Multiscale Pedestrian Walking Preference Model in Historic Area Based on Second-Order Markov Chain

Chenlei Liu[1] ⓘ, Nan Wu[2] ⓘ, Zhisheng Huang[3] ⓘ and Shaosen Wang[4] ⓘ

[1]Huaqiao University, Xiamen Key Laboratory of Integrated Application of Intelligent Technology for Architectural Heritage Protection, Xiamen University, 22013085009@stu.hqu.edu
[2]Huaqiao University, archt_wn@qq.comhzs_
[3]Huaqiao University, mahakala@foxmail.com
[4]Xiamen Key Laboratory of Integrated Application of Intelligent Technology for Architectural Heritage Protection, Xiamen University, ymcai@xmu.edu.cn

Corresponding author: Nan Wu, archt_wn@qq.com

Abstract. The road space within the historic district serves as a vital carrier for pedestrian activities. The distribution of pedestrian density and the patterns of route choice are typically influenced by environmental factors, particularly the features of the roads. However, the same distribution of pedestrian density may result from various route choices. Therefore, comprehensively considering both the distribution of pedestrian density and the underlying patterns of route choices, as well as their combined relationship with road environment features, becomes crucial for describing pedestrian walking preferences. This study proposes a pedestrian walking preference model: it maps the road network into a graph structure and employs a second-order Markov chain to describe pedestrian walking processes, thereby predicting pedestrian density distribution and route choices. In this process, an NGO-BP neural network is utilized to learn the route choice probabilities under the influence of road features. Additionally, a route choice feature characterization method based on stroke cluster is proposed to adapt to the irregularities and small-scale road network features commonly found in historic districts. Finally, using Quanzhou Ancient City as a case study, data obtained from both online networks and field surveys are used for simulation experiments. By comparing the simulation results with actual data, this method is demonstrated to assist planners in regulating pedestrian distribution and walking trajectories through the transformation of the built environment within the historic district. This enables a more refined approach to the preservation and revitalization of historic districts.

Keywords: Second-Order Markov Chain, Multiscale, Historic District, Walking Preference Model, Street Vitality

1 INTRODUCTION

The primary objective of urban public environment renewal decision-making is to enhance the pedestrian experience in public spaces by altering urban environmental factors, thereby attracting crowds and providing positive impacts on various aspects of society, economy, and culture. Within historic district, the road space serves as the primary carrier for the transmission of historical culture and the occurrence of pedestrian activities in urban public spaces. Spontaneous congregation of people occurs due to their preferences for the environment. However, unplanned pedestrian flows may lead to an imbalanced distribution of density in the roads. Therefore, it is necessary to consider regulating pedestrian movement to promote the sustainable development of historic district. With the increasing importance of historic cultural heritage preservation in the concept of smart cities, and the demand for refined urban design [1], urban planners face challenges in identifying pedestrian walking preference patterns within historic district with complex environments. They require more comprehensive and accurate algorithms to assist in the renovation of historic district.

In recent years, within the research scope of urban planning, street vitality has often been the focus of attention concerning the spatial distribution of the macro density or its specific route choices in relation to environmental factors. However, these two directions of inquiry are often studied independently, with fewer studies addressing the mechanisms behind the influence of pedestrian route choices on pedestrian density distribution and their combined relationship with environmental factors. This may obscure the true pedestrian intentions behind density distributions and lack practicality. Additionally, determining effective scales and corresponding evaluation method in complex, geographically heterogeneous historic district remains a challenge.

Hence, we aim to devise a method to assist in the planning decisions for historic district revitalization from a more flexible multi-scale perspective. This study introduces a model based on second-order Markov chains to predict pedestrian density and trajectory distribution influenced by the environment within the historic district, where the measurement cost of transition probabilities is significant. To address this, we propose a method based on stroke cluster to characterize the route choice features combined with deep learning to derive pedestrian route choice preference patterns, which can predict transition probabilities based on environmental factors. This approach can aid planners in guiding pedestrian trajectories by altering environmental factors, thereby achieving density distribution regulation objectives.

The paper consists of the following five sections, each covering the following content:

1. Introduction: This section provides an overview of the research background and significance, research objectives, content, and the structure of the paper.

2. Literature Review: This section reviews existing research in the field of historic districts, focusing on pedestrian distribution, route choice, and their relationships with the road environment features.

3. Main Idea: This section develops a pedestrian walking dynamics model based on second-order Markov chains. It proposes a method of route choice feature characterization based on stroke clusters and utilizes NGO-BP neural networks to explore pedestrian route choice preference patterns.

4. Experimental Validation and Analysis: This section encompasses research scope selection, experimental data acquisition, and processing, as well as validation and analysis of experimental results.

5. Conclusion and Outlook: This section provides a retrospective review of the main research content and summarizes the research conclusions and innovations.

2   LITERATURE REVIEW

Studying pedestrian activities has always been a focal point for the preservation and revitalization of historic districts. Excessive pedestrian traffic may exert pressure on historic heritage [2], while limited visitation may lead to the abandonment of historic heritage [3].

Influenced by environmental factors such as road network structure and built environment [4-7], research on pedestrian activities encompasses macro-level pedestrian density distribution and micro-level route choice, among which route choice is the underlying cause of density distribution changes [8]. Therefore, these studies typically approach the topic from two perspectives:

(a) Various environmental factors indicators influence the distribution of pedestrian density. These studies often employ graph theory, complex networks, and other theoretical frameworks to describe the metrics, geometry, and topological properties of urban road networks, such as accessibility, centrality, and connectivity [9, 10]. For instance, Zou et al. [11] utilized road network data and Baidu heat map data, employing spatial and statistical analysis methods to establish an associative model between the road building environment (SBE) and population density distribution in historical areas of Wuhan. They concluded that small businesses in historical areas play a pivotal role in fostering street vitality. Zhang et al. [12] evaluated the architectural and landscape features of roads, and then employed a discrete choice model to investigate the impact of environmental indicators in historical roads on the distribution intensity of pedestrian trajectory points. They found that road width and road aspect ratio significantly influence tourist distribution.

