WHIOCE PUBLISHING PTE. LTD.
PROVIDING
FIRST-CLASS SCIENTIFIC INFORMATION
FOR TOP SCHOLARS

2026 Volume 4, Issue 3
Contemporary Education Frontiers
ISSN:3029-1879 (Online)
ISSN: 3029-1860 (Print)

An In-depth Analysis of Olympic Medal Number Prediction and Influencing Factors Based on Multi-model Fusion

Jinger Gao*
Neusoft Institute Guangdong, Foshan 528225, Guangdong, China

Author to whom correspondence should be addressed.

Abstract: People often predict the final medal count of the Olympics based on factors such as the number of events and the host country information. This paper will predict the final medal table data based on models like Markov and Logistic Regression. Based on the Markov model, the number of gold medals and the total number of medals are each represented by four states. A model is established according to these four states, and the state transition probability matrix is calculated. Then, the Monte Carlo algorithm is used to solve the model. The results show that the number of medals and gold medals of countries such as CHN and USA are in state 4, while those of countries such as ZAM and DOM are in state 1. Indicators such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) are also used to measure the model's performance. The results show that the prediction model performs relatively well in predicting the total number of medals. A 95% confidence interval is also provided for all prediction results. Regarding the issue of taking into account countries that haven't won any medals yet, by analyzing the dataset, three factors are counted: the number of participations NC , the number of participating events NL, and the average ranking A . Based on these, a logistic regression model is established to predict the countries that will win their first medal in the next Olympic Games. Finally, the specific countries predicted to win their first medal in the next Olympics are identified as BOL, ESA, LIE, MLT, MYA, and NCA. Moreover, a feasibility analysis of the established model is carried out. Metrics such as accuracy, precision, recall, and the F1 - score are calculated. The results indicate that the model performs well overall and is feasible. Regarding the impact of the "Great Coach" effect on a country's Olympic performance, the Propensity Score Matching (PMS) method is selected. According to the model solution results, for national teams with the "Great Coach" effect, the difference in the winning situation under the influence of other winning-related factors is small. Therefore, it can be considered that the "Great Coach" effect has a significant impact on the winning situation of national teams. Regarding the problem that requires considering the competition events of a specific Olympic Games. The results show that the model can explain approximately 72.97% of the variation in the dependent variable, indicating that this quadratic fitting model has a certain feasibility.

Keywords: Markov prediction model; Logistic regression; Propensity score matching method

Online publication: March 26, 2026

1. Introduction

1.1. Background

The Olympic Games are one of the world's most influential sports events, with their medal table drawing global attention.

In recent years, Olympic medal table prediction has become a hot topic [1], which requires considering multiple factors: historical data, sports development trends, athlete status, event settings, host country advantages, as well as non-sports factors like the “excellent coach effect”.

2. The description of the problem

2.1. Problem statement

Considering the above-mentioned background information, this paper aims to address the following core issues:

2.1.1. Medal count prediction model construction

Construct a medal prediction model with uncertainty/accuracy estimation and performance evaluation.

2.1.2. Prediction of the medal table for the 2028 Los Angeles Olympics

Predict the 2028 Los Angeles Olympics medal table with prediction intervals, and analyze countries’ performance changes compared to 2024.

2.1.3. Analysis of the relationship between events and medal counts

(1) Take into account the number and types of events in the Olympics in the model, and explore the connection between them and the number of medals won by various countries.

(2) Identify the most important sports events for different countries and the reasons, and analyze the impact of the events selected by the host country on the competition results.

2.1.4. Analysis of the “Excellent Coach Effect”

(1) Evaluate the impact of the “excellent coach effect” on the number of medals by analyzing the data.

(2) Select three countries, determine in which sports events these countries should consider hiring “excellent” coaches, and estimate the impact.

2.1.5. Exploration of Other Insights Revealed by the Model

(1) Through model analysis, discover other unique insights of the model regarding the number of Olympic medals.

(2) Explain how these insights can provide references for national Olympic committees.

3. Models

3.1. The use and analysis of the construction of a model for predicting the medal table

In the first question, we model and analyze the Olympic medals. First, we preprocess the data and use a Markov model to predict the medal range for each country in 2028, evaluate the model performance, and analyze the progress or regression of each country. Then, we use a logistic regression model to predict the probability of winning for countries that have not yet won a medal, which is verified to be reliable by a variety of metrics; and then we analyze the relationship between the events of the competition and the number of medals with the help of data visualization and a quadratic regression model to assess the importance of the events to each country using the “Winning Score” model [2].

3.1.1. Modeling the predictive medal table

3.1.1.1. Data pre-processing

(1) Comprehensive checking and processing of missing values

For the summerOly_programs.csv file, we found that there are many missing values in the number of events, and after calculating and checking the total number of events, we filled them with 0. For summerOly_programs.csv, there is no data for some countries in some years, so we set the number of gold, silver and bronze medals and the total number of medals to 0 to avoid the errors caused by missing data in the subsequent calculation.

