The 2028 Olympic Medal Prediction Model Based on Entropy Weight – TOPSIS

The prediction of the Olympic medal table holds significant implications for strategic planning in sports development, resource allocation, and the evaluation of event influence. This study, centered on the 2028 Summer Olympics in Los Angeles, constructed two medal prediction models, namely the information entropy weighted model and the TOPSIS method, and analyzed the "star coach" effect.

For Problem 1, we propose two models: the **Information Entropy Weighting Model (IEWM)** and the **Long Short-Term Memory (LSTM) model** (used as our Benchmark Model / Comparison Model). In IEWM, the **ID3 algorithm** and the **entropy weight method** are integrated to determine the weight of each influencing factor, with the athlete's status value and the standard project value set as the core independent variables. For LSTM, we only use historical data of countries' medal wins for training. The model predicts that China, Japan, South Korea, Australia, and other countries may experience a decline in their medal table rankings, while Britain, France, the Netherlands, Germany, and others are expected to improve. Our results are displayed in table 5.

In Problem 2, our **TOPSIS Prediction Model** analyzes and predicts which countries are likely to win their first gold medal by considering the number of participants in 2024 and the number of times they have participated in the Olympic Games since 2000 as independent variables. Our results show that MLI, GUI, and ANG have a very high probability of winning their first gold medal in the next Olympic Games.

In terms of Problem 3, we used the **association rule method** to identify the unique sports of countries from the perspectives of support and confidence. At the same time, we demonstrated that when a host country selects its unique sports, it can leverage its maximum advantage. The results are shown in xxx.

With regard to Problem 4, we propose a model to quantify the influence of "great coaches" on athletes' medal-winning abilities using two metrics: **the Linear Metric and the T metric**. The Linear Metric quantifies the coach's influence by fitting a linear model and obtaining the normalized coefficient of the great coach parameter and the T metric . Additionally, we conducted an ablation experiment using **ARIMA(5, 1, 1)**. The results are shown in xxx.

As for Problem 5, through **EDA data analysis** and our aforementioned **IEWM**, we revealed some special phenomena of the Olympic games.

At the very last, we analyze the , we analyze the strengths and weaknesses of our model as well as its sensitivity.We also evaluated the model.

Keywords: Olympic medal table prediction; Information entropy weighted model; TOPSIS method;LSTM; Star coach effect; ARIMA;association rule method

删除[几]:

设置格式[几]: 段落间距段前: 8 磅, 段后: 8 磅

设置格式[几]: 段落间距段前: 8 磅

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

删除[几]: First, we compare the two models by having th

删除[几]: Figure XXX

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

删除[几]:

设置格式[几]: 正文, 两端对齐, 缩进: 首行缩进: 2 字

设置格式[几]: 字体: 加粗

设置格式[几]: 段落间距段前: 8 磅

设置格式[沐]: 字体颜色: 红色

设置格式[沐]: 字体: 加粗, 字体颜色: 红色

设置格式[沐]: 字体颜色: 红色

删除[几]: We calculated the influence of Lang Ping on th

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

设置格式[几]: 字体: 加粗

设置格式[几]: (中文) 中文(简体)

删除[几]: the following

设置格式[几]: (中文) 中文(简体)

删除[几]: : the host country has a significant impact on

删除[几]: Eventually, robustness and sensitivity analysis

设置格式[几]: 字体: (中文) 等线, (复杂文种)

设置格式[几]: (中文) 中文(简体)

设置格式[几]: 字体: (中文) 等线, 小四, (复杂文种)

设置格式[几]: 正文, 两端对齐, 缩进: 首行缩进: 8.5

设置格式[几]: 字体: (中文) 等线, 小四, (中文) 中

Contents

1.Introduce . . . 1
  1.1Problem Background . . . 1
  1.2Restatement of the Problem . . . 1
  1.3 Literature Review . . . 1
  1.4 Our Work . . . 2
2 Assumptions and Justifications . . . 3
3 Notations . . . 4
4. Information entropy weighted prediction model . . . 5
  4.1 Data analysis and establishment of prediction model . . . 5
    4.1.1 Discussion of independent variables affecting the number of MEDALS . . . 5
    4.1.2 Selection of algorithm and overview of model . . . 6
  4.1 Solving the model . . . 7
    4.2.1 Calculation of specific parameters . . . 7
    4.2.2 Predictions for the medal table in 2028. . . . 8
    4.2.3 Evaluation of the importance of each country's events . . . 10
  4.3 Comparision Test . . . 11
  4.4 Reliability Test . . . 11
5 TOPSIS method prediction model . . . 12
  5.1 Data analysis and establishment of prediction model . . . 12
    5.1.1 Discussion of independent variables and selection of algorithms . . . 12
  5.2 Overview and solution of the model . . . 14
  5.3 Solving the model . . . 14
6 Star coach discussion on the influence of the number of MEDALS . . . 15
  6.1 Data analysis and algorithm selection . . . 15
  6.2 Formulating Hypotheses . . . 18
  6.3 Making Decisions . . . 19
7 Insights . . . 21
  7.1 Winning the award for inefficient countries and efficient countries . . . 21
  7.2 Gender and the Olympiad . . . 21
  7.3 Host distribution . . . 23
  7.4 The difference between the awards of the sports powerhouses in traditional and emerging sports . . . 错误！未定义书签。
8 Sensitivity Analysis . . . 25
9 Model Evaluation and Further Discussion . . . 25
  9.1 Strengths . . . 25
  9.2 Weaknesses . . . 26
  9.3 Further Discussion . . . 26
Conclusion . . . 26
References . . . 27

1.Introduce

1.1 Problem Background

In recent years, the global attention to the Olympic Games has grown, with far-reaching impacts on politics, economy, society, and culture. Fans closely follow the Olympic medal table, and some small countries have made breakthroughs in winning medals, makingpredictions about Olympic results a hot topic. Olympic results reflect a country's competitive strength, and nations are increasingly focused on prediction models to gain insights. Predicting Olympic medal outcomes can provide valuable information to Olympic committees, helping them optimize resource allocation, design targeted training programs, and make data-driven decisions to improve overall performance.

设置格式[几]: 字体: （中文）Times New Roman

设置格式[几]: 缩进: 首行缩进: 7.4 毫米，段落间距段前: 8 磅，段后: 8 磅，行距: 单倍行距，正文文本，允许西文在单词中间换行，不压缩句首标点符号，不使用中文规则控制首尾字符

删除[几]:

1.2 Restatement of the Problem

In this problem, we are given the data of Information about the Summer Olympics, including: national medal tables, athlete information, Summer Olympics hosts, Summer Olympics program information. We will solve the following problems based on the historical Olympic Data:

1. Predict the medal table of each country at the 2028 Los Angeles Olympics in the United States, predict which countries are relatively progressive, and which countries are relatively regressive;
2. Which countries will win their first medal at next year's Olympics, with an estimated probability;
3. The relationship between the project and the country's medal award, and explore how the project chosen by the host country will affect the outcome of the award;
4. Explore the phenomenon of great coaching and list projects that have invested in three countries that have introduced a great coach;
5. Provide additional insights into the number of gold medals won at the Olympics based on the model.

删除[沐]: In recent years, the global attention to the Olympic Games has been continuously rising, and its hosting has far-reaching impacts in politics, economy, society and culture. To ensure the smooth running of the Games, the host city will adjust policies, budgets and resource allocation, improve ...

设置格式[几]: 段落间距段前: 13.45 磅

设置格式[几]: 段落间距段前: 8 磅，段后: 8 磅

删除[几]:

1.3 Literature Review

Muller, M., Gollner, B., & Tschernutter, C. In 2020, International Journal of Sports Science & Coaching,[1] "Predicting Olympic Success: "A Machine Learning Approach" features national historical performance, athletes' personal records and background information, the intensity of competition in the sports, and the resources invested (such as training facilities, financial support). Linear regression, decision tree, random forest, support vector machine (SVM), neural network and other models are used to conduct cross-validation. This paper discusses how to accurately predict the number and category of MEDALS (such as gold, silver, bronze) for each country in the Olympic Games, and what factors can affect the success rate of athletes and countries.

设置格式[几]: 段落间距段前: 8 磅，段后: 8 磅

Julia Bredtmann, Carsten J. Crede, and Sebastian Otten published in Significance, Issue

3, 2016, "Olympic medals: does the past predict the future?" Taking the total medal count of the past four Olympic[2] Games, sports participation in each country, and government investment and support for sports as important features, the model is evaluated by regression analysis, time series analysis, classification model, random forest and other methods such as cross-validation to study whether unequal resource allocation between different countries leads to differences in the chances of winning MEDALS.

Christoph Schlembach, Sascha L. Schmidt, Dominik Schreyer, and Linus Wunderlich in Technological Forecasting and Social "Forecasting the Olympic medal distribution - A socioeconomic machine learning model[3]" in Change, Volume 175, 2022 to GDP (GDP), per capita income, population size, education level as important features, through a variety of machine learning algorithms to build forecasting models, Cross-validation and other statistical evaluation methods (such as mean square error, R-squared value) were used to evaluate the performance of the model. Socioeconomic factors such as the economic strength of the country, the investment in sports facilities, and the proportion of the population participating in sports activities had a significant impact on medal allocation.

