ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Application of reinforcement Learning in smart Tourism Personalized Route Dynamic Planning and tourist experience optimization

Yanmin Lai[1,2], *and Yanjun Lai*[1,*]

[1] Shunde Polytechnic University, Foshan, Guangdong, 528300, China
[2] Graduate Business School, UCSI University, Kuala Lumpur 56000, Kuala Lumpur, Malaysia

SUMMARY: *Aiming at the problem that static route recommendation in smart tourism scenarios is difficult to adapt to changes in traffic, weather, passenger flow and tourist preferences, this paper constructs a personalized route dynamic programming model based on reinforcement learning. The model takes tourists' historical behavior, real-time context, spatial location and itinerary constraints as state input, and combines dynamic action space screening, tourist experience-oriented reward function and asynchronous advantage actor-critic training mechanism to realize the continuous optimization of route recommendation. Experimental results show that the HR@5 and HR@10 of A3C-RL model reach 56.8% and 82.5%, the comprehensive score of route rationality is 87.6, the success rate of dynamic event transfer is 88.0%, and the average replanning delay is 1.12 s. The results show that the proposed method can improve the accuracy of route recommendation, the personalized matching degree and the real-time response ability in unexpected situations, and provide a feasible calculation method for intelligent tourism service optimization.*

KEYWORDS: *Reinforcement learning; Smart tourism; Personalized route planning; Visitor experience optimization*

1 Introduction

With the continuous development of mobile Internet, location-based services, digital management of scenic spots and intelligent terminals, tourism route planning has gradually shifted from static scenic spot ranking to real-time context-oriented intelligent decision-making process. In the smart tourism scenario, tourists are no longer satisfied with the itinerary generated by popular attractions, fixed routes or artificial experience, but pay more attention to whether the route conforms to their personal interests, time budget, traffic state, scenic congestion, weather changes, and instant experience feedback. Traditional recommendation methods mostly rely on historical ratings, collaborative filtering or rule matching, which can complete basic recommendation in a stable environment. However, in dynamic situations such as sudden increase in passenger flow, traffic congestion, temporary closure of attractions, and changes in tourist interests, problems such as lagging route adjustment, insufficient personalization and decline in experience evaluation often occur. Therefore, building a personalized route planning model with real-time perception, continuous decision-making and adaptive optimization capabilities has become an important direction in the research of smart travel recommender systems.

*ymmj2026@163.com
https://doi.org/10.65102/is2026912
Published: 30 April, 2026

Dalla Vecchia et al. [1] introduced deep reinforcement learning into attraction sequence recommendation, emphasizing the self-learning ability of the model in the generation of sustainable tourism routes, which provided important ideas for dynamic itinerary optimization. Massimo and Ricci[2] systematically discussed the problem of effective construction of tourism recommender systems, and pointed out that user preferences, contextual information and explanatory feedback are the key factors affecting the quality of recommendations. Massimo and Ricci[3] further combined reinforcement learning with spatial proximity exploration to improve the cold start problem in the recommendation of new users and new POIs. Ghobadi et al. [4] built a comprehensive recommendation system for multi-day travel itinerary to improve the feasibility of complex itinerary combination. Pitakaso et al. [5] discussed the balance between tourist preferences and resource protection from the perspective of multi-objective sustainable tourism path design. Kolaee et al. [6] used adaptive large neighborhood search to deal with the group tourism planning problem, indicating that the traditional optimization algorithm still has strong application value in complex constrained scenes. Adamo et al. [7] constructed a multimodal tourism planning method integrating road network and walking network to make route recommendation closer to the real travel process. Ruiz-Meza and Montoya-Torres[8] pointed out through a review that the tourism itinerary design problem is gradually expanding from single-objective shortest path to multi-constraint, multi-objective and dynamic situational decision-making. Divsalar et al. [9] put forward optimization methods around green tourism route design, emphasizing the role of low-carbon constraints in tourism planning. Ruiz-Meza et al. [10] adopted the GRAP-VND algorithm to solve the fuzzy and sustainable group tourism route design problem, which provided inspiration for route search under complex constraints. The above studies have laid a methodological foundation for smart tourism recommendation, but most of the models still focus on offline planning or heuristic search, and the joint utilization of tourists' state changes, real-time event shocks and experience feedback is still insufficient.

Table 1 summarizes the related research directions. It can be seen that the existing research has covered scenic spot sequence recommendation, multi-day trip planning, sustainable path optimization and context-aware recommendation, but there is still room for further expansion in reinforcement learning driven continuous decision-making, dynamic re-planning and visitor experience closed-loop optimization.

Table 1: Comparison of related studies

Research Direction	Representative Literature	Main Method	Main Contribution	Remaining Problem
Attraction sequence recommendation	Dalla Vecchia et al. [1]	Deep reinforcement learning	Improved the adaptability of attraction sequence recommendation	Response to real-time emergencies still needs to be enhanced
Tourism recommendation system construction	Massimo and Ricci [2]	Recommendation system modeling	Emphasized the fusion of user preferences and context	Insufficient personalized feedback loop
Recommendation for new users and new attractions	Massimo and Ricci [3]	Reinforcement learning and spatial exploration	Improved cold-start recommendation performance	Insufficient dynamic continuous route planning
Multi-day tourism itinerary planning	Ghobadi et al. [4]	Integrated recommendation and optimization	Suitable for complex multi-day itinerary combinations	Limited use of real-time state changes
Sustainable tourism route design	Pitakaso et al. [5], Divsalar et al. [9]	Multi-objective optimization	Balanced preferences, cost, and sustainability	Insufficient modeling of tourist experience rewards
Group tourism route optimization	Kolaee et al. [6], Ruiz-Meza et al. [10]	Heuristic optimization algorithms	Suitable for multi-constraint route search	Difficult to form a real-time policy learning mechanism
Multimodal route planning	Adamo et al. [7]	Integration of road and walking networks	Improved route executability	Insufficient coupling with individual tourist preferences

Based on the above research basis, this paper regards smart tourism personalized route planning as a continuous decision-making problem, and introduces reinforcement learning method to construct a dynamic programming model. In the model, the state space is composed of tourists' interest portrait, current location, stay time, budget constraint, peer structure, scenic spot heat, traffic state and weather conditions, and the scenic spot selection, visit sequence adjustment, route replacement, stay time allocation and re-planning trigger are taken as the action space. The reward function is constructed by preference matching degree, route rationality, time cost, congestion avoidance, experience feedback and trip completion degree. At the policy learning level, we design a deep policy network for travel route dynamic planning, and combine experience replay with asynchronous advantage actor-critic training mechanism, so that the model can continuously revise the recommendation policy from historical trajectories and real-time feedback. This paper focuses on answering two questions: first, whether reinforcement learning can improve the accuracy and adaptability of smart travel route recommendation in dynamic environments; The second is whether the tourist experience-oriented reward design can improve the satisfaction of personalized matching and

the efficiency of route re-planning. With the above design, this paper aims to provide a computable, trainable, and scalable technical path for personalized route generation, dynamic incident response, and visitor experience optimization in smart tourism scenarios.

2 The reinforcement learning driven dynamic planning model of smart tourism personalized route

2.1 Construction of tourist state space

Tourist state space is the basis of reinforcement learning model for route dynamic decision. Its role is to transform scattered tourist preferences, real-time environment, spatial location and trip progress into computable state vectors. Smart tourism route planning is not a single scenic spot recommendation, but a continuous decision-making process that is constantly updated over time. If the state representation is too simple, the model can only select popular attractions or distance, and it is difficult to identify the influence of tourists 'interest changes, the congestion degree of scenic spots and traffic fluctuations on route experience. Therefore, this paper constructs a joint state space containing tourist preference features, real-time context features, current location features and itinerary constraints features, so that the deep reinforcement learning model can obtain a more complete description of the environment at each decision moment.

In this paper, we denote the tourist state at the t-th decision moment as S_t. The state is composed of four kinds of vectors: tourist preference vector u_t, real-time situation vector c_t, spatial location vector p_t and trip constraint vector q_t. Among them, the tourist preference vector is mainly derived from historical browsing, collection, rating, length of stay and the type of visited attractions. The system first organizes the historical behavior sequence of tourists in chronological order, and then maps the attraction number, theme category and behavior intensity into a low-dimensional dense representation through the embedding layer, and inputs the gated recurrent unit for temporal coding. Its update process can be expressed as follows.

$$u_t = \mathrm{GRU}(e_1, e_2, \ldots, e_n) \tag{1}$$

where, e_i represents the embedding vector of the ith historical tourism behavior, and u_t is used to depict the long-term preferences of tourists on different topics such as natural landscapes, cultural attractions, leisure consumption, and parent-child activities.