(b) Research on the relationship between pedestrian route choice patterns and environmental factors. Ben-Akiva and Bierlaire discussed the application of various discrete choice models in route choice decision-making, assuming that based on the principle of maximum utility, decision-makers probabilistically select travel plans from a set of route choices [13]. Foltete et al. indicated that tall residential buildings along the route decrease the likelihood of the route being chosen [14]. A study in San Francisco suggests that excessively high vehicle speeds and flow on roads can reduce the probability of pedestrian route choice [15].

The aforementioned two aspects of research are often treated as independent study objects. It is worth noting that different route choice patterns may lead to the same pedestrian density distribution, which can be mathematically explained: multiple solutions can be obtained by inversely solving the transition probability matrix from the stable distribution of a dynamical system [16]. This implies that focusing solely on density distribution may result in overlooking valuable information. Therefore, this paper attempts to comprehensively consider both the population density distribution and the underlying route choice patterns in the context of historic district, along with their collective relationship with environmental evaluation indicators.

In fact, Markov chains can effectively describe pedestrian walking behavior, enabling the construction of more accurate pedestrian mobility models. Jiang and Jia utilized Markov chains to simulate human movement in large-scale road networks [17]. Jiang suggested that simulating pedestrian random walks on graph-structured networks can reflect the complex network properties of road networks, such as scale-free and small-world characteristics. It is believed that the accuracy of constructing walking models using higher-order Markov chains can significantly improve [18]. Additionally, the irregular and small-scale road networks of historic district may lead to inaccuracies in single-scale assessment indicators calculation methods.

In summary, considering the memory-based nature of pedestrian activities, this paper proposes a multi-scale walking preference model based on second-order Markov chains to adapt to the context of historic district.

3 MAIN IDEA

3.1 Pedestrian Density and Trajectory Distribution Model Based on Second-Order Markov Chain

3.1.1 Walking process and Markov chain

One can map road segments in the road network as edges and map intersections as vertices through the original graph representation method [19] to establish a graph structure $G = (V, E)$, where V is a set of n vertices and E is a set of edges. This facilitates the preservation of the topological morphology and connectivity of the road network to simulate a more realistic pedestrian walk.

In the road network G, there are multiple possible trajectories for a pedestrian to traverse from vertex A in a block to vertex B further away, which can be described using chains in graph theory, as shown in Figure 1(a). A chain $(v_1, e_1, v_2, e_2, \ldots, v_k)$ is a sequence formed by a group of alternating vertices and edges, where $e_i = (v_i, v_{i+1})$.

A Markov chain is a stochastic process model that describes a series of possible events. Its characteristic is that the probability of the system being in each state at the current moment depends on the states of events that have occurred within a set memory duration [20]. For example, in an m-order Markov chain, there is

$$
\begin{aligned}
P\ X_n = x_n &|\ X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1 \\
&= P\ X_n = x_n |\ X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-m} = x_{n-m} \quad \text{for } n > m
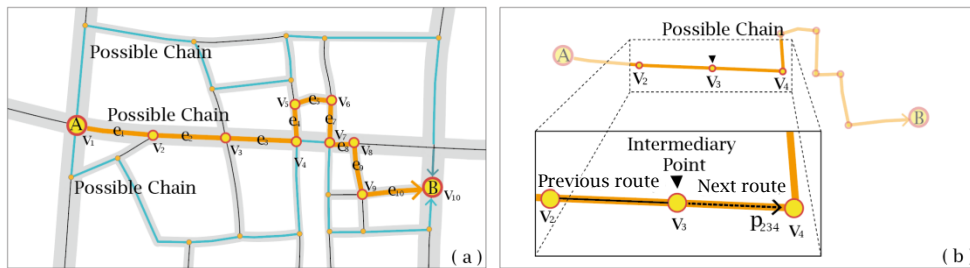\end{aligned}
\tag{1}
$$

Figure 1: (a) Several possible trajectories from vertex A to vertex B, (b) The walking process described by second-order Markov chains.

3.1.2 Dynamic modeling of walking

To establish a dynamic model describing pedestrian walks in historic districts, it is necessary to consider that when a pedestrian crosses from one road segment to another through an intersection, they are simultaneously influenced by both the previous route and several potential next routes. This means that the choice of the next vertex is jointly determined by the previous vertex and the current vertex, reflecting the memory-based nature of pedestrian walking [21]. Therefore, the pedestrian walking process can be described using a second-order Markov chain, as shown in Figure 1(b). The second-order transition probability involves the states of three time points, which is denoted as

$$
P_{kij} = P(X_{t+1} = j \mid X_t = i, X_{t-1} = k), \text{ for } i, j, k \in V
\tag{2}
$$

where X_t represents the vertex index where a pedestrian is located at time t. The specific values of the second-order transition probabilities can be obtained from the pedestrian walking preference mode constructed later in the text. Specially, if there is no connection between two vertices, the corresponding transition probability is recorded as zero. The second-order distribution $\rho_{ki}^{(t)}$

represents the proportion of pedestrians who were at vertex k at time $t-1$ and are now at vertex i at time t, relative to the total number of pedestrians, thus there is

$$\rho_{ij}^{(t+1)} = \sum_{k=1}^{n} \rho_{ki}^{(t)} P_{kij}, \text{ for } i,j,k \in V \tag{3}$$

which describes the evolution of second-order distribution over time in a block with a constant total number of pedestrians.

Given the initial density distribution on each vertex, i.e., the first-order distribution, solely relying on second-order transition probabilities is insufficient to determine the subsequent evolution of the system. This is because it necessitates employing a first-order transition probability acting on the first-order distribution to obtain the initial second-order distribution. In the scenario where second-order transition probabilities are known while first-order transition probabilities are unknown, considering that the same system should be able to evolve to the same steady-state distribution via first- and second-order transition probabilities $\rho_i^{(\infty)}$, one can utilize second-order transition probabilities to derive the uniquely corresponding first-order transition probabilities.

