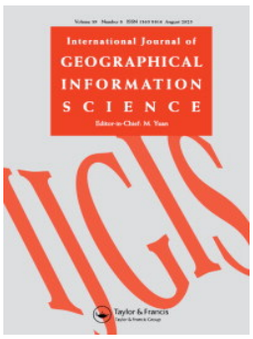

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Striking a balance between diversity and regularity: a preference-guided transformer for individual mobility prediction

Guangyue Li, Yang Xu, Zhipeng Gui, Xiaogang Guo & Luliang Tang

To cite this article: Guangyue Li, Yang Xu, Zhipeng Gui, Xiaogang Guo & Luliang Tang (28 Jul 2025): Striking a balance between diversity and regularity: a preference-guided transformer for individual mobility prediction, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2025.2534159

To link to this article: https://doi.org/10.1080/13658816.2025.2534159

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2025.2534159

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Striking a balance between diversity and regularity: a preference-guided transformer for individual mobility prediction

Guangyue Li[a], Yang Xu[a,b], Zhipeng Gui[c], Xiaogang Guo[d] and Luliang Tang[d]

[a]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China; [b]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China; [c]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China; [d]State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China

ABSTRACT

Human mobility modeling and prediction are central research topics in GIScience. Although deep learning has led to significant advances in these fields, existing trajectory prediction models still face challenges in capturing the complexity of individual mobility behavior. Regression-based models often overestimate the diversity of human mobility, whereas classification models tend to underestimate it. This study attributes these biases to the models' limitations in recognizing the spatial relationships among activity locations and mobility heterogeneity across individuals. To address these challenges, we propose the Spatial Preference Map-based Transformer (SPM-Former), explicitly integrating spatial proximity and mobility heterogeneity to enhance trajectory sequence prediction. To capture individual mobility characteristics, SPM-Former utilizes the Spatial Preference Map (SPM) to represent individuals' spatial visitation preferences and adjacency relationships between locations. Then, we introduce two encoding modules to decode the information hidden within the SPM: one for encoding trajectory-level spatial-temporal information and another for embedding individual-level overall mobility features. Furthermore, we propose a novel optimization method, SPM-Loss, to assess prediction accuracy from the global spatial distribution perspective. Experimental results on a large-scale dataset from Japan demonstrate that SPM-Former outperforms state-of-the-art classification-based models, achieving approximately 3% and 20% improvements in trajectory sequence similarity and overall spatial feature similarity, respectively.

ARTICLE HISTORY
Received 29 November 2024
Accepted 10 July 2025

KEYWORDS
Human mobility; sequence prediction; mobility heterogeneity; spatial preference; GeoAI

1. Introduction

Human mobility prediction has long been a fundamental topic of geographical information science (GIScience) (Barbosa *et al.* 2018). By accurately forecasting individual

CONTACT Yang Xu yang.ls.xu@polyu.edu.hk

trajectories, researchers can reveal the spatial and temporal patterns of human movement, crucial for a range of applications including traffic modeling (Li *et al*. 2024b), urban planning (Yang *et al*. 2025), and epidemic simulation (Perofsky *et al*. 2024). One potential application involves integrating mobility predictions with disease transmission theories to more precisely identify hotspots and develop more effective prevention and control strategies. Over the past two decades, the widespread adoption of global positioning systems (GPS), rapid advancements in communication technologies, and the extensive use of mobile phones have contributed to an exponential increase in human mobility data (Huang *et al*. 2021). As these data continue to accumulate, large-scale and open-source human mobility datasets have emerged (Yabe *et al*. 2024b), providing unprecedented opportunities for long-term large-scale individual human mobility prediction.

In the field of individual prediction, two primary tasks are identified: next location prediction (Feng *et al*. 2018) and trajectory sequence prediction (Yabe *et al*. 2024a). The former task predicts an individual's next probable location, whereas the latter involves forecasting longer trajectory sequences, such as weekly movements. While sequence prediction provides deeper insights into the regularity and periodicity of human mobility, it also poses greater challenges compared to next location prediction.

Nevertheless, whether focusing on next location or longer trajectory sequences, existing human mobility prediction models typically rely on two main frameworks: regression and classification (Pappalardo *et al*. 2019). In regression tasks, the goal is to predict the latitude and longitude of possible locations, aiming to minimize the deviation between the predicted and actual coordinate sequences (Brébisson *et al*. 2015). In classification tasks, each visited location is treated as a distinct class, and the model predicts the most probable locations from all potential classes (Feng *et al*. 2018, Gambs *et al*. 2012). However, when researchers apply these two architectures to human mobility prediction tasks, they discover that each architecture poses distinct challenges. In Figure 1, we illustrate three typical trajectories and their overall spatial distributions to demonstrate these deviations, originating from real-world scenarios, predictions from regression-based models, and predictions from classification-based models, respectively. First, in regression tasks, Guo *et al*. (2023) find that such models

