Journal of Intelligent Computing

DLINE JOURNALS

Print ISSN: 0976-9005
Online ISSN: 0976-9013

JIC 2025: 16 (1)
https://doi.org/10.6025/jic/2025/16/1/28-35

The Combination of Red Tourism Policy Tools Based on K-Means Clustering Algorithm

Gang Xu
Foreign Languages and Tourism Department
Wuxi Institute of Technology, Wuxi. China
tigerxg2023@163.com

ABSTRACT

Red tourism is a form of tourism with red tourist attractions and related historical culture as its theme. It is significant for promoting historical and cultural preservation and economic development of tourism. The configuration of the government's red tourism policy tool combination is crucial for developing red tourism. This study explores how to effectively configure the combination of red tourism policy tools to achieve sustainable development of red tourism based on the K-means clustering algorithm. Firstly, red tourist attractions are analyzed and clustered into different categories. Then, combining relevant historical and cultural data and tourism economic data, the K-means clustering algorithm is applied to optimize the configuration of red tourism policy tools. Through systematic research, we found that the configuration of red tourism policy tools using the K-means clustering algorithm can effectively meet the needs of different categories of red tourist attractions, thus promoting the sustainable development of red tourism. This research provides a valuable reference for the government and relevant tourism agencies in configuring red tourism policy tools.

Keywords: K-Means Clustering Algorithm, Red Tourism, Policy Tools

Received: 22 September 2024, Revised 4 December 2024, Accepted 13 December 2024

1. Introduction

Red tourism is a form of tourism with red tourist attractions and related historical culture as its theme, which is significant for promoting historical and cultural preservation and tourism economic development [1]. With the rapid rise of red cultural tourism in China, the government's efforts to promote it have increased in recent years. They have adopted a series of practical measures to ensure its long-term and stable development. They consider them an effective sightseeing industry to meet people's demand for beautiful nature. Formulating appropriate red tourism policy measures is essential for promoting the sustainable development of China's red tourism [2].

The government needs to configure policy tools reasonably based on the characteristics and needs of red tourist attractions to achieve the sustainable development of red tourism. However, due to the diversity and complexity of red tourist attractions, the government faces specific challenges in policy tool configuration. Therefore, practical algorithms are needed to assist the government in policy tool configuration [3]. The K-means clustering algorithm is a commonly used clustering analysis algorithm that can divide a dataset into different categories, helping to understand and interpret the data [4]. In the research on the configuration of the combination of red tourism policy tools, the K-means clustering algorithm can help the government classify red tourist attractions into different categories and configure policy tools rationally according to the needs of different categories. This can help the government meet the needs of different categories of red tourist attractions more accurately and promote the development of red tourism. This study explores how to effectively configure the combination of red tourism policy tools to achieve sustainable development of red tourism based on the K-means clustering algorithm [5]. Specifically, this study will achieve its objectives through the following steps: Clustering analysis of red tourist attractions: Through clustering analysis of relevant data and features of red tourist attractions, red tourist attractions will be divided into different categories. Clustering analysis can help us understand the characteristics and needs of attractions in different categories. Optimization of red tourism policy tool configuration: Combining the characteristics and needs of red tourist attractions and relevant historical cultural and tourism economic data, the K-means clustering algorithm is applied to optimize the configuration of red tourism policy tools [6]. By considering the needs of attractions in different categories, we can better meet the requirements for the development of red tourism. Evaluation of policy tool configuration schemes: The impact of varying configuration schemes on the development of red tourism will be evaluated by comparing the effects of different configuration schemes. This can help the government choose the optimal policy tool configuration scheme and promote the sustainable development of red tourism [7]. This study will combine the actual situation and needs of red tourism development and use the K-means clustering algorithm to optimize the configuration of the combination of red tourism policy tools. By implementing this study, the government and relevant tourism agencies can better formulate and implement red tourism policies and promote the sustainable development of red tourism [8]. In addition, this study also provides theoretical and practical guidance for the configuration of red tourism policy tools, which is of great significance for promoting the role of red tourism in historical and cultural preservation and tourism economic development.