Based on the relationship between the first-order distribution and the second-order distribution as

$$\rho_i^{(t)} P_{ij} = \rho_{ij}^{(t+1)}, \ \rho_i^{(t)} = \sum_{j=1}^{n} \rho_{ij}^{(t+1)}, \tag{4}$$

one can derive that the first-order transition probabilities under steady-state distribution satisfy:

$$P_{ij} = \frac{\rho_{ij}^{(\infty)}}{\sum_{j=1}^{n} \rho_{ij}^{(\infty)}} \tag{5}$$

Where $\rho_i^{(\infty)}$, in principle, can be obtained by repeatedly multiplying the second-order transition probability matrix infinitely. However, in practice, we cannot perform infinite calculations. Therefore, consider the following approach:

By defining the element in the $[(l-1)n+j]$-th row and $[(k-1)n+i]$-th column of the second-order transition probability matrix $\mathbf{P}$ as

$$P_{lj,ki} = \begin{cases} P_{kij}, \text{ for } l = i \\ 0, \quad \text{ for } l \neq 0 \end{cases} \tag{6}$$

there is $\sum_{j=1}^{n} P_{lj,ki} = 1$, causing matrix $\mathbf{P}$ to have spectral radius $\max(\lambda_i) = 1$. Diagonalize matrix $\mathbf{P}$ in the manner of $\mathbf{P} = \mathbf{F}\mathbf{\Lambda}\mathbf{F}^{-1}$, resulting in

$$\mathbf{P}^\infty = \mathbf{F}\mathbf{\Lambda}^\infty \mathbf{F}^{-1} = \mathbf{F}f(\mathbf{\Lambda})\mathbf{F}^{-1}, \text{ where } f(x) = \begin{cases} 1, \text{ for } |x| = 1 \\ 0, \text{ for } |x| < 1 \end{cases} \tag{7}$$

Based on the above conclusion, one can take any column of $\mathbf{P}^\infty$ to obtain $\rho_i^{(\infty)}$, and then from Equation (5) one can obtain the first-order transition probability.

3.1.3 Discussion on open systems

Furthermore, this model can also be used to simulate the dynamic evolution of open systems, corresponding to scenarios where the number of pedestrians in the road network can vary. We

introduce v_0 in the vertex set V to represent the external environment, while $\rho_{i0}^{(t)}$ and $\rho_{0i}^{(t)}$ respectively represent the proportion of pedestrians going from vertex i to the outside and entering vertex i from the outside to the total number at time t. It should be noted that $\rho_{i0}^{(t)}$ is determined by the second-order distribution from the previous time step, whereas $\rho_{0i}^{(t)}$, due to its unpredictability, can only be treated as a known condition. Therefore, in open systems, the vertex indices in Equation (2) should satisfy $i \neq 0$.

Thus, when the second-order transition probabilities are known, simulating the second-order distribution of pedestrian counts at each moment in a real open neighborhood only requires specifying the initial distribution $\rho_{ki}^{(0)}$ and the current $\rho_{0i}^{(t)}$.

3.2 Predicting Pedestrian Trajectory Selection Preference Modes Based on Environmental Factors

In pedestrian movement simulated by Markov chains, it is assumed that a pedestrian can transition from one vertex to adjacent vertices within a unit time interval. When a pedestrian reaches the vertex v_i and needs to choose the next direction to move; this process is referred to as the route choice process, which can be described using second-order transition probabilities in Equation (2). Here, v_i represents the intermediary point at this moment, the edge e_{i-1} that the pedestrian has just passed through is considered the previous route, and the edge e_i that the pedestrian will take next is considered the next route. Influenced by the overall road environment factors, including both the previous route and all available next routes, the route choice process exhibits a pattern known as route choice preference, as illustrated in Figure 2(a).
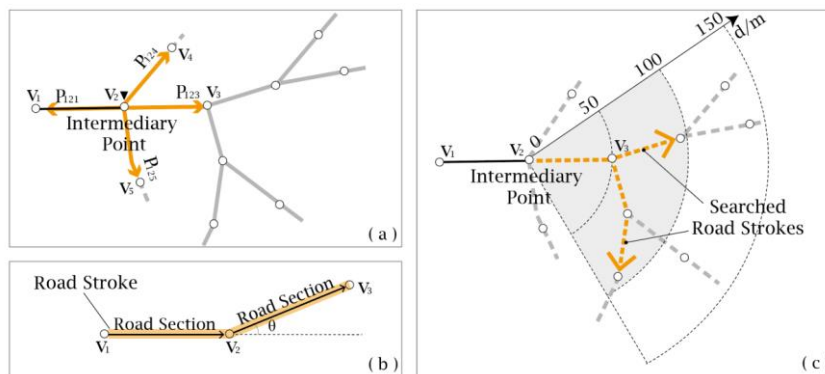

Figure 2: (a) Route Choice Preferences, (b) A stroke cluster is formed by two smoothly connected road segments, where smaller turning angles between segments indicate higher overall smoothness, (c) With a network scale of $d = 50\mathrm{m}$ and a truncation length of $L = 100\mathrm{m}$, all strokes found from edge (v_2, v_3) are included in the search results.