The real-time context vector c_t is used to describe the changes of the external environment when the route is generated, which mainly includes weather level, road congestion index, passenger flow ratio of scenic spots, holiday marks and the current time period. The weather information is coded by category, the traffic congestion index and the proportion of tourist flow in scenic spots are normalized, and the time variable is coded by period to avoid the order deviation caused by the direct input of hours. The current location vector p_t is composed of the spatial relationship between the current longitude and latitude of the tourist, the stay time of the last scenic spot and the distance from the candidate scenic spot. The itinerary constraint vector q_t records the remaining tour time, budget margin, acceptable walking distance and the number of visited attractions, which is used to constrain the model to avoid generating routes that are not feasible in time or have too high experience burden. The four types of state information are concatenated and standardized to form a joint state vector:

$$S_t = [u_t; c_t; p_t; q_t] \tag{2}$$

In order to ensure the scale consistency of features from different sources in network training, Z-score normalization is used for continuous variables and one-hot encoding or embedding representation is used for categorical variables. The standardization process is as follows:

$$x' = \frac{x - \mu}{\sigma} \tag{3}$$

where, x represents the original feature value, μ and σ are the mean and standard deviation of the feature in the training set, respectively. The above parameters are only statistically obtained from the training set and remain unchanged in the validation and test sets to reduce the impact of data leakage on the model evaluation results. The tourist state space construction process is shown in Figure 1.

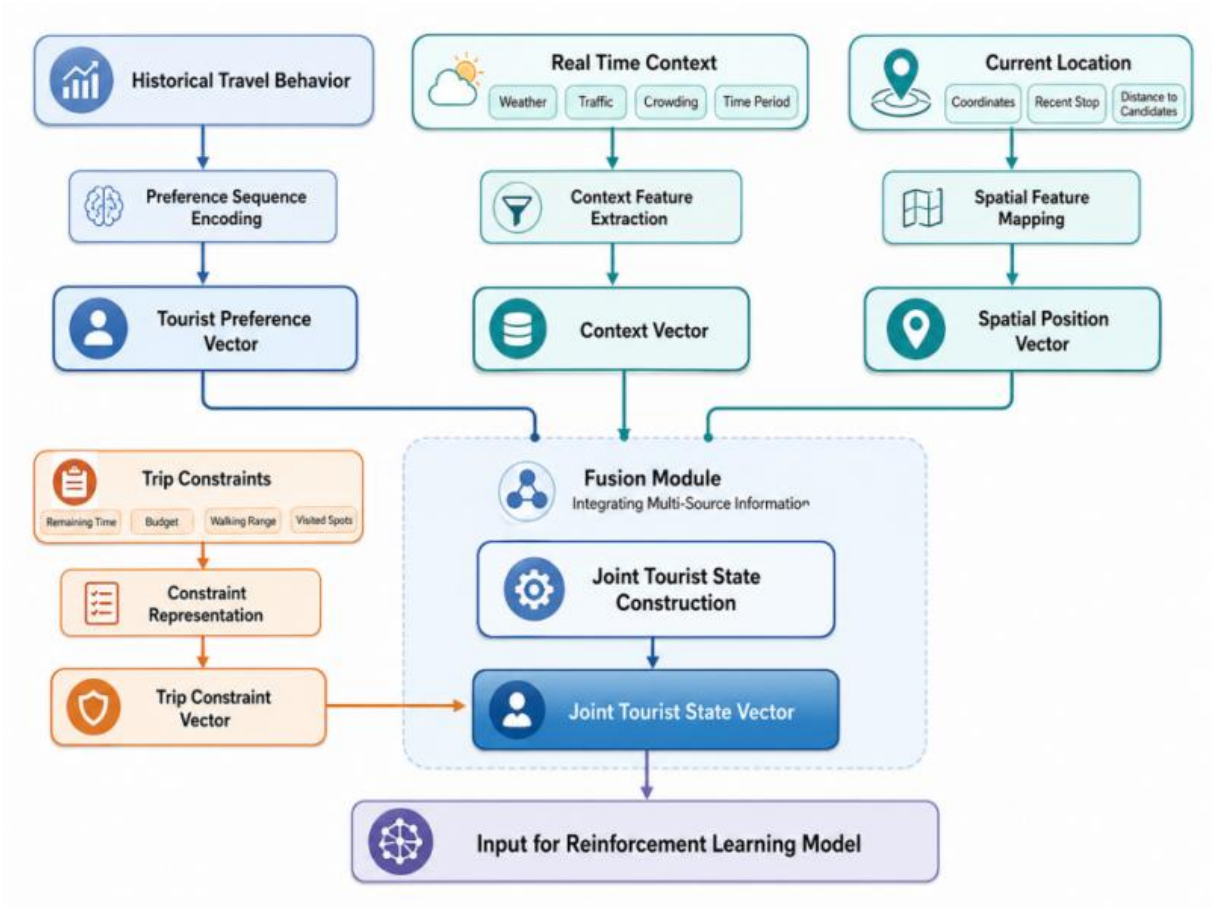

Figure 1: Tourist state space construction process

2.2 Definition of travel route action space

The action space of travel route determines the route adjustment method that the reinforcement learning model can choose at each decision moment, and also directly affects the executability and real-time adaptability of the recommendation results. Different from static travel recommendation, the action in smart tourism scenario is not equivalent to simply selecting a popular scenic spot, but to determine the next accessible node or route adjustment

scheme under the common constraints of tourists' current location, remaining time, interest preference, traffic state and scenic load. Therefore, in this paper, the action at the TTH decision moment is defined as the selection of the next visit target from the set of candidate attractions, which is denoted as follows.

$$a_t = j, \quad j \in A_t \tag{4}$$

where, a_t represents the route action output by the model in the current state, represents the number of candidate attractions, and A_t represents the set of feasible actions formed after dynamic screening. The set is not fixed, but adjusted in real time with the change of tourists' location, the update of scenic spot status and the progress of the trip.

The construction of action space mainly focuses on five conditions: space reachability, time feasibility, interest matching, budget constraint and access control. Spatial accessibility is used to exclude attractions that are too far away or have inconvenient transportation connections. According to the current longitude and latitude coordinates of tourists, combined with the road network distance and real-time traffic index, the system screens the scenic spots within a reasonable travel radius to avoid the cross-regional skip route recommended by the model. The time feasibility is judged according to the current time, the opening time of the attraction, the estimated arrival time and the suggested length of stay. If a certain attraction is close to closing or unable to complete the basic tour after the arrival of the tourists, it will not enter the current action set.

Interest matching is used to ensure that the action space still revolves around the visitor portrait. According to the historical browsing, collection, rating, stay time and theme preference of tourists, the system retains the attractions with high correlation with the current tourists' interests, and reserves the exploration space for a small number of potential interest attractions to avoid excessive concentration of recommendation results. The budget constraint mainly considers the cost of admission, transportation and consumption, and when the predicted comprehensive cost of the candidate attraction exceeds the residual budget of the tourists, the action is eliminated. Visit deduplication is used to avoid the model repeatedly recommending attractions that have been visited or rejected by tourists in a short period of time, so as to improve route continuity and user acceptance. The above screening criteria are shown in Table 2.

Table 2: Travel route action space screening conditions

Screening Dimension	Input Information	Processing Method	Effect on Action Space
Spatial Accessibility	Current location, road distance, traffic index	Calculate reachable range and filter distant attractions	Ensure natural route connection
Time Feasibility	Current time, opening hours, estimated arrival time	Determine whether visiting time conditions are satisfied	Avoid invalid recommendations
Interest Matching	Tourist profile, attraction theme, historical behavior	Calculate preference relevance	Improve personalization
Budget Constraint	Remaining budget, ticket price, transportation cost	Filter attractions exceeding the budget	Ensure plan executability
Visit Deduplication	Visited records, rejected records	Remove repeated or temporarily invalid recommendations	Reduce route redundancy

Taking the above constraints into account, the dynamic action space is expressed as follows in this paper.

$$A_t = C_{spa} \cap C_{time} \cap C_{pref} \cap C_{cost} \cap C_{novel} \quad (5)$$

where, C_{spa} represents the candidate set satisfying spatial reachability, C_{time} represents the candidate set satisfying temporal conditions, C_{pref} represents the candidate set matching tourists' interests, C_{cost} represents the candidate set satisfying budget constraints, and C_{novel} represents the candidate set after excluding repeated visits. When the candidate attraction j satisfies the following conditions, it can enter the policy network as the current action:

$$T_{arr}(j) + T_{stay}(j) \leq T_{rem} \quad (6)$$

Here, $T_{arr}(j)$ represents the estimated time to reach attraction j from the current location, $T_{stay}(j)$ represents the proposed stay time of the attraction, and T_{rem} represents the current remaining disposable time of the visitor. If the scale of the action set is too small, the system relaxes the space radius or the interest similarity threshold moderately. If A_t is empty, the trip end or return to rest point action is triggered. The model recalculates the action space after each state update, and generates a list of candidate routes according to the action values output by the policy network, so that the travel route recommendation can remain continuous, feasible and personalized in a dynamic environment.

2.3 Tourist experience-oriented reward function design

The reward function determines the optimization direction of the reinforcement learning model, and is also the key to whether the personalized travel route can take into account preference satisfaction, reasonable time, environmental adaptation and experience stability. Smart tourism route planning does not take "the largest number of arriving attractions" as a single goal, but needs to comprehensively consider whether tourists are willing to accept recommendations, whether the route is smooth, whether the state of the scenic spot is appropriate, and whether the burden of the trip is controllable. If the reward signal is too single, the model is easy to favor popular attractions or short-distance attractions, resulting in the lack of individual differences in the recommendation results. If the punishment constraint is insufficient, there may be problems such as repeated recommendation, too many detours, or recommended visits near the closure. Therefore, this paper constructs a tourist experience-oriented composite reward function, which incorporates preference matching, route efficiency, situational comfort, feedback satisfaction and ineffective behavior punishment into the immediate feedback signal. The reward function structure is shown in Figure 2.