1.4 Our Work

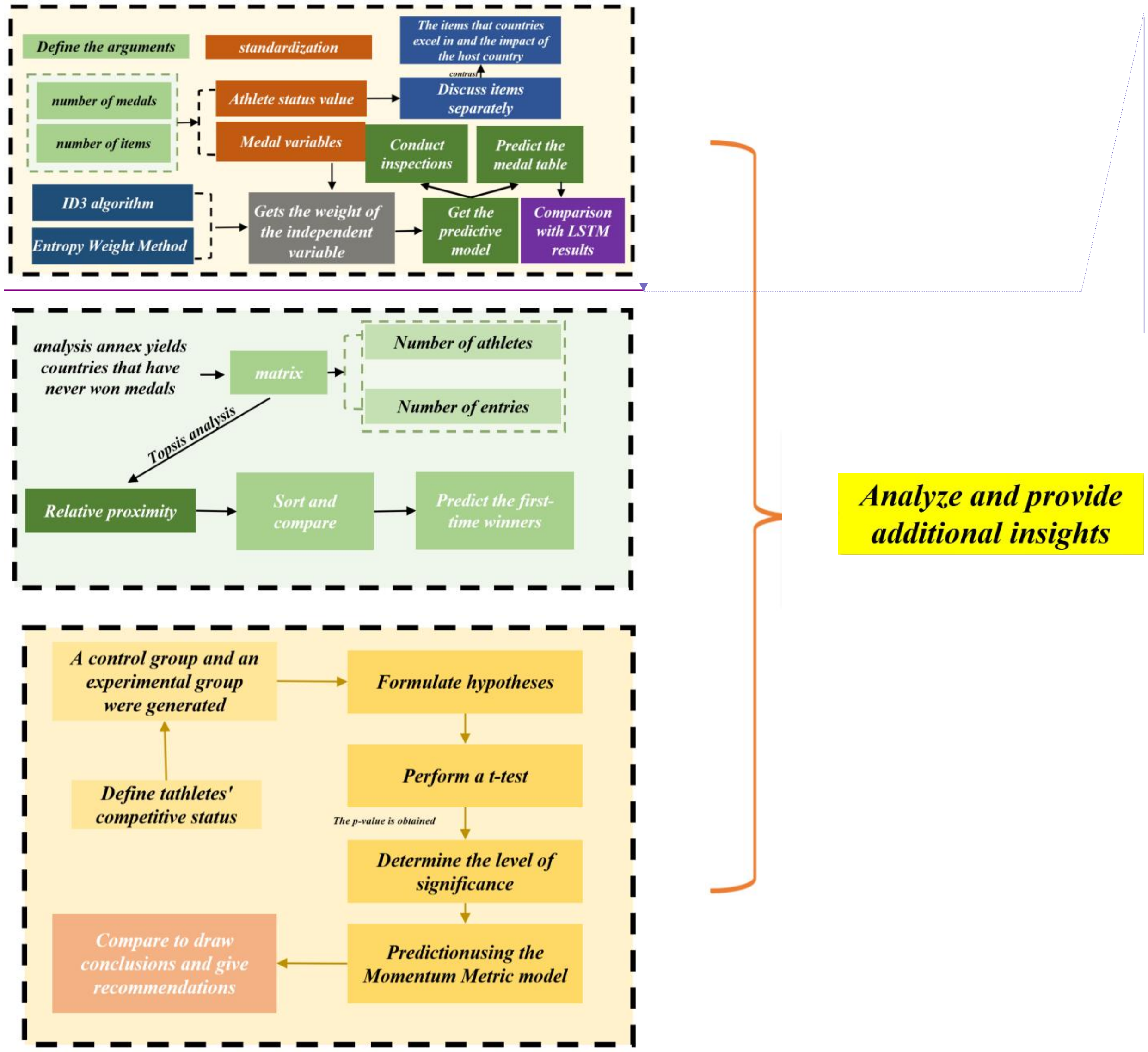
We propose a model framework as shown in Figure 2, which mainly includes three models, namely **IEW Model**, **TOPSIS Prediction Model**, **Impact Quantification Model**, and some evaluation criteria, such as **association rules**, **linear correlation criteria**.

The **IEW Model** is used to predict the number of national medals, the **TOPSIS Prediction Model** is based on the improvement of the TOPSIS algorithm to finely evaluate whether a country can win the next medal.

Regarding the impact of great coaches, we use the **Influence Quantification Model**, which combines the **momentum metric** and the **linear metric**, to quantify the influence of great coaches.

In addition, we also conducted comparative experiments to prove the effectiveness of the model, such as using **LSTM** to predict the number of national medals directly based on time compared to our **IEW Model**, and using ***ARIMA(p, d, q)*** (specifically *ARIMA(2,1,1)*)to compare data excluding the great coach factor with data containing the great coach factor.

设置格式[几]: 段落间距段前: 8 磅, 段后: 8 磅
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
删除[沐]: , etc
删除[沐]: .
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
删除[沐]: **m**
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 加粗, 倾斜
删除[沐]: ***5***
删除[沐]: ***1***
删除[沐]: ***1***
设置格式[沐]: 字体: 加粗
设置格式[沐]: 字体: 非加粗
设置格式[沐]: 字体: 非加粗, 倾斜
设置格式[沐]: 字体: 非加粗
设置格式[几]: 段落间距段前: 13.45 磅, 段后: 13.45 磅



删除[沐]:

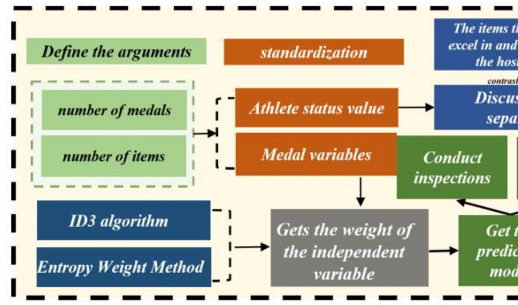


Figure2 our work

2 Assumptions and Justifications

Assumptions related to athlete status: When constructing the information entropy weighted prediction model, it was assumed that the athlete status was only affected by the performance of the past Olympic Games, and the influence of the innovation of training methods, changes in diet and nutrition, and fluctuations in psychological factors on the current competitive status of athletes was ignored. At the same time, it is assumed that the state of athletes in different sports is independent of each other. For example, the state change of swimming athletes will not interfere with the state of track and field athletes, and the transmission effect of cross-sport training or event experience exchange on athlete state is not considered.

设置格式[几]: 段落间距段前: 8 磅, 段后: 8 磅

The **change hypothesis of the number of** events: The change trend of the number of events in the Olympic Games is assumed to be not affected by major adjustments in sports organization policies, changes in global sports culture, and the rise of new sports. The change is only based on the rule described by the Logistic population model. It is believed that the adjustment of the number of sports in the future Olympic Games will be smooth and predictable, and there will be no sudden increase or decrease of large-scale sports or subversive change of rules.

Data Representative hypothesis: This study uses the data of the Olympic Games since 2000, and assumes that these data can comprehensively and accurately reflect the development trend of sports in various countries and the change law of athletes' competitive level. The impact of missing data, recording errors or special years (such as the special circumstances of the host country affecting the fairness of the event) on the data quality is ignored, and it is regarded as a fully representative sample for the construction and validation of the model.

Weight assumption of TOPSIS method: In the prediction model of TOPSIS method, it is assumed that the influence weight of the number of entries and the number of participants on the probability of winning MEDALS is equal, which is 0.5. Because in the countries that often participate in the Olympic Games and often win MEDALS, there are a large number of athletes who are never absent, and it is impossible to quantify the weight, and through daily experience, the two have little difference in the degree of influence on the results.

Besides

删除[沐]: **"Great Coach" effect hypothesis**: When studying the "great coach" effect, it is assumed that the coach has a direct and unique influence on the athletic state of the athletes, and the synergistic or countervative effects of other coaches, training team members, or changes in the effort level of the athletes themselves on performance during the same period are not considered. At the same time, it is assumed that the coaching effect of "great coach" is consistent in different countries and different programs, and the influence of cultural differences and sports foundation differences on coaching results is not considered.

The influence of innovations in training methods, changes in diet and nutrition, and is assumed to remain constant, denoted as a constant value C.

Different countries exhibit varying degrees of resistance or inertia towards the same great coach, and this resistance reduces the coach's influence on a particular sport in that country.

3 Notations

Symbol	Description
G	Number of gold MEDALS (with duplicates)
S	Number of silver MEDALS (with duplicates)
B	Number of bronze MEDALS (with duplicates)
SUM	Total number of games
p_i	Medal win percentage
X	Athlete status value
E	Total number of projects
M_i	Number of MEDALS
d	Index of variation
e	Information entropy
m_G	Medal variables
φ	Standard item ratio

K	Maximum ambient capacity
r	Natural growth rate
A	Residual growth space ratio
PI	imprecision
D	Distance
C	Relative approximation
sc	The average score of the athlete
nae	The number of athletes in the project
P	The ability of a great coach himself
R	Inertia of team capabilities
O	The number of sessions of the Olympiad

Some of the symbols in the table above illustrate: One event can have multiple athletes, and each athlete can also participate in multiple competitions, so the total number of games here does not refer to the total number of events, nor the number of athletes. Suppose N events, on average, n athletes participate in each event, SUM = N × n. Because athletes participate in teams, and a champion only corresponds to a medal in the medal table, G, S and B represent the number of MEDALS in this repeated situation, and represent the number of MEDALS recorded in the medal table.

4. Information Entropy Weighted Prediction Model

4.1 Data analysis and establishment of prediction model

4.1.1 Discussion of independent variables affecting the number of MEDALS

To analyze the performance of athletes from a specific country, we refer to their historical performance and conduct regression analysis based on the country's results in previous Olympic Games to reflect the current state of its athletes. As shown in the attachment, athletes participate in every Olympic Games. The gold medal winning rate is calculated by dividing the number of gold medals won by the total number of events. Since the status of the three types of medals varies, the gold, silver, and bronze medal winning rates are weighted and summed, with weights of 3, 2, and 1 respectively, to obtain a new variable. We define this as the athlete status value X, as follows:

$$p_1 = \frac{G}{SUM},\quad p_2 = \frac{S}{SUM}, p_3 = \frac{B}{SUM} \#(4.1)$$

$$X = 3p_1 + 2p_2 + p_3 \#(4.2)$$

Where the value is between 0 and 1, X can directly affect the final medal count.

In addition, the more sports there are in the Olympic Games, the more MEDALS will be distributed throughout the Games, and the higher the probability that each country will win a medal. Obviously, the total number of Olympic Games will also affect the number of MEDALS for a country, and the total number of Olympic Games directly determines the total

设置格式[沐]: 字体: （默认）Cambria Math, 非倾斜

设置格式[沐]: 字体: （默认）Cambria Math

设置格式[几]: 段落间距段前: 8 磅, 段后: 8 磅

删除[沐]: <math>

删除[沐]:

设置格式[沐]: 左, 段落间距段前: 0 磅, 段后: 0 磅, 正文文本

删除[沐]: **e**

删除[沐]: **w**

删除[沐]: **p**

删除[沐]: **m**

设置格式[沐]: 字体: 加粗

删除[沐]: The number of Olympic MEDALS is generally not referred to the total number of historical MEDALS, so the number of historical MEDALS is not considered, and it is used as the dependent variable of the function to test the model. In most cases, the current status of the athletes can directly affect the final results of the number of MEDALS.

设置格式[几]: 段落间距段前: 8 磅, 段后: 8 磅

number of Olympic Games MEDALS, but the contribution degree of the two to the number of MEDALS is obviously different. The total number of events E and X are too different, so E should be standardized. When E=300, φ=1 and

$$\varphi = \frac{E}{300}\#(4.3)$$

Therefore, the medal variable can be defined as follows: m_G

$$m_G = k_1 X + k_2 E = k_1 (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + k_2 \varphi \#(4.4)$$

4.1.2 Selection of algorithm and overview of model

设置格式[沐]: 字体: 加粗

设置格式[几]: 段落间距段前: 8 磅, 段后: 8 磅

Decision tree model ID3[4], as a greedy algorithm, plays the role of constructing decision trees. ID3 algorithm originates from Concept Learning System (CLS). At each node, it selects the attribute with the highest information gain as the partition criterion, and then continues this process until the generated decision tree can perfectly classify the training examples.