3.2.1 Road feature assessment indicators

Drawing upon previous research [11, 22, 23], this study employs commonly used environmental factors to evaluate pedestrian preferences for road features. In terms of the built environment, indicators for evaluating road features include road width, length, tortuosity, width-to-height ratio (W/H), plot ratio, building density, POIs (Points of Interest) diversity, as well as the density of POIs such as dining, shopping, daily necessities, public facilities, accommodation, and scenic spots. Regarding the inter-road relationships, evaluation indicators encompass road deflection degree and

the (co-)counter flow. Specifically, the (co)counter flow is represented using one-hot encoding. We refer to this collection of road feature assessment indicators as the feature set. The selection of indicators and their specific contents are as follows:

road Feature Assessment Indicators	Indicators for Road	Meaning	Calculation Formula
Metrics	Road Width	Road average width	W
	Road Length	Road axis length	L_c
Geometric	Tortuosity	The curvature of a road is expressed as the ratio of road length to straight-line distance.	$\left|\vec{L}_c\right| / L_c$
Built Environment	W/H	The ratio of the average width of roads to the average height of buildings, where i denotes the building index.	$W / \langle h_i \rangle$
	Building Density	The ratio of the total building footprint area to the road buffer area.	$\sum_i s_i / S_r$
	Plot Ratio	The ratio of the total built-up area to the road buffer area.	$\sum_i S_i / S_r$
	Density of POIs	The ratio of the number of POIs to the length of road.	N / L_c
	Diversity of POIs	The diversity of POI types, where p_i represents the proportion of the POI i-th type.	$-\sum_i p_i \log_2 p_i$
Road Relationship	Road Deflection Degree	The deflection angle between road a and b.	θ_{ab}
	(Co)Counter Flow	The alignment of route choice with the direction of pedestrian flow.	-

Table 1: The road feature assessment indicators.

3.2.2 Route choice feature characterization method based on stroke cluster

The characterization of road features at appropriate scales is crucial for accurately reflecting road information. In small-scale road networks, features of short roads are often characterized using stroke-based methods and shared evaluation indicators with other strokes, while in large-scale networks, long roads are often divided into smaller segments to capture road detail features. To investigate pedestrian walking preferences, it is necessary to quantify the road features that influence pedestrian route choice. However, historic district road networks, unlike regular grid systems, consist of irregular and fragmented road segments, making it challenging to find appropriate scales for computing assessment indicators. Additionally, Previous methods have not adequately captured all the road branch features that must be taken into account in route choice.

We propose a route choice feature characterization method based on stroke cluster to compute the feature set for each route choice. A stroke [24] is a naturally extended road formed by connecting multiple road segments smoothly, as illustrated in Figure 2(b), where smaller turning angles between segments indicate higher overall smoothness. The smoothness of a stroke is defined as the sum of turning angles between segments. In this study, we define the feature set of a stroke as the collection of average features of all segments it contains. In small-scale road networks, strokes exhibit superior feature representation capabilities compared to individual segments, while in large-scale networks, constraining the length of strokes can effectively describe road details. Thus, strokes can flexibly adapt to multi-scale requirements.

As shown in Figure 2(c), when a cluster of strokes extends from an intermediary point, the edge from which these strokes originate is termed as the starting edge of the stroke cluster. Depending on the desired scale, one can flexibly set the truncation length L of strokes. By employing a Depth-First Search (DFS) method, one can obtain all strokes extending outward from a central point with a length not exceeding L, which can be grouped into multiple clusters based on different starting edges. The comprehensive feature set of a stroke cluster, where each stroke's feature set is weighted by its straightness, is defined as the integrated feature set of the starting edge. During a route choice process, all available next edges are in competition. Hence, all edges connected to the intermediary point should be considered, not limited to just one pair of previous and next edges. Consequently, the comprehensive feature set that fully describes a route choice process is formed by merging the integrated feature sets of all edges connected to a central point. It is important to note that at a central point, multiple pairs of previous-next edges may undergo route choice processes, distinguished by the placement sequence of the integrated feature sets in the route choice feature set, which follows the order of the previous edge, next edge, and others.

This approach avoids redundant computations of data at different scales, thereby enhancing efficiency. Moreover, the introduction of a stroke cluster not only reflects the feature of individual roads but also captures the overlay of attributes from underlying potential roads, thereby providing a more comprehensive perspective.

3.2.3 Using NGO BP neural network to predict route choice preferences

The NGO-BP neural network (Northern Goshawk Optimization BP Network), based on intelligent swarm optimization techniques, aims to mimic the cognitive processes and foraging tendencies exhibited by the Northern Goshawk. Renowned for its remarkable optimization accuracy and consistency, this algorithm has found widespread application in fitting nonlinear relationships [25-27]. The selection of the fitness function involves using the mean square error between the output results and the expected values as the objective function:

$$F = \left\langle y_i - o_i^{\ 2} \right\rangle \tag{8}$$

where n represents the number of output nodes in the neural network, y_i denotes the expected output of the i-th node of the BP neural network, and o_i represents the predicted output of the i-th node.

To uncover pedestrian route choice preference, the NGO-BP neural network can be employed to establish the mapping relationship between the route choice feature set determined by the road environment and its corresponding second-order transition probabilities. Subsequently, predictions of unknown route preferences can be made based on the new road environment.
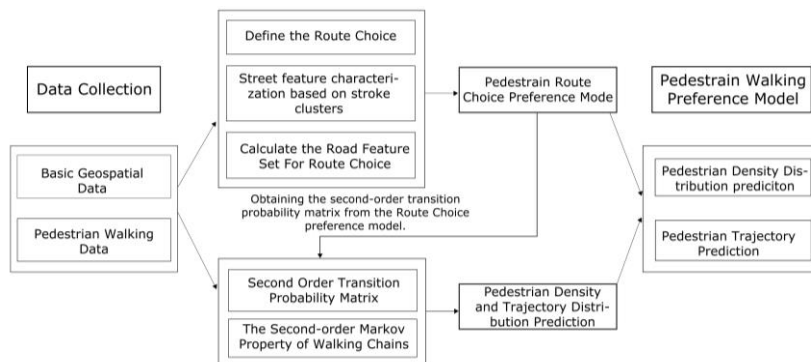

Figure 3: Flow chart of the technical method.

4　CASE STUDY

The technical method of this study is shown in Figure 3.