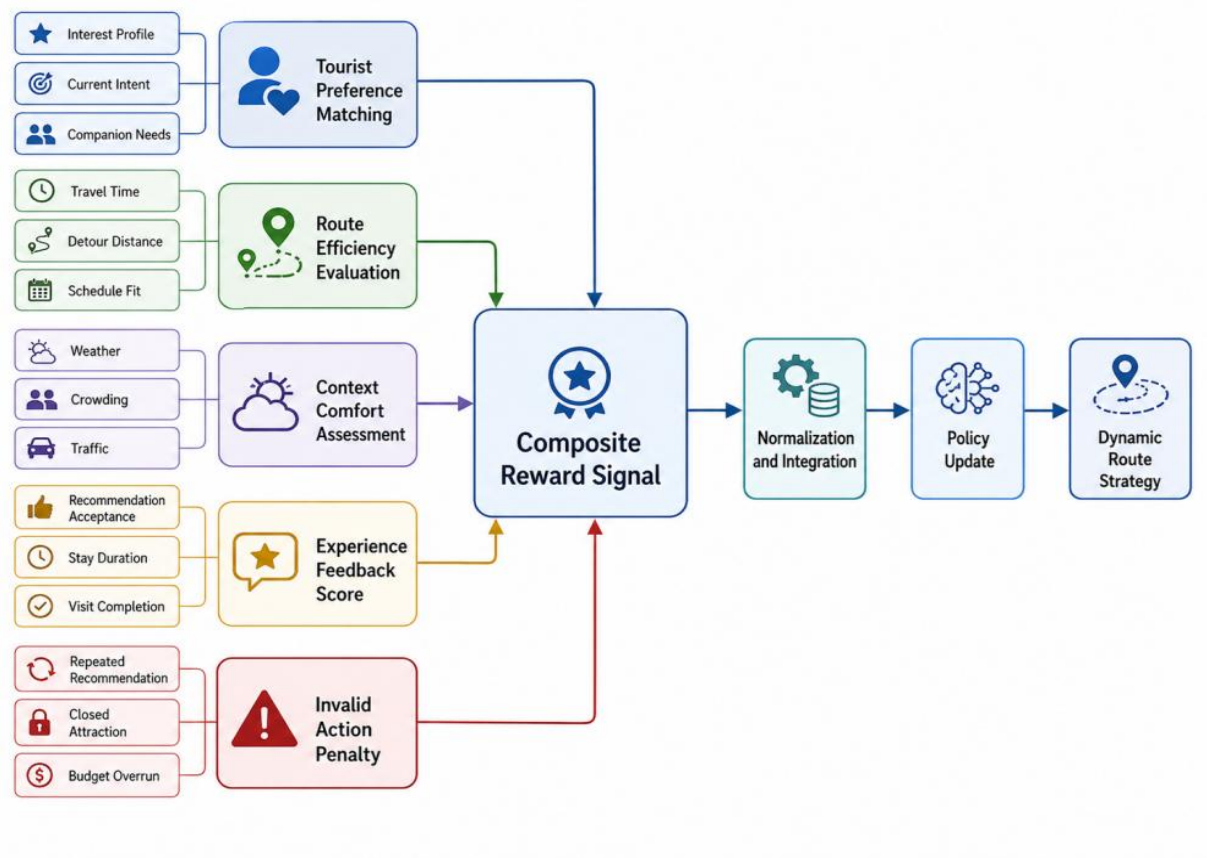

Figure 2: Structure of composite reward function for tourist experience orientation

Let the immediate reward obtained by the model after performing action a_t the TTH decision moment be R_t, which is calculated as follows.

$$R_t = \lambda_1 R_{match} + \lambda_2 R_{eff} + \lambda_3 R_{env} + \lambda_4 R_{sat} - \lambda_5 R_{invalid} \tag{7}$$

where, R_{match} represents the reward for tourists' preference matching, R_{eff} represents the reward for route efficiency, R_{env} represents the reward for environment adaptation, R_{sat} represents the reward for tourists' experience feedback, and $R_{invalid}$ represents the penalty term for invalid actions. λ_1 to λ_5 are the weights of each reward component. In this paper, we set $\lambda_1 = 0.25, \lambda_2 = 0.20, \lambda_3 = 0.22, \lambda_4 = 0.23, \lambda_5 = 0.30$ to highlight the role of experience feedback and invalid behavior constraints in dynamic route planning. The preference matching reward is used to measure the degree of agreement between the recommended attraction and the visitor interest profile. The system inputs the tourists' preference vector and the candidate attractions' topic vector into the similarity calculation module. When the recommended attractions are close to the tourists' long-term interest, current browsing intention and peer demand, the model gets higher rewards. This part can be expressed as follows.

$$R_{match} = \frac{u_t \cdot g_j}{\|u_t\| \|g_j\|} \tag{8}$$

Here, u_t represents the preference vector in the visitor state, and g_j represents the topic feature vector of candidate attraction j. Route efficiency rewards mainly characterize the impact of recommended actions on travel time and spatial coherence. When the arrival time of

the candidate attractions is short, the detour distance is small, and the subsequent trip is not compressed, the system gives positive feedback. If the route leads to significant redirection or exceeds the remaining time, the reward value is reduced. Environmental adaptation rewards are used to deal with changes in weather, passenger flow, and traffic, such as preferentially recommending indoor venues on rainy days, reducing the score of popular scenic spots during high congestion periods, and reducing jumps across areas in traffic congestion areas.

Visitor experience feedback rewards come from ratings, length of stay, adoption of recommendations and tour completion. If the tourist accepts the recommendation and completes the tour, and the length of stay is close to the recommended tour time of the scenic spot, the route arrangement has high experience value. The invalid action penalty term mainly focuses on repeated recommendation of visited attractions, recommendation of closed attractions, over-budget routes, and over-detour routes to inhibit the formation of myopic choices in the strategy network. All reward components are unified and normalized before entering the policy update, and are used as the immediate feedback signal of the reinforcement learning model to guide the policy network to generate a dynamic route that is more in line with the tourist experience in the process of maximizing the long-term cumulative revenue.

2.4 Structure design of deep policy network for route dynamic planning

In order to realize the continuous optimization of smart tourism routes in dynamic situations, this paper constructs a deep strategy network for personalized route planning. The network is composed of a tourist preference temporal encoding layer, a real-time context feature extraction layer, a joint feature fusion layer and a strategy decision output layer, which can simultaneously deal with historical behavior sequences, real-time environment variables and current location constraints. Different from the single route scoring model, the structure does not directly give a fixed scenic spot ranking, but outputs the selection probability of the executable route action according to the tourist state S_t and the action set A_t at each decision moment, so as to support the subsequent reinforcement learning training and dynamic re-planning. The overall network structure is shown in Figure 3.

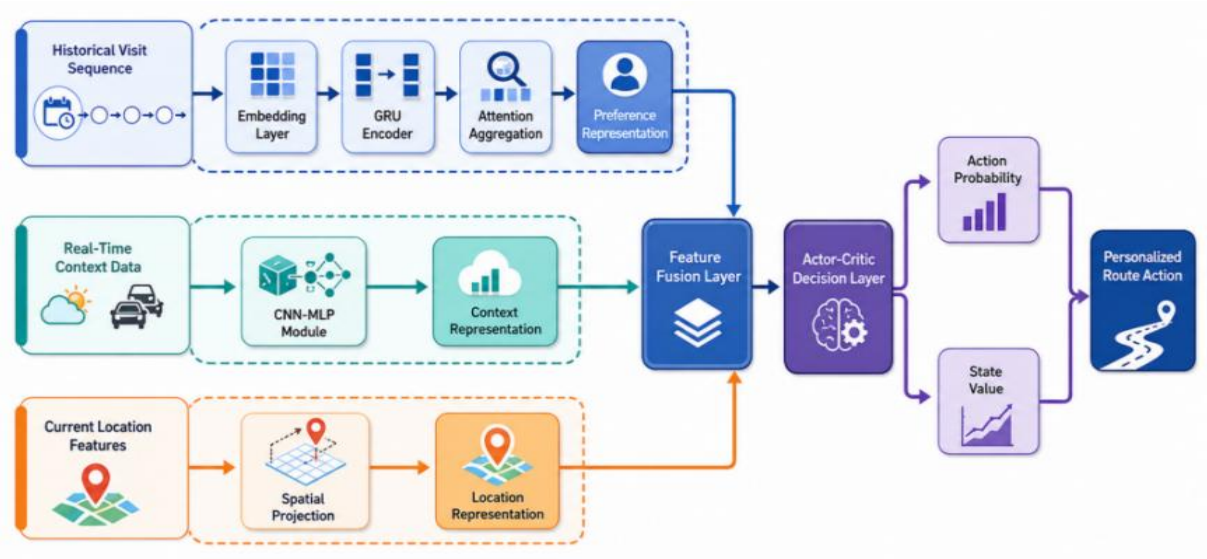

Figure 3: Deep policy network structure for route dynamic planning

The tourist preference temporal coding layer is used to capture the interest evolution law

in the historical visit behavior of tourists. The system organizes the last several views, collections, ratings and visits of tourists into time series, and converts them into low-dimensional dense vectors by embedding layer. Let the sequence of historical actions be $\{b_1, b_2, \ldots, b_n\}$, the embedded vector is fed into the GRU encoder to obtain the hidden state hi at different time steps. Considering that the correlation between recent behaviors and current route decisions is not completely consistent, this paper introduces an attention aggregation mechanism to assign differentiated weights to historical behaviors:

$$\eta_i = \frac{\exp(w^\top \tanh(W_h h_i + W_s s_t))}{\sum_{r=1}^{n} \exp(w^\top \tanh(W_h h_r + W_s s_t))} \tag{9}$$