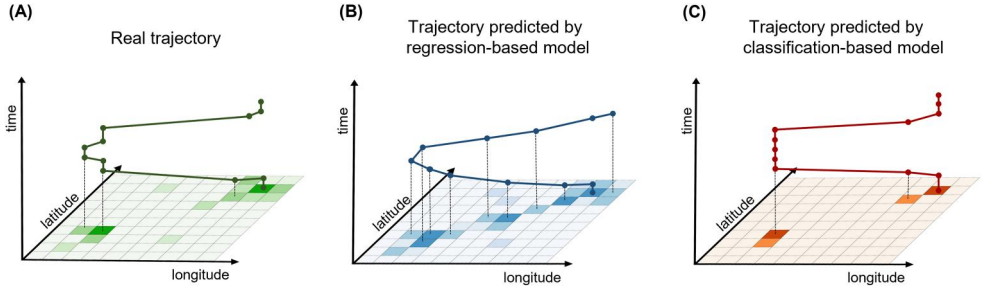

Figure 1. Representative trajectory examples, including the true trajectory (shown in green), a regression-based prediction (blue), and a classification-based prediction (orange). The underlying grids show the spatial distributions of these trajectories, with darker colors indicating higher visit frequencies. For clarity, the trajectories are standardized to a 12-hour period, with points marking the hourly locations. Note that the spatial distributions below reflect a longer time period and are not strictly aligned with the displayed trajectories.

often overestimate the diversity in human mobility while underestimating its regularity. As shown in Figure 1(B), predictions generated by regression-based models are often scattered across the spatial area of the actual trajectory and may even cluster in regions between frequently visited locations, where real trajectories rarely pass. Second, in classification tasks, Terashima *et al*. (2023) observe that these models tend to overestimate the regularity of travel behavior while underestimating its diversity. As depicted in Figure 1(C), predictions from classification-based models tend to focus on a few frequently visited hotspots, showing minimal exploratory behavior in surrounding areas (Pappalardo *et al*. 2015).

We believe these biases arise from the inherent complexities of human trajectories as well as the underlying principles of the two task architectures. Human mobility behavior often exhibits a tendency to return to familiar locations, such as home or workplace (Song *et al*. 2010a). This tendency results in several prominent, frequently visited areas within an individual's spatial distribution. At the same time, human movement patterns are complex and dynamic, rather than uniformly distributed across space. Regression-based models attempt to learn individual mobility patterns by minimizing the overall distance between predicted and actual sequences. However, this approach fails to evaluate the precision of predictions at each time step. For instance, when predictions are close to frequently visited areas, the overall loss remains low—even if the predictions do not perfectly align with the actual trajectory. In some cases, predictions located between two frequently visited locations may yield even smaller distance discrepancies. Consequently, regression models often struggle to capture true spatial visitation preferences, thereby overestimating the diversity of human mobility behavior and ultimately reducing prediction accuracy.

In classification tasks, models view spatial locations as distinct categories rather than a continuous space. This categorical format enables them to identify frequently visited locations and model their relationships. However, the inherent spatial segmentation in classification-based models limits their ability to capture spatial proximity. Because of their categorical nature, each location is treated as an isolated unit, thereby ignoring the spatial relationships among adjacent areas. As a result, predictions tend to focus on the most probable categories rather than being distributed across all commonly visited areas, which does not reflect the exploratory nature of human mobility (Song *et al*. 2010a). Although some data-driven methods have partially captured the relationships between locations (Terashima *et al*. 2023), these relationships remain indirect and limited. Consequently, the insufficient spatial proximity awareness causes classification-based models to underestimate the inherent diversity of human mobility patterns.

Moreover, another contributing factor is the models' limited ability to recognize individual mobility heterogeneity. Pappalardo *et al*. (2015) highlight significant differences in mobility patterns among individuals: some, known as explorers, tend to venture into new areas, while others, referred to as returners, frequently revisit familiar places such as home or the workplace. Notably, even explorers exhibit a tendency to return to familiar locations. Basic classification-based models often struggle to distinguish this individual heterogeneity, typically favoring common return patterns. Thus,

these models tend to predict trajectories that consistently return to familiar locations rather than venturing into new, unexplored areas.