4.1　Research Scope

The study focuses on the historic district of West Street and its surrounding areas in Quanzhou Ancient City, Fujian Province, China. The research will include samples from the streetscape of West Street, a historical and cultural pedestrian road, and adjacent lively roads, excluding main traffic arteries and residential roads within closed residential areas. The study will consider pedestrian groups primarily comprising tourists, including local residents.

4.2　Data Acquisition and Processing

Geospatial data during the National Day holiday and regular holidays in October 2023 and pedestrian walking behavior data were collected to obtain route choice features assessment indicators and probability distributions.

4.2.1　Obtaining route choice features from geospatial data

Geospatial data for the National Day holiday and regular holidays in October 2023 were collected from the basic geographic database of the Baidu API (Application Programming Interface).

The building data includes height, building footprint area, shape vector, etc., Baidu heat map data, provided in vector format, consists of 32 heat map data sets for October 1st and 22nd, with hourly intervals and precision ranging from levels 17 to 19.

POIs data, in text format, is categorized into six types based on route choice feature assessment indicators: dining, shopping, daily necessities, public facilities, accommodation, and scenic spots. During the National Day holiday, a large number of mobile stalls were set up in Xiaoxicheng and Suqingmen Square, so data of mobile stalls on Quanzhou West Street was updated in the commercial POIs data.

Road data comprises road segment names, grades, lengths, types, etc. The average length of roads in the statistical research area is 121.89m, with a minimum length of 13.65m for natural road sections and a maximum length of 483.81m.

The above basic geospatial data is imported into ArcGIS. Based on the average length of the road in the area, the road network data is processed into scales of 50m, 100m, 150m, and 200m. In other words, roads longer than the specified scale are divided into smaller segments, ensuring that the average length of the road network approaches the specified value. Subsequently, the road features of the road segments are computed for the road network on these four scales during the National Day holiday and regular holidays in October 2023.

4.2.2　Obtaining probability distribution of route choice from pedestrian walking behavior data

(a) Tourist Trajectory Data

Tourist trajectory data was obtained from the Six Feet GPS Travel Community website (http://www.foooooot.com/) and converted to vector data in ArcGIS. A trajectory recognition algorithm was designed to calculate the probability distribution of each path selection, resulting in a collection of 620 path selection data.

(b) Self-collected Pedestrian Walking Video Data

The trajectory data of online tourists is more concentrated on the main roads of the West Street, and there is less data on the internal streets and alleys. Therefore, it is necessary to supplement the path selection data within the streets and alleys. From October 1st to 5th, 2023, high-definition cameras were installed at 47 intersections within the study area, as shown in Figure 4, to capture pedestrian walking video data from 9:00 AM to 6:00 PM. The video data was cleaned, integrated, and processed for resolution. A total of 3256 video data were obtained.

The YOLOv5-DeepSORT [28, 29] multi-target tracking algorithm was used to identify pedestrian trajectories in the videos. Trajectory mining algorithms were employed to remove offset, short, or misidentified trajectories that may be vehicles or errors [30]. The trajectories were processed into probability distributions of route choice. Selected for 1-hour data from 10 intersections randomly, and manual calculations were conducted to validate the accuracy of the algorithm in extracting probability distributions. The results showed that the accuracy of the data obtained using video mining technology was 90.135%.
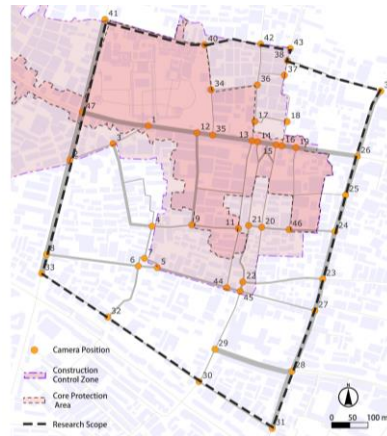

Figure 4: Research scope and camera position

4.3   Experimental Results and Analysis

4.3.1   Density distribution prediction results and analysis

Multiple sets of road networks with different scales of road segments are configured, each paired with varying truncation lengths L. Utilizing a route choice feature characterization method based on stroke cluster, partial route choice features and their corresponding probabilities are extracted from the basic geographic spatial data and pedestrian walking data during the National Day holiday. These are then utilized to train an NGO-BP neural network to establish the mapping relationship between feature data and probability data, i.e., pedestrian route choice selection preferences, to predict other unknown route choice probabilities. The training ratio is set at 0.85.

To mitigate the impact of random data, each dataset is iteratively computed 20 times. The mean squared error averaged across all datasets is 0.031. After multiple experiments, it was found that the dataset with a 50-scale graph structure and a truncation length of 200 meters yielded the best mean square error performance, with an average mean square error of 0.026. Therefore, this scale is selected to predict the probabilities of all route choice in the entire historic district, and then to conduct dynamic simulations to obtain the value of pedestrian density and trajectory distribution. The density values were imported into QGIS and the Inverse Distance Weighted (IDW) interpolation method was used to generate raster graphics. After vectorization, the average density values were linearly assigned to the road elements. The final result is shown in Figure 5.

Comparing the predicted density distribution during the 2023 National Day period with the actual average density data, as shown in Figure 5(a) and Figure 5(b), demonstrates the model's ability to accurately predict areas of high pedestrian concentration in the road network due to favorable road features. It also effectively reflects the overall density distribution.

The area centered around the Suqingmen Square and Xiaoxicheng is considered a high-density region, as depicted in Box 1 in Figure 5(b). Suqingmen Square, a historical heritage site along the West Street, features a spacious layout and numerous mobile stalls, contributing to the high foot traffic. Additionally, Xiaoxicheng, which also hosts numerous mobile stalls like Suqingmen, exhibits

relatively lower heat intensity. This could be attributed to the narrowness of the alley leading to Xiaoxicheng, measuring only 2.5 meters wide, indicating that in similar business environments, road width is an important indicator affecting pedestrian flow.