Here, η_i represents the influence of the ith historical action on the current decision weight s_t represents the current state summary, W_h、W_s and w are learnable parameters. The visitor preference representation is obtained by weighted summation:

$$\bar{h}_t = \sum_{i=1}^{n} \eta_i h_i \tag{10}$$

The real-time context feature extraction layer mainly deals with structured variables such as weather, traffic congestion index, passenger flow ratio of scenic spots, current time period and holiday marks. In this part, the combination structure of one-dimensional convolution and multi-layer perception machine is used to extract the combination relationship of local variables, and then form the situation representation f_t through nonlinear mapping. For example, there may be complex interactions between high temperature weather and outdoor scenic spots, congested roads and cross-area mobility, holidays and popular scenic spots, which are difficult to be fully expressed by simple linear weighting. The current location feature, on the other hand, processes the longitude and latitude, the road distance and the proximity relationship of the candidate scenic spots through the spatial mapping layer to form the location representation l_t.

In the joint feature fusion layer, the network stitches three representations of tourist preference, real-time context and spatial location, and inputs them into the fully connected layer to complete the high-level abstraction:

$$z_t = \phi(W_z[\bar{h}_t; f_t; l_t] + b_z) \tag{11}$$

where z_t is the fused route decision feature and $\phi(\cdot)$ represents the ReLU activation function. The output layer of policy decision adopts the actor-critic structure. The Actor branch outputs the selection probability of each candidate action according to z_t, and the Critic branch estimates the current state value to improve the stability of policy update. The action probability distribution is defined as follows.

$$\pi_\theta(a|S_t) = \frac{\exp(g_\theta(z_t, a))}{\sum_{a' \in A_t} \exp(g_\theta(z_t, a'))} \tag{12}$$

Here, $g_\theta(z_t, a)$ represents the policy score of action a. Through the network structure, the model can establish a joint mapping relationship among tourist preferences, external environment and route constraints, so that the route recommendation is no longer a static similarity ranking, but a learnable, updable and interpretable dynamic decision-making process.

2.5 Asynchronous Advantage actor-critic Training mechanism with experience replay

In order to improve the learning efficiency and strategy stability of the smart tourism route dynamic planning model in complex environments, this paper adopts the asynchronous advantage actor-critic training mechanism based on experience replay. Tourist route recommendation has obvious continuous decision-making characteristics, and each choice of tourist attractions will affect the subsequent route, remaining time, budget consumption and experience feedback. Single-step supervised learning is difficult to depict this long-term benefit relationship. Compared with the recommendation model that only relies on static sample fitting, the asynchronous advantage actor-critic structure can update the strategy in parallel in multiple simulated travel trajectories, which makes the model obtain more full exploration ability in scenarios such as traffic fluctuations, scenic congestion, weather changes and tourist preference shifts. The training flow is shown in Figure 4.

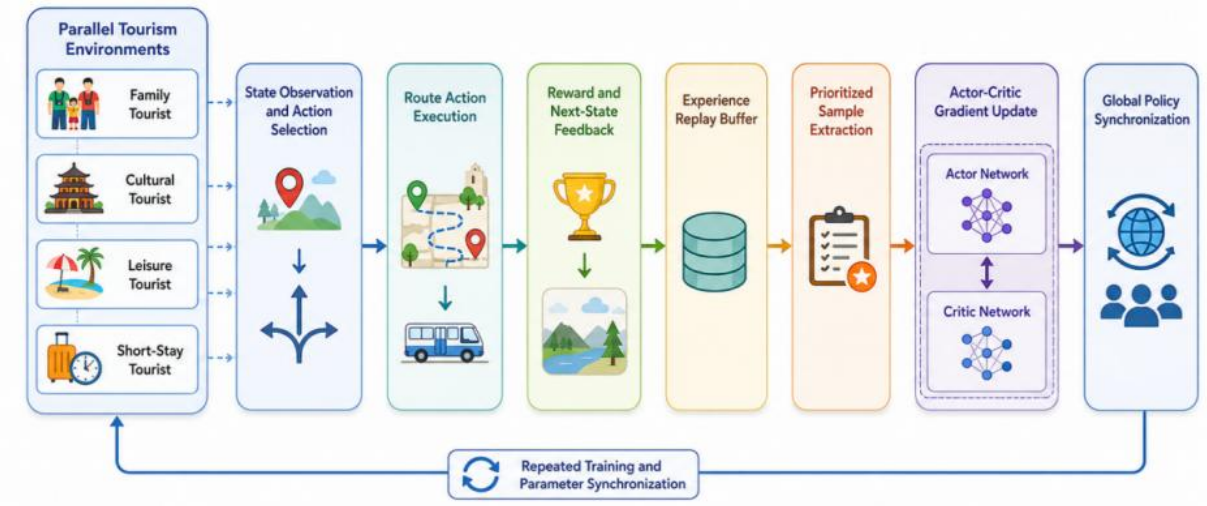

Figure 4: Asynchronous advantage actor-critic training flow based on experience replay

In the training process, the system sets multiple parallel environment threads, and each thread simulates the route planning process of a class of tourists, including parent-child tourists, cultural tourists, leisure tourists and short-term tourists. Each thread copies the parameters from the global policy network, selects the action a_t in the local environment according to the current state S_t, executes it, gets the immediate reward r_t and the next state S_{t+1}, and stores the experience samples in the local replay pool:

$$E_t = (S_t, a_t, r_t, S_{t+1}, d_t) \tag{13}$$

where, d_t denotes whether the current trajectory ends or not. d_t is set as a termination marker when the tourist completes the preset itinerary, the remaining time is insufficient, or the action space is empty. In order to avoid excessive dependence of training samples on continuous trajectories at adjacent moments, the experience replay pool will shuffle the order of samples, so that the network can learn more stable strategy rules from different tourists, different time periods and different route situations.

Considering the uneven distribution of high-value samples in smart tourism scenarios, this paper introduces a priority experience replay mechanism to give higher sampling probability to samples such as route failure, tourist rejection recommendation, sudden replanning and completion with high satisfaction. The priority of experience samples is determined by the

temporal difference error:

$$\Delta_i = |r_i + \gamma V(S_{i+1}) - V(S_i)| \tag{14}$$

where γ is the discount factor and $V(S_i)$ represents the critic network's estimate of the state value. The higher the priority, the more valuable the sample is for the current policy correction. Sampling probability is defined as follows.

$$P(i) = \frac{(\Delta_i + \varepsilon)^\alpha}{\sum_k (\Delta_k + \varepsilon)^\alpha} \tag{15}$$

where, ε is used to avoid zero probability sampling and α controls the priority sampling intensity. Through this mechanism, the model is able to focus on learning the key decision fragments under dynamic tourism events, instead of dilating the training signal by a large number of stationary samples.

The critic network is responsible for evaluating the long-term value of the current route state, and the actor network is responsible for output action selection strategies. In order to balance the immediate experience and subsequent revenue, we use multi-step return estimation:

$$G_t^{(n)} = \sum_{k=0}^{n-1} \gamma^k r_{t+k} + \gamma^n V(S_{t+n}) \tag{16}$$

The critic loss function is defined as the state value estimation error:

$$L_c = \left(G_t^{(n)} - V(S_t)\right)^2 \tag{17}$$

The actor network adjusts the action probability according to the advantage function, so that high-payoff route actions are enhanced and low-payoff actions are suppressed:

$$L_a = -\log \pi(a_t|S_t)\left(G_t^{(n)} - V(S_t)\right) - \beta H(\pi) \tag{18}$$

Among them, $H(\pi)$ is the policy entropy term, which is used to maintain a certain exploration ability and avoid premature convergence of the model to the local route pattern. Each thread uploaded the gradient to the global network after completing a number of steps of sampling, and the server side updated the shared parameters using RMSprop or Adam optimizer, and regularly synchronized the latest policy to the local thread. The training mechanism can reduce the strategy deviation caused by a single trajectory while ensuring the sample utilization, and provide a stable parameter basis for subsequent personalized tour route generation and dynamic re-planning.

2.6 Personalized travel route generation mechanism

The personalized travel route generation mechanism is used to transform the trained reinforcement learning strategy model into an online callable recommendation service, so that the system can complete the closed-loop operation of "state perception - route reasoning - route distribution - feedback update" in the actual travel process of tourists. Different from offline trip planning, route generation in smart tourism scenarios needs to continuously receive feedback from tourists' location changes, attraction congestion, traffic state and instant preference, and complete candidate attraction rearrangement and route re-planning in a

relatively short time. Therefore, this paper designs the personalized route generation mechanism as a dynamic recommendation chain oriented to real-time services, and its overall process is shown in Figure 5.