2. Related Work

Research on the Combination and Configuration of Red Tourism Policy Tools has become an essential topic in red tourism, and many research results in this area have been widely applied. This section will review some relevant works and analyze their shortcomings to provide background for the innovation points of this research. The importance of studying red tourism policy is self-evident, as it can help us better understand the development trend of the tourism industry and evaluate the promotion of various policies on the tourism industry [9]. However, due to the uniqueness and complexity of red tourism, we rarely delve into how to combine multiple tourism policies to achieve optimal results. In the research on the configuration of the combination of red tourism policy tools, the classification and evaluation of red tourist attractions are key issues. Classifying attractions based on their characteristics and attributes is important for better understanding and managing red tourism resources [10]. For example, some studies divide red tourist attractions into different categories, such as revolutionary history, red culture, and war memorials. However, these classification methods often lack objectivity and accuracy, relying solely on subjective judgment and experience, which cannot meet the needs of the government for policy tool configuration. Establishing an optimization model for red tourism policy tools is essential for their rational allocation. For instance, using multi-objective programming methods and considering multiple indicators such as economic benefits, social benefits, and environmental benefits, an optimization model for red tourism policy tools has been developed. However, these models often lack detailed analysis of the needs and characteristics of red tourist attractions and cannot provide specific policy tool configuration schemes. The *K*-means clustering algorithm has wide applications in the tourism field. For example, some studies use the *K*-means clustering algorithm to divide tourist destinations for better tourism resource management and development. Additionally, it has been applied to group tourists to better understand their needs and behaviors. However, using the *K*-means clustering algorithm in the research on configuring the combination of red tourism policy tools has not been fully exploited. In summary, there are some shortcomings in the current research on configuring the combination of red tourism policy tools. The lack of in-depth research and analysis of the configuration hinders effectively meeting the needs of red tourism development. Therefore, this study will classify and analyze red tourist attractions based on the *K*-means clustering algorithm and optimize the configuration of the combination of red tourism policy tools to achieve sustainable development of red tourism. By introducing data-driven methods, this research aims to explore the configuration of red tourism policy tools more accurately and effectively and provide a scientific basis for red tourism's sustainable development.

3. Algorithm Design Based on K-Means Clustering Algorithm

In the research on the configuration of the combination of red tourism policy tools, the design of the classification algorithm is a critical step. This section will propose the algorithm design based on the *K*-means clustering algorithm and introduce its implementation steps and advantages. Data preprocessing is required before applying the *K*-means clustering algorithm. Firstly, the characteristics and attributes of red tourist attractions, such as historical background, geographical location, number of visitors, and cultural connotations, are collected and organized. Then, data cleaning and normalization are conducted to eliminate outliers and differences in data scales. In the *K*-means clustering algorithm, the number of clusters *K* needs to be determined. The elbow or silhouette methods can be used to determine the optimal value of *K*. The Elbow Method selects the best *K* value by plotting the relationship curve between *K* values and the sum of squared errors of clustering. The Silhouette Method selects the *K* value corresponding to the most significant silhouette coefficient, which calculates the closeness and dispersion of samples in clusters. Once the K value is determined, *K* initial cluster centers need to be initialized. *K* samples can be randomly selected as the initial cluster centers, or more complex initialization methods like the *K*-means++ algorithm can be adopted. Clustering is a type of unsupervised learning, and the *K*-means clustering algorithm is the most basic and commonly used clustering algorithm. Its basic idea is to find a partition scheme of *K* clusters through iteration, minimizing the loss function corresponding to the clustering result. The loss function can be defined as the sum of squared errors of each sample's distance to its cluster center.

$$J(c, \mu) = \sum_{i=1}^{M} x_i - \mu_{c_i}{}^2 \qquad (1)$$

Where Xi represents the i-th sample, Ci is the cluster to which Xi belongs, Uci represents the center point of the cluster, and M is the total number of samples. In the *K*-means clustering algorithm, the clustering result is continuously optimized by iteratively updating the cluster centers. As the number of clusters *K* increases, the sample partition becomes finer, and the sum of squared errors (SSE) for all samples gradually decreases.

$$SSE = \sum_{i=1}^{k} \sum_{x \in D_i} \left| x - \mu_i \right|^2 \qquad (2)$$

When K is smaller than the true number of clusters, increasing K will significantly impact the clustering effect, leading to a large decrease in SSE. However, when K is larger than the true number of clusters, increasing K will not significantly impact the clustering effect, resulting in a flattening of the SSE decrease.

The better the clustering effect, the closer the samples within a class and the farther the samples between classes. This can be measured using the average silhouette coefficient. The within-cluster dissimilarity: average ai, reflects the cohesion. ai represents the average distance of sample *xi* to other samples within the same class. Smaller ai indicates lower within-class dissimilarity and higher cohesion, suggesting that the samples within the class should be clustered together. The between-cluster dissimilarity, minimum bi, reflects the separation. bi represents the average distance of sample xi to all samples in other classes. Larger bi indicates higher between-class dissimilarity and higher separation, suggesting that the samples should not be clustered together. The nearest cluster:

$$C_j = \min_{D_k} \frac{1}{n} \sum_{x \in D_k} \left| x - x_i \right|^2 \qquad (3)$$

Dk represents the nearest cluster, *x*represents all the samples in the nearest cluster, and *n* represents the total number of samples in the closest cluster. After completing the iterative update of the cluster centers, the clustering results need to be evaluated and analyzed. Various indicators can be calculated, such as the average distance within each category and the comparison of intra-cluster and inter-cluster distances, to assess the quality of the clustering results. Additionally, the clustering results can be visualized, for example, by plotting a scatter plot of the sample points and using different colors or markers to represent different clusters.

The classification algorithm based on the K-means clustering algorithm has the following advantages in the research of red tourism policy tool combination configuration:
1. Simplicity and effectiveness: The K-means clustering algorithm is a simple and commonly used clustering algorithm that is easy to implement and understand. It can provide reasonable and effective solutions to the classification problem of red tourism sites.

2. Strong scalability: The K-means clustering algorithm can handle large-scale datasets and flexibly adjust the clustering results by increasing or decreasing the number of cluster centers.

3. Strong interpretability of results: The K-means clustering algorithm assigns samples to different categories, each with a cluster center as a representative sample. Such results are easy to interpret and understand, aiding government decision-makers in configuring red tourism policy tools.

The design of the classification algorithm based on the K-means clustering algorithm is an integral part of the research on red tourism policy tool combination configuration. By preprocessing data, determining the number of clusters, initializing cluster centres, iteratively updating cluster centres, and evaluating clustering results, a reasonable and feasible classification scheme for red tourism sites can be obtained, providing a scientific basis for government decision-makers. Moreover, this algorithm has the potential to be widely applied due to its simplicity, effectiveness, scalability, and strong interpretability of results.

4. Experimental Design and Analysis

In the research of red tourism policy tool combination configuration, experimental design and analysis are crucial as they help evaluate the impact of different combination configurations on red tourism policies and provide reasonable decision recommendations. This paper will explain the experimental design and analysis based on the K-means clustering algorithm.

Before starting the experimental design, the objectives and hypotheses of the experiments need to be clarified. For example, the aim might be to determine an optimal combination configuration of red tourism policy tools that maximizes the number of visitors and revenue at red tourism sites. The hypothesis could be that different combination configurations will impact the number of visitors and revenue differently. In the experimental design, the key factors affecting the configuration of red tourism policy tools and their levels need to be determined. For example, the types, quantities, and investment of red tourism policy tools can be selected as experimental factors and set at different levels, such as three types of policy tools (A, B, C) and three quantities (1, 2, 3).

The appropriate experimental design should be selected based on the experimental objectives and hypotheses. Common types of experimental designs include utterly randomized, randomized block, and factorial designs. In the research of red tourism policy tool combination configuration, a factorial design can be adopted, treating different combinations of policy tools as experimental factors and observing the corresponding changes in the number of visitors and revenue by controlling and adjusting the levels of experimental factors.

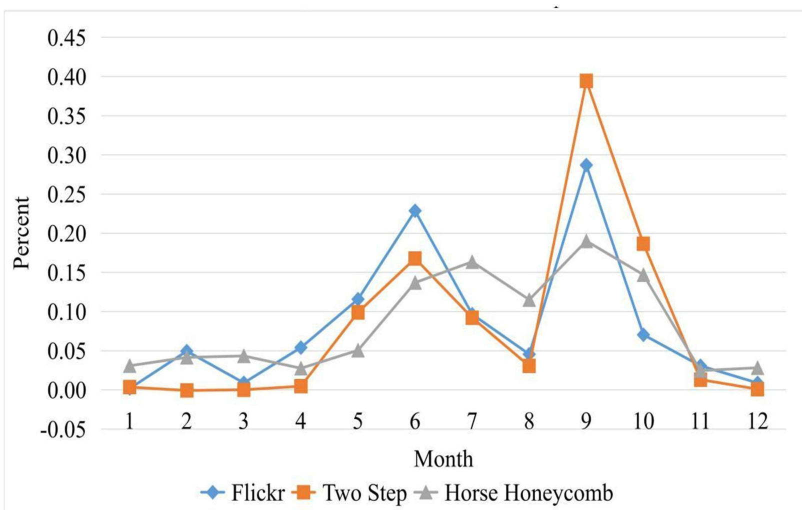

Figure 1. Proportion of Monthly Provided Photos on Three Social Media Platforms from 2007 to 2022

Based on Figure 1, the monthly analysis of collected geotagged photos shows a bimodal distribution with two peaks. September and June have the highest proportion of photos taken, accounting for 29.05% and 17.82%, respectively. The reason behind this observation is closely related to the variation in tourist hotspots within the Everest Nature Reserve and the local climate. The Everest Nature Reserve belongs to a temperate semi-arid continental plateau climate, with an average annual temperature of 2.1°C. From a temperature perspective, the weather is freezing from November to April, making it unsuitable for tourism activities. Visitors tend to choose May after the spring season in the Tibetan area until October before the onset of severe cold for tourism activities. In terms of precipitation, it is concentrated in July and August. Consequently, many tourists choose to visit

Tibet between May and June, as well as between September and October, avoiding the peak of the rainy season. During the experiment, relevant data needs to be collected and appropriately processed. For example, the number of visitors and revenue data corresponding to each combination configuration of red tourism sites can be recorded. Additionally, some control variables, such as season and weather, can be collected to eliminate interference with the experimental results.

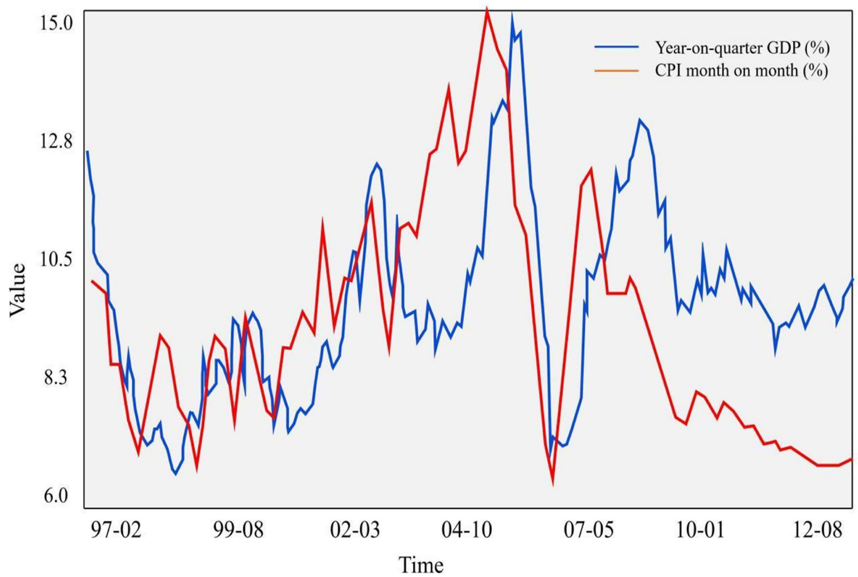

Figure 2. Year-on-Year Growth of Tourism GDP and GPI in Mount Everest Nature Reserve

Based on Figure 2, July has the highest year-on-year growth in tourism GDP and GPI in the Mount Everest Nature Reserve. In spring, the most frequently visited areas, including Everest Base Camp (Rongbuk Monastery), Everest Peak, Gyawu La Pass, and Kama Valley, are identified as high tourist hotspots with a point distribution covering 8 grids (0.09%). In summer, 13 grids are identified as high tourist hotspots (0.14%), and Everest Peak is no longer considered a hotspot area, while Kama Valley's popularity significantly increases. In autumn, high tourist hotspots are similar to the overall distribution in summer. It is worth mentioning that Kama Valley exhibits a polygonal clustering pattern, with high tourist hotspots covering 16 grids (0.18%). In winter, high tourist hotspots only occupy 3 grid areas (0.03%), with Everest Base Camp (Rongbuk Monastery) and Gyawu La Pass remaining hotspot areas. After completing the data collection, the experimental data needs to be analysed to draw conclusions and provide decision recommendations. The following analysis methods can be used for the experimental design based on the K-means clustering algorithm: Cluster analysis: Use the K-means clustering algorithm to cluster the experimental data, dividing the red tourism sites into different categories. Based on the clustering results, analyze the characteristics and attributes of various categories of sites and evaluate the effectiveness of different combination configurations. Analysis of variance: Use analysis of variance to statistically analyze the data on the number of visitors and revenue corresponding to different combination configurations, assessing the significant differences among combinations.

Data visualization

Use visual methods such as bar charts and scatter plots to intuitively display the number of visitors and revenue corresponding to different combinations.

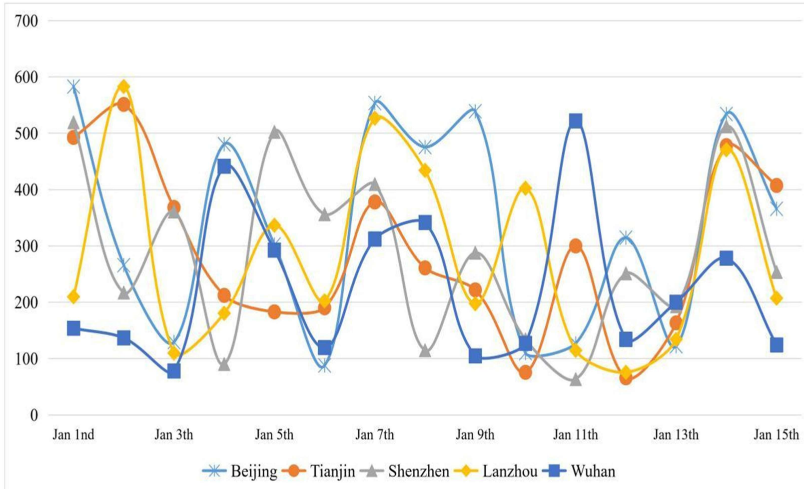

Figure 3. Red Tourism Popularity in Different Cities

Based on Figure 3, early January is the peak period for red tourism, with explosive growth in tourism popularity in many places. According to the experiment and analysis results, the experimental data is interpreted, and corresponding decision recommendations are proposed. Based on the clustering results and analysis of variance, the optimal combination of red tourism policy tools can be determined to achieve the goal of maximizing the number of visitors and revenue. The experimental design and analysis based on the K-means clustering algorithm can help us evaluate the effectiveness of different combinations of red tourism policy tools and provide decision recommendations. By clarifying the experimental objectives and hypotheses, determining the experimental factors and levels, selecting suitable experimental design types, conducting data collection and processing, using clustering analysis and analysis of variance, scientific and reliable experimental conclusions can be drawn to provide a reference for decision-making on red tourism policies.

5. Conclusion

The conclusions of the red tourism policy tool combination configuration research based on the K-means clustering algorithm are as follows: The K-means clustering algorithm can effectively divide red tourism sites into different categories. Based on the clustering results, we can observe that sites under different combination configurations have different characteristics and attributes. This helps us understand the impact of different combinations of red tourism policy tools on the sites. In selecting experimental factors, the type, quantity, and investment of policy tools are key factors influencing the combination configuration of red tourism policy tools. Different combination configurations may significantly impact the number of visitors and revenue. Through the analysis of the variance method, we found significant differences in the number of visitors and revenue corresponding to different combination configurations. Some combination configurations can significantly increase the number of visitors and revenue, while others are relatively lower. Based on the experiment and analysis results, we suggest the following strategies for the combination configuration of red tourism policy tools:Consider increasing the types and quantities of policy tools to increase the number of visitors and revenue. By introducing new red tourism policy tools, more tourists and investments can be attracted, further promoting the development of the red tourism industry.Increase investment in key red tourism sites to improve their facilities and services, increase tourist satisfaction and visitation frequency, and increase the number of visitors and revenue. Tailor targeted policy tool combination configurations based on the characteristics and attributes of different red tourism sites. Different sites may require different combinations of policy tools, so adjustments and optimizations must be made based on specific circumstances.

References

[1] Huang, J. (2022). Design of Tourism Data Clustering Analysis Model Based on K-Means Clustering Algorithm. International Conference on Multi-modal Information Analytics. Cham: *Springer International Publishing, : 373-380.*

[2] Dou, S., Yao, Z., Shi, X., et al. (2021). Research on location and path planning of distribution centre based on improved K-means clustering algorithm and improved ant colony algorithm. LISS 2020: Proceedings of the 10th International Conference on Logistics, Informatics and Service Sciences. *Springer Singapore 797-809.*

[3] Zhang, L., Shu, Y. (2021). Research on K-Means Clustering Algorithm Based on Improved Genetic Algorithm. Big Data and Security: Second International Conference, ICBDS 2020, Singapore, Singapore, December 20–22, 2020, Revised Selected Papers 2. Springer Singapore 561-571.

[4] Garde, A H., Hansen, J., Hansen, S M., et al. (2021).O-269 Self-rostering and sickness absence – a Danish cohort study on payroll data.*Occupational and Environmental Medicine,78(Suppl 1):A28-A28.*

[5] Adhikary, R., Das, A. (2022). Atomistic Pictures of Self-Assembled Helical Peptide Nanofibers.*The journal of physical chemistry.B,126(46):9476-9492.*

[6] Liu, Y., Chen, W., Chen, Y., et al. (2021). Ocean front reconstruction method based on k-means algorithm iterative hierarchical clustering sound speed profile. *Journal of Marine Science and Engineering 9(11): 1233.*

[7] Freddy Varona, Domínguez. (2023). Paulo Freire's educational ideas. Reflections from higher education.*Region - Educational Research and Reviews 5(2):13-23.*

[8] Gao, M., Yiu, K F C. (2023). On the simultaneous design of broadband beamformer filters and configuration. *Numerical Algebra, Control and Optimization 13(2):328-339.*

[9] Levesque, W., André Bégin-Drolet, Julien Lépine. (2023). Effects of Pavement Characteristics on Rolling Resistance of Heavy Vehicles: A Literature Review. *Transportation Research Record, 2677(6):296-309.*

[10] Wei, K. (2021). Simulation research of nodes target tracking algorithm based on clustering in wireless sensor network.*International journal of internet protocol technology (2):14.*