Additionally, In Figure 5(b), at the location marked by Box 2, the model's predicted relative population density is lower than the actual value. This discrepancy may be due to the presence of a large hospital with a high concentration of people in the area. As a result of sampling, the model tends to reflect the distribution of tourist-dominated groups.

Comparing the predicted density distribution during the National Day holiday and regular holidays, as depicted in Figure 5(b) and Figure 5(e) respectively, during regular holidays, the heat intensity around Suqingmen Square and Xiaoxicheng notably decreases, as shown in Box 3, resulting in a more uniform heat distribution across the area, without the presence of high-density zones observed during the National Day holiday. This is primarily attributed to the closure of temporary mobile stalls that were set up during the National Day period, leading to the removal of their corresponding POIs from the dataset. Consequently, this area exhibits reduced attractiveness, as illustrated in Figure 5(f). Conversely, there is a relative increase in crowd density in other areas, indicating that highly attractive regions can quickly attract crowds, but once the high attractiveness diminishes, there is a higher probability of crowd dispersion to surrounding areas.

When road networks feature well-designed characteristics and highly attractive areas, they may attract significant pedestrian flows that exceed the carrying capacity of historic districts. In such cases, timely adjustments to the road features within historic districts are necessary to ensure their sustainable development. Conversely, if certain historic heritage sites experience low visitation rates, interventions to enhance environmental features are warranted to attract pedestrian traffic.
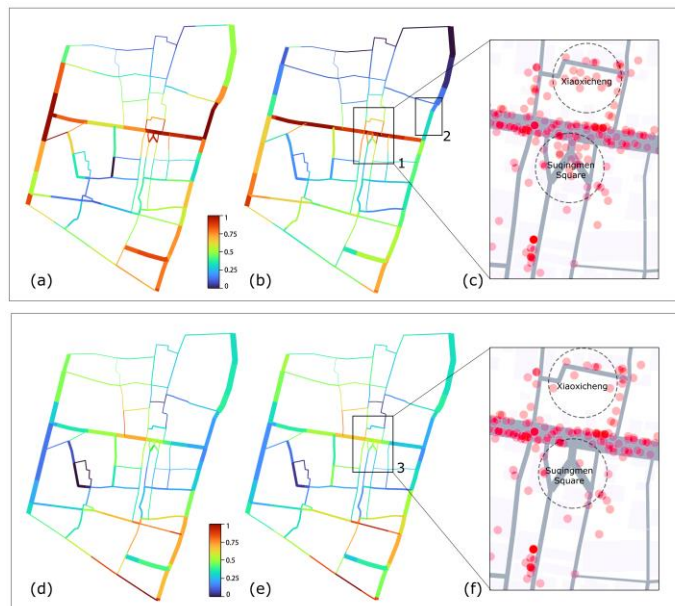

Figure 5: (a) Real average density distribution map during the National Day holiday, (b) Predicted density distribution map during the National Day holiday, (c) Distribution map of POIs in Xiaoxicheng and Suqingmen Square during the National Day holiday, (d) Real average density distribution map during regular holidays, (e) Predicted density distribution map during regular holidays, (f) Distribution map of POIs in Xiaoxicheng and Suqingmen Square during regular holidays.

In (a), (b), (d), and (e), the colors in the road network represent relative density, as indicated in the legend, with values ranging from 0 to 1. In (c) and (f), the red dots represent POI locations, with each dot having the same transparency. The apparent color intensification in certain areas is due to overlapping dots, and the transparency itself holds no actual significance.

4.3.2 Trajectory prediction results and analysis

On the other hand, the scope of trajectory prediction encompasses the diverse environmental settings, concentrated road networks, and ample samples of West Street. Upon specifying the previous route and trajectory length, a trajectory set can be obtained, where each trajectory is associated with a selection probability, as illustrated in Figure 6(a). For a given length of trajectory set, the predictive probabilities of individual trajectories and their actual probabilities are represented by vectors $\vec{q}$ and $\vec{p}$ respectively. The prediction accuracy can then be quantified using the Kullback-Leibler Divergence (KLD)

$$D_{\mathrm{KL}} = \sum_i p_i \log \frac{p_i}{q_i} \tag{9}$$

which is shown in Figure 6(b) and (c). The trajectory prediction performs well within the initial 10 steps. Moreover, the KLD approximately follows an exponential function of the trajectory length.
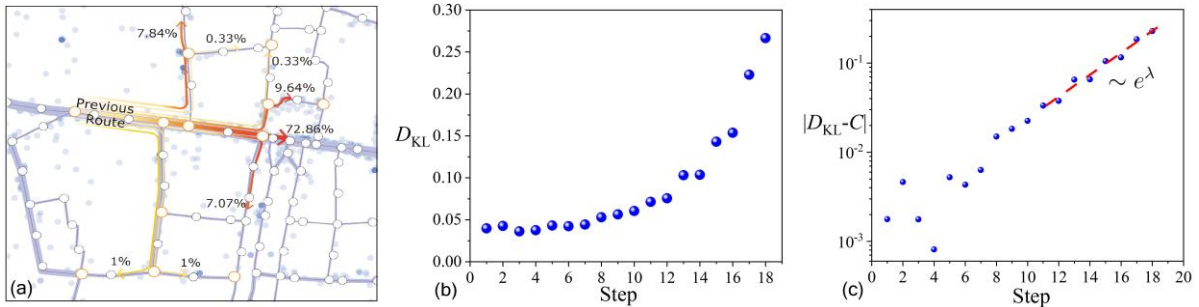

Figure 6: (a) In a 50-meter-scale road network, with a truncation length of L=200 meters, and given the previous direction, the probability distribution of pedestrian trajectories for eight steps, (b) The relationship between (KLD) and the number of steps between predicted probability and actual probability under the setting conditions of Figure 6(a), (c) The KLD in Figure 6(b) is approximately an exponential function of the number of steps, where $C = 0.0038$ and $\lambda = 1.22$. Please note that selecting a logarithmic vertical axis will amplify the fluctuations in the image within fewer steps.

