

Research on the Optimization of Cultural Tourism Resource Allocation and Tourist Flow by Integrating Reinforcement Learning

Meili Huo

School of Culture and Tourism, Suzhou University, Suzhou 234000, Anhui, China

Corresponding author: Meili Huo (e-mail: sunny3266@163.com)

ABSTRACT Efficient allocation of cultural tourism resources and balanced tourist-flow distribution are essential for improving regional tourism sustainability and service quality. To address the dynamic and multi-objective characteristics of resource allocation and tourist guidance, this study proposes an intelligent optimization framework integrating deep reinforcement learning. A dynamic environment model incorporating resource attractiveness, capacity constraints, and tourist behavioral prediction is first established. An improved Deep Q-Network combined with a multi-agent collaborative mechanism is then employed to optimize resource supply, traffic regulation, and route planning simultaneously. Furthermore, a reward function considering tourist satisfaction, regional balance, and resource utilization is designed to support adaptive decision-making. Experimental results based on five major tourist cities indicate that resource utilization increases from 68.7% to 93.3%, peak tourist concentration decreases to 40.9%, and overall satisfaction improves to 0.84. The proposed framework provides an effective solution for intelligent tourism management and offers methodological references for dynamic decision optimization, information propagation analysis, and intelligent service networks.

INDEX TERMS reinforcement learning, tourist-flow optimization, resource allocation, information propagation, intelligent service networks

I. INTRODUCTION

THE cultural tourism business occupies a significant niche in regional economic arrangements; however, the ineffective allocation of resources and uneven spatial distribution, including the underutilization of historical industrial zones, have created an imbalance in tourist influx that limits the sustainable development of the tourism sector. Popular scenic attractions experience the strain of negative influence, and the resources in the least demanded regions are not used, which reduces the corresponding level of tourist flow, infrastructure, and service networks, and it is challenging to enhance the overall gains and tourist experience [1,2]. Conventional approaches to optimization require only constant parameter values and fixed models to solve these problems which fail to address the multi-temporal and spatial dynamic variations and non-linear feedback of the tourist behaviour. As the level of data collection and intelligent computing technologies increases, the creation of a decision-making framework capable of responding to shifts in the environment in real-time and adapting to changes in strategies has become one of the primary directions towards enhancing the operational efficiency of the cultural tourism system.

In order to handle the dynamic complexity and the distributed decision-making nature of the cultural tourism systems, the implementation of intelligent reinforcement learning techniques can transform the resources allocation and tourist directions to the proactive optimization and regulation. Such approach provides the collaborative decision-making in resource supply, traffic diversion, and predictive tourist behavior by the way of continuous interactive learning and thus, increases the flexibility and stability of the system [3,4]. This paper presents a joint optimization problem of cultural tourism resource allocation and tourist diversion, integrating the spatial revitalization of industrial heritage to study methods for balancing resources and traffic while achieving synergistic satisfaction under multi-objective deep reinforcement learning constraints.

II. RELATED WORK

Optimization of cultural tourism resources, diversion of tourists, has been the subject of research attention in the academic and industrial spheres, the main focus being on three domains, namely, the need to develop better models of resource allocation, forecast tourist behavior, and the

development of intelligent control instruments. The initial spatial and temporal balance of resources was reached by relying on the initial resource allocation methods using linear programming and multi-objective optimization in resource allocation research. Nevertheless, all these approaches are not enough to react to the rapid influx of tourists and active behavior. Others have also proposed the use of metaheuristic algorithms like genetic algorithms, particle swarm optimization and ant colony optimization to enhance the efficiency of model search, although they are also characterized by slow convergence and high dependency on parameters when applied to large-scale complex problems. Tourist diversion studies have progressively changed their focus towards behavioral modeling as opposed to the empirical statistics. Discrete choice theory and stochastic utility model approaches can explain the preference of tourists in time, route, and destination selection, although these approaches have limitations in explaining the interaction nature of time and real-time interventions. As the technologies of big data and artificial intelligence have advanced, the machine learning techniques have come to the cultural tourism industry where the decision trees, support vector machines, and neural network models can be used to predict the flow of tourists and estimate the demand on resources, enhancing the adaptability of the models. Most of the methods, however, cannot go further than prediction or static optimization, and do not have global feedback and dynamic decision-making.

