Paper Type: Original Article

# A Study on Olympic Medal Table Prediction Based on LSTM and DBILSTM

**Xuanzhi Zhu[1*#], Shuo Wang[1#], Xiaoyang Li[1], Qirui Wu[1], Zisheng Hong[1], Mingyan Shao[1]**
[1]*School Of Mechanical And Electrical Engineering, Anhui Jianzhu University, Hefei, China*

***Corresponding Author**: Zhu Xuanzhi, **E-mail**: 17714048197@163.com
**#Co-first author:** Zhu Xuanzhi, Wang Shuo

## abstract

This paper focuses on the prediction of Olympic medals list, using Long Short-Term Memory (LSTM) and two-order bi-directional long short-term memory neural network (DBILSTM) models. The study collates and cleans the data related to several Olympic Games, selects 10 variable indicators such as the gender of athletes, consecutive awards, the proportion of national participation and awards, etc., and applies the LSTM model to predict the number of medals of some countries in 2028, and passes the tests of MAE, RMSE, and MAPE indexes, and the results show that the model accuracy is good. Meanwhile, the innovative DBILSTM model is used to effectively integrate the information before and after the sequence to calculate the probability of multiple countries winning the first medal in 2028. The results of this research provide valuable references for countries to formulate sports development strategies and commercial layout of events, and also contribute new methods and ideas to the field of sports event forecasting.

**Keywords:** Olympic Medal Table; LSTM; DBILSTM; Prediction Model; Time Series Analysis.

## 1 | Introduction

In the field of sports event research, the Olympic Games medal table prediction has always been a hot topic that attracts a lot of attention. With the global development of sports, the influence of the Olympic Games is increasing day by day, and the belonging of the medal table has touched the hearts of countless sports fans. Accurately predicting the distribution of Olympic medals can not only greatly satisfy the curiosity and expectation of sports fans for the results of the event, but also has far-reaching significance in many aspects. For national sports administrations, accurate medal predictions can provide a key reference for the formulation of scientific and reasonable sports development strategies, helping them to rationally plan their resources, identify key development projects, and thus enhance the comprehensive competitiveness of their countries in the international sports arena.

In recent years, deep learning technology has made rapid progress, showing strong application potential in many fields. In the research direction of Olympic medal list prediction, many researchers have tried to use advanced deep learning algorithms, expecting to break through the limitations of traditional prediction methods and improve the accuracy of medal list prediction. In this context, this paper focuses on the use of Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) models to explore more accurate Olympic medal table prediction methods. Through the innovative application and optimisation of these two models, we expect to contribute new ideas and results to the research in this field, and promote the Olympic medal table prediction research to a new level.

## 2 | Literature Review

Sports forecasting improves foresight in all endeavours. Forecasting is the basis for decision-making, and without scientific forecasting, research and grasp of the future situation, it is impossible to take correct decisions and actions. Scientific sports forecasting is needed to ensure the healthy and steady development of sports in China, so that we can understand the situation, correctly predict the development trend of sports programmes and the development trend of various countries, and take a series of effective countermeasures.

Many scholars have actively explored and achieved rich results in the research of Olympic medal forecasting model, Bermard and Busse (2000) regarded the population size and economic resources measured by GDP as factors of production, and researched the distribution of Olympic medals with the help of Cobb-Douglas production function. This study provides an important perspective for understanding the distribution of medals from the perspective of macroeconomic and demographic factors, and makes people realise that a country's economic strength and population base will, to a certain extent, affect the number of medals it obtains at the Olympic Games [6].

Johnson and Ali (2000), on the other hand, propose the use of empirical modelling for Olympic-related forecasting. They constructed a participation model for the total number of athletes and the number of female athletes from all countries in the world participating in the Olympic Games, taking into account a variety of influencing factors such as GDP per capita, population size, home advantage (represented by the dummy variable for whether the country is connected to the host country or not), the political system, and the historical background (the dummy variable for whether the country is a colony or not). The model fully explores the potential links between social, political, economic and other factors and the number of participants in the Olympic Games, providing new ideas and methods for an in-depth study of participation in the Olympic Games [7].