4.3.3 Discussion on the accuracy of density and trajectory prediction

As the number of dynamic evolution steps increases, the accuracy of the model in predicting pedestrian density distribution gradually improves, while conversely, the accuracy of trajectory distribution prediction gradually decreases. The reason behind this phenomenon is that trajectory distributions contain more details, and under steady-state conditions, the same density distribution can correspond to multiple different trajectory distributions. A simple example is when two trajectories intersect at a certain point. If the predictions of their probabilities are respectively biased towards overestimation and underestimation, this will lead to a reduction in the prediction density error at the intersection point due to mutual cancellation. Below, this conclusion will be explained in more rigorous mathematical terms.

Let the actual second-order transition probability be denoted as p_{kij}, and the corresponding predicted value as p'_{kij}. Then, the true probability of the trajectory determined by vertices $i_1, i_2, \cdots, i_s$ is

$$p_{i_1,i_2,\cdots,i_s} = \prod_{j=1}^{s-2} p_{i_j,i_{j+1}i_{j+2}} \tag{10}$$

According to Equation (9), the KLD between the predicted probability and the actual probability of a t-step trajectory departing from a certain road is given by:

$$D_{\mathrm{KL}}(t) = \sum_{\substack{\text{All the } t-\text{step} \\ \text{trajectories}}} \left(\prod_{j=1}^{t-1} p_{i_j,i_{j+1}i_{j+2}} \right) \log \left(\frac{\prod_{j=1}^{t-1} p_{i_j,i_{j+1}i_{j+2}}}{\prod_{j=1}^{t-1} p'_{i_j,i_{j+1}i_{j+2}}} \right) \tag{11}$$

The $D_{\mathrm{KL}}(t)$ approximately follows an exponential function of the step t, as shown in Figure 6(c), possibly originating from the relationship between the number of trajectory types and the step t in the graph structure involved in this study. On the other hand, denoting the actual second-order transition probability matrix as $\mathbf{P}$ and its corresponding prediction as $\mathbf{P}'$, considering that the actual density distribution approximates the steady state of dynamic evolution, the prediction error of the density distribution $\rho^{(\infty)} \sim \left\| \mathbf{P}^\infty - \mathbf{P}'^\infty \right\|$, is a finite value, which can be obtained through Equation (7).

5   CONCLUSIONS

Considering both trajectory and density distributions can better reveal the microscopic essence of pedestrian movements within the historic district and obtain more comprehensive pedestrian walking information. This facilitates direct guidance on altering environmental factors to influence pedestrian trajectories, thereby regulating the density distributions of the historic district.

We propose a multi-scale pedestrian preference model based on a second-order Markov chain to predict pedestrian density and trajectory distributions. By considering the influence of stroke clusters, we mitigate model distortion when the road network scale changes. Incorporating memory-based route preference patterns enables a more accurate simulation of pedestrian route choices influenced by the historic districts. Predicting and comparing density and trajectories during the National Day holiday in 2023 with regular holidays demonstrate the model's ability to capture variations in environmental factors within historic districts, reflected in density and trajectory distributions. This approach assists planners in regulating the pedestrian distribution and walking trajectory by revitalizing environments in historic districts, facilitating more nuanced preservation and revitalization efforts.

Chenlei Liu, https://orcid.org/0009-0007-5726-0259
Nan Wu, https://orcid.org/0009-0000-6349-6347
Zhisheng Huang, https://orcid.org/0009-0006-5688-9023

REFERENCES

[1]   Angelidou, M.; Stylianidis, E.: CULTURAL HERITAGE IN SMART CITY ENVIRONMENTS: THE UPDATE, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020, V-2-2020. https://doi.org/10.5194/isprs-archives-XLII-2-W5-27-2017

[2]   Meetiyagoda, L.: Pedestrian safety in Kandy Heritage City, Sri Lanka: Lessons from World Heritage Cities, SUSTAIN CITIES SOC, 2018, 38, 301-308. https://doi.org/10.1016/j.scs.2018.01.017

[3]   Micelli, E.; Pellegrini, P.: Wasting heritage. The slow abandonment of the Italian Historic Centers, J CULT HERIT, 2018, 31, 180-188. https://doi.org/10.1016/j.culher.2017.11.011

[4]   Crawford, D.-W.; Ashihara, Y.: The Aesthetic Townscape, J AESTHET ART CRITIC, 1987, 44, (4), 416-417. https://doi.org/10.2307/429797

[5]   Feuillet, T.; Commenges, H.; Menai, M.; Salze, P.; Perchoux, C.; Reuillon, R.; Kesse-Guyot, E.; Enaux, C.; Nazare, J.-A.; Hercberg, S.; Simon, C.; Charreire, H.; Oppert, J.-M.: A massive geographically weighted regression model of walking-environment relationships, J TRANSP GEOGR, 2018, 68, 118-129. https://doi.org/10.1016/j.jtrangeo.2018.03.002

[6]   Li, J.; Lin, S.; Kong, N.; Ke, Y.; Zeng, J.; Chen, J.: Nonlinear and Synergistic Effects of Built Environment Indicators on Street Vitality: A Case Study of Humid and Hot Urban Cities, Sustainability, 2024, 16, (5), 1731. https://doi.org/10.3390/su16051731

[7]   Lin, J.; Zhuang, Y.; Zhao, Y.; Li, H.; He, X.; Lu, S.: Measuring the Non-Linear Relationship between Three-Dimensional Built Environment and Urban Vitality Based on a Random Forest Model, International Journal of Environmental Research and Public Health, 2023, 20, (1), 734. https://doi.org/10.3390/ijerph20010734

[8]   Lorch, B.-J.; Smith, M.-J.: Pedestrian movement and the downtown enclosed shopping center, J AM PLANN ASSOC, 1993, 59, (1), 75-86. https://doi.org/10.1080/01944369308975846.