Reinforcement learning (RL) has a good performance in recent years in the traffic scheduling, urban computing and resource allocation scenarios. RA can also attain self-learning and policy evolution in a continuous feedback system through state-action interaction. In comparison with the classical algorithms, RL possesses an adaptive decision-making framework, which can flexibly adjust the strategies of resource allocation in complicated situations. Nonetheless, the current studies on its implementation in cultural tourism systems are still few and far between, which do not comprise of a systematic research incorporating the aspect of resource scheduling, traffic control and the guidance of the tourists. Thus, the discussion of the problem of the implementation of reinforcement learning in the optimization of cultural tourism resources presents great theoretical and practical importance.

III. METHODS

A. CONSTRUCTION OF DYNAMIC ENVIRONMENT MODEL

The fundamental data of the model include monitoring of the flow of visitors in real-time, records of the flow of traffic, log of transactions in the ticketing system, and the presence of interest in social media. Multi-source data fusion is a representation grounded on updates of the state of the environment. The system evaluates the resource carrying capacity, visitor density, and accessibility of every tourist node synchronously within each observation period using

the scenic area attraction index as a dynamic variable in the decision-making input. The environmental state is updated by the event-driven mechanism, and when the inflow or load of resources reaches the safety level, the model independently changes the resource allocation and guidance policies [5,6]. The visitor behavior modeling section presents the simulated decision-making by sequence preferences, which predicts the likelihood of destination change according to historical routes and the current place, and rectifies the response patterns of a visitor according to a feedback of the situation. Policy restrictions and emergency response modules are also a dynamic environment using the signals of traffic control, temporary openings, and price adjustment to create the real-time interventions, to maintain the stability of scheduling and strategy sustainability in high volatility environments.

B. DEEP Q-NETWORK IMPROVEMENT STRATEGIES

The model structure adopts a dual-network collaborative architecture and sets a delayed update mechanism during training. By synchronizing parameters at fixed time intervals, the instability caused by policy oscillations is reduced. The network input comes from spatial distribution data, tourist behavior sequences and resource carrying capacity in the dynamic environment. Multi-dimensional features are extracted through the attention encoding layer, enabling the model to identify potential high-pressure areas and potential resource redundancy areas. The hidden layer introduces a multi-scale convolutional structure to capture cross-regional tourist flow migration features and resource supply and demand change patterns [7,8]. The policy output end is designed with a hierarchical action space, treating resource allocation, traffic flow scheduling and tourist guidance strategies as independent branches, and achieving multi-objective coordinated decision-making through weighted fusion. During training, the experience replay mechanism does not use completely random sampling, but adopts a priority strategy, using the fluctuation range of tourist flow and the difference in resource utilization as the sampling basis to strengthen the learning intensity of key states. The reward function consists of three-dimensional indicators of tourist satisfaction, spatial balance and resource utilization rate. After normalization, the weight coefficients are dynamically adjusted to ensure that the model maintains stable policy convergence and regional adaptability in long-term learning. The parameter settings are detailed in Table 1.

This structure, through the fusion of multi-scale perception and attention, enables the deep Q-network to have the collaborative perception capabilities of tourism resource distribution, tourist flow characteristics and dynamic environmental changes, and forms a multi-dimensional intelligent decision mapping at the strategy layer.