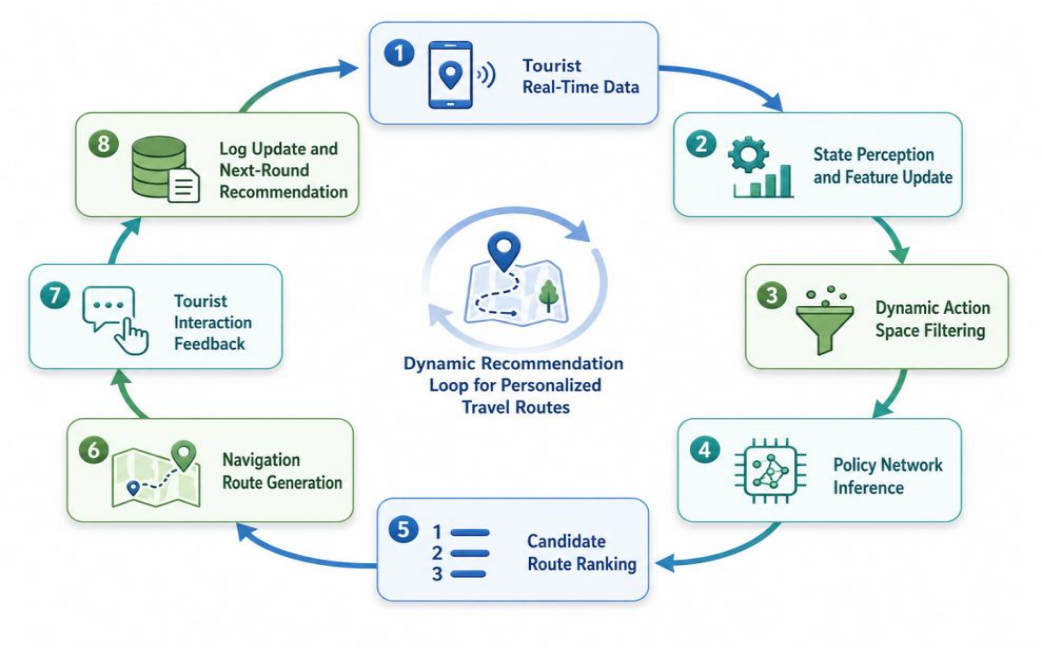

Figure 5: Closed-loop process of personalized tour route generation

In the online reasoning phase, the system constructs the state vector S_t according to the current location, historical preference, remaining time, budget constraint and real-time context data of tourists, and generates the legal action set A_t synchronously. After reading the state input, the policy network evaluates the value of the candidate scenic spots and their corresponding route actions, and outputs the optimal visit target at the current time. In order to avoid excessive concentration of recommendation results on a single scenic spot, this paper introduces a candidate route ranking mechanism in the action selection, and integrates the strategy score, route cost and experience constraint into the route generation process. Let the personalized route sequence generated by the system at time t be as follows.

$$\mathcal{P}_t = (v_t^1, v_t^2, \ldots, v_t^K) \tag{19}$$

where, $\mathcal{P}_t$ represents the list of candidate routes pushed to tourists, and v_t^K represents the KTH candidate attraction or route node. The candidate routes are not simply arranged in ascending order according to the distance, but sorted comprehensively according to the action value output by the strategy network, the matching degree of tourists' interest, the estimated travel time and the real-time state of the scenic spot. The system keeps the top K route schemes and generates walking, public transit, or mixed transit paths through the map navigation interface, while returning the estimated arrival time, proposed length of stay, and recommended reason.

In order to enhance the interpretability of the recommendation results, the client gives a brief explanation synchronously when displaying the route, such as "this scenic spot has a high matching degree with tourists 'recent cultural preferences", "the current passenger flow pressure is low, and it is suitable to arrange visits", "this route can reduce the cross-regional movement time", etc. This kind of explanation information does not change the output of the strategy, but it can improve the understanding of the recommended route and reduce the

abruptness caused by dynamic adjustment.

Visitor interaction feedback is an important source of continuous optimization of route generation mechanism. When the tourist clicks the navigation, collects the attractions, completes the check-in or the actual stay time is close to the suggested tour time, the system records the positive feedback. Weak negative feedback was recorded by the system when visitors skipped recommendations, frequently cancelled navigation, or stayed significantly short. After the feedback information enters the log queue, it is used to update the short-term preferences of tourists and subsequent offline training data. The visitor status update process can be expressed as follows.

$$S_{t+1} = F(S_t, a_t, r_t, \Delta l_t, \Delta e_t) \tag{20}$$

where $F(\cdot)$ represents the state update function, Δl_t represents the change of tourists' location, and Δe_t represents the environmental change information, including traffic, weather, passenger flow, and scenic spot opening status. The system can trigger a new round of recommendation according to a fixed time interval or a location change threshold. For example, when tourists move more than a certain distance, complete the current scenic spot visit, the tourist flow of the scenic spot suddenly increases or the traffic time consumption increases significantly, the model reconstructs the state and performs route inference.

3 Experiment and verification

3.1 Experimental Design

In order to verify the effectiveness of reinforcement learning model in smart tourism personalized route dynamic planning and tourist experience optimization, this paper constructs a simulation experiment environment including tourist behavior trajectory, scenic spot attributes, real-time context and route feedback. The experiment does not simply investigate the prediction ability of the model for the next scenic spot, but comprehensively evaluates the performance of the model from the aspects of route recommendation accuracy, route executability, personalized satisfaction and dynamic event response efficiency. The experimental platform uses Ubuntu 22.04 operating system, the core program is implemented based on Python 3.11 and PyTorch 2.2, and the database uses PostgreSQL 14 and PostGIS extension to store the spatial information of attractions. Real-time caching and log queues are supported by Redis 7.0. The hardware environment is Intel Core i7-12700 processor, 32 GB memory, and NVIDIA RTX 4080 16 GB GPU.

The experimental data consists of three parts: historical behavior data of tourists, basic attribute data of scenic spots and dynamic context data. The visitor behavior data includes browsing, collection, rating, check-in, length of stay and route adoption records. A total of 526 valid tourists, 38,740 visit records and 1,286 candidate points of interest (POI) were collated. Attraction attribute data contains attraction category, opening hours, suggested length of stay, admission price, geographic coordinates, and maximum capacity. The dynamic contextual data includes weather type, road congestion index, real-time passenger flow proportion of scenic spots and holiday marks, in which the weather information is updated hourly, and the traffic and passenger flow information is updated at 10 min intervals. In order to ensure the continuity of the trajectory, this paper deleted the abnormal records with stay time less than 5 min, and eliminated the tourist samples with visit times less than 6 times. Subsequently, the tourist route sequence is reconstructed in chronological order, so that each trajectory can reflect the preference change and spatial transfer relationship in the real travel

process.

The dataset is divided into training, validation, and test sets in chronological order with a ratio of 8:1:1. The first 80% of trajectories are used for policy network training, the middle 10% for parameter adjustment and early stop judgment, and the last 10% for model generalization testing. This division method avoids the future behavior information entering the training stage, which is more in line with the actual operation logic of smart tourism online recommendation. PageRank, GRU4Rec, DQN and the proposed A3C-RL model are selected as comparison models. Among them, Popularity generates recommendations according to the popularity of scenic spots. PageRank calculates the importance of nodes based on the tourist-attraction interaction graph. GRU4Rec predicts the next scenic spot through sequence modeling. DQN adopts the same state and action space, but does not use asynchronous actor-critic structure. The model in this paper is run under the same training set, the same candidate attraction library and the same evaluation index to ensure that the experimental results are comparable. The main experimental parameters are shown in Table 3.

Table 3: Experimental environment and key parameter Settings

Parameter Category	Parameter Name	Setting Value
Data Scale	Number of valid tourists	526
Data Scale	Visit behavior records	38,740 records
Data Scale	Number of candidate points of interest	1,286
Data Split	Training/validation/test set	80%/10%/10%
Training Parameter	Learning rate	3×10^{-4}
Training Parameter	Discount factor	0.90
Training Parameter	Number of parallel environment threads	8
Training Parameter	Batch size	32
Training Parameter	Early stopping patience	10
Dynamic Update	Traffic and tourist-flow refresh interval	10 min
Dynamic Update	Route replanning trigger interval	30 s

During training, all models progressively generate route recommendations on the same test trajectory. When the tourist accepts the recommendation and completes the check-in, the system records a valid hit. When the recommended attraction is closed, exceeds the remaining time, or is continuously rejected by visitors, it is recorded as an invalid recommendation. The model was trained by early stopping mechanism. If the validation set loss did not decrease for 10 consecutive rounds, the training was stopped and the optimal parameters were retained. Through the above experimental design, the dynamic programming ability of the model can be tested in an environment closer to the real smart tourism service, which provides a stable data basis for the subsequent analysis of recommendation accuracy, route rationality, satisfaction and replanning delay.

3.2 Comparison of route recommendation accuracy

In order to quantify the accuracy of the model in smart tourism personalized route recommendation, the experiment uses Top-K hit rate as the core evaluation index, and calculates HR@5 and HR@10 respectively. This index is used to measure whether the top K candidate attractions generated by the model contain the real next visit attraction of tourists, and it can reflect the consistency between the recommended list and the actual route choice of tourists. The experiments are conducted on the test set constructed in Section 3.1, and all models use the same initial state of tourists, set of candidate attractions, and dynamic context

data to ensure fairness in comparison of results.