[9]   Hajrasouliha, A.; Yin, L.: The impact of street network connectivity on pedestrian volume, URBAN STUD, 2015, 52, (13), 2483-2497. https://doi.org/10.1177/0042098014544

[10]   Lerman, Y.; Rofè, Y.; Omer, I.: Using Space Syntax to Model Pedestrian Movement in Urban Transportation Planning, GEOGR ANAL, 2014, 46, (4), 392-410. https://doi.org/10.1111/gean.12063

[11]   Zou, H.; Liu, R.; Cheng, W.; Lei, J.; Ge, J.: The Association between Street Built Environment and Street Vitality Based on Quantitative Analysis in Historic Areas: A Case Study of Wuhan, China, SUSTAINABILITY-BASEL, 2023, 15, (2), 1732. https://doi.org/10.3390/su15021732

[12]   Zhang, L.; Zhang, R.; Yin, B.: The impact of the built-up environment of streets on pedestrian activities in the historical area, ALEX ENG J, 2021, 60, (1), 285-300. https://doi.org/10.3390/su15021732

[13]   Ben-Akiva, M.; Bierlaire, M.: Discrete choice models with applications to departure time and route choice, 2003, 7-37. https://doi.org/10.1007/0-306-48058-1_2

[14]   Foltête, J.; Piombini, A.: Deviations in Pedestrian Itineraries in Urban Areas: A Method to Assess the Role of Environmental Factors, Environment and Planning B: Planning and Design, 2010, 37, (4), 723-739. https://doi.org/10.1068/b35015

[15]   Sevtsuk, A.; Basu, R.; Li, X.; Kalvo, R.: A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco, TRAVEL BEHAV SOC, 2021, 25, 41-51. https://doi.org/10.1016/j.tbs.2021.05.010

[16]   Gzyl, H.; Velásquez, Y.: Reconstruction of transition probabilities by maximum entropy in the mean, 2002, 617, 192-203. https://doi.org/10.1063/1.1477048

[17]   Jiang, B.; Jia, T.: Agent-based simulation of human movement shaped by the underlying street structure, INT J GEOGR INF SCI, 2011, 25, (1), 51-64. https://doi.org/10.1080/13658810802022822

[18]   Persson, C.; Bohlin, L.; Edler, D.; Rosvall, M.: Maps of sparse Markov chains efficiently reveal community structure in network flows with memory, arXiv preprint arXiv:1606.08328, 2016. https://doi.org/10.48550/arXiv.1606.08328

[19] Khoshraftar, S.; An, A.: A Survey on Graph Representation Learning Methods, ACM T INTEL SYST TEC, 2024, 15, (1), 19. https://doi.org/10.1145/3633518

[20] Gagniuc, P.-A.: Markov chains: from theory to implementation and experimentation, John Wiley & Sons, USA, UJ, 2017.

[21] Mou, N.; Jiang, Q.; Zhang, L.; Niu, J.; Zheng, Y.; Wang, Y.; Yang, T.: Personalized tourist route recommendation model with a trajectory understanding via neural networks, INT J DIGIT EARTH, 2022, 15, (1), 1738-1759. https://doi.org/10.1080/17538947.2022.2130456

[22] Saelens, B.-E.; Handy, S.-L.: Built environment correlates of walking: a review, Medicine and science in sports and exercise, 2008, 40, (7 Suppl), S550. 10.1249/MSS.0b013e31817c67a4

[23] Wang, B.; Lei, Y.; Xue, D.; Liu, J.; Wei, C.: Elaborating Spatiotemporal Associations Between the Built Environment and Urban Vibrancy: A Case of Guangzhou City, China, CHINESE GEOGR SCI, 2022, 32, (3), 480-492. https://doi.org/10.1007/s11769-022-1272-6

[24] Thomson, R.-C.; Richardson, D.-E.: The 'good continuation' principle of perceptual organization applied to the generalization of road networks, 1999, 1215-1223.

[25] Wang, C.; Wang, Y.; Dong, L.; Yao, F.: SOC estimation of lithium battery based on the combination of electrical parameters and FBG non-electrical parameters and using NGO-BP model, OPT FIBER TECHNOL, 2023, 81, 103581. https://doi.org/10.1016/j.yofte.2023.103581

[26] Gargominy, O.; Butaud, J.; Fontaine, B.; Prié, V.; Zuccon, D.; Tercerie, S.; Hauata, J.; Hauata, M.: RESEARCH NOTE Discovery of the only living population of Pupoidopsis hawaiensis (Gastropoda: Pupillidae) in the last 50 years. https://mnhn.hal.science/mnhn-03084456

[27] Huang, J.; Guo, W.; Wei, R.; Yan, M.; Hu, Y.; Qin, T.: Short‐term power forecasting method for 5G photovoltaic base stations on non‐sunny days based on SDN‐integrated INGO‐BP and RGAN, IET RENEW POWER GEN, 2024. https://doi.org/10.1049/rpg2.12943

[28] Lin, L.; He, H.; Xu, Z.; Wu, D.; Raja, S.-P.: Realtime Vehicle Tracking Method Based on YOLOv5 + DeepSORT, COMPUT INTEL NEUROSC, 2023, 2023, 7974201. 1 https://doi.org/0.1155/2023/7974201

[29] Bai, T.: Multiple Object Tracking Based on YOLOv5 and Optimized DeepSORT Algorithm, Journal of Physics: Conference Series, 2023, 2547, (1), 12022. https://doi.org/10.1088/1742-6596/2547/1/012022

[30] Basiri, A.; Jackson, M.; Amirian, P.; Pourabdollah, A.; Sester, M.; Winstanley, A.; Moore, T.; Zhang, L.: Quality assessment of OpenStreetMap data using trajectory mining, GEO-SPAT INF SCI, 2016, 19, (1), 56-68. https://doi.org/10.1080/10095020.2016.1151213