To address these challenges, we propose a novel Spatial Preference Map-based Transformer (SPM-Former) to achieve a balance between diversity and regularity in individual mobility prediction. Given that classification-based models currently demonstrate higher accuracy in mobility sequence prediction (Yabe *et al*. 2024a), SPM-Former extends this framework by explicitly integrating spatial proximity (Mai *et al*. 2022) and individual mobility heterogeneity into the model. We introduce the Spatial Preference Map (SPM), which quantifies mobility characteristics by recording visit frequencies at various spatial locations, reflecting overall travel behavior and mobility heterogeneity. Moreover, the SPM encompasses the entire study area, providing the adjacency relationships among all locations. We then decompose the spatial proximity and individual heterogeneity information encoded in the SPM by means of both sequence-level and holistic-level encoding, and explicitly integrate this information into the attention mechanism to bridge existing research gaps. Finally, we propose a new evaluation criterion, SPM-Loss, which assesses prediction accuracy based on the overall spatial distribution of trajectories. The main contributions of this paper are summarized as follows:

1. To enable classification-based models to be aware of spatial proximity and individual mobility preferences, we propose a novel SPM-based Spatial Encoding module for trajectory sequence representation comprising two sub-modules: Proximity Encoding, which encodes absolute spatial relationships, and Tendency Encoding, which integrates visit frequency information. By tightly integrating the SPM with trajectory sequences, the Spatial Encoding module explicitly represents spatial relationships among all locations while effectively capturing individual mobility preferences.
2. To explicitly account for individual mobility heterogeneity, we introduce the Individual Mobility Feature Embedding (IME), which leverages the SPM to capture comprehensive travel behavior of individuals. The process begins with feature engineering, where classical mobility indicators—such as the radius of gyration and activity entropy—are extracted from SPM to quantify individual mobility range and diversity. These features are then embedded into a high-dimensional representation through embedding layers, thereby enhancing the classification-based model's ability to distinguish individual mobility heterogeneity.
3. Given SPM's effectiveness in representing overall mobility characteristics, we introduce a SPM-based loss function, SPM-Loss, to assess the similarity between the spatial distributions of predicted and actual trajectory sequences. The combined use of SPM-Loss and cross-entropy loss evaluates the alignment between sequences from the perspectives of spatial distribution and sequence similarity, respectively, providing a more comprehensive assessment.

Experiments conducted on a real-world open-source human mobility dataset (Yabe *et al*. 2024b) from a metropolitan area of Japan, demonstrate that the proposed SPM-Former outperforms other classification-based baseline models (Terashima *et al*. 2023, Kim *et al*. 2023) in both trajectory sequence similarity and overall trajectory feature

similarity. Furthermore, ablation studies confirm the effectiveness of each module within SPM-Former.

The article is structured as follows: Section 2 reviews human mobility concepts, related prediction models, and current challenges. Section 3 describes the dataset and the specific trajectory sequence prediction problem addressed. Section 4 details the construction steps of SPM and the proposed SPM-Former, including its internal modules. Section 5 presents the experiments conducted to validate the proposed model. Finally, Section 6 provides conclusions and outlines future research directions.

2. Research background

2.1. Human mobility: concepts and characteristics

Human mobility refers to the movement behavior of individuals or groups across spatial and temporal dimensions (Barbosa *et al*. 2018). The extensive accumulation of geolocation data has enabled numerous quantitative analyses of human mobility patterns (Xu *et al*. 2021, Shaw and Sui 2021), uncovering several valuable characteristics. One of the most notable findings is the high degree of regularity in human mobility behavior. Song *et al*. (2010b) demonstrated that although individuals may visit multiple locations, most of their time is spent in a few key places, such as home and workplace. Building on this, they proposed a mechanistic model that simulates the two primary human mobility behaviors: exploration and preferential return. This model further supports the notion that human trajectories are not random but follow statistical laws (Song *et al*. 2010a). Such regularity endows human mobility behavior with high predictability, forming the theoretical basis for various mobility prediction models.

In addition, Pappalardo *et al*. (2015) identified two groups with distinct individual mobility behaviors: returners and explorers. Returners' mobility predominantly centers around a few fixed locations, with their activity range constrained by these places. In contrast, explorers visit a greater number of different locations more frequently, displaying a broader activity range and more random behavior patterns. This highlights the significant heterogeneity in human mobility, evident in the variations in individuals' movement ranges, frequencies, and distances. Addressing and representing this heterogeneity remains a challenge in human mobility prediction (Zhou *et al*. 2024).

2.2. Human mobility prediction: next-location and sequence models

Although next-location prediction and trajectory sequence prediction are two distinct tasks, they share many similarities. Both tasks involve predicting an individual's future locations based on their historical movement data, and the methods used to address these tasks often share similar model structures, such as recurrent neural networks (RNNs) and attention mechanisms. This section provides a comprehensive overview of the models and methods used to address these two types of tasks.