In the specific implementation process, the system gradually slides the evaluation according to the real trajectory of tourists. For each tourist in the test set, the model generates a Top-K recommendation list according to the current state, and matches it with the actual check-in attractions of the tourist in the next step. If the actual visited attraction appears in the recommendation list, it is counted as a hit. HR@K is computed as follows:

$$\mathrm{HR@K} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{I}\left(\mathrm{g_i} \in \mathrm{Rec_i^K}\right) \tag{21}$$

where, N represents the number of effective evaluation steps, $\mathrm{g_i}$ represents the actual scenic spots visited by tourists in the ith decision step, $\mathrm{Rec_i^K}$ represents the set of top K recommended scenic spots generated by the model, and $\mathrm{I}(\cdot)$ is the indicator function. It takes the value 1 if the condition is true and 0 otherwise.

In this paper, the proposed A3C-RL model is compared with DQN, GRU4Rec and PageRank models. DQN uses the same state space and action space, but lacks asynchronous advantage actor-critic training mechanism. GRU4Rec mainly relies on tourists' historical visit sequence to predict the next attraction. PageRank generates recommendations based on the importance of nodes in the tourist-attraction interaction graph, which is easy to bias towards the global popular attractions. The route recommendation accuracy results of different models are shown in Table 4.

Table 4: Comparison of route recommendation accuracy of different models

Model	HR@5 (%)	HR@10 (%)	Mean Rank
A3C-RL Model	56.8	82.5	3.21
DQN	44.2	70.4	4.37
GRU4Rec	40.6	66.8	4.82
PageRank	35.1	58.7	5.76

Table 4 shows that the A3C-RL model achieves the highest results on both HR@5 and HR@10. Among them, HR@5 reached 56.8%, which was 12.6, 16.2 and 21.7 percentage points higher than DQN, GRU4Rec and PageRank, respectively. HR@10 reaches 82.5%, which is 12.1, 15.7 and 23.8 percentage points higher than the three baseline models, respectively. The results show that the proposed model can not only accurately capture the next visit intention of tourists in a short recommendation list, but also maintain a high hit level after expanding the candidate range. The Mean Rank results further show that the A3C-RL model can rank the real visited attractions in the top position, and the effectiveness of the recommendation list is stronger.

In order to avoid that the accuracy improvement only comes from the repeated recommendation of popular attractions, this paper further investigates the coverage of candidate attractions of different models. The coverage rate represents the proportion of the number of different scenic spots recommended by the model to the total number of candidate scenic spots, which is used to measure whether the recommendation system has good exploration ability and diversified recommendation ability. The coverage results of the different models are shown in Figure 6.

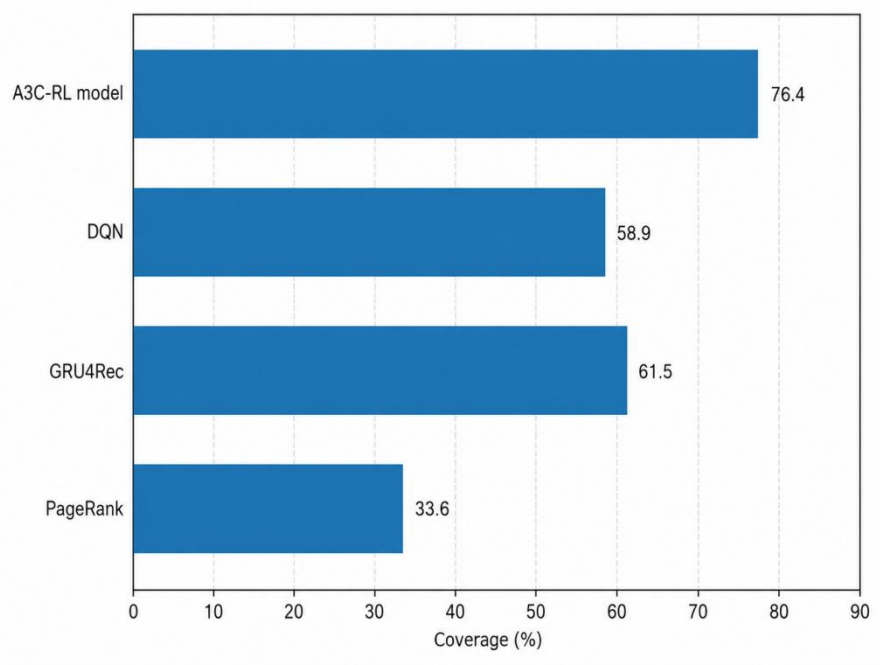

Figure 6: Comparison of coverage of candidate scenic spots for different models

Figure 6 shows that the coverage of the A3C-RL model reaches 76.4%, which is higher than 58.9% of DQN, 61.5% of GRU4Rec, and 33.6% of PageRank. This result shows that the proposed model does not rely solely on a few high-popularity scenic spots to achieve a high hit rate, but can combine tourists' interests, real-time environment and route constraints to generate more differentiated candidate routes. GRU4Rec can use historical sequence information, but it is insufficient to respond to real-time traffic, passenger flow changes and remaining time constraints. DQN has a certain exploration ability, but the strategy fluctuation is obvious in the dynamic action space. PageRank is greatly affected by the global visit popularity, and the recommendation results are concentrated on a few core attractions. On the whole, the proposed model achieves a good balance between the accuracy of route recommendation and the coverage breadth of recommendation, which provides a stable foundation for the subsequent route rationality and tourist satisfaction evaluation.

3.3 Tourism route rationality assessment

In order to further test whether the recommended route has spatial coherence and time enforceability, this paper evaluates the rationality of the route from three aspects: average travel time, cross-area jump rate and the proportion of effective tour time. The average travel time is used to measure the traffic burden between adjacent recommended attractions. The cross-region jump rate is used to determine whether there is unnecessary large-scale spatial transfer of the route. The proportion of effective tour time reflects the proportion of the total travel time that tourists spend on actual visiting, experiencing and staying. Different from focusing solely on the recommendation hit rate, route rationality evaluation emphasizes whether the recommendation results can actually be transformed into executable travel itinerary.

The experiment calculates the estimated travel time between adjacent attractions based on the road network distance and real-time traffic index, and compares the total time of the recommended route with the remaining disposable time of tourists. If the recommendation results frequently jump across regions, the detour distance is too long or exceeds the time budget, it means that although the model may hit tourists' interests, it is not suitable for real

travel. In order to uniformly compare the comprehensive performance of different models, this paper constructs the comprehensive score of route rationality after normalizing the three types of indicators, and the results are shown in Figure 7.

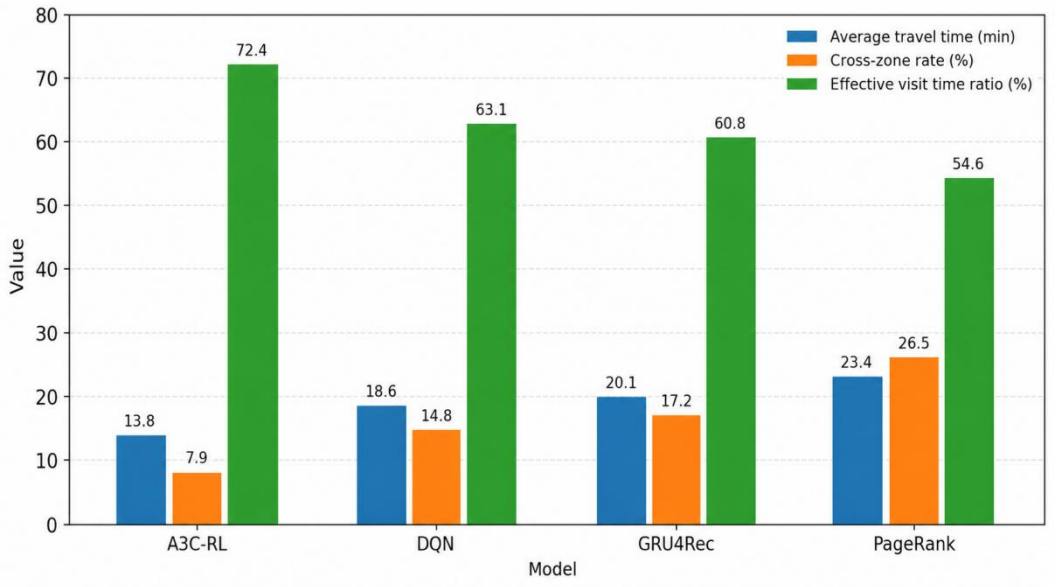

Figure 7: Comprehensive score of route rationality for different models

The experimental results show that the average travel time of A3C-RL model is 13.8 min, which is lower than that of DQN (18.6 min), GRU4Rec (20.1 min) and PageRank (23.4 min), indicating that the proposed model makes good use of the current location, traffic state and action space constraints in the recommendation process. In terms of the jump rate across slices, the A3C-RL model is 7.9%, which is significantly lower than that of DQN (14.8%), GRU4Rec (17.2%) and PageRank (26.5%). This result shows that the proposed model is able to reduce the cross-region skip recommendation and make the route more spatially continuous.