C. REWARD FUNCTION DESIGN

Depending on the level of distribution of tourists, traffic flow and consumption of resources, the model computes

TABLE 1. Parameter Settings for Deep Q-Network Structure Design

Network layer	Type	Parameter Scale	Activation function	Output Dimension	Update frequency (steps)
Input layer	State coding layer	256×256	ReLU	1024	—
Hidden layer 1	Multi-scale convolutional layers	3×3 convolution kernel × 64	ReLU	512	—
Hidden layer 2	Feature fusion layer	512×256	ReLU	256	—
Attention layer	Region weight mapping	256×128	Softmax	128	—
Strategy layer	Action value estimation	128×64	Linear	64	—
Dual network branches	Master-target network synchronization	—	—	—	Every 500 steps
Output layer	Strategy Selection Module	64×3	Softmax	3	—

the real-time feedback signals per decision step. They include resource allocation benefit rewards, tourist satisfaction rewards, and order maintenance penalties. The allocation of the benefits is computed with respect to the capacity of the scenic area that corresponds to the level of the resources utilization; positive incentives are provided in the case when the utilization of the resources is in the range of reasonable values and penalties are imposed in the situation when they are overloaded or not used [9,10]. To gauge tourist satisfaction it is measured in real-time according to the smoothness of the journey, the waiting time and accessibility and the feedbacks are transformed to a uniform scale by a smoothing function. The order penalty element establishes negative feedback gradation of traffic congestion or high-density gathering phenomena to limit overly centered approaches.

The weight configuration in Table 2 was determined based on multiple rounds of simulation verification. When the model was trained in a high-density tourist area, the system mainly relied on the resource allocation benefits and satisfaction terms for positive guidance, while traffic and environmental penalties were used to stabilize local congestion zones, so that the overall regulation had variable adaptability and robust balance.

D. MODEL TRAINING AND IMPLEMENTATION PROCESS

The simulation environment consists of multiple virtual tourism nodes, each dynamically generating status data based on tourist flow, resource supply, and transportation accessibility. Upon training startup, the model initializes its experience pool with historical records and real-time monitoring data, generates behavioral sequences through policy exploration, and feeds the behavioral results back to the environment module. Experience samples are prioritized, filtered, and normalized before being input into the network. The main network and the target network synchronize parameters at a fixed update interval to maintain policy convergence stability.

The gradient update algorithm uses an automated learning rate control system, which automatically increases/decreases the step size in response to the policy error and reward offset to decrease the chance of entrapment in local optima. The choice of action applies a hierarchical model: The resource sub-network chooses the supply strategy, the guiding sub-network produces tourist diversion schemes

and the final layer, the fusion layer, completes global coordination instructions. At each iteration in system, the system keeps checking resource consumption ratio, tourist density balance as well as network loss variations. When the three indicators become stable, the strategy is frozen, and optimum parameters are obtained. The model training and implementation process will guarantee that the distribution of cultural tourism resources and tourist diversion strategies are optimized steadily at a long-term to adapt flexibly to a complex environment that is dynamic and volatile.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL AREA AND DATA DESCRIPTION

The experimental site has chosen 5 common cultural tourism cities in southwest China which includes Chengdu, Chongqing, Kunming, Guiyang, Nanchong and 3 types of scenic spots were covered: historical and cultural sites, mountain ecological sites, and urban integrated tourism areas. The sources of data were publicly available statistical data of municipal cultural and tourism bureaus, real time monitoring data of scenic attractions, road network traffic data of transportation departments and the social media geographical check-in data. The data were collected over one year, with a temporal resolution of 15 minutes and a spatial resolution corresponding to administrative regions and scenic-spot-level grid units. We used anonymous location tracking to process tourist behavior data, and create continuous visit sequences, which were used to implicate tourist movement and stay patterns. The resource data were scenic spot capacity, reception intensity, and facility service index and normalized in such a way that it was based on climate and holiday factors. It produced an approximate of 820,000 valid records after cleaning and interpolation repair to be used in model training, validation, and testing to make sure the data distribution is representative and timely between the various types of scenic spots.