In next-location prediction, traditional approaches are typically based on probabilistic models. Gambs *et al*. (2012) introduced a Mobility Markov Chain (MMC), where the states represent points of interest (POIs), and the transitions correspond to movements between these POIs. By learning the transition probabilities between consecutive

locations, MMCs can predict the next visited location with relative accuracy. Although effective on small-scale datasets, these methods require substantial feature engineering and struggle to capture long-term spatial-temporal dependencies (Luca *et al*. 2023).

With advancements in artificial intelligence, researchers have adopted deep learning models such as Recurrent Neural Networks (RNNs) (Chu *et al*. 2024), Graph Neural Networks (GNNs) (Li *et al*. 2021), and attention mechanisms (Li *et al*. 2024a) to capture complex spatial-temporal dependencies in human mobility. More recently, Transformer-based models have gained significant attention due to their ability to efficiently model long-range dependencies. Unlike traditional RNNs, which process data sequentially, Transformers leverage self-attention mechanisms to assign different importance weights to time steps in a trajectory (Vaswani 2017). This advantage has led to the adaptation of Transformer architectures for mobility prediction tasks, enabling more effective feature extraction and improving predictive accuracy. For instance, DeepMove (Feng *et al*. 2018) is a notable model that employs an attentive recurrent network to predict human movement from long and sparse trajectory data. By integrating a multimodal embedding module, DeepMove combines historical and current trajectory data to generate dense representations of spatial-temporal and individual features. This allows DeepMove to effectively consider historical characteristics of individual trajectories, resulting in more accurate next-location predictions.

Compared to next-location prediction, trajectory sequence prediction requires longer time-series data and larger-scale human mobility datasets, making it a less explored research area. A significant milestone in this field was the release of the YJMob100K dataset (Yabe *et al*. 2024b). Developed from mobile location data, this open-source, anonymized longitudinal dataset provides comprehensive coverage of the mobility trajectories of 100,000 individuals over a 90-day period in metropolitan areas. Then, Yabe *et al*. (2024a) hosted the HuMob Challenge 2023 at the ACM SIGSPATIAL conference, which aimed to benchmark the accuracy of various approaches for human mobility sequence prediction tasks, specifically focusing on two-week trajectory sequence predictions.

In this competition, various trajectory sequence prediction methods were proposed, including statistical models, machine learning techniques, and deep learning approaches. Guo *et al*. (2023) introduced a statistic-based prediction method that models human mobility periodicity with time decay, enabling efficient and generalized trajectory predictions. Suzuki *et al*. (2023) developed a machine learning approach using Support Vector Regression (SVR) to create personalized prediction models tailored to each user's travel characteristics. The highest accuracy was achieved by Terashima *et al*. (2023) with the Location Prediction BERT (LP-BERT) model, adapted from the original BERT model for natural language processing (Devlin 2018). LP-BERT employed a unique masking technique that allowed the model to perform continuous long-term predictions, significantly enhancing prediction accuracy.

2.3. The need for spatial preference map-based transformer

In the field of human mobility prediction, despite the variations in the deep learning models employed, their core tasks typically fall into two categories: regression or

Table 1. Classic models in human mobility prediction.

Task	Reference	Name	Modules	Task
Next-Location Prediction	Feng *et al.* (2018)	DeepMove	Attention, GRU	Classification
	Lv *et al.* (2018)	T-CONV	CNN	Regression
	Yang *et al.* (2020)	Flashback	Attention, RNN	Classification
	Tang *et al.* (2021)	CLNN	LSTM, Embedding	Regression
	Gui *et al.* (2021)	LSI-LSTM	LSTM, Attention	Regression
	Hong *et al.* (2023)	MHSA	Attention, Embedding	Classification
Sequence Prediction	Li *et al.* (2020)	HTAED	LSTM, Attention	Classification
	Guo *et al.* (2023)	P-Former	Attention, Embedding	Regression
	Terashima *et al.* (2023)	LP-BERT	Attention, Embedding	Classification
	Solatorio (2023)	GeoFormer	GPT2	Classification
	Wang and Deng (2023)	MSCNN	RNN, Late Fusion	Classification

classification (Luca *et al.* 2023). Table 1 summarizes representative models for next-location prediction and trajectory sequence prediction, detailing the deep learning modules they utilize and their task type (i.e., classification or regression). These studies predominantly select one of these two approaches as their basis.

When mobility prediction is approached as a regression task, the predicted outputs typically consist of geographic coordinate sequences (Brébisson *et al.* 2015). In this context, the model aims to minimize the Euclidean or Haversine distances between the predicted and actual coordinate sequences (Pappalardo *et al.* 2019). Unlike classification tasks, which predict discrete categories, regression tasks give continuous numerical predictions. Regression tasks are commonly applied in various time series prediction scenarios, such as traffic forecasting (Li *et al.* 2024b), weather prediction (Huang *et al.* 2013), and stock market forecasting (Kumar *et al.* 2021). A commonly used loss function in regression tasks is the Mean Squared Error (MSE) loss, as defined in Equation (1):

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2 \tag{1}$$

where L represents the MSE loss value, N is the number of samples, y_i is the true value of the i-th sample, and $\widehat{y_i}$ is the predicted value of the i-th sample.