The entropy weight method[5] is used as the weight determination method. Based on the information entropy, the weight of each index is determined by calculating the entropy value of each index and comparing the degree of difference between the indexes. The entropy value is inversely correlated with the difference between the indicators, which can show that the index is more important in decision-making.

In this paper, the two methods are combined. Firstly, the algorithm of information entropy and information gain in ID3 algorithm is used to obtain the disorder degree of decision attribute and the influence degree of each attribute on decision attribute. This method can combine the advantages of the two methods to make the weight determination more reasonable and comprehensive.

We divide the athlete state X into three parts. Taking the United States as an example, the state is set as follows. The total number of Olympic events E is divided into three levels: more, medium and less. The number of gold MEDALS won by the United States is divided into three levels: more, medium and less. The proportion of each level is calculated and the information gain of E and X is calculated by ID3 algorithm. $M_G M_G$

About the specific use of the idea of entropy weight method: the process of entropy weight method to standardize the sample index is to control the value of the described sample in [0,1], and use it to calculate the entropy value, and the scope of the information entropy algorithm is, this paper uses the entropy weight method to calculate the weight of the information entropy, and in the entropy weight method to calculate the variation index of the entropy value: $[0, \log_2 3]$d= 1-E, where "1" is the maximum value of the value of the sample after standardization, then when we calculate the variation index of the information entropy, the formula should be:

$$d = \log_2 3 - e\#\#(4.5)$$

The corresponding weights k1 and k2 are calculated from the formula of weight obtained by processing the variation index in the entropy weight method:

$$k_i = \frac{d_i}{\sum_{i=1}^{n} d_i} \#(4.6)$$

Finally, the formula for the total number of gold MEDALS in the United States is obtained:

$$m_G = \frac{\log_2 3 - e_1}{\sum_{i=1}^{n} d_i} \times \left(3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}\right) + \frac{\log_2 3 - e_2}{\sum_{i=1}^{n} d_i} \times \varphi \#(4.7)$$

4.1 Solving the model

4.2.1 Calculation of specific parameters

Because of the rapid development of all countries in the world, the results calculated by the data of the Olympic Games since 2000 have more reference value for today. Through the screening and calculation of the attachment, we obtain the table:

Table 1 Calculation of athlete status in the United States in recent years

Year	SUM	Gold	Silver	Bronze	X
2000	776	130	61	51	0.725515
2004	735	117	75	71	0.778231
2008	768	127	110	80	0.886719
2012	698	145	57	46	0.852436
2016	726	139	54	71	0.820937
2020	856	113	110	75	0.740654
2024	854	131	96	94	0.795082

Use the United States itself as a reference to classify athlete status, specifying that an athlete status value between 0.75 and 0.8 is fair, greater than 0.8 is good, and less than 0.75 is bad. By observing the number of Olympic events since 2000, the number of stipulated events between 310 and 320 is medium, less than 310 is less, more than 320 is more, and the following table is obtained. Similarly, the number of stipulated gold MEDALS between 37 and 45 is medium, less than 37 is less, and more than 45 is more:

Table 2 Decision attribute Table

Year	X	E	M_G
2000	Bad	less	General
2004	General	less	less
2008	Good	less	less
2012	Good	less	more
2016	Good	less	more
2020	Bad	more	General
2024	General	more	General

According to the above table, the calculation results of information gain and information entropy can be calculated:

Table 3 takes the gold medal of the United States as an example, and the data results of all indicators

Metrics	Results obtained
Unconditional entropy	1.5566567074628228
Information gain and entropy of athlete states	(1.138724635395457, 0.27709636376654934)
Information gain and information entropy of the total amount of the project	(0.18150882310232208, 1.2343121760596842)
k_1, k	(0.6756010197686305, 0.32439898023136954)

The results are as follows:

$$m_G = 0.6756 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.3244 \times \varphi\#(4.8)$$

The medal constant is defined based on the number of gold MEDALS in the United States in 2000Q_G

$$m_G \times Q_G = M_G\#(4.9)$$

$$m_G \times Q_G = 37 \rightarrow Q_G = 45.4\#(4.10)$$

$$M_G = 45.4 \times (0.6756 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.3244 \times \varphi)\#(4.11)$$

Plug in the data to find the number of medals.Similarly, we can calculate:

Table 3 Data results of all indicators for silver and bronze MEDALS

Indicators	Silver Medal	Bronze
Unconditional entropy	1.556656707	
k_1, k	0.45730441975063163, 0.5426955802493684	0.826933753115614, 0.17306624688438585
Q	27.4	41.4

$$M_S = 27.4 \times (0.4573 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.5427 \times \varphi)\#(4.12)$$

$$M_B = 41.4 \times (0.8270 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.1731 \times \varphi)\#(4.13)$$

4.2.2 Predictions for the medal table in 2028

The two main independent variables in equations (4.11), (4.12) and (4.13) are the athlete status value X and the standard event value φ. We can reflect the φ value by two-dimensional image analysis of the total number of events E and the year:

设置格式[沐]: 字体: 加粗

删除[沐]: .

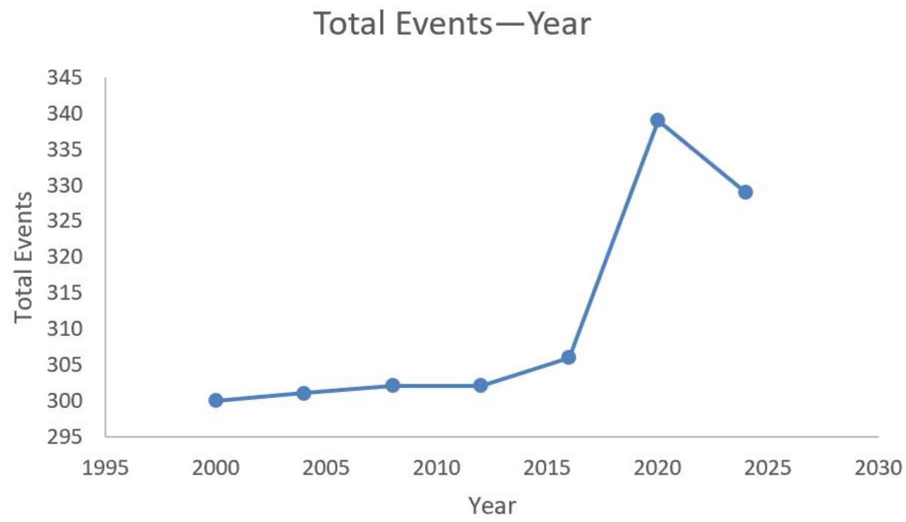


Figure 3 Relationship between total number of events and year

As shown in the figure, starting from the year 2000, the first segment of the image is similar to the exponential image, and the second half begins to decline but is relatively gentle. Such images can be approximately described by the Logistic population model[8], and the expression of the Logistic population model is:

$$E - 300 = \frac{K - 300}{1 + Ae^{-r\,(t-2000)}}, \quad A = \frac{K - E\,(0)}{E\,(0)} \#(4.14)$$

It is reasonable to set K=350 and E(0)=300 for the observation image, then A=0.17, and r=0.0105 can be calculated by substituting the value (2016,306) into equation (4.14). namely$E = \frac{50}{1{+}0.17e^{-0.0105\times\,(t-2000)}} + 300\#(4.15)$

Substituting t=2028 into equation (15), E=344, and P=1.1467.

Next, take the top ten MEDALS table in 2024 as the sample to calculate the athlete status value of each country:

Table 4 Athlete status values for each country

Year / Country	2000	2004	2008	2012	2016	2020	2024
United States	0.725515	0.778231	0.886719	0.852436	0.820937	0.740654	0.795082
China	0.380368	0.389068	0.473236	0.524345	0.427562	0.460317	0.543296
Japan	0.23416	0.374396	0.246696	0.35589	0.254587	0.39136	0.279264
Australia	0.472081	0.552413	0.446208	0.363813	0.312741	0.362069	0.355932
France	0.297872	0.227766	0.339326	0.403756	0.373047	0.581784	0.484395
Netherlands	0.552901	0.509363	0.53169	0.630631	0.273556	0.409207	0.616927
Britain	0.284337	0.322222	0.402878	0.368421	0.686192	0.485401	0.515947
South Korea	0.342618	0.313953	0.519757	0.352564	0.209125	0.215976	0.45968
Italy	0.269147	0.407173	0.156863	0.319372	0.320802	0.26145	0.278689
Germany	0.358025	0.466216	0.342342	0.413725	0.559701	0.249581	0.307573

Using the least squares[7] method, we can predict and plot the athlete status of each country in 2028, taking France as an example:

删除[几]:

设置格式[几]: 左, 缩进: 首行缩进: 7.4 毫米

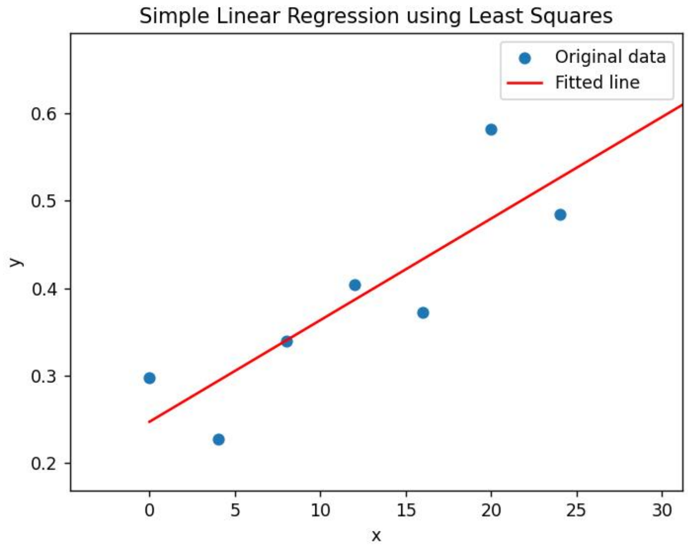


Figure 4 Regression analysis of athlete status values for France

The abscissa of the figure above is the year -2000, and the ordinate is the athlete status value.

After this calculation, the new medal table is obtained:

Table 5 Medal tally forecast for 2028

2024 Rankings	Projected 2028 Rankings	M_G	M_S	M_B
United States	United States	42	27	37
China	Britain	36	24	31
Japan	France	34	24	29
Australia	China	33	23	27
France	Netherlands	31	22	25
Netherlands	Germany	27	21	20
Britain	Japan	27	21	20
South Korea	Australia	26	21	19
Italy	Italy	25	20	18
Germany	South Korea	22	19	15

A comparison of the new medal tally with the old one shows that China, Japan, South Korea and Australia will fall back in 2028, while Britain, France, the Netherlands and Germany will improve.