B. RESULTS AND DATA DISPLAY

The experiment constructs a simulation environment to dynamically input operational data from scenic spots in five cities. A traditional linear programming model and an improved deep reinforcement learning model are used to perform 1000 rounds of resource allocation and tourist diversion tasks respectively for comparative testing.

TABLE 2. Reward Function Weight Parameters and Scheduling Range Settings

Reward Category	Initial weights	Dynamic adjustment range	Update step size	Average impact percentage (%)
Resource allocation efficiency	0.45	0.35–0.55	0.01	41.2
Tourist satisfaction	0.30	0.25–0.40	0.01	32.7
Traffic congestion penalties	0.15	0.10–0.20	0.005	15.8
Environmental load penalty	0.05	0.05–0.10	0.005	6.4
Spatial balance	0.05	0.05–0.15	0.002	3.9
Satisfaction Correction Factor	—	0–0.05	0.001	—
Dynamic normalization factor	—	0–1	0.01	—

TABLE 3. Comparison of results between traditional methods and improved models

Indicator Categories	Traditional linear programming	Improved model (%)	DRL	Relative improvement	Test mean difference
Resource utilization rate (%)	68.7	93.3		24.6	±1.8
Peak tourist concentration (%)	72.4	40.9		-31.5	±2.1
Overall satisfaction	0.71	0.84		18.2	±0.03
Environmental load index	0.63	0.48		-15.0	±0.02
Average configuration latency (s)	14.8	13.1		-35.4	±0.4

Table 3 shows that in the comprehensive test across five cities, the improved DRL model increased the average resource utilization rate from 68.7% to 93.3%. Peak tourist concentration decreased from 72.4% to 40.9%. Overall satisfaction rose from 0.71 to 0.84. The environmental load index decreased from 0.63 to 0.48, reflecting the system's ability to suppress energy consumption and visitor density in high-pressure scenic areas. The average configuration latency decreased from 14.8 seconds to 13.1 seconds, with a response speed improvement of approximately 1.7 seconds, indicating that the model possesses faster adaptive computation and resource scheduling capabilities in multi-node decision-making. Overall, the results demonstrate significant improvements in the model's coordination, efficiency, and sustainability.

C. MODEL VALIDITY

The three principal areas where the model can be effective include policy stability, environmental adaptability as well as task coordination. The model is capable of optimizing the distribution of resources and guiding tourists constantly in dynamic environments, which attain the control of the whole decision-making to worldwide profits. The policy network offers stability in convergence during training in response to variations in multi-dimensional inputs and balanced weight updating trends and a lack of oscillating decision-making behavior, implying that the algorithm has the self learning and steady-state maintenance properties. In front of the environmental disruptions and unexpected changes in the tourist flows, the model is capable of modifying its course of policy depending on the real-time status and sustaining the dynamic balance between the use of resources and the distribution of tourists. The fact that it can be generalized is confirmed by crossregional migration experiments, the policy can continue to develop quality scheduling behavior

rapidly in new environments without the need to make further adjustments to the parameters. Resource scheduling and tourist diversion approaches can also offset when it comes to multi-task collaborative operation, in which case, the approaches will provide a collaborative optimization, closed loop, eliminating the loss of efficiency as with the traditional approaches to the conflict of multi-objectives.

V. CONCLUSION

The reinforcement learning model for cultural tourism optimization effectively plans resource supply and traffic regulation by integrating the spatial revitalization of industrial heritage into a multi-dimensional coordination framework that dynamically manages tourist diversion. The model is more robust to endure complex environments and it is well-adapted to the environment and has strong spatial mobility, as well as, the capability to make a real-time correction and optimization of the resource allocation in the presence of uncertain surroundings. With the implementation of reinforcement learning, the system can be designed to possess the self-learning and constant improvement, which makes the decision making process of cultural tourism resource allocation an intelligent and sophisticated way of making a decision. This approach is useful in addressing the challenges of the traditional optimization models slow reaction to the dynamics of changing demand, non-uniform use of resources, and the lack of human control over tourist congestion. Nevertheless, there are a few shortcomings like the accuracy of model predictions as the real-time nature

of external environmental data is restricting, and the necessity to work on the further enhancement of synchronization of multi-agent collaborative learning in cross-regional contexts. The next round of research will employ spatiotemporal big data and multimodal perceptions to enhance widespread collaboration and long-term strategy development by integrating the spatial layout of historical textile industrial sites into the urban tourism network.