On the other hand, when mobility prediction is approached as a classification task, each possible location is treated as an independent class, and the model selects the most probable outcome from these classes (Feng *et al.* 2018). Classification tasks are widely applied in various domains, including image classification (Law *et al.* 2020), text classification (Nilsson and Delmelle 2023), and disease diagnosis (Aggarwal *et al.* 2021). To optimize the classification-based models, the cross-entropy loss function is typically employed, as defined in Equation (2). This function can effectively measure the divergence between the predicted class distribution and the true class distribution, applying greater penalties to samples where there is a significant difference between the predicted probabilities and true labels:

$$L = -\sum_{i=1}^{N} y_i \log\left(\widehat{y_i}\right) \tag{2}$$

where L represents the cross-entropy loss value, N is the number of classes, y_i is the indicator of the true class label (i.e., $y_i = 1$ if class i is the true class, otherwise $y_i = 0$), and $\widehat{y_i}$ is the probability that the model predicts class i.

Despite numerous methods achieving relatively accurate predictions, both regression- and classification-based tasks exhibit inherent biases when applied to trajectory sequence prediction. Regression-based models often overestimate the diversity of human movement and generate intermediate points between frequently visited locations (Guo *et al*. 2023). This arises from the MSE loss function's tendency to tolerate small average errors, even if the predicted points do not coincide with real trajectories. In contrast, classification-based models typically overestimate regularity, emphasizing a few primary hotspots while neglecting exploratory behavior (Terashima *et al*. 2023). This limitation stems from treating each location as an independent class while overlooking individual heterogeneity, causing most predictions to align with returner-like patterns (Pappalardo *et al*. 2015).

Given current technological limitations, completely discarding these two task architectures in favor of entirely new methods is impractical. Classification-based models generally outperform regression-based models in terms of prediction accuracy (Terashima *et al*. 2023), and many powerful large language models developed in recent years are based on classification task frameworks (Brown 2020). Therefore, continuing research within the classification paradigm seems a more practical approach. The main challenge lies in enhancing classification models to integrate spatial proximity and individual heterogeneity, enabling a better understanding and prediction of human mobility.

To address this challenge, we propose the Spatial Preference Map-based Transformer (SPM-Former). By quantifying individual visit frequencies across different locations and clearly representing relationships between adjacent locations, SPM-Former effectively compensates for the limitations of traditional classification-based models in capturing complex spatial-temporal dependencies and individual heterogeneity, leading to improved prediction accuracy and diversity.

3. Data and preliminary

The dataset used in this study is the official dataset from the HuMob Challenge 2023 (Yabe *et al*. 2024a), provided by Yahoo Japan Corporation. It comprises synthetic but realistic human mobility data for 100,000 individuals over a 75-day period within an urban area (Yabe *et al*. 2024b). The dataset is collected from a mid-sized, densely populated urban region in Japan. It can be observed from Figure 2(A), visualization of the trajectory distribution indicates that the spatial characteristics of this area closely resemble those of the Nagoya metropolitan region. As illustrated in Figure 2(B), the entire urban area is divided into a grid of 500×500 m cells, resulting in a total of 200×200 grid units. The human mobility dataset includes the movement trajectories of 100,000 individuals over 75 days of regular activity, recorded at 30-minute intervals and represented using the 500 m grid cells. Each individual's trajectory at any given time is represented by four values (d, t, x, y), where d denotes the day of sampling, t indicates the specific time point (with 48 time points per day due to the 30-minute

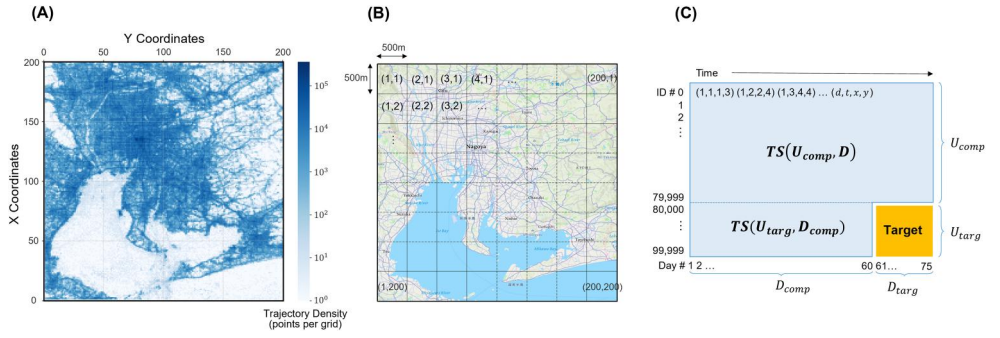

Figure 2. (A) Visualization of the distribution of all trajectory points in the dataset (rotated 90° clockwise); (B) Grid partitioning method of the study area; (C) Illustration of the prediction task. Target stands for $TS(U_{targ}, D_{targ})$.