4.2.3 Evaluation of the importance of each country's events

设置格式[沐]: 字体: 加粗

For the athlete state value X defined by us, athletes in different countries have different performances in different kinds of sports. That is to say, by discussing the athlete state value of one kind of sports separately and comparing the X value of each sport in a country, the sport corresponding to the largest athlete state value is the most important sport in a country. Analyzing the data, here is a list of the most important sports for a subset of countries:

Table 6 table of events that each country excels in

Countries	Project	X
United States	Swimming	1.244094

China	Swimming	0.57732
Japan	Judo	1.5
Australia	Swimming	0.87395
France	Judo	1.896552
Netherlands	Hockey	3
Britain	Rowing	1.714286
South Korea	Archery	2.357143
Italy	Volleyball	1.5
Germany	Athletics	0.04

If a host country wants to improve its medal ranking, hosting more events it is good at will bring other countries that are good at these events up with it, while countries that are not good at these events will cause its ranking to fall.

4.3 Comparision Test

We use LSTM to make predictions based on historical national award data, and compare them with our proposed IEW model, and the experimental results are as follows:

Table 7 Comparison of prediction results

Actual situation	IEWM	LSTM
United States	United States	United States
China	China	China
Netherlands	Japan	Britain
Japan	Australia	France
France	France	Australia
Australia	Netherlands	Japan
South Korea	Britain	Germany
Germany	South Korea	Italy
Britain	Italy	South Korea
Italy	Germany	Netherlands

带格式表格[沐]

Experiments have shown that our IEW Model can model the state of athletes in a more refined manner, so as to make more accurate predictions.

4.4 Reliability Test

According to the meaning of unconditional entropy, the number of decision attributes selected by the model is 3, so the maximum value of unconditional entropy is according to the formula. Compared with the maximum value of unconditional entropy, the ratio is about 98.2%, which shows that the random sampling results of the established model are successful, which can better reflect the actual situation and contain more information. $\log_2 3 \approx 1.5850$.
The information gain of the condition attribute is the reflection of the degree of influence on

the decision attribute, and the information entropy of the condition attribute is the reflection of its irrelevance to the decision attribute. It is observed that the contribution of athlete status value and the total number of events to the decision attribute is obvious, because their information gain values are relatively high, which means that they are important classification features.

The model of medal quantity is tested on the sample of China since 2000: The following table is obtained by calculation:

Table 8 Actual indicators of China

Year	**X**	M_G	M_S	M_B	**E**
2000	0.380368	28	16	14	300
2004	0.389068	32	17	14	301
2008	0.473236	48	22	30	302
2012	0.524345	39	31	22	302
2016	0.427562	26	18	26	306
2020	0.460317	38	32	19	339
2024	0.543296	40	27	24	329

Substitute X into the model and obtain the result:

Table 9 Predicted results of China's MEDALS

M_G	M_S	M_B
26	20	20
27	20	21
30	21	23
31	22	25
29	21	22
29	23	24
33	23	26

Comparing the two tables, it can be seen that the model has certain reference value for the prediction of this sample, but there are still some inaccuracies. We take the quotient of the difference between the predicted value and the actual value and the actual value to reflect the inaccuracy of the result:

$$\pi = \frac{\sum_{i=1}^{21}(x_i - x)}{21} = 0.19\#(4.16)$$

The error of the model is small, which has a good reference value. The model has passed the reliable test.

5 TOPSIS method prediction model

5.1 Data analysis and establishment of prediction model

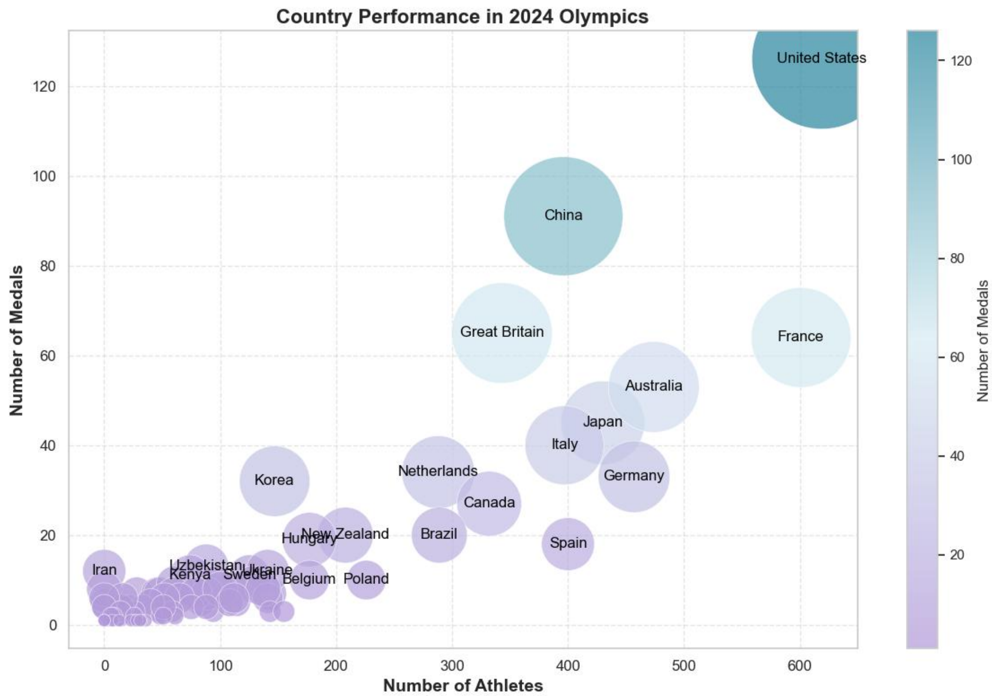
5.1.1 Discussion of independent variables and selection of algorithms

The medal-winning efficiency varies across countries. The figure below shows the

删除[沐]: , but there is a small deviation from the actual situation. This is not considering the situation of data outside the attachment given by the title. When other sample data are obtained, the model can be improved to achieve smaller error.

设置格式[沐]: 字体: 加粗

删除[沐]: Observe the table, in the top of the Olympic Games medal list are the strong comprehensive strength of the big countries, not only are regular visitors to the Olympic Games, there are many participants, more participants, naturally have greater competitiveness in the competition, participate in the Olympic Games more times, relationship between the number of participants and the number of medals in each country:



删除[沐]:

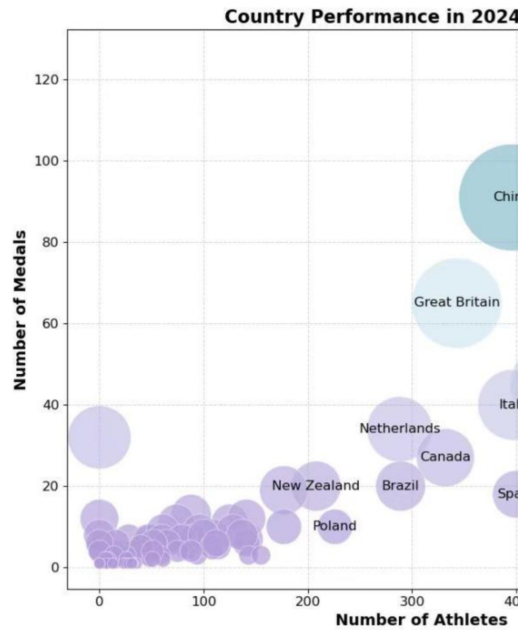


Figure 5 relationship between the number of participants and the number of medals in each country

Additionally, greater familiarity with the rules, resulting from frequent participation, can also improve the medal-winning rate. Therefore, the number of participants in 2024 and the number of participating in the Olympic Games since 2000 of each country are used as independent variables to establish a prediction model for 2028 through the TOPSIS comprehensive evaluation method[6].

删除[沐]: , so,

Firstly, we select the countries that have never won a medal from the table, and then count their number of participants in the 2024 Olympics as well as the number of times they have participated in the Olympics since 2000. There are 68 countries that have not won a medal but will participate in the 2024 Olympics. By combining the number of participants and the number of Olympic appearances, we can create a 68x2 matrix. Below is a sample of this matrix:

设置格式[沐]: 字体: 小四

设置格式[沐]: 缩进: 首行缩进: 2 字符

Table 10 Number of Olympic Games participated by countries and number of athletes participating

删除[沐]: the

NOC	Number of appearances	Number of participants
LIE	19	1
MAD	14	7
MAW	12	3
MDV	10	5
MHL	5	4
MLI	15	24
MLT	18	5
MTN	11	2
MYA	19	2

NCA	14	7
NEP	15	7
NRU	8	1

5.2 Overview and solution of the model

The basic principle of TOPSIS method is to construct the positive ideal solution and negative ideal solution of the decision problem, that is, the optimal solution and the worst solution of each attribute index, and then calculate the distance between each alternative and the positive ideal solution and the negative ideal solution, and determine the order of the pros and cons of each alternative by comparing the size of these distances. The closer the solution is to the positive ideal solution and the farther the solution is to the negative ideal solution, the higher the relative superiority is. The following is the calculation procedure

Positive transform the data

$$x_{ij}^{'} = \frac{1}{x_{ij}} \#(5.1)$$

Re-normalize the data

$$\tilde{x}_{ij} = \frac{x_{ij}^{'}}{\sqrt{\sum_{i=1}^{m} (x_{ij}^{'})^2}} \#(5.2)$$

For the case that countries that often win MEDALS are almost absent in the Olympic Games, and these countries that often win MEDALS will have a large number of athletes to participate in each time, it is difficult to measure the contribution value of these two indicators, so it is assumed that the contribution value of the number of entries and the number of participants is the same, that is, the weight is 0.5, the ideal solution is the maximum value, and the negative ideal solution is the minimum value, then there is a distance d$max_j min_j$

$$D^{+} = \sqrt{\sum_{j=1}^{m} (0.5\tilde{x}_{ij} - max_j)^2} = \sqrt{\sum_{j=1}^{m} \left(\frac{0.5}{x_{ij}\sqrt{\sum_{i=1}^{n} \left(\frac{1}{x_{ij}}\right)^2}} - max_j\right)^2} \#(5.3)$$

$$D^{-} = \sqrt{\sum_{j=1}^{m} (0.5\tilde{x}_{ij} - min_j)^2} = \sqrt{\sum_{j=1}^{m} \left(\frac{0.5}{x_{ij}\sqrt{\sum_{i=1}^{n} \left(\frac{1}{x_{ij}}\right)^2}} - min_j\right)^2} \#(5.4)$$

设置格式[几]: 字体: 小五

It can be concluded that the relative closeness C=, comparing the size of the relative approximation can get the most likely to get the medal of the country.$\frac{D^{-}}{D^{+}+D^{-}}$

5.3 Solving the model

The relative approximation of each country is calculated. Here are the top ten relative approximation countries:

Table 11 Relative approximation for each country

NOC	MLI	GUI	ANG	SAM	LBR	ESA	NEP	MAD	NCA	GAM
C	0.001869	0.001863	0.001855	0.001854	0.001848	0.001844	0.001838	0.001835	0.001835	0.001828

By analyzing the annex, we can get that in recent decades, the minimum number of countries winning MEDALS for the first time in the Olympic Games is 8, and usually more than 10. An analysis of the situation of the first award-winning country in recent years was carried out and it was found that nearly 100 percent of the top three countries in relative proximity were the first winners of the yearthe probability of the ten countries in the above table winning MEDALS for the first time in the next Olympic Games is very high, and the probability of the first three countries winning MEDALS can be almost 100%.

6. Great Coach Impact

In the study of the impact of great coaches on a team, we first use PACF and ACF to determine the parameters of the ARIMA model, and then apply the ARIMA(p, d, q) model to predict the data before the coach's tenure, which helps to exclude the influence of the great coach, allowing for a comparison with the actual data that accounts for the coach's impact. Next, we construct the Influence Quantification Model, which uses linear and momentum criteria to quantify the coach's impact on the team. Finally, we use the t-distribution test to assess the accuracy of the model.

6.1 Model Establishment

6.1.1 Establishing the model equation

Considering the volatility of medal counts, we use the variation in athletes' average scores to reflect the influence of the great coach on the team. First, we define the average score of athletes as follows, with the convention that the weight for gold medals is 3, silver medals is 2, and bronze medals is 1:

$$score_{ave}=\frac{3\times G+2\times S+1\times B}{nae}$$

where $score_{ave}$ represents the average score of the athletes and Num represents the number of athletes. Clearly, $0 <= score_{ave} <= 3$.

We model the impact of a great coach on a team by examining their influence on the team's average score (considering that the number of medals varies with the number of events). The coach's effect on the team's score can be quantified as follows:

$$Impact=\frac{P}{R}$$

Here, P represents the coach's own ability, and R represents the team's performance inertia, which refers to the resistance to the coach's influence.

6.1.2 ARIMA Parameter Selection

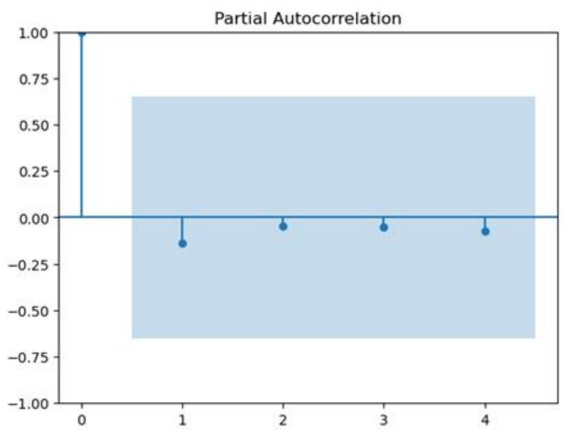
We use the ARIMA model to predict the data without considering the influence of the great coach, in order to compare it with the real data. Taking the influence of Béla Károlyi on

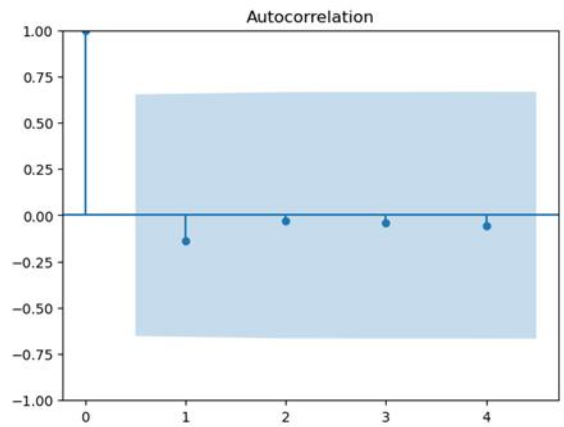
设置格式[几]: 居中, 缩进: 左侧: 8.5 毫米, 悬挂缩进:
设置格式[几]: 项目符号和编号
删除[沐]: **Star**
删除[沐]: **coach discussion on the influence of the num**
设置格式[沐]: 字体: 小四, 非加粗
设置格式[沐]: 缩进: 首行缩进: 2 字符
设置格式[沐]: 与下段不同页, 段中分页, 正文文本
设置格式[沐]: 左, 孤行控制
设置格式[沐]: 字体: 非加粗
删除[几]: $score_{ave}=\frac{3\times G+2\times S+1\times B}{E}$
删除[沐]: $score_{ave}=\frac{3\times G+2\times S+1\times B}{E}$
设置格式[沐]: 与下段不同页, 段中分页, 正文文本
设置格式[沐]: 字体: 倾斜
设置格式[沐]: 字体: 倾斜, 下标
设置格式[沐]: 字体: 非倾斜
设置格式[沐]: 非上标/ 下标
设置格式[沐]: 字体: 倾斜
设置格式[沐]: 非上标/ 下标
设置格式[沐]: 字体: 非倾斜
设置格式[沐]: 缩进: 首行缩进: 2 字符, 与下段不同页
设置格式[沐]: 字体: (默认) Times New Roman
设置格式[沐]: 字体: (默认) Times New Roman, 倾斜
设置格式[沐]: 字体: (默认) Times New Roman
设置格式[沐]: 字体: (默认) Times New Roman, 倾斜
设置格式[沐]: 字体: (默认) Times New Roman
设置格式[沐]: 字体: 非加粗

the U.S. women's artistic gymnastics team as an example, we first select the ARIMA parameters based on PACF and ACF:

Based on the PACF plot, we choose the first significant cutoff point as p, as shown in the figure, with p = 2. And based on the ACF plot, we select the first significant cutoff point as q, as shown in the figure, with q = 1.

设置格式[沐]: 字体: （中文）等线, 非加粗





Additionally, since the data is a non-stationary series, we set d = 1.

Therefore, we will use the ARIMA(2, 1, 1) for prediction in the next section.

设置格式[沐]: 字体: （默认）Times New Roman

6.1.3 Metrics

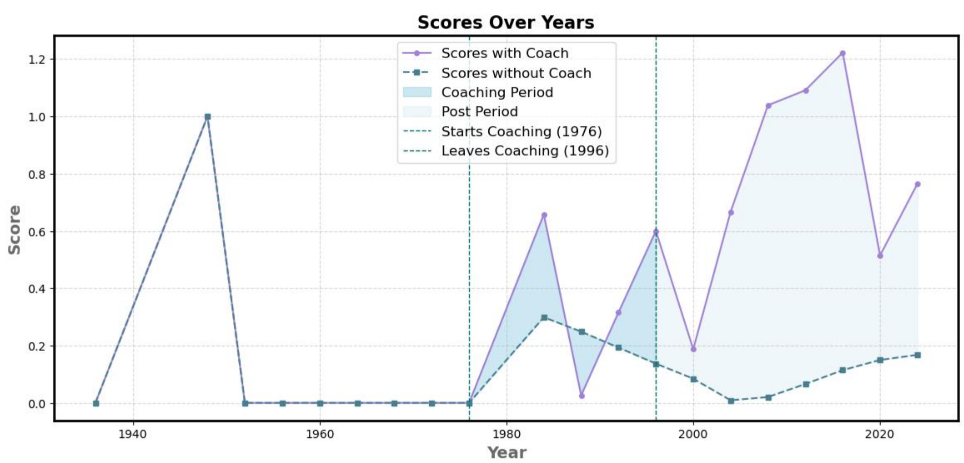
设置格式[沐]: 段落间距段前: 0 磅, 段后: 0 磅, 行距: 单倍行距, 正文文本

First, we use the ARIMA(2, 1, 1) model to predict the data based on the performance before the coach's tenure, and the difference between the ARIMA forecast and the actual data is shown in the figure below, illustrating the changes in the average performance of the U.S.

设置格式[沐]: 字体: （默认）Times New Roman, （中文）Times New Roman, 小四, 非加粗, （中文）

设置格式[沐]: 缩进: 首行缩进: 2 字符

设置格式[沐]: 字体: 小四, 非加粗



women's artistic gymnastics team:

设置格式[沐]: 字体: 小四, 非加粗

As seen in the figure, Béla Károlyi brought a significant improvement to the average score of the U.S. women's artistic gymnastics team. Next, we will use two criteria to quantify

设置格式[沐]: 字体: （默认）Times New Roman

this impact:

Linear Metric: We believe that the average performance of the team members in the current Olympic Games is related to the performance of athletes in the previous Games, the team's historical performance, the coach's influence, and other constant factors, such as nutrition. Therefore, we model the problem with the following equation:

设置格式[沐]: 字体: 加粗

设置格式[沐]: 两端对齐, 缩进: 首行缩进: 3 字符, 段落间距段前: 0 磅, 段后: 0.3 行, 行距: 单倍行距, 不对齐到网格, 正文文本, 定义网格后不调整右缩进

$$score_{ave_i} = C + \alpha\, score_{ave_{i-1}} + \beta O_i + \gamma T_i$$

Where C represents constant factors, Oi represents the historical performance of the team in the i-th Olympic Games, and Ti is an indicator variable representing whether there was a great coach in the i-th Games.

设置格式[沐]: 字体: 小四, 非加粗

Using the above equation, we fit a linear regression model to the average score, which allows us to calculate the coefficient of the indicator variable Ti, representing the influence of the coach on the team. And in the above linear relationship, the team's resistance to the influence of the great coach, denoted as *re*, is:

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: 缩进: 首行缩进: 2 字符

设置格式[沐]: 字体: 倾斜

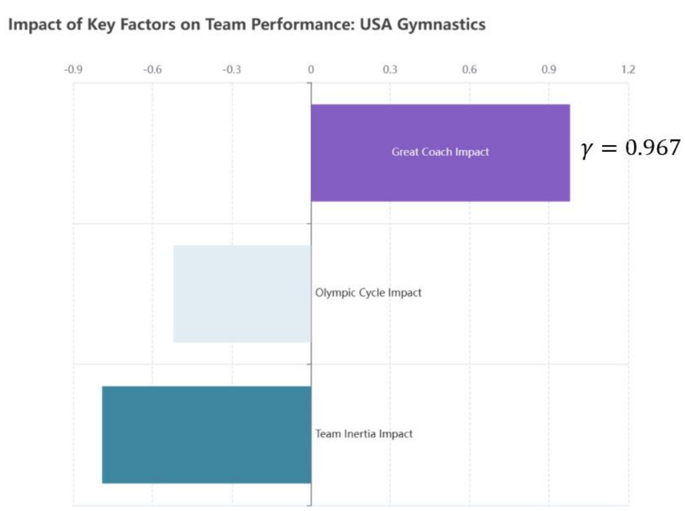
$$re = |\alpha + \beta|$$

We apply a linear regression model to the average performance data of the U.S. women's gymnastics team and obtain the parameters as shown in the figure:

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: 字体: 非加粗

设置格式[沐]: 字体: 小四, 非加粗



It is clear that the coach's influence is positive and significantly higher than the other two coefficients, indicating that Coach Béla Károlyi indeed brought a tremendous improvement to the U.S. women's gymnastics team. This aligns with the facts and validates the effectiveness of our model.

设置格式[沐]: 缩进: 首行缩进: 2 字符

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: 字体: 非加粗

T metric: 待完成

as can be seen from Figure 6,7 and annex, in some Olympic Games and major competition events, some countries will shine on the events they were not good at and win gold MEDALS, silver MEDALS and bronze MEDALS. In addition, for the four cases of bronze MEDALS, silver MEDALS, gold MEDALS and not winning MEDALS, we set the value of athletes' competitive status y_i based on the winning situation in the first game of the athletes, where the gold medal is won, The value is 3 when winning the gold medal, 2 when winning the silver medal and bronze medal, and 1 when not winning the medal.

Because the status values were compared across multiple games, we $Y = \{y_i | \mathrm{i} = 1, \ldots, n\}$ calculated the difference between each pair of observations using a paired sample t-test, and then tested whether the mean of these differences was significantly greater than zero.

Selection and overview of test methods

Paired Sample t-test is a common statistical method, which can be used to compare the effects of different treatment conditions, and can also be used to compare the differences between different individuals in the same treatment condition to compare whether there is a significant difference in the mean of two groups of related samples.

In this article, the data is collected first, and the data of each paired sample is organized into two columns, one column is the value before the treatment and the other is the value after the treatment, or one column is the value of the experimental group and the other is the value of the control group.

6.2 Formulating Hypotheses

Null hypothesis: It is usually assumed that the means of two paired samples are not significantly different, that is, where is the population mean of the difference between the two paired samples.$H_0: \mu_d = 0 \mu_d$

Alternative Hypothesis: Depending on the purpose of the study, the one-tailed alternative hypothesis or the two-tailed alternative hypothesis may be chosen. The one-tailed alternative hypothesis is often used to test whether a treatment is effective-for example, whether the mean of the experimental group is greater than the mean of the control group; The two-sided alternative hypothesis is often used to test whether two group means differ regardless of the direction of the difference.

Calculate the test statistic

Calculate the difference: Calculate the difference for each pair of samples, where and are the two observations for the second pair, respectively. $d_i = X_{1i} - X_{2i} X_{1i} X_{2i} i$

Calculate the mean and standard deviation: Calculate the sample mean and sample standard deviation of the difference as follows:$\bar{d} s_d$

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i, \quad s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2} \# (6.1)$$

Where, is the number of paired samples.n

设置格式[沐]: 字体颜色: 红色

设置格式[沐]: 字体: 非加粗, 字体颜色: 红色, (中文)

删除[沐]: **6.1 Data analysis and algorithm selection**

In the Olympic Games, the role of coaches is not only reflected in technical guidance and tactical development, but also in stimulating the potential of athletes, improving team cohesion, and optimizing training plans. Unlike

Calculate the test statistic: Calculate the statistic t of the paired sample t-test according to the formula:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \#(6.2)$$

Determine significance level and degrees of freedom

significance level: The significance level is usually chosen; a common value is 0.05 or 0.01. αThe significance level indicates the maximum probability that you are allowed to reject the null hypothesis.

Degrees of freedom: The degrees of freedom for a paired samples t-test are, where is the number of paired samples. $df = n - 1n$

Make a decision to

find the critical value: Based on the significance level and degrees of freedom, look up the t-distribution table or use statistical software to find the critical value.$\alpha df t_{\alpha/2}(df)$

Compare statistics and critical values: Compare the calculated test statistics to the critical values.$t \;\; t_{\alpha/2}(df)$

For the national team, the arrival of the star coach will strengthen the strength of the team and make the team's competitive ability stronger in the competition. In order to quantify the strength change brought by the star coach, we let $|Y|$ be the value of the team's competitive state, and $\hat{Y}$ be the growth rate of the team's competitive state, so as to measure the influence of the star coach on the national team's competitive state.

$$|Y| = \sum_{i=1}^{n} y_i \#(6.3)$$

$$\hat{Y} = \frac{|Y_2| - |Y_1|}{|Y_1|} \times 100\% \#(6.4)$$

Momentum Metric

The criterion introduces the idea of momentum from the Adam optimizer, first using Arima(5, 1, 1) to predict the subsequent average score based on the average score before the introduction of the great coach, and then quantifying the coach's influence on the team's project by calculating the difference in the momentum change (ignoring dimensions) of the athlete's average score (α=0.5 in the actual calculation):

$$Impact = \Delta Momentum_{in} + \alpha \Delta Momentum_{after} \#(6.5)$$

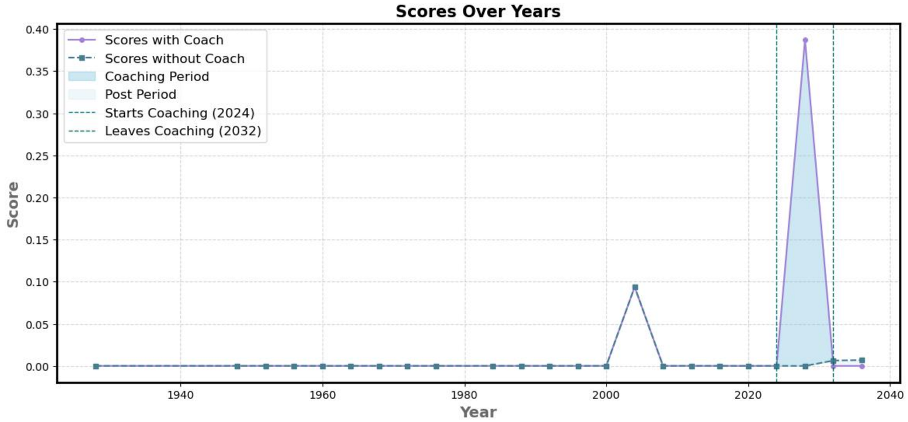
That is, the influence is the sum of the change in momentum during the coach's tenure and the change in momentum after leaving the coach after scaling.

6.3 Making Decisions

If, reject the null hypothesis that the means of the two paired samples are significantly different.$|t| > t_{\alpha/2}(df)$

If, the null hypothesis is not rejected and the means of the two paired samples are not significantly different.$|t| \leqslant t_{\alpha/2}(df)$

The competitive status values of the US women's volleyball players before and after



Lang Ping coached the US women's volleyball team were respectively

$$Y_1 = \{1,1,1,2,2,1,2,1\}$$
$$Y_2 = \{2,2,2,2,3,2,3,2\}$$

The p value obtained by t test is 0.0011, which rejects the null hypothesis and shows a significant difference.

The competitive status values of Chinese gymnastics team athletes before and after Nabieva coached the Chinese gymnastics team were respectively

$$Y_1 = \{1,1,1,1,1,1,1,1,1\}$$
$$Y_2 = \{2,3,2,3,2,2,2,2,2\}$$

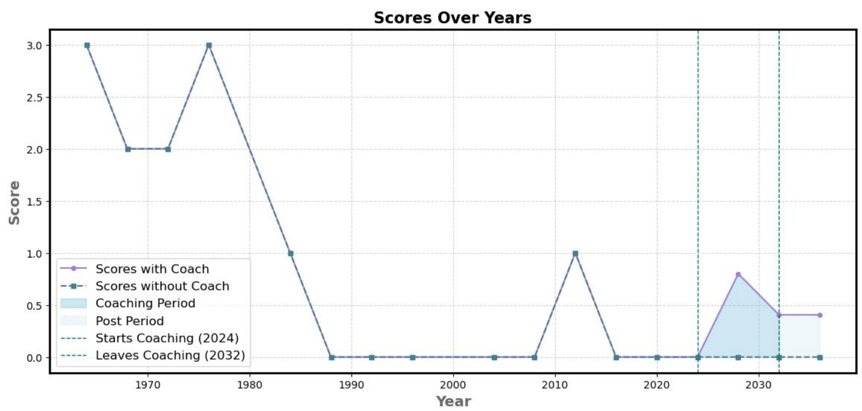
p value of t test: 0.0000, the result of rejecting the null hypothesis, there is a significant difference.

For Bela Karolyi before and after coaching the US women's gymnastics team, the competitive status values of the US women's gymnastics team athletes were

$$Y_1 = \{1,1,1,1,1,1\}$$
$$Y_2 = \{3,2,2,2,3,3\}$$

The p value of t test was 0.0002, which rejected the null hypothesis and showed a significant difference.

Through calculation, it can be concluded that the competitive state value of the American women's volleyball team when Lang Ping $\hat{Y}$100%coached it is, the competitive state value of the Chinese women's volleyball team when $\hat{Y}$122.22% Nabieva coached it is, and the



competitive state value of the American women's gymnastics team when Bela Karolyi coached it is.$\hat{Y}$150%

We used the Momentum Metri model to predict the effect of the "Great Coach" on the Japanese women's volleyball, the French women's rhythmic gymnastics team, and the French women's rhythmic gymnastics team, respectively, to reveal that the introduction of great coaches in these events will bring excellent improvements:

删除[沐]: <sp>

The influence of the "Great Coach"

The improvement of the team by the "great coach" is obvious. The competitive state of the athletes is almost doubled, so the influence of the "great coach" on the single project is doubled, and the influence on the overall medal is almost a direct addition of a medal. All countries should invest in "great coaches". This is a universal situation.

7 Key Insights and Implications for Olympic Medal Predictions

According to the model in this paper, it is found that the Olympic medals are related to some variables, and we put forward the following insights for the prediction of the Olympic medals:

7.1 Winning the award for efficient countries and inefficient countries

We use our IEW Model to predict the medal outcomes for each country in the 2028 Olympics, compare them with the results from the 2024 Olympics, and fit the data using a linear regression model. The results are as follows:

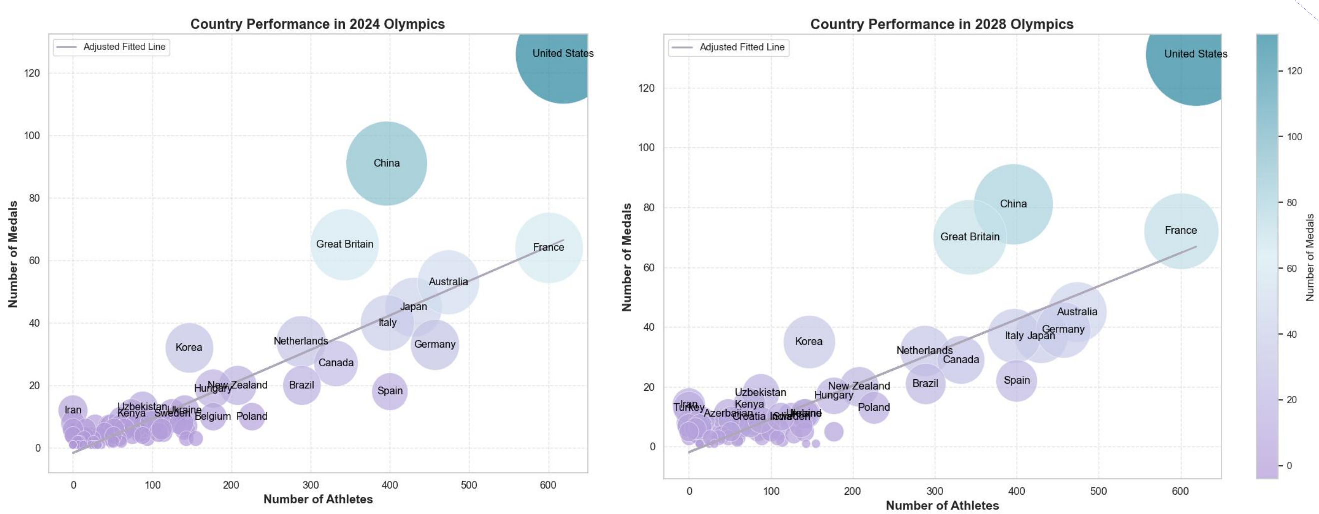


Figure 9 Relationship between Number of Athletes and Number of Medals

We define countries above the line as high-efficiency medal-winning countries and those below the line as low-efficiency medal-winning countries (of course, medal count is not the primary goal of the Olympics, this is only for research purposes)

We found that countries like the United Kingdom and France have seen an improvement in medal-winning efficiency, while countries like China have experienced a decline in their efficiency.

Additionally, we observed that most smaller countries have seen an increase in their medal-winning efficiency. This reflects the Olympics' shift towards greater diversity. This

设置格式[沐]: 缩进: 左侧: 0 毫米, 无项目符号或编号

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: 缩进: 左侧: 0 毫米, 首行缩进: 2 字符

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: 字体: 小四, 非加粗

删除[沐]: <sp>**Winning the award for inefficient countries and efficient countries**

删除[沐]: p

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: 缩进: 左侧: 0 毫米, 首行缩进: 2 字符, 无项目符号或编号

could be due to the introduction of new events in the Olympics, allowing some smaller countries to break the medal monopoly of sporting powerhouses in emerging events, leading to an increase in medal-winning efficiency. We will discuss this point further below.

7.2 Gender and Olympic

The Olympic spirit embodies the values of excellence, respect, friendship, peace, and fair play, promoting unity and understanding among nations through the power of sport.

The Olympics advocates the spirit of equality, which is why we have studied the number of male and female athletes participating in the Games, as shown in the figure below.

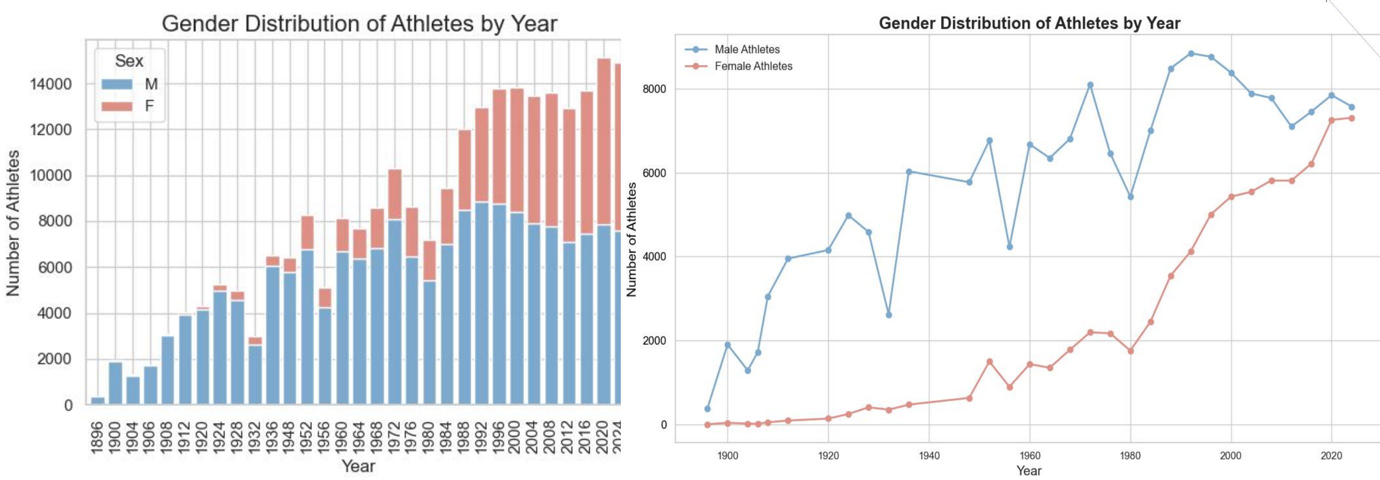


Figure 10 Number of men and women participating

We found that the number of female athletes participating in the Olympics is steadily increasing and is beginning to match that of male athletes.

Additionally, we also examined the number of medals won by male and female athletes, as shown in the figure below:

设置格式[沐]: 缩进: 左侧: 0 毫米, 无项目符号或编号

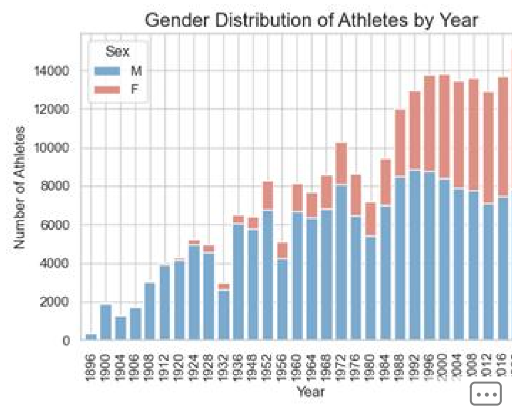
删除[沐]: **the Olympiad**

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: List Paragraph, 左, 缩进: 首行缩进: 2

设置格式[沐]: 字体: （默认）Times New Roman, 小四

删除[沐]:

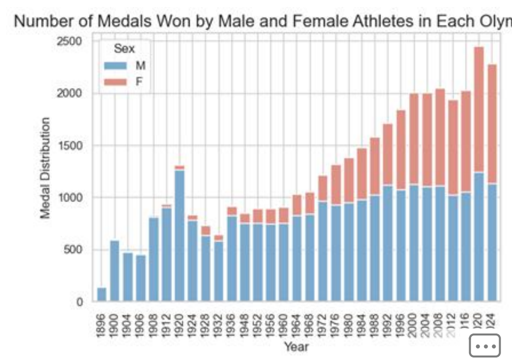


设置格式[沐]: 字体: （默认）Times New Roman, 小四

设置格式[沐]: 缩进: 首行缩进: 7.4 毫米

设置格式[沐]: 字体: （默认）Times New Roman, 小四

删除[沐]:



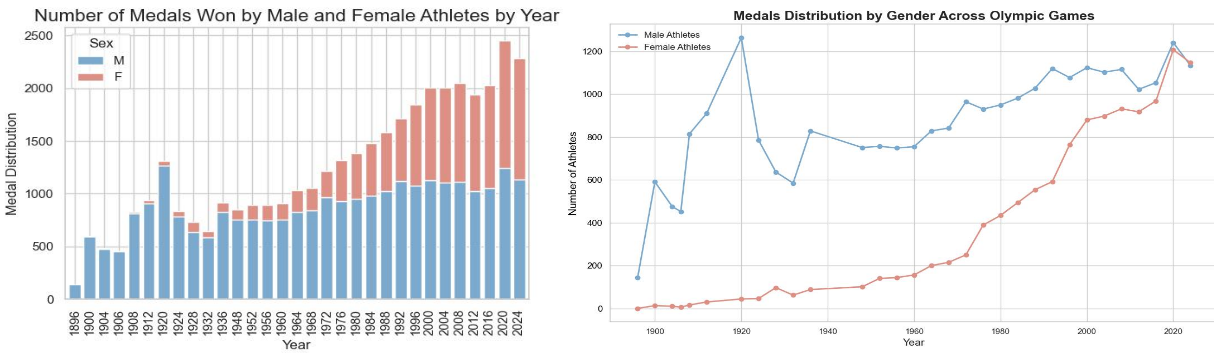


Figure 11 Male and female winning rates

We found that the number of medals won by female athletes is also rising and gradually approaching that of male athletes. This is a positive trend, signaling the increasing gender equality in sports and reflecting the growing recognition and opportunities for female athletes in the Olympic Games.

设置格式[沐]: 字体: 小四, 非加粗

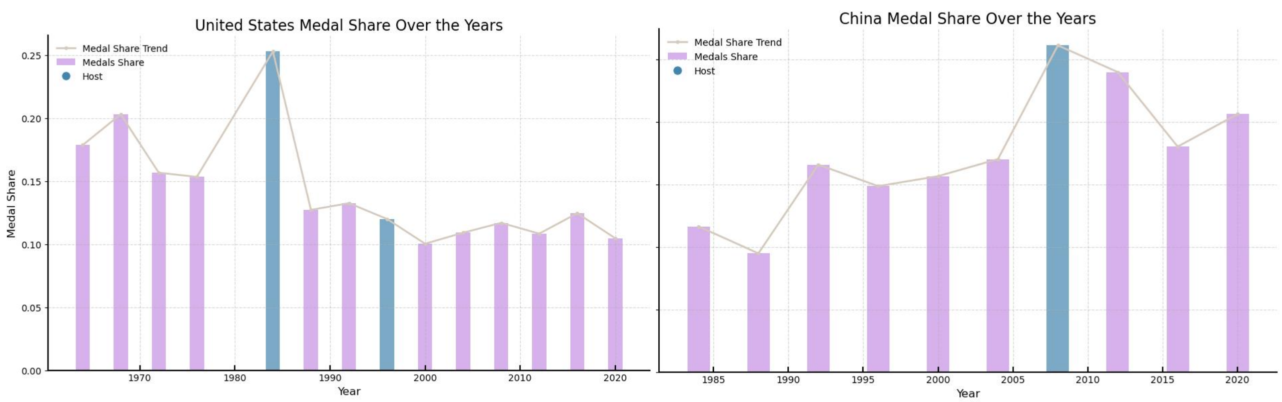
设置格式[沐]: 缩进: 左侧: 0 毫米, 首行缩进: 2 字符, 无项目符号或编号

设置格式[沐]: 缩进: 左侧: 0 毫米, 无项目符号或编号

7.3 Host Country as a Key Medal Factor

Our model analyzes whether being the host country is an important factor for a nation in winning more medals at a particular Olympics. We first visualized the medal data for the United States and China, as shown in the figure below:

设置格式[沐]: 字体: 小四, 非加粗



It is clear that the host country has a significant impact on the nation's medal share. In Section 4, we used the TOPSIS method to predict the medal share of countries, and the results are shown below:

删除[沐]: <sp>

设置格式[沐]: 字体: （默认）Times New Roman, 小四

设置格式[沐]: 字体: （默认）Times New Roman, 小四

设置格式[沐]: 缩进: 左侧: 0 毫米, 无项目符号或编号

7.4 The difference between the awards of the sports powerhouses in traditional and emerging sports

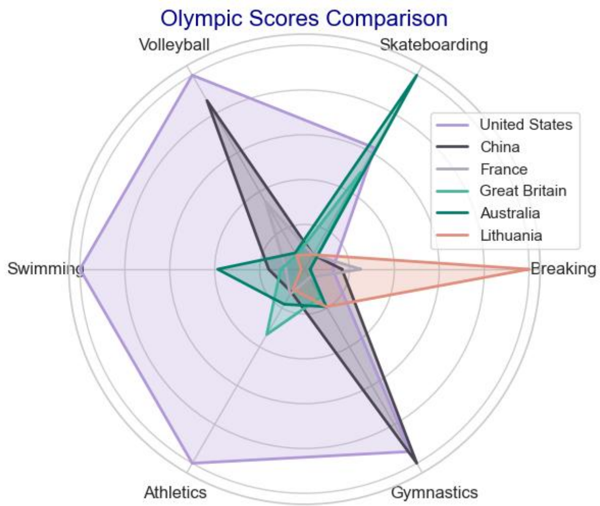
In this section, we examine the medal performance of several countries in both

设置格式[沐]: List Paragraph, 左, 缩进: 首行缩进: 9.9 毫米, 段落间距段前: 12 磅, 段后: 3 磅, 行距: 多倍行距 1.3 字行, 制表位: 6.86 字符, 左对齐, 2 级

traditional and emerging events at the 2024 Olympics and present the radar chart below(The data has been normalized).

The selected countries are the U.S., China, France, the U.K., Australia, and Lithuania, with the first five being sporting powerhouses and Lithuania having fewer medals.

The chosen events are volleyball, swimming, athletics, gymnastics, breakdancing, and skateboarding, with the first four being traditional sports and the latter two being emerging sports.



Interestingly, we found that although Lithuania's performance in traditional events is indeed not as strong as that of other sporting powerhouses, its performance in emerging events, especially breakdancing, surpasses that of other countries. This clearly demonstrates that in emerging events, traditional sporting powerhouses do not hold as significant an advantage over smaller countries, allowing smaller nations to break the monopoly of medals traditionally held by larger sports powers.

Additionally, we used our IEW Model to predict the medal outcomes for countries in 2028. We also plotted the radar chart for the same countries and events (with normalized data), and the results, shown in Figure XXX, further validate the conclusions above.

设置格式[沐]: 字体: 小四, 非加粗

设置格式[沐]: 缩进: 左侧: 0 毫米, 首行缩进: 0 毫米

设置格式[沐]: List Paragraph, 左, 缩进: 首行缩进: 9.9 毫米, 段落间距段前: 12 磅, 段后: 3 磅, 行距: 多倍行距 1.3 字行, 制表位: 6.86 字符, 左对齐, 2 级

8 Sensitivity Analysis

We use MATLAB to give the influence of different k1 values on the image, here +=1, so it changes with the change of the other side, the stability test of one of them is the stability test of the other, we change the value to make a diagram, the step flow chart is as follows:$k_1k_2k_1k_2k_1$与k_2k_1.

Here the value is (0,0.25,0.5,0.75,1), let a country capacity value is 0.8, the number of items is 300, and several relationship images with M are obtained:k_1y_1, y_2

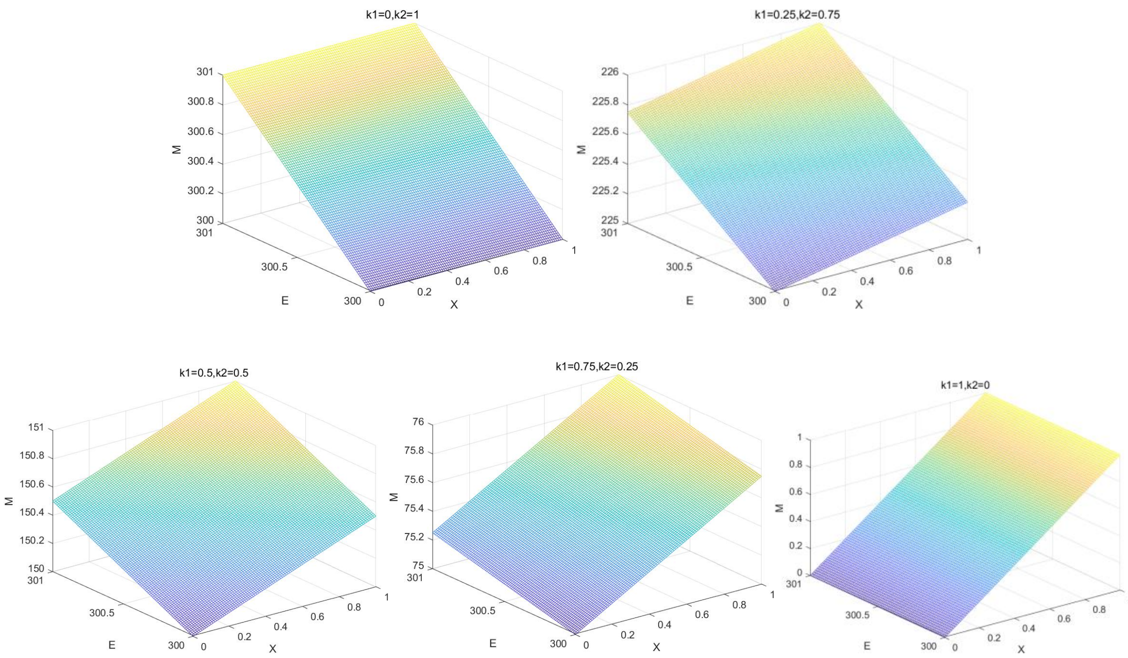


Figure 8 Sensitivity analysis of the model

By analyzing the five images, it is found that in non-extreme cases, the value of does not have a great influence on the value of M. However, when the difference between, and is too large, it still has a great influence on M. Therefore, the model has local stability, but has a certain sensitivity in a large range.$k_1k_2k_1k_2$

9 Model Evaluation and Further Discussion

9.1 Strengths

1. The combination of ID3 algorithm and entropy weight method is innovative, and the results obtained by this calculation are objective and reliable.

2. A lot of data is standardized, and more variables are defined, which not only makes the calculation simple, but also does not affect the final results, and the analysis of the problem is more clear and convenient to describe.

3. Using regression analysis and Logistic population model to process the data, making the prediction results more credible.

4. Using TOPSIS comprehensive analysis method to process the data and predict the results, the results are intuitive, easy to understand, and can deal with a large number of data.

5. Use t test to verify the "great coach" effect, making the results intuitive.

9.2 Weaknesses

1. The weight cannot be estimated in TOPSIS method, which may bring bias.

2. The idea of entropy weight method is strongly dependent on data and may be disturbed by outliers.

3. The current low number of "great coaches" worldwide may result in an insufficient sample size for the data analysis here.

9.3 Further Discussion

This model can be further improved by considering the introduction of more evaluation factors, such as the level of national economic development and the difference in athletes' living habits, in order to comprehensively assess the medal competition ability of a country and predict the medal table more comprehensively. At the same time, more complex mathematical models, such as neural network models, can be used to more accurately describe the value of athletes' status and the relationship between events and MEDALS. The model in this paper can not only be used in medal prediction, but also has universality. As long as the actual situation can be quantified, such as "athlete status value" and "standard project ratio", the weight can be calculated to obtain the prediction model. After adding enough variables, more accurate predictions can be made in weather prediction, stock market, and employment situation.

Conclusion

Focusing on the prediction of the Olympic medal table, this study constructed the information entropy weighted prediction model and the TOPSIS method prediction model, and analyzed the "great coach" effect, and obtained the following conclusions:

Medal list prediction results: The information entropy weighted prediction model determines the weight by innovatively combining ID3 algorithm and entropy weight method, and uses the athlete state value and standard item value as the key independent variables to predict the medal list of the 2028 Olympic Games. The reliability test of the model shows that although there is a certain error (the inaccuracy is 0.19), it still has good reference value. The prediction results show that in the 2028 Olympic Games, China, Japan, South Korea, Australia and other countries may decline in the medal table rankings, while Britain, France, the Netherlands, Germany and other countries may improve. At the same time, the important sports of each country are identified, which provides a reference for each country to formulate the competition strategy.

First time winning country prediction: The TOPSIS method prediction model takes the number of participants in 2024 and the number of countries participating in the Olympic Games since 2000 as independent variables. By calculating the relative proximity, it is predicted that MLI, GUI, ANG and other countries have a high probability of winning a medal for the first time in the next Olympic Games. Among them, the relative proximity is that the probability of the first three countries winning the first medal is close to 100%., which provides a new perspective for studying the diversity of Olympic medal distribution.

Verification of "Great Coach" effect: The paired sample t-test was used to analyze the effect of "great coach". The results showed that after Lang Ping coached the US women's volleyball team, Napiyewa coached the Chinese gymnastics team, and Bela Karolyi coached the US women's gymnastics team, the athletes 'competitive status was significantly improved, which proved that the "great coach" had an obvious effect on improving the team's performance. It also has a positive effect on the increase of the overall medal number.

References

[1] Muller, M., Gollner, B., & Tschernutter, C. (2020). Predicting Olympic success: A machine learning approach. *International Journal of Sports Science & Coaching*, 15(3), 345—359.

[2] Bredtmann, J., Crede, C. J., & Otten, S. (2016). Olympic medals: does the past predict the future? *Significance*, 13(3), 22—25.

[3]Schlembach, C., Schmidt, S. L., Schreyer, D., & Wunderlich, L. (2022). Forecasting the Olympic medal distribution – A socioeconomic machine learning model. *Technological Forecasting and Social Change*, 175, 121—134.

[4]Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81—106.

[5] Zhang, X., & Wang, Y. (2019). An entropy weight method for evaluating the performance of sustainable development goals. *Sustainability*, 11(22), 6265.

[6] Mardani, A., Jusoh, A., Zavadskas, E. K., & Raut, R. D. (2015). Application of TOPSIS method for supplier selection in supply chain management: A case study. *Journal of Business Economics and Management*, 16(6), 1134—1155.

[7] Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis* (2nd ed.). Wiley.

[8] Verhulst, P. F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathematique et Physique*, 10, 113—121.