In terms of the proportion of effective tour time, A3C-RL model reaches 72.4%, which is higher than DQN (63.1%), GRU4Rec (60.8%) and PageRank (54.6%). The PageRank model is mainly based on the global popularity of attractions, which is easy to include popular but distant attractions into the recommendation list, thereby increasing the traffic time. GRU4Rec is able to exploit historical access sequences, but it is insufficient to handle real-time traffic and remaining time constraints. DQN has dynamic decision-making ability, but it is weaker than the proposed model in terms of value estimation stability and route continuous control. It can be seen from Figure 7 that the comprehensive score of route rationality of A3C-RL model reaches 87.6, which is higher than that of DQN, GRU4Rec and PageRank respectively, indicating that it can further improve the geographical coherence, time utilization efficiency and actual execution value of routes on the basis of recommendation accuracy.

3.4 Satisfaction evaluation of personalized matching

In order to evaluate the matching degree between the recommended routes and the individual preferences of tourists, this paper further carries out personalized satisfaction evaluation. Different from route recommendation accuracy, satisfaction evaluation not only focuses on whether tourists arrive at a certain scenic spot, but also pays attention to whether the recommended route meets tourists' interest type, tour rhythm, and situational needs. The experiment was scored on a 5-point scale, where 1 indicates a serious discrepancy between

the recommended route and tourist preferences, and 5 indicates a high degree of agreement between the recommended route and tourist preferences. The evaluation object is the personalized route plan generated in the test set. The evaluators are composed of tourism management researchers, smart tourism system developers and tourists with multiple independent travel experience, a total of 30 people, and the routes generated by different models are blindly evaluated in an anonymous manner.

The evaluation dimension includes five aspects: interest type matching, tour rhythm adaptation, preference intensity response, dynamic interest tracking and recommendation logic interpretability. Interest type matching is used to determine whether the route is consistent with tourists 'preferences for natural landscape, cultural history, leisure consumption or parent-child activities. Tour rhythm adaptation focuses on whether the route arrangement is consistent with tourists 'previous stay habits and physical consumption characteristics. The preference strength response is used to measure whether the model can prioritize the types of scenic spots that tourists frequently pay attention to. Dynamic interest tracking investigates the ability of the model to adjust the recommended direction under weather changes, traffic congestion and fluctuations in tourist flow. The interpretability of recommendation logic evaluates whether the system can provide clear recommendation basis. The satisfaction score results of different models are shown in Figure 8.

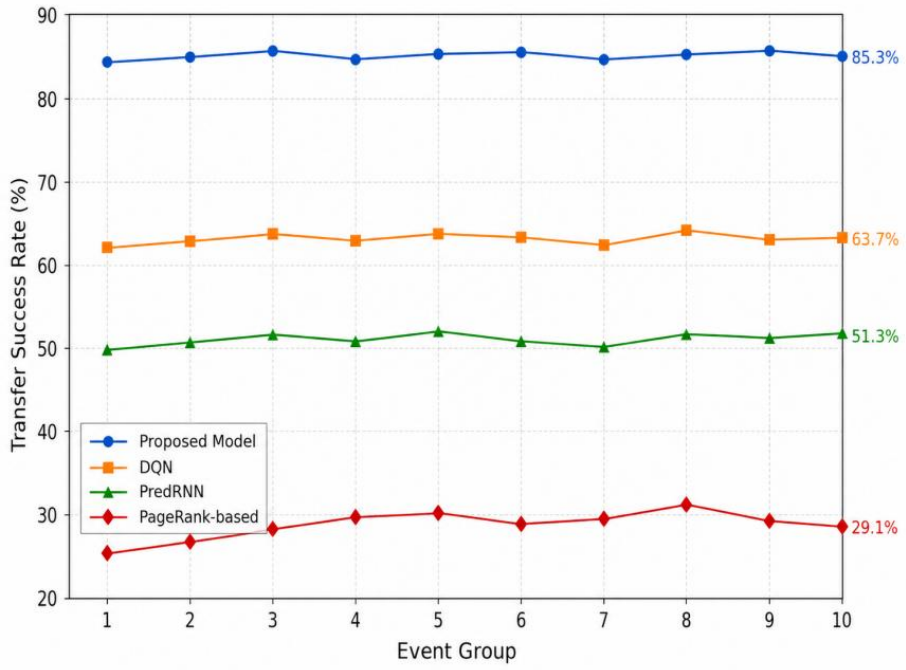

Figure 8: Personalized matching satisfaction evaluation of different models

Figure 8 shows that the A3C-RL model achieves high scores in all five dimensions. Among them, the scores of interest type matching, tour rhythm adaptation, preference intensity response, dynamic interest tracking and recommendation logic interpretability are 4.62, 4.51, 4.58, 4.47 and 4.36, respectively, and the average satisfaction is 4.51. The average score of the five items of the DQN model is 3.76, indicating that although it has a certain dynamic decision-making ability, it is not stable enough to integrate tourists 'long-term preferences and real-time situations. The average score of GRU4Rec is 3.51, and its advantages are mainly reflected in the modeling of historical behavior sequence, but its ability to adjust routes in unexpected situations is insufficient. The average score of PageRank model is 2.93, and the recommendation results are more dependent on the global popularity of

attractions, which is difficult to reflect the individual differences of tourists.

3.5 Dynamic travel event response and route replanning delay testing

In order to test the robustness and online replanning ability of the model under sudden tourism events, this paper sets up a dynamic tourism event response experiment. The test scenarios include temporary closure of scenic spots, overrun of tourist flow in scenic spots, aggravated road congestion and sudden deterioration of weather. The experiment focuses on whether the system can update the status of tourists in time, eliminate infeasible or unsuitable attractions, and generate new executable routes after the original recommended route fails. The test can reflect the fault tolerance ability and service continuity of the smart travel recommender system in the real travel environment.

During the experiment, the system randomly triggered dynamic events when tourists finished visiting the current scenic spot and were ready to go to the next recommended scenic spot. If the target attraction is marked as "temporarily closed" or "overrun", the attraction is immediately deleted from the action space. If the road congestion index exceeds the preset threshold, the system recalulates the travel time and reduces the priority of cross-area recommendation. If the weather worsens, the weight of outdoor attractions decreases, and indoor venues, commercial blocks and exhibition Spaces are preferentially included in the candidate set. After the event is triggered, the model reconstructs the state vector, updates the set of feasible actions, and invokes the policy network to output alternative routes. If the system generates a new route within 60 s that is legal, non-repetitive and satisfies the remaining time constraint, a transfer is judged to be successful. In this paper, a total of 50 independent dynamic events are simulated, and the transition success rates of different models are recorded. The transition success rate variation of different models during event accumulation is shown in Figure 9.

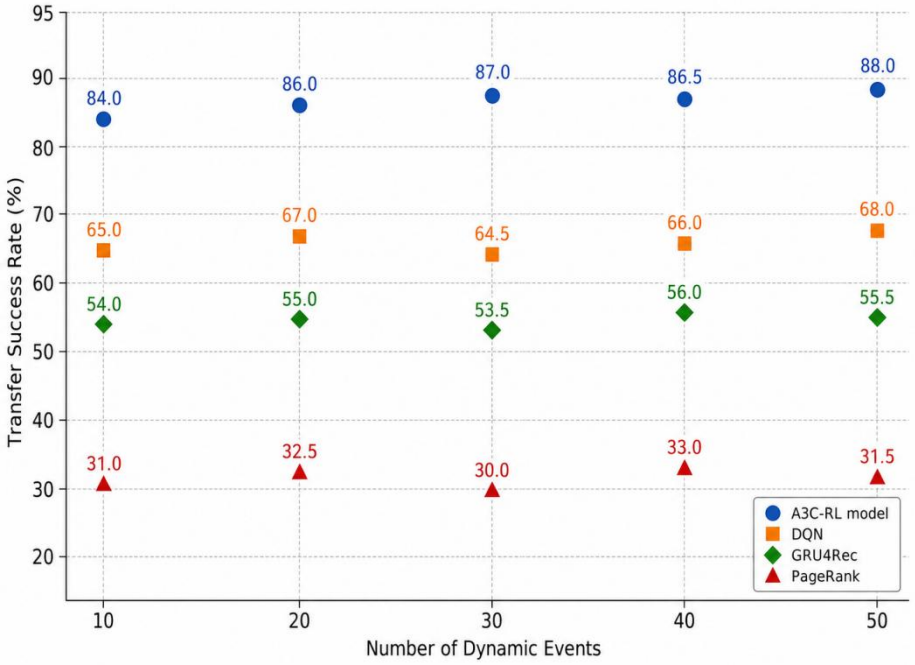

Figure 9: Change in transfer success rate under dynamic tourism events

As can be seen from Figure 9, the final transfer success rate of A3C-RL model reaches 88.0% in 50 dynamic event tests, and the overall curve fluctuates slightly, indicating that the

model can still maintain a relatively stable ability to generate alternative routes under event disturbances. The final success rate of the DQN model is 68.0%. Although it has a certain dynamic adjustment ability, it is prone to unstable value estimation when the action space changes rapidly. The final success rate of GRU4Rec is 55.5%. Gru4rec mainly relies on historical visit sequences, and lacks an effective real-time constraint correction mechanism when attractions are closed or traffic changes. The final success rate of PageRank model is 31.5%, which indicates that the static sorting method based on global popularity is difficult to deal with the route failure problem under emergencies.