Table 2. Summary of notations and descriptions.

Notations	Descriptions	Notations	Descriptions
U_i	User with ID i	d	The day of sampling, $d \leq 75$
t	The time of sampling, $t \leq 48$	x, y	The grid coordinates of sampling, $x, y \leq 200$
$TS(U_i, D)$	U_i's trajectory sequence over a time period	U_{comp}	Users with complete trajectory sequences
U_{targ}	Target users with partially missing trajectories	D_{comp}	Days when U_{targ} have complete trajectories
D_{targ}	Days when U_{targ} have missing trajectories	$mask$	Mask set for the pre-trained model
h	Hidden size	T	Length of the trajectory sequence

intervals), and x and y are the horizontal and vertical coordinates of the grid cell where the trajectory point is located.

The objective of this study aligns with that of the HuMob Challenge 2023, which is to predict long-term trajectory sequences for a large number of individuals. As shown in Figure 2(C), the dataset includes 100,000 participants, with complete trajectory data provided for 80,000 individuals, denoted as $TS(U_{comp}, D)$, for the purpose of training models to learn general human mobility patterns. For the remaining 20,000 individuals, trajectory data for the first 60 days is provided, denoted as $TS(U_{targ}, D_{comp})$, allowing models to capture the individual heterogeneity of these users. Finally, the trajectories of these 20,000 individuals for days 61 to 75 are masked, serving as the target for prediction and denoted as $TS(U_{targ}, D_{targ})$. The primary task of this study is to develop an effective prediction method F to uncover human mobility patterns from large-scale trajectory data. For clarity, the main symbols used in this paper are summarized in Table 2.

4. Method

4.1. Overview of spatial preference map-based transformer

In this study, we propose a novel trajectory sequence prediction method called Spatial Preference Map-based Transformer (SPM-Former), as illustrated in Figure 3. SPM-Former randomly masks 15-day sequences within each user's historical trajectory data in the training set during the data preprocessing stage. This duration was selected to align with the HuMob Challenge 2023, which mandated a 15-day prediction horizon. Furthermore, a half month period is considered inherently suitable as it typically covers two full weekly cycles, enabling the model to capture and learn distinct human

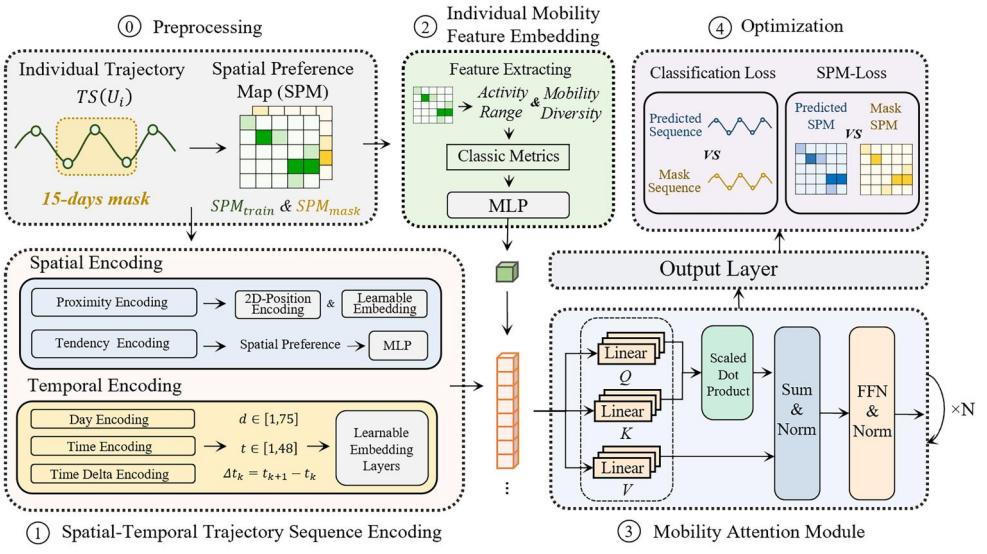

Figure 3. Spatial preference map-based transformer (SPM-Former).

mobility patterns associated with weekdays and weekends. The unmasked sequences serve as training data, while the masked sequences are used as prediction targets. Subsequently, SPM-Former constructs the Spatial Preference Map (SPM) for both the training data and prediction targets, representing each individual's spatial movement characteristics. SPM-Former consists of three main functional modules and a loss function for optimization, as illustrated in Figure 3.