(2) Outlier detection and handling

After a comprehensive outlier check on the four datasets, it was found that there were no outliers.

(3) Data type checking and conversion

For the summerOly_hosts.csv file, it is found that the name of the country used subsequently is not the full name, but the capitalized abbreviation of the English letter, so it is necessary to add a new column to add the capitalized abbreviation of the host country in each year.

3.1.1.2. Modeling

For the prediction of the number of medals of each country in the future, considering that the number of medals given in the question is not related to the historical number of medals restriction, a Markov prediction model is used to solve the problem[3], Markov chain is a stochastic process with no posteriority, that is, the state of the system at a certain moment in the future depends only on the current state, and has nothing to do with the past historical state. In this model, the change of the state of the medal count of each country is regarded as a Markov chain, and this property makes it possible to use the Markov prediction model for the prediction of the medal count. The following is the analysis of the process of model building.

(1) Determine the current state

Obtain the number of medals won by each country in the most recent edition (e.g., Paris 2024) from the dataset and determine the current state of the medal count that each country is in according to the pre-defined state classification rules. It is assumed that a country is in the state of Gold Medal Count in 2024, and the total number of medals is in the state. The number of gold medals and the total number of medals are used here as key indicators to portray a country's status in the Olympic medal competition.

(2) State definition

Define the states S_n^{gold} and S_n^{medal} as two four-dimensional vectors corresponding to the state of the number of gold medals and the state of the total number of medals for each country in the n-th Olympics:

(a) Divided by gold medal status: To analyze the trend of gold medals in each country in a more detailed way, the number of gold medals is divided into 4 statuses, status 1 is the number of gold medals greater than or equal to 5, status 2 is the number of gold medals within 0-15, status 3 is the number of gold medals within 15-30, and status 4 is the number of gold medals greater than 30.

(b) Division by Medal Totals Status: Similarly: the medal totals are divided into 4 statuses in order to fully assess the country's overall performance in the Olympics, with status 1 being medals greater than or equal to 5, status 2 being medals within 5-15, status 3 being medals within 15-30, and status 4 being medals greater than or equal to 3.

Based on the above medal categorization, each of the four states of gold and total medals is used as an element of the vectors S^{gold} and S^{medal}. The final state vectors for the number of gold medals, as well as total medals, are as follows:

$$\begin{cases} S_n^{gold} = \left(S_{n1}^{gold}, S_{n2}^{gold}, \ldots, S_{n4}^{gold}\right) \\ S_n^{medal} = \left(S_{n1}^{medal}, S_{n2}^{medal}, \ldots, S_{n4}^{medal}\right) \end{cases} \tag{1}$$

3.1.1.3. State transition probability matrix

State probability describes the possibility that the object under study, when in a certain state of the state space at time t, will be in the next state at time $t+1$. By analyzing the changes in the medal - number states of various countries between different Olympic Games in historical data, the state probability matrix can be calculated. The specific model established is as follows [4].

$$\begin{cases} P_{ij}(t) = \left(X(t) = S_j \mid X(t-1) = S_i \right) & (i, j = 1, 2, .., n) \\ P_{ij}(t) = \dfrac{N_{ij}}{\sum_{j=1}^{n} N_{ij}} & (i, j = 1, 2, .., n) \\ 0 \le P_{ij}(t) \le 1 & (i, j = 1, 2, .., n) \\ \sum_{j=1}^{n} P_{ij}(t) = 1 & (i, j = 1, 2, .., n) \end{cases} \tag{2}$$

4. Summary and recommendations

This study analyzes Olympic medal prediction and key factors using four models. Countries should focus on advantageous events, hire elite coaches, and adjust strategies to improve medal-winning efficiency.

Disclosure statement

The author declares no conflict of interest.

References

[1] Shi HM, Zhang DY, Zhang YH, 2023, Can Olympic Medals Be Predicted? — From the Perspective of Interpretable Machine Learning. Journal of Shanghai University of Sport, 47(12): 1–12.

[2] Zhang YH, 2013, Prediction of China's 31st Olympic Medal Count Based on Linear Regression Dynamic Modeling. Journal of Henan Normal University (Natural Science Edition), 41(2): 24–26 + 60.

[3] Kvam P, Sokol JS, 2006, A Logistic Regression/Markov Chain Model for NCAA Basketball. Naval Research Logistics, 53(8): 788–803.

[4] Zhu MN, Peng T, Chen K, 2017, Mathematical Analysis of Influencing Factors of Olympic Medal Table. Contemporary Sports Technology, 7(27): 239–243.