Since then, other scholars have expanded their research from different perspectives. For example, Smith and Jones introduced sports input factors, including sports infrastructure construction investment, athlete training capital investment, etc., and constructed a more complex medal prediction model, which further refined the factors affecting medal acquisition [8][9]. Schlembach C, on the other hand, used machine learning algorithms to deeply mine a large amount of historical Olympic data, combined with the dynamic trend of sports development in various countries, to achieve the dynamic prediction of the Olympic medal list, so that the prediction results are more timely and accurate. These research results have continuously enriched and improved the theoretical and practical system in the field of Olympic medal prediction, providing valuable reference and reference for subsequent research.

## 3 | LSTM-based Prediction of the Number of Medals

In the field of sports, the Olympic Games, the world's premier sporting event, has a much-anticipated competition for the medal table. Predicting Olympic medals is not just an activity to satisfy the curiosity of the public, but it is of vital significance at many levels. From the perspective of sports strategic planning, national sports administrations can rely on the results of the medal table prediction to rationally allocate sports resources, clearly identify key development projects, formulate long-term sports development strategies, and enhance the competitiveness of the country in international sports events. For the sports business sector, accurate medal table prediction helps event sponsors, advertisers and other relevant parties to plan their market strategies in advance, make more effective commercial layouts, and improve the return rate of resource investment. Meanwhile, in the scope of academic research, predicting the Olympic medal table provides a valuable practice scenario for studying the development law of sports and exploring the time series data prediction method, and promotes the development of theories and technologies of related disciplines.

Predicting the Olympic medal table is a multidimensional and complex task involving many factors, such as the status of athletes, the development trend of sports in various countries, and the environment in which the event is held, which change dynamically over time to form a complex time series of data. Traditional prediction methods are difficult to effectively capture the long-term dependencies, resulting in limited accuracy of prediction results. The Long Short-Term Memory (LSTM) model, as a powerful deep learning model, can effectively solve the gradient disappearance and gradient explosion problems faced by traditional models when dealing with long time series data, accurately capture the long-term dependency features in the data, and provide a reliable technical means for the Olympic medal list prediction, so this model will be used in this study.

We consider four editions as the career of an elite athlete, so we can use the data from the first three editions to predict the athlete's winnings in the fourth edition. We set 10 variable indicators using the gender of the athlete, the athlete's three consecutive awards, all competitions in the country and the ratio of male to female winners, and whether or not they were in the host city as the 0/1 variables, and analysed each country individually.

Long Short-Term Memory (LSTM for short) is a special Recurrent Neural Network (RNN) architecture designed to solve the gradient vanishing and gradient exploding problems faced by traditional RNNs when dealing with long time-series data, and its core principle is partly an input gate, a forgetting gate, and an output gate, and the cellular state, the conceptual diagram is shown in Fig. 1.

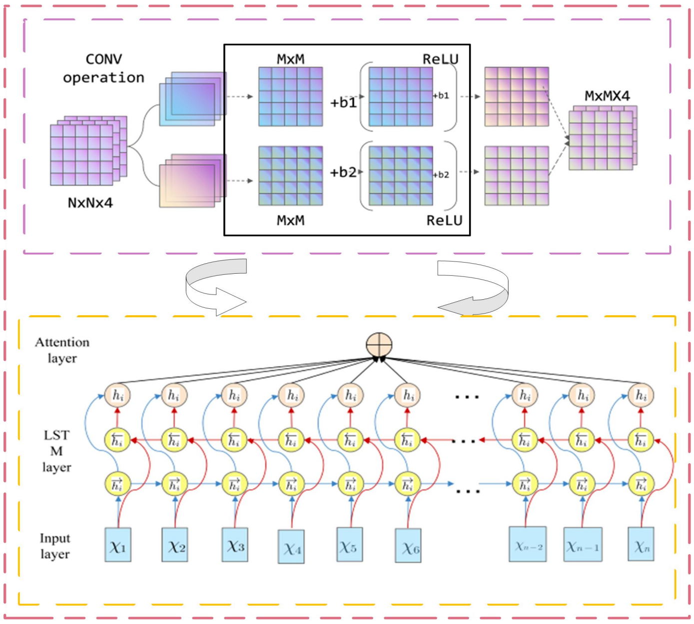


**Figure 1 LSTM conceptual diagram**

In the field of time series forecasting, "lag characteristics" is a very critical technical tool. Its core principle is to take the past observation as the input basis for current prediction, so as to improve the accuracy and reliability of prediction with the value of historical data. In this study, "lagged features" also plays an important role in the prediction of Olympic medals list. After several training sessions, the iterative graph of LSTM residuals is obtained, as shown in Fig. 2, which shows that the loss value decreases rapidly at the beginning, and at about the 100th epoch, the loss has a big spike, and after the spike, the loss value gradually decreases and tends to be stable, which shows a more ideal state.