AUTHOR CONTRIBUTIONS

Meili Huo designed, collected and analyzed the data, and drafted the manuscript. Meili Huo conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Meili Huo participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in

ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by, Anhui Province Social Science Innovation and Development Research Key Project: Research on Cultural Tourism Development and Cultural Heritage Protection of the “Wanjiang Poetry Road”(2021CX135)

University-Level Key Research Project of Suzhou University: Research on the Integration and Creative Development of Chu-Han Cultural Resources in Northern Anhui from the Perspective of Holistic Tourism (2019yzd15)

Suzhou University’s Horizontal Project: Research on the Integrated Development of Cultural Industry and Ecological Agriculture in the Context of Rural Revitalization(2023xhx176)

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] M. Zhao, X. Yan, J. Zhang, et al., “An exploration of the allocation of health resources in rural tourist areas by ‘tidal’ medical assistance,” *Chinese Rural Health Service Management*, vol. 43, no. 8, pp. 586-589+608, 2023, doi: 10.19955/j.cnki.1005-5916.2023.08.009.
- [2] H. Rong, “The Dilemma and Solutions of Industry-Education Integration in Applied Undergraduate Tourism Majors from the Perspective of Resource Integration,” *Journal of Hubei Open Vocational College*, vol. 36, no. 4, pp. 24-26, 2023, doi: 10.3969/j.issn.2096-711X.2023.04.010.
- [3] G. Zhong, C. Feng, and H. Li, “Research on the Development Path of Sports-related Tourism Resources of Southern Intangible Cultural Heritage in Southeast Chongqing under the New Development Concept,” *Contemporary Sports Technology*, vol. 13, no. 30, pp. 81-84, 2023, doi: 10.16655/j.cnki.2095-2813.2023.30.022.
- [4] L. Song, D. Li, and X. Xu, “A review of inverse reinforcement learning algorithms, theories and applications,” *Acta Automatica Sinica*, vol. 50, no. 9, pp. 1704-1723, 2024.
- [5] Y. Chen, J. Huo, T. Ding, and Y. Gao, “A review of meta-reinforcement learning research,” *Journal of Software*, vol. 35, no. 4, pp. 1618-1650, 2024.
- [6] X. Wang, S. Wang, X. Liang, et al., “Deep reinforcement learning: A survey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 5064-5078, 2022, doi: 10.1109/TNNLS.2022.3207346.
- [7] T. Moerland, J. Broekens, A. Plaat, et al., “Model-based reinforcement learning: A survey,” *Foundations and Trends in Machine Learning*, vol. 16, no. 1, pp. 1-118, 2023, doi: 10.1561/22000000086.
- [8] A. Charpentier, R. Elie, and C. Remlinger, “Reinforcement learning in economics and finance,” *Computational Economics*, vol. 62, no. 1, pp. 425-462, 2023, doi: 10.1007/s10614-021-10119-4.
- [9] H. Wang, N. Liu, Y. Zhang, et al., “Deep reinforcement learning: a survey,” *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 12, pp. 1726-1744, 2020, doi: 10.1016/j.apenergy.2020.115036.
- [10] S. Gu, L. Yang, Y. Du, et al., “A review of safe reinforcement learning: Methods, theories, and applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 11216-11235, 2024, doi: 10.1109/TPAMI.2024.3457538.