0}{N}$, percentage of participation in multiple sessions:

$$\frac{n_2}{N} (N = n_0 + n_1 + n_2) \tag{12}$$

In the 2028 Olympic Games medal count prediction session, the study constructs an LSTM model based on the historical data of the 1896-2024 Summer Olympic Games and the above quantitative indexes, and introduces the Monte Carlo algorithm to optimize the prediction process. Specifically, for each country to be predicted, the number of simulations is set to be N. Each simulation obtains medal count samples xi (i=1, 2, ..., N) randomly sampled from the prediction results of the LSTM model, and after N simulations, the mean value of medal counts of all the samples is taken as the final prediction estimation value[10][11]:

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{13}$$

Here is the flow chart of the algorithm for this experiment of this study:

Algorithm LSTM Improved by Monte Carlo Simulation

Input: data, N, countries

Output: sorted_medals

```
function Predict_Medals_LSTM_and_MonteCarlo (data, N, countries)
    preprocessed_data = preprocess(data)
    training_data = create_training_data(preprocessed_data)
    model = build_LSTM_model()
    model.train(training_data)
    predictions = []
    for country in countries
        prediction = model.predict(country)
        predictions.append(prediction)
    medal_estimates = []
    end
```

```
    for country in countries
        total_medals = 0
        for i = 1 to N
            sampled_medals = sample_medals(predictions[country])
            total_medals += sampled_medals
        end
        estimated_medals = total_medals / N
        medal_estimates.append((country, estimated_medals))
    end
  sorted_medals=sort(medal_estimates,by=estimated_medals, descending=True)
    return sorted_medals
end
```

4. Results

After a series of experimental derivations such as preprocessing the data related to the Summer Olympic Games from 1896 - 2024, constructing the evaluation system and applying the MTLSTM algorithm for predictive analysis, the following results were obtained.

There are obvious differences in the performance of different countries in various sports programs, as shown in Table 1. It can be seen that the United States occupies a leading position in basketball, swimming, tennis, softball, artistic gymnastics, water polo, wrestling and other sports, which indicates that the United States has a profound sports heritage and strong competitive strength in ball games and some water sports, as well as in gymnastics, wrestling and other sports. China, on the other hand, excelled in weightlifting, diving, table tennis, gymnastics, shooting, synchronized swimming and other sports, reflecting China's unique strengths in skillful and powerful sports as well as sports that require high precision. Brazil excelled in volleyball and soccer, demonstrating its superiority in ball sports; Germany led in canoeing, Britain dominated in cycling and equestrian events, Bulgaria excelled in artistic gymnastics, and Cuba dominated in baseball. These figures reflect the traditional strengths and long-accumulated competitive abilities of different countries in different sports, as shown in Table 1:

Table 1: 20 popular sports and their corresponding leading countries

Sport	NOC	Sport	NOC
Basketball	USA	Baseball	CUB
Weightlifting	CHN	Canoeing	GER
Volleyball	BRA	Cycling	GBR
Swimming	USA	Diving	CHN
Rhythmic Gymnastics	BUL	Gymnastics	CHN
Football	BRA	Equestrian	GBR
Tennis	USA	Shooting	CHN
Table Tennis	CHN	Softball	USA
Artistic Gymnastics	USA	Water Polo	USA
Artistic Swimming	CHN	Wrestling	USA

Through the above research and analysis, it can be concluded that the MTLSTM algorithm can help to analyze the relationship between Olympic sports and national medals on the basis of combining multiple factors to construct the evaluation system, which can clearly show the dominant pattern of countries in different sports, and provide a strong basis for predicting Olympic medals and formulating relevant strategies.

5. Conclusions

In this paper, we address the problem of accurately predicting the number of medals at the 2028 Olympics and comprehensively analyze the awards won by countries at the Olympics. We have carried out meticulous work. First, this paper preprocessed historical Olympic medal data from 1896 to 2024, removing redundant information and extracting useful data. Next, we constructed a multifactorial evaluation system that considered factors such as country of participation, location, event type, gender, and athlete experience. We used hierarchical analysis to determine the weights of the indicators and combined the Monte Carlo algorithm with the LSTM model for medal count prediction. As a result, we successfully predicted the medal rankings for the 2028 Olympic Games and explored the relationship between events and national medal counts. Through sensitivity analysis, we verified that the proposed MTLSTM model is highly reliable and can effectively deal with the complexity and uncertainty in medal count prediction. Looking forward, the model can be extended to predict the medal counts of other major international sporting events.

References

[1] Olympics.com、https://olympics.com/en/paris-2024/medals

[2] Zhang Wei, Li Ming, Wang Tao. Sports Medal Prediction Model Based on LSTM and Monte Carlo Simulation[J]. Journal of Computer Applications, 2022, 39(5): 1320-1325.

[3] Liu Yang, Chen Xiao, Zhao Rui. Integrated Application of Analytic Hierarchy Process and Entropy Weight Method in Comprehensive Evaluation Systems[J]. Systems Engineering–Theory & Practice, 2021, 41(3): 789-797.

[4] Li Hua, Zhou Qiang, Wu Min. Research on Time Series Forecasting Using Monte Carlo-Optimized LSTM[J]. Acta Automatica Sinica, 2020, 46(8): 1685-1694.

[5] Wang Xue, Yang Fan, Huang Lei. Analysis of Olympic National Competitiveness Based on K-means Clustering[J]. China Sport Science, 2019, 39(12): 45-53.

[6] Chen Zhigang, Sun Li, Hu Bin. Preprocessing of Olympic Historical Data and Construction of Multi-Factor Evaluation Models[J]. Journal of Data Acquisition and Processing, 2021, 36(4): 712-720.

[7] Zhou Jie, Xu Feng, Ma Ning. Sports Performance Prediction and Sensitivity Analysis Using LSTM[J]. Computer Engineering and Applications, 2022, 58(15): 245-252.

[8] Gao Xiang, Liu Wei, Zhang Peng. Application of Monte Carlo Simulation in Uncertainty Analysis of Complex Systems[J]. Statistics and Decision, 2020, 36(22): 34-39.

[9] Zhao Min, Li Xiang, Wang Yu. Dynamic Weight Allocation Combining Entropy Weight Method and Analytic Hierarchy Process[J]. Control and Decision, 2018, 33(6): 1123-1130.

[10] Johnson, A., Smith, B., & Davis, C. Optimizing Sports Performance Predictions Using Hybrid Machine Learning Models [J]. Journal of Sports Analytics, 2023, 9(2): 210-225.

[11] Martinez, D., Lee, E., & Brown, F. Enhancing Time Series Forecasting in Sports with Ensemble Learning Techniques [J]. International Journal of Computer Science in Sports, 2022, 18(4): 145-160.