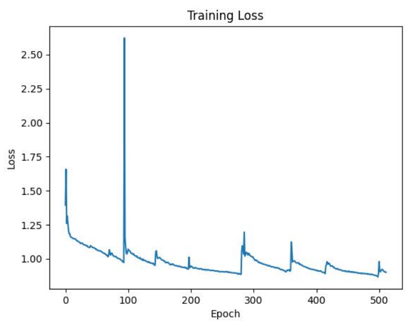


**Figure 2 Residual Iteration Chart**

The resulting 2028 medal table - including the number of gold medals predicted, the number of silver medals, the number of bronze medals, as well as the total number of these three predicted medals, and the total number of predicted medals. The results are shown in Table 1

**Table 1 Medal table predictions**

| Country | Gold | Silver | Bronze | Summed_Total | Predicted_Total |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| United States | 59 | 64 | 69 | 192 | 188 |
| China | 50 | 48 | 69 | 102 | 102 |
| Germany | 48 | 22 | 34 | 104 | 105 |
| Great Britain | 33 | 31 | 29 | 93 | 91 |
| Italy | 28 | 54 | 26 | 108 | 103 |
| Australia | 28 | 28 | 26 | 82 | 70 |

The data was then visualised for all 2028 medal totals and the results are shown in Figure 3.

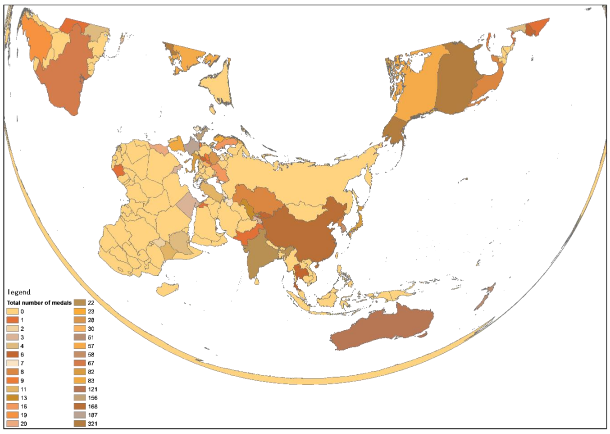

**Figure 3  Total number of medals visualised**

Meanwhile, the accuracy of the model is examined by the RMSE indicator, as shown in Fig. 4, which shows that the RMSE is kept around 1, indicating that the model accuracy is good. Comparing the actual predictions of the training set and the test set, the sample points are distributed on the diagonal, which further indicates that the model fits well.

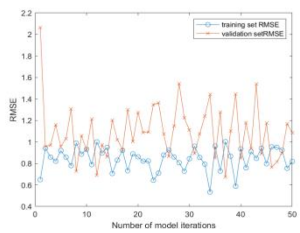


**Figure 4  Model RMSE indicator test**

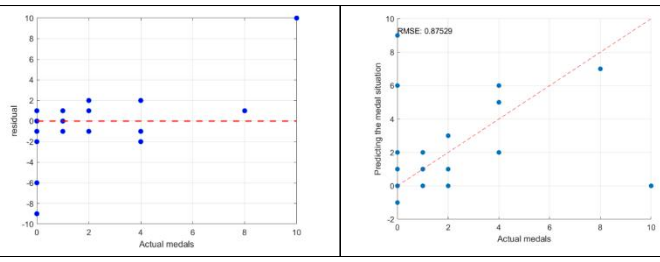


**Figure 5  Actual predictions for the training and test sets**

Next, we compare the actual predictions of the training set and the test set. As shown in Fig. 5, the sample points of the training set and the test set are distributed on the diagonal, which indicates that the model fits well.

## 4|DBILSTM Predicts the Probability of a Country Winning the Award for the First Time

In the study of Olympic sports events, predicting the probability of a country winning a medal for the first time can help to understand the development trend and potential of sports in various countries, and can also provide valuable reference for the rational allocation of sports resources and the formulation of event strategies. For countries that have yet to win a medal at the Olympic Games, knowing the probability of winning a medal for the first time can help them plan their sports development path in a more scientific way, focus their resources on the development of advantageous sports, and enhance their competitiveness in the international sports arena. In order to accurately predict the probability of a country's first medal, this study adopts an innovative model, the two-order bidirectional short-term and long-term memory neural network (DBILSTM).

