Proceedings of Machine Learning Research vol 278:1–7, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

SDE for Olympic selection Based on Dynamic Bayesian Network

Si Chen* CB622@163.COM and **Xiang Peng** and **Shixuan Xu**
Shenzhen Nanshan Foreign Language Senior High School, Shenzhen 518052, China
**Corresponding author*

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

This paper concentrates on the evaluation of Sports, Disciplines, or Events (SDEs) for Olympic selection. It presents a comprehensive approach that integrates multiple methods. The Dynamic Bayesian Network (DBN) is at the core, supplemented by data collection, normalization, and the TOPSIS method. This approach allows for a systematic assessment of SDEs, taking into account various criteria such as popularity, gender equity, and sustainability. The model's outcomes provide valuable predictions for future Olympic SDE selection, and sensitivity analyses confirm its stability. The research proposes a data-centric approach for the International Olympic Committee (IOC) to refine and enhance the Olympic sports program, leveraging insights from AI and analytics.

Keywords: Dynamic Bayesian Network; Olympic sports selection; SDE evaluation; TOPSIS; multi - criteria decision - making

1. Introduction

The Olympic Games, with a long - standing history dating back to ancient Greece (Hu et al., 2025), are a symbol of global harmony and athletic achievement. The selection of Sports, Disciplines, or Events (SDEs) for the Olympics is a crucial process, involving a comprehensive set of criteria. The IOC currently employs six key criteria to choose SDEs, which include the global participation and popularity of the sport, gender equality among participants, and the sport's alignment with the Olympic values and traditions. However, accurately anticipating which SDEs will be included or excluded in future Games continues to pose a significant challenge. Existing methods often fail to comprehensively consider the dynamic evolution and causal relationships of SDEs. This research aims to develop a more effective evaluation model based on a Dynamic Bayesian Network (DBN) to address these issues and provide more reliable recommendations for the IOC.

2. Theoretical Foundation

2.1. Dynamic Bayesian Network

A Dynamic Bayesian Network is a probabilistic graphical model that can handle time - series data (Zhang et al., 2025). It is designed to capture the dynamic relationships between variables. In the context of Olympic SDE selection, it can model the relationship between a sport's inclusion status at different times and various associated factors. For example, variables like Years Since First Inclusion, Inclusion Consistency, Times Removed, Reinclusion Count, Host Specialization, and Previous Inclusion Status can be incorporated into the DBN (Slavic et al., 2025). These variables are discretized into different levels (Low, Medium, High) based on their characteristics and significance to the selection process. The DBN then uses conditional probability tables to quantify the dependencies between nodes, enabling the prediction of a sport's inclusion probability in future Olympics.

2.2. Other Related Theories

The Entropy Weight Method (EWM) and Analytic Hierarchy Process (AHP) are used to determine the weights of evaluation indicators (Wei et al., 2025). EWM objectively assesses the significance of indicators using data-derived information, whereas AHP integrates subjective expert opinions. The Technique for Order of Preference by Similarity to Ideal Solution is a multi - criteria decision - analysis method (Huang et al., 2023). It constructs a decision matrix, standardizes the data, calculates weighted normalized values, and determines the Positive Ideal Solution (PIS) and Negative Ideal Solution (NIS). The closeness of each SDE to the PIS and NIS is calculated to obtain a comprehensive score, which reflects how well an SDE meets the Olympic criteria.

3. Methodology

3.1. Data Collection and Preparation

Data on SDEs were sourced from various databases (Kokolakakis and Lera-Lopez, 2020), including the official records from the Chinese Olympic Committee and the analytical insights from Google Trends. The data covered multiple aspects such as popularity, cost, carbon footprint, and gender ratio. Different types of indicators, including macro-indicators, micro-indicators, and intermediate indicators, were collected. Macro indicators, such as market value and average popularity, exhibit a positive correlation with the evaluation score, indicating the broader market perception and general appeal. Conversely, micro indicators, including cost per capita, provide a more granular perspective on efficiency and resource allocation, which can also influence the overall evaluation.carbon footprint needs to be transformed. Intermediate indicators, such as the male - female ratio, have an optimal value close to 1. Data normalization was performed to make all indicators comparable. For micro-indicators, methods were used for positive transformation.transformation. The weights of the evaluation indicators (Shi and Bairner, 2022) were determined by combining EWM and AHP.