In addition to the transfer success rate, this paper further counts the response delay between the event triggering and the output of the new route. The lower the delay, the better the system can complete the route replanning with a short waiting time for tourists. The statistical results of response delay for different models are shown in Table 5.

Table 5: Dynamic event response delay statistical results

Model	Average Latency (s)	Median Latency (s)	Standard Deviation (s)	Minimum Latency (s)	Maximum Latency (s)
A3C-RL Model	1.12	1.09	0.21	0.68	1.71
DQN	2.47	2.52	0.31	1.96	3.34
GRU4Rec	2.86	2.91	0.37	2.18	3.96
PageRank	4.63	4.71	0.72	3.42	6.35

Table 5 shows that the A3C-RL model has an average response delay of 1.12 s and a maximum delay of 1.71 s, which can meet the requirements of low-delay replanning for online tourism services. The average latency of DQN and GRU4Rec is 2.47 s and 2.86 s, respectively, which is mainly affected by action value re-estimation and historical sequence re-calculation. The average delay of PageRank model reaches 4.63 s, and the maximum delay is 6.35 s, which may affect the tourist experience in tourism scenarios with frequent emergencies. Comprehensive transfer success rate and response delay results show that the proposed model can quickly generate alternative plans after the original route failure through real-time update of state, dynamic filtering of action space and fast reasoning of policy network, so as to improve the stability and practical application value of smart tourism route recommendation system.

4 Discussion

The reinforcement learning smart tourism personalized route dynamic planning model constructed in this paper shows good comprehensive advantages in recommendation accuracy, route rationality, tourist satisfaction and dynamic event response. The experimental results show that the HR@5 and HR@10 of the A3C-RL model reach 56.8% and 82.5% respectively, which are higher than those of the DQN, GRU4Rec and PageRank models, indicating that the proposed method can more accurately capture the next visit intention of tourists. Its performance improvement does not rely solely on historical behavior sequences, but comes from the joint modeling of tourist preferences, real-time context, spatial location and trip constraints, which enables the model to form a more fine-grained state understanding in different tourism scenarios. From the perspective of route rationality, the average travel time of the model in this paper is 13.8 min, the cross-area jump rate is 7.9%, and the proportion of effective tour time reaches 72.4%. It shows that the dynamic action space filtering and reward

function constraint can effectively reduce the cross-area jump recommendation, and make the route planning more in line with the real travel logic. In the evaluation of personalized satisfaction, the average score of the model reaches 4.51, which indicates that the tourist experience-oriented reward design not only improves the recommendation hit rate, but also enhances the adaptation degree between the route and tourists' interests, tour rhythm and temporary context. In the dynamic travel event test, the transfer success rate of A3C-RL model reaches 88.0%, and the average response delay is 1.12 s, which reflects the strong online replanning ability. In contrast, DQN has insufficient stability when dynamic action space changes, GRU4Rec has weak response to sudden environmental changes, and PageRank is obviously limited by static heat ranking. In general, through the synergy of state space representation, composite reward function, asynchronous actor-critic training and closed-loop route generation mechanism, the proposed model achieves the unity of high accuracy, good executability and strong dynamic adaptability.

5 Conclusion

Aiming at the problems of insufficient real-time adaptability, insufficient personalized expression and delayed response to emergencies in traditional route recommendation methods in smart tourism scenarios, this paper constructs a personalized route dynamic programming model based on reinforcement learning. The model takes tourists' preferences, real-time context, spatial location and itinerary constraints as state inputs, ensures the accessibility and time feasibility of recommended attractions through dynamic action space screening, and designs a composite reward function guided by tourists' experience, so that the strategy learning process can simultaneously focus on interest matching, route efficiency, environment adaptation and feedback satisfaction. Experimental results show that the A3C-RL model is superior to DQN, GRU4Rec and PageRank in terms of route recommendation accuracy, route rationality, personalized satisfaction and dynamic replanning ability. Among them, the scores of HR@5 and HR@10 reached 56.8% and 82.5%, respectively. The comprehensive score of route rationality was 87.6, and the average score of personalized satisfaction was 4.51. In 50 dynamic event tests, the transfer success rate of the model reaches 88.0%, and the average response delay is 1.12 s, which indicates that the method can adapt to complex situations such as scenic spots closing, passenger flow exceeding limit, traffic congestion and weather changes. This study provides a trainable, updatable, and deployable computing framework for personalized route recommendation in smart tourism. Limited by the scale of the experimental data and the simulation environment, the model still has room for improvement in the cold start of new tourists, cross-city generalization and high concurrent service deployment. Future research can combine hybrid recommendation, federated learning, model compression and event-triggered re-planning mechanism to further enhance the scalability and stability of the system in real tourism platforms.

Author's Profile

Lai Yanmin (1990–), female, PhD candidate, Assistant Researcher. She works at Shunde Polytechnic University, with research interests focus on management science, tourism management and user behavior research.

Lai Yanjun (1981–), female, Bachelor of Engineering, Associate Professor. She works at Shunde Polytechnic University, with research interests in mechanical design and manufacturing.

References

[1] Dalla Vecchia A, Migliorini S, Quintarelli E, et al. Promoting sustainable tourism by recommending sequences of attractions with deep reinforcement learning[J]. Information Technology & Tourism, 2024: 1-36.

[2] Massimo D, Ricci F. Building effective recommender systems for tourists[J]. AI Magazine, 2022, 43(2): 209-224.

[3] Massimo D, Ricci F. Combining reinforcement learning and spatial proximity exploration for new user and new POI recommendations[C]//Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization. 2023: 164-174.

[4] Ghobadi F, Divsalar A, Jandaghi H, et al. An integrated recommender system for multi-day tourist itinerary[J]. Applied Soft Computing, 2023, 149: 110942.

[5] Pitakaso R, Srichok T, Khonjun S, et al. Multi-objective sustainability tourist trip design: An innovative approach for balancing tourists' preferences with key sustainability considerations[J]. Journal of Cleaner Production, 2024, 449: 141486.

[6] Kolaee M H, Jabbarzadeh A, Al-e S M J M. Sustainable group tourist trip planning: An adaptive large neighborhood search algorithm[J]. Expert Systems with Applications, 2024, 237: 121375.

[7] Adamo T, Colizzi L, Dimauro G, et al. A multi-modal tourist trip planner integrating road and pedestrian networks[J]. Expert Systems with Applications, 2024, 237: 121457.

[8] Ruiz-Meza J, Montoya-Torres J R. A systematic literature review for the tourist trip design problem: Extensions, solution techniques and future research lines[J]. Operations Research Perspectives, 2022, 9: 100228.

[9] Divsalar G, Divsalar A, Jabbarzadeh A, et al. An optimization approach for green tourist trip design[J]. Soft computing, 2022, 26(9): 4303-4332.

[10] Ruiz-Meza J, Brito J, Montoya-Torres J R. A GRASP-VND algorithm to solve the multi-objective fuzzy and sustainable tourist trip design problem for groups[J]. Applied Soft Computing, 2022, 131: 109716.

[11] Piya S, Triki C, Al Maimani A, et al. Optimization model for designing personalized tourism packages[J]. Computers & Industrial Engineering, 2023, 175: 108839.

[12] Derya T, Atalay K D, Dinler E, et al. Selective clustered tourist trip design problem with time windows under intuitionistic fuzzy score and exponential travel times[J]. Expert Systems with Applications, 2024, 255: 124792.

[13] Shambour Q, Abualhaj M, Abu-Shareha A, et al. Personalized tourism recommendations: Leveraging user preferences and trust network[J]. Interdisciplinary Journal of Information, Knowledge, and Management, 2024, 19: 017.

[14] Huda C, Heryadi Y, Budiharto W. Smart tourism recommender system modeling based on hybrid technique and content boosted collaborative filtering[J]. IEEE Access, 2024,

12: 131794-131808.

[15] Belussi A, Cinelli A, Vecchia A D, et al. Forecasting POI occupation with contextual machine learning[C]//European Conference on Advances in Databases and Information Systems. Cham: Springer International Publishing, 2022: 361-376.

[16] Migliorini S, Quintarelli E, Gambini M, et al. Sequence recommendations for groups: A dynamic approach to balance preferences[J]. Information Systems, 2022, 108: 102023.

[17] Sarkar J L, Majumder A, Panigrahi C R, et al. Tourism recommendation system: a survey and future research directions[J]. Multimedia tools and applications, 2023, 82(6): 8983-9027.

[18] Afsar M M, Crump T, Far B. Reinforcement learning based recommender systems: A survey[J]. ACM Computing Surveys, 2022, 55(7): 1-38.

[19] Rahmani H A, Aliannejadi M, Baratchi M, et al. A systematic analysis on the impact of contextual information on point-of-interest recommendation[J]. ACM Transactions on Information Systems (TOIS), 2022, 40(4): 1-35.

[20] Rahmani H A, Deldjoo Y, Di Noia T. The role of context fusion on accuracy, beyond-accuracy, and fairness of point-of-interest recommendation systems[J]. Expert Systems with Applications, 2022, 205: 117700.