Basic Bidirectional Long Short-Term Memory (BiLSTM for short) is a special type of Recurrent Neural Network (RNN) specifically designed to process sequence data. It improves on LSTM by taking into account both past and future information, enabling the model to better capture contextual relationships in sequence data. And we innovatively improve on BILSTM again by adopting a two-order bi-directional long and short-term memory neural network, which is named DBILSTM.The BILSTM model is able to integrate the information from both directions, which significantly improves the accuracy of the sequence modelling task. DBILSTM, on the other hand, processes the data from the beginning to the end of the sequence in the first order, which can obtain the historical information before the current time step; the second reverse run processes from the end to the beginning, which can obtain the future information after the current time step. By fusing the results of the two runs, the model will be able to simultaneously grasp the sequence's before and after information, thus providing a more comprehensive understanding of the sequence data, while improving the model's accuracy, generalisation capability and ability to handle complex structures. Its conceptual diagram is shown in Figure 6

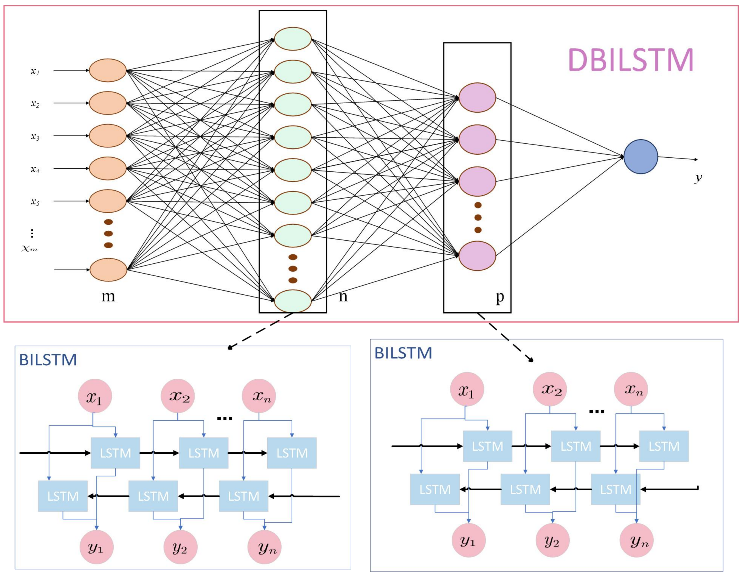


**Figure 6 DBILSTM conceptual diagram**

After several training sessions, an iterative plot of the DBILSTM residuals is obtained. It is shown in Fig. 7. The above figure is the first order BILSTM model, the loss value in the figure has a large fluctuation at the beginning of the training, and there is a large spike in the loss value at the first 100 epochs. The loss then gradually decreases and levels off, indicating that the model is starting to converge and stabilise. The second order BILSTM also shows a large fluctuation at the beginning and gradually decreases in subsequent training. Compared to the first order, the loss of E-BiLSTM decreases faster and smoother during the preceding training process, indicating that the error correction step helps to improve the training process quickly. This indicates that DBILSTM training is more effective.

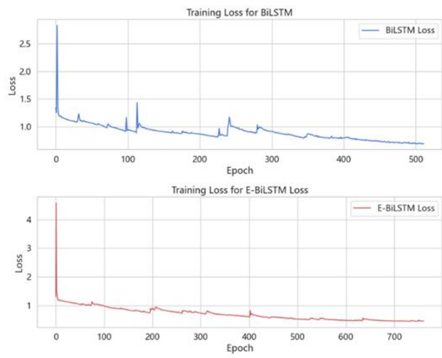


**Figure 7 DBILSTM iteration diagram**

In the first order BILSTM the parameters were set to 128, 32, 128 neurons, after 4, 256 iterations, with a learning rate of 0.001; in the second order BILSTM the parameters were set to 128 , 32, 128 neurons, after 4,762 iterations, with a learning rate of 0.001.After obtaining the medals data, the next step was to compute the countries' first medals likelihood. Prediction intervals are used to estimate the probability, and this method has a solid statistical basis. The prediction intervals reflect the range of uncertainty in the model predictions and provide a quantitative reference for assessing the probability of winning a medal. When the lower bound of the prediction interval is greater than 0, the model predicts that the country is very likely to win a medal in 2028, and it is almost certain that the country will be able to win at least one medal in that year. Even if the prediction interval contains zero, the possibility of the country winning a medal cannot be ruled out, e.g. a prediction interval of [-1,2] would imply that there is a probability of winning the first medal.

This study focuses on calculating a 95% prediction interval, which means that a 95% scenario gives a reliable assessment of the probability of a country winning its first medal. The results of the rigorous process of calculating the probability of each country winning its first medal are summarised and finally presented in Table 2, which gives a visual representation of the probability of each country winning its first medal in 2028.

**Table 2 Probability of winning for the first time**

| Country | Gold | Probability | Silver | Probability | Bronze | Probability | Prdicted Total | Probability |
|---|---|---|---|---|---|---|---|---|
| Angola | 1 | 61.75 per cent | 0 | 21.96 per cent | 0 | 0.09 per cent | 0 | 59.04 per cent |
| Mali | 0 | 18.72 per cent | 0 | 28.07 per cent | 1 | 53.31 per cent | 1 | 63.52 per cent |
| Guinea | 0 | 11.58 per cent | 0 | 28.54 per cent | 1 | 61.72 per cent | 1 | 64.32 per cent |
| South Sudan | 0 | 13.08 per cent | 0 | 11.73 per cent | 1 | 60.40 per cent | 1 | 65.33 per cent |
| Honduras | 1 | 64.13 per cent | 1 | 50.70 per cent | 1 | 67.49 per cent | 1 | 71.73 per cent |
| Iraq | 0 | 40.66 per cent | 0 | 25.00 per cent | 2 | 73.21 per cent | 1 | 65.59 per cent |

After obtaining the first time win probabilities for each country, at this point, we assessed the performance of the model by looking at how well the model was trained. It can be clearly seen from Fig. 8 that the model's fit to the training set is generally good. It is able to effectively capture and fit the subtle fluctuations present in the data, which indicates that the model has a certain degree of sensitivity and is able to uncover the more hidden features and patterns in the data. However, when facing some rapidly changing data, the model's fitting effect is somewhat lacking.

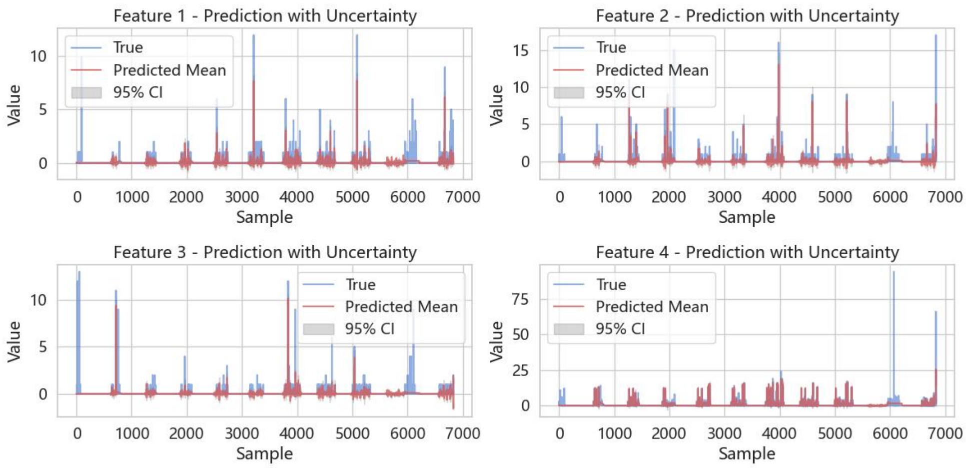


**Figure 8 Map of model training level**

In order to gain further insight into how well the model fits the data from different countries and different Olympic Games, we then used the trained model to fully train the data from each country for each Olympic Games. In this process, the quality of the model's fit was assessed more accurately by calculating two key metrics, the mean and variance, resulting from the model's fit to the training data.