1. Spatial-Temporal Trajectory Sequence Encoding: This module encodes the spatial-temporal features of trajectory sequences. For spatial encoding, information such as spatial proximity and location visit frequency abstracted from the SPM, is explicitly integrated into the trajectory sequence tensor. For temporal encoding, variables such as number of days, time of day, and the duration of missing data are included.
2. Individual Mobility Feature Embedding: This module extracts and quantifies overall individual characteristics from the SPM using classic metrics in human mobility research. These features, serving as global factors, help SPM-Former better capture individual mobility heterogeneity.
3. Mobility Attention Module: This module leverages the attention mechanism to capture spatial-temporal dependencies within trajectory sequences.
4. Spatial Preference Map-based Loss: During the model optimization phase, we introduce a novel loss function, SPM-Loss, to evaluate the similarity between predicted and actual sequences at the overall spatial feature level. This loss function is combined with the traditional multi-class loss function to create a multi-loss evaluation criterion.

4.2. Preprocessing of individual trajectories

As shown in Figure 3, the trajectory of a user U_i in the training dataset can be represented as: $TS(U_i) = ((d_1, t_1, x_1, y_1), (d_2, t_2, x_2, y_2) \ldots, (d_T, t_T, x_T, y_T))$, where T denotes the

length of the trajectory sequence. To enhance the model's long-term prediction capability, we divide each individual's trajectory sequence and randomly select a continuous 15-day segment for masking. This approach allows the model to infer the masked portion based on the remaining complete trajectory sequence. During masking, we employ a random time segment selection method instead of fixing the mask to the last 15 days, which helps improve the model's generalization capability. Fixing the mask at the end of the sequence might lead the model to learn temporal biases. Specifically, the coordinates x and y of the masked sequence are set to -1, while d and t remain unchanged to enable the model to leverage temporal information for inferring the current location. This process yields a masked trajectory sequence for each individual, denoted as $TS(U_i, mask)$, and the complete remaining sequence used for training, denoted as $TS(U_i, train)$.

Subsequently, we generate corresponding Spatial Preference Maps (SPMs) for both the masked and training sequences, as shown in Figure 3. The SPM quantifies each user's mobility characteristics at the overall spatial level, capturing visit frequencies at different spatial locations. For instance, the SPM of an explorer typically contains a larger number of visited locations, while that of a returner is concentrated in a few high-frequency areas. Integrating SPM into the deep learning model allows for explicit embedding of individual mobility heterogeneity. The SPM generation process is straightforward. Given a trajectory sequence, each trajectory point's coordinates are mapped to the corresponding position in an initially zero-valued matrix, with the value at that position incremented by one for each visit. This frequency matrix, denoted as FM, is updated as follows:

$$FM(x_k, y_k) = FM(x_k, y_k) + 1 \tag{3}$$

where (x_k, y_k) represents the coordinates of each trajectory point in $TS(U_i)$. After accumulating the visit frequencies, the matrix FM is normalized using min–max normalization to scale its value range to [0,1]. The normalization formula is as follows:

$$SPM = FM_{norm}(x, y) = \frac{FM(x, y) - FM_{\min}}{FM_{\max} - FM_{\min}} \tag{4}$$

where $FM_{\max}$ and $FM_{\min}$ represent the maximum and minimum values in the matrix FM, respectively. The normalized matrix $FM_{norm}(x, y)$ forms the spatial preference map, which is a 200×200 matrix consistent with the study area's size and represents individual mobility characteristics. To prevent information leakage, we generate two separate SPMs using the above approach. SPM_{train} is generated from $TS(U_i, train)$, which contains only the unmasked trajectory points, allowing the model to learn meaningful mobility patterns. In contrast, SPM_{mask} is computed based on the original values of the masked sequence $TS(U_i, mask)$, preserving its spatial distribution for later validation.

4.3. Spatial-temporal trajectory sequence encoding

This section introduces the Spatial-Temporal Trajectory Sequence Encoding (TSE) module, as depicted in Figure 3, which is designed to explicitly integrate spatial-temporal information at the trajectory sequence level into the model. The TSE module comprises two main components: Spatial Encoding and Temporal Encoding. Spatial

Encoding is based on the spatial preference map, and embeds spatial proximity relationships and location visit preferences into the sequence to reflect individual spatial mobility characteristics. Temporal Encoding incorporates the time of data collection and information about missing periods into the sequence to capture the temporal dependencies of the trajectory. The following sections provide a detailed explanation of these two components.