3.2. Dynamic Bayesian Network Modeling

The DBN was constructed with the six key variables. The construction of a Dynamic Bayesian Network (DBN) centers on the conditional dependencies expressed in a time series format, designed to capture the dynamic relationships between the inclusion status (S_t) of Olympic sports and associated variables. The objective at each time step t is to forecast (S_t), that is, whether a sport is included in the Olympics at time step t. The process of building the DBN model is as follows. The core variables of the DBN include the following six features, represented as the status at time step t:

(1) X_1^t (Years Since First Inclusion):

This variable measures the time elapsed since a sport was first included in the Olympic Games, reflecting its historical depth and traditional status. A longer history may indicate a stable influence within the Olympic context. The reason for considering this variable is that sports with a more extended history are more likely to be reconsidered for inclusion by the Olympic Committee.

(2) X_2^t (Inclusion Consistency):

This metric assesses whether a sport has been continuously included in multiple editions of the Olympic Games, indicating its stability and level of preference. Sports that have been included consecutively often enjoy higher visibility and recognition.

(3) X_3^t (Times Removed):

This count reflects the number of times a sport has been discontinued in the Olympic history, highlighting its volatility and the uncertain attitude of the Olympic Committee towards it. Frequent removals may suggest a sport's contentious nature, potentially affecting its chances of future inclusion.

(4) X_4^t (Reinclusion Count):

This indicates the number of times a sport has been reinstated after being canceled, demonstrating its resilience and importance of continued advocacy. If a sport has been removed multiple times but was also reinstated, it may hold appeal under certain conditions.

(5) X_5^t (Host Specialization):

This factor denotes whether the host country has a particular strength or interest in a sport. Under the IOC regulations, within the reform framework of 'Olympic Agenda 2020+5', the host organizing committee can propose up to five additional sports for their Olympic Games. Host countries tend to select sports that showcase their capabilities or appeal.

(6) S_{t-1} (Previous Inclusion Status):

This indicates whether the sport was included in the immediately preceding Olympic Games, capturing the inertia of continuity in selection. There is a tendency for sports that were included in the previous edition to be selected again.

In Dynamic Bayesian Networks, variables often require discretization to meet the model's requirements and enhance inference efficiency (**?**). For the six key variables in this study, the discretization criteria are based on the characteristics of their actual distribution and their significance to the selection of Olympic sports, dividing the variables into three levels (Low, Medium, High) (See Table 1).

Table 1: The discretization rules for variables

Variables	Low	Medium	High
X_1^t Years Since First Inclusion	< 10	10-50	> 50
X_2^t Inclusion Consistency	0	0-5	> 5
X_3^t Times Removed	0	0-3	¿3
X_4^t Reinclusion Count	0	0-3	> 3

After discretization, we can determine the probability of a sport's inclusion status in the 2032 Olympic Game by using the six core variables. The process is shown in Figure 1.

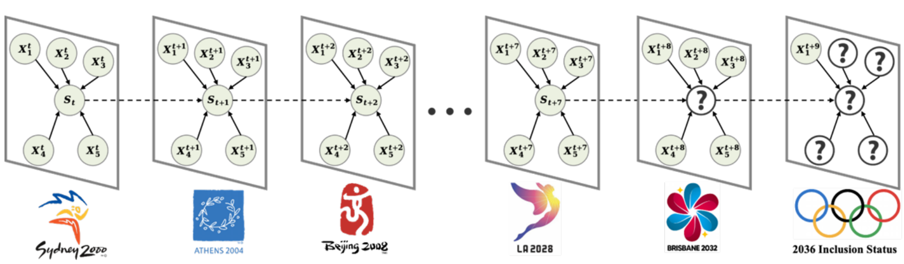

Figure 1: The process of Dynamic Bayesian Network (DBN)

3.3. Dynamic Bayesian Network Training

A Dynamic Bayesian Network (DBN) is a probabilistic graphical model that deals with time series data, focusing primarily on learning the **Conditional Probability Tables (CPTs)**. The parameters in a Dynamic Bayesian Network are categorized into two types: Static Conditional Probability Table and Time-Dependent Conditional Probability Table.

3.3.1. STATIC CONDITIONAL PROBABILITIES

In the current Olympic Games, how do variables such as the 'Years Since First Inclusion,' 'Inclusion Consistency,' and 'Reinclusion Count' collectively determine a sport's inclusion? This involves describing the interdependencies among these variables at the present time step.

(1) Probability Formula

$$P\left(X_i^t \mid S_t\right)=\frac{P\left(S_t, X_i^t\right)}{P\left(S_t\right)} \tag{1}$$

where $P\left(S_t, X_i^t\right)$ represents the joint probability, which is obtained by normalizing the joint counts from the data. And $P\left(S_t\right)$ is a marginal probability, which is calculated as follows,

$$P\left(S_t\right)=\Sigma_{X_i^t} \cdot P\left(S_t, X_i^t\right) \tag{2}$$

(2) Training Process

For each variable X_i^t:

Calculate the frequency of each value of X_i^t under different states of S_t from historical data (after discretization of the value range). Then normalize these frequencies to form the conditional probability table $\left(X_i^t \mid S_t\right)$.

3.3.2. TIME-DEPENDENT CONDITIONAL PROBABILITIES

By training the Time-Dependent Conditional Probability Table, we aim to calculate the conditional probability $P\left(S_t \mid S_{t-1}, X_1^t, \ldots, X_5^t\right)$, which represents the likelihood of the current state S_t given the previous state S_{t-1} and the static variables $X_1^t, \ldots, X_5^t$.

(1) Probability Formula

The time-dependent conditional probability is based on the following formula:

$$P\left(S_t \mid S_{t-1}, X_1^t, \ldots, X_5^t\right)=\frac{P\left(S_t, S_{t-1}, X_1^t, \ldots, X_5^t\right)}{P\left(S_{t-1}, X_1^t, \ldots, X_5^t\right)} \tag{3}$$

where $P\left(S_t, S_{t-1}, X_1^t, \ldots, X_5^t\right)$ is a joint probability, obtained by normalizing the counts from historical data. $P\left(S_{t-1}, X_1^t, \ldots, X_5^t\right)$ is a marginal probability, calculated through the data distribution.

(2) Training Process

Step 1. Extract data pairs $\left(S_{t-1}, S_t\right)$ and their corresponding static variable values $X_1^t, \ldots, X_5^t$ from the historical dataset.

Step 2. For each combination of S_{t-1} and S_t, calculate the conditional frequencies.

Step 3. Normalize these frequencies to construct the TCPT.

3.4. Inference in Bayesian Networks

For the 2032 Olympic Games, we are already aware of the static variables X_1^{2032}, X_2^{2032}, X_3^{2032}, X_4^{2032}, X_5^{2032}, and the inclusion status S_2030 from the previous edition. Using Dynamic Bayesian Network, we can infer the probability distribution $P(S_2030)$ of the inclusion status S_2032 for the 2032 Olympic Games. The steps are as follow:

Step 1. Joint Probability Decomposition:

According to the Dynamic Bayesian Network, the joint probability is factored into:

$$P\left(S_{2032}, X_1^{2032}, \ldots, X_5^{2032}, S_{2030}\right) = \\ P\left(S_{2032} \mid S_{2030}, X_1^{2032}, \ldots, X_5^{2032}\right) \cdot P\left(S_{2030}\right) \cdot \prod_{i=1}^{5} P\left(X_i^{2032} \mid S_{2032}\right) \tag{4}$$

Step 2. Conditional Probability:

The time-dependent conditional probability $P\left(S_{2032} \mid S_{2030}, X_1^{2032}, \ldots, X_5^{2032}\right)$ is provided by the TCPT constructed during the training phase. In actual calculations: S_2030 is a known value (the inclusion status from the previous edition), and $X_1^{2032}, \ldots, X_5^{2032}$ represents the known static variables for 2032.

Table 2: The inclusion probabilities for the 2032 Olympics Game of Some SDEs

SDE	probability of inclusion
Swimming	0.91
Virtual Cycling	0.12
Australian Rules Football	0.81
Track and Field	0.88
Rock Climbing	0.56

Table 2 presents the derived inclusion probabilities, we can see that Virtual Cycling have a low probability because it has never been included. In contrast, Swimming has a high inclusion probability due to its consistent presence in nearly every Olympic Games. Also, although Australian Rules Football only appeared once during the 1956 Melbourne Olympic Games, its probability of inclusion is still relatively high since it is a specialty sport for the 2032 Olympic host.

However, for the 2036 Olympic Games, We do not have knowledge of X_2, X_3, X_4, X_5 and S_t, hence there is insufficient data to calculate the time-dependent conditional probability.

3.5. Sensitivity Analysis

In our decision-making models, the weight distribution of subjective and objective indicators is crucial for the final decision outcomes. Different weight configurations may have varying impacts on the model's output, hence understanding these influences is essential for optimizing model performance.

We considered two weight configurations: 3:7 and 7:3, representing the weight ratio of subjective to objective indicators. Subjective indicator data is derived from expert evaluations, and objective indicator data is derived from historical records and statistical data. The score calculation formula is:

$$w^{\text{final}} = w_s \cdot S + w_o \cdot O \tag{5}$$

Where w_S and w_0 are the weights of subjective and objective indicators, respectively, and S and O are the scores of subjective and objective indicators, respectively.

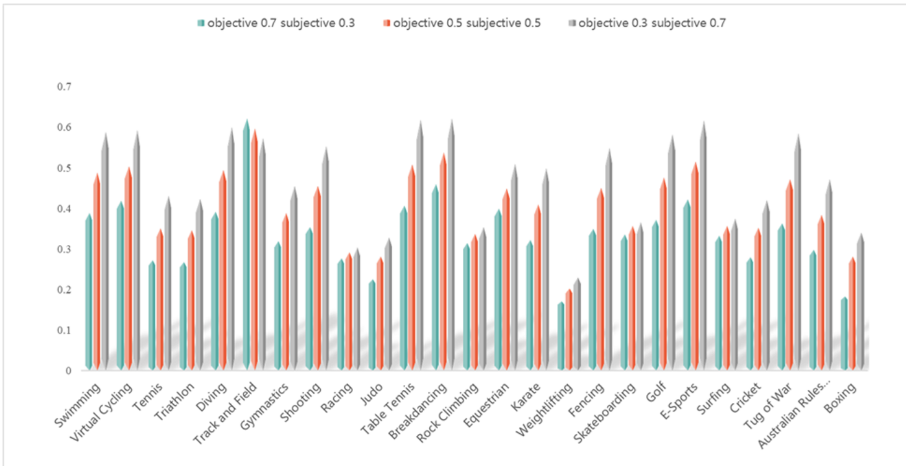

Figure 2: Comparison chart before and after weight adjustment

With a weight ratio of 3:7, the objective indicators contribute more to the score. The score calculation results indicate that high scores in objective indicators significantly enhance the total score. As we can see in Figure 2, the only sport that reaches a slightly higher score is track and field, at above 0.6 in green.

When the weight ratio is adjusted to 7:3, shown in grey in Figure 2, the influence of subjective indicators is enhanced. At this point, the weights of subjective indicators have a more critical impact on the total score, reflecting the importance of personal judgment in decision-making. It can be seen that all the other sports have a higher weight because it is quantitative and qualitative.

4. Conclusion

This research developed an evaluation model for Olympic SDE based on a Dynamic Bayesian Network. The combination of multiple methods such as DBN, EWM, AHP, and TOPSIS enabled a comprehensive and in-depth assessment of SDEs. The model's predictions for future Olympic SDE selection provide valuable references for the IOC. Sensitivity analyses also verified the importance of model structure and weight settings. However, there are still some limitations, such as the one - sidedness of data in calculating sustainability indicators. Future research could focus on improving data collection methods and further optimizing the model to make more accurate predictions for Olympic SDE selection.

References

S. Hu, L. Wang, Y. Chen, and et al. Sleep disturbances, mental health symptoms, and chronotype in chinese elite athletes: insights from the beijing 2022 winter olympics preparatory period. *BMC Psychiatry*, 25(1):400, 2025. doi: 10.1186/s12888-025-06839-8.

Z. Huang, C. Yang, X. Zhou, and et al. An improved topsis-based multi-criteria decision-making approach for evaluating the working condition of the aluminum reduction cell. *Engineering Applications of Artificial Intelligence*, 117:105599, 2023. doi: 10.1016/j.engappai.2022.105599.

T. Kokolakakis and F. Lera-Lopez. Sport promotion through sport mega-events. an analysis for types of olympic sports in london 2012. *International Journal of Environmental Research and Public Health*, 17(17):6193, 2020. doi: 10.3390/ijerph17176193.

P. Shi and A. Bairner. Sustainable development of olympic sport participation legacy: A scoping review based on the pager framework. *Sustainability*, 14(13):8056, 2022. doi: 10.3390/su14138056.

G. Slavic, P. Zontone, L. Marcenaro, and et al. Vehicle localization in an explainable dynamic bayesian network framework for self-aware agents. *Information Fusion*, 122:103136, 2025. doi: 10.1016/j.inffus.2025.103136.

X. Wei, M. Wang, Q. Wei, and et al. Piping material selection in water distribution network using an improved decision support system. *Water*, 17(3):342, 2025. doi: 10.3390/w17030342.

F. Zhang, X. Pu, X. Huang, and et al. Risk assessment of flammable liquid transportation on waterways: An ontology-driven dynamic bayesian network approach. *Journal of Loss Prevention in the Process Industries*, 93:105504, 2025. doi: 10.1016/j.jlp.2024.105504.