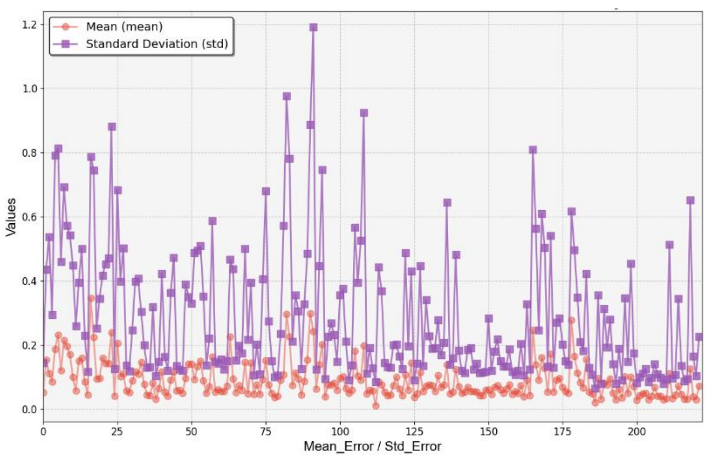


**Figure 9 The error generated by the model fit for each Olympic Games for each country**

It can be analysed from Fig. 9 that the variance of the model fit to the data is basically maintained within 1.0, with only a few exceeding 1.0, but the mean variance of the data fit is basically maintained within 0.4, which indicates that the model fits the data better, and that the variance exceeding 1.0 can be within the acceptable range.

## 4| Conclusion

This paper focuses on the Olympic medal table prediction, using LSTM and DBILSTM models to carry out the research. In the LSTM-based medal count prediction, by organising and cleaning the data and setting 10 variable indicators to effectively capture the time series data features, the model accuracy is good and accurately predicts the number of medals of some countries in 2028. In DBILSTM prediction of countries' first medal probability, the innovative two-order structure improves the model performance, and calculates the probability of multiple countries to win their first medals in 2028, which provides a reference for the development of sports in various countries.

However, although this study has tried its best to mine the key indicators in data processing, there may still be potential factors affecting the medal table that have not been included, resulting in the model not reflecting the complex reality comprehensively enough. As for the models, although LSTM and DBILSTM showed some advantages, the fitting effect was poor when dealing with rapidly changing data, and there is still room for improvement in the generalisation ability. Future research can further expand the data dimensions and incorporate more diversified influencing factors, such as athletes' injuries and illnesses, and the development trend of emerging sports. At the same time, we can explore the optimisation of the existing model or introduce other advanced models, combined with techniques such as transfer learning and reinforcement learning, to enhance the model's adaptability and prediction accuracy to complex data.


## Acknowledgement

Zhu Xuanzhi and Wang Shuo are co-first authors.

## Funding

Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant number XXX" and "The APC was funded by XXX". Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant XXX" and "The APC was funded by XXX".


## References


[1] Nagpal P, Gupta K, Verma Y, et al. Paris Olympic (2024) Medal Tally Prediction[C]//International Conference on Data Management, Analytics & Singapore: Springer Nature Singapore, 2023: 249-267.

[2] Zhao S, Cao J, Steve J. Research on Olympic medal prediction based on GA-BP and logistic regression model[J]. F1000Research, 2025, 14: 245.

[3] He Z, Wang Z. Prediction of olympic medal count for USA based on robust time series model and computer implementation[C]//Third International Conference on Electronic Information Engineering and Data Processing (EIEDP 2024). SPIE, 2024, 13184: 1361-1369.

[4] Schlembach C, Schmidt S L, Schreyer D, et al. Forecasting the Olympic medal distribution-a socioeconomic machine learning model[J]. Technological Forecasting and Social Change, 2022, 175: 121314.

[5] Huimin S H I, Dongying Z, Yonghui Z. Can Olympic Medals Be Predicted? Based on the Interpretable Machine Learning Perspective[J]. Journal of Shanghai University of Sport, 2024, 48(4): 26-36.

[6] Bernard A B, Busse M R. Who wins the Olympic Games: economic resources and medal totals[J]. Review of economics and statistics, 2004, 86(1): 413-417.
[7] Johnson D K N, Ali A. Coming to play or coming to win: Participation and success at the Olympic Games[J]. Wellesley College Dept. of Economics Working Paper, 2000 (2000-10).
[8] Jones G J, Edwards M B, Bocarro J N, et al. An integrative review of sport-based youth development literature[J]. Sport in Society, 2017, 20(1): 161-179.
[9] Smith D J. A framework for understanding the training process leading to elite performance[J]. Sports medicine, 2003, 33: 1103-1126.
[10] Schlembach C, Schmidt S L, Schreyer D, et al. Forecasting the Olympic medal distribution-a socioeconomic machine learning model[J]. Technological Forecasting and Social Change, 2022, 175: 121314.