4.3.1. Spatial encoding

Previous prediction models (Terashima *et al*. 2023) often employ data-driven approaches to capture spatial relationships between locations, using trainable embedding modules to learn inter-location dependencies from a large volume of trajectory sequences. While this method can establish spatial dependencies to some extent, it does not explicitly represent spatial proximity. After training, such models can only infer the visit intensity or dependency between different locations without clearly identifying adjacent spatial units.

To effectively capture spatial dependencies and individual movement patterns, we propose a Spatial Encoding module based on SPM, which explicitly integrates absolute spatial location information and mobility characteristics into trajectory representations. To achieve this, we design two complementary encoding mechanisms: Proximity Encoding, which extracts spatial adjacency information to ensure that nearby locations maintain similar feature representations; and Tendency Encoding, which embeds an individual's location visit preferences into trajectory modeling to account for mobility heterogeneity.

1) Proximity Encoding: This module is designed to explicitly encode spatial proximity relationships, ensuring that adjacent spatial locations have similar hidden features after encoding (Mai *et al*. 2022). As a comprehensive spatial representation, the spatial preference map provides adjacency relationships among 200×200 locations. Therefore, the Proximity Encoding module can construct a meaningful encoding based on the positional relationships in the SPM.

To achieve this, we draw inspiration from the positional encoding methods used in attention mechanisms (Vaswani 2017). Positional encoding provides high-dimensional representations of positions in a one-dimensional sequence, maintaining similarity for adjacent positions in the encoding space and possessing directional characteristics. As shown in Figure 4, we visualize the features of positional encoding with a length of 50 and a hidden dimension of 64, which clearly demonstrates both directionality and proximity. We extend this method to two-dimensional space by separately encoding the *x*-axis and *y*-axis positions, then concatenating them to form a 2D Position Encoding (2*DPE*), as depicted in Figure 4. The 2*DPE* captures absolute positional relationships in two-dimensional space and ensures the similarity of adjacent regions in the encoding, effectively reflecting spatial proximity. The detailed calculation process is as follows.

$$pe_{(x,2i)} = \sin\left(\frac{x}{10000^{\frac{2i}{h/2}}}\right), pe_{(x,2i+1)} = \cos\left(\frac{x}{10000^{\frac{2i}{h/2}}}\right), pe_{(x)} \in R^{T\times h/2} \tag{5}$$

$$pe_{(y,2i)} = \sin\left(\frac{y}{10000^{\frac{2i}{h/2}}}\right), pe_{(y,2i+1)} = \cos\left(\frac{y}{10000^{\frac{2i}{h/2}}}\right), pe_{(y)} \in R^{T\times h/2} \tag{6}$$

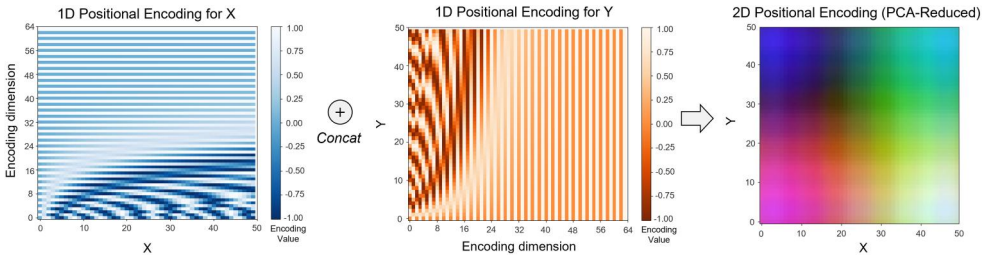

Figure 4. Visualization of 1D Positional Encodings for the X and Y Axes, and PCA-Reduced 2D Positional Encoding Heatmap.

$$2DPE_{(x,y)} = Concat(pe_{(x)}, pe_{(y)}), 2DPE_{(x,y)} \in R^{T\times h} \tag{7}$$

where $x, y \in Z$ represent the horizontal and vertical coordinates of each position in the SPM, $i \in [0, h/2]$ indicates the index of the encoding dimension, and h denotes the total dimension of Proximity Encoding. The functions $pe_{(x)}$ and $pe_{(y)}$ independently encode positional information for the X and Y axes. By concatenating these embeddings, the final $2DPE$ ensures that each spatial location maintains both absolute positioning information and local spatial coherence.

To illustrate the effectiveness of our proposed $2DPE$, we visualize it in Figure 4. We compute the $2DPE$ encoding for each location within a 50×50 spatial grid for better interpretability, with the encoding dimension h set to 64 as a standard choice in deep learning models. For better visualization, we apply principal component analysis (PCA) to reduce the 64-dimensional encoding to three principal components, which are then mapped to the RGB color channels, as shown in Figure 4. The 1D positional encodings of the X and Y axes exhibit a monotonic variation along both spatial dimensions, indicating that the encoding effectively captures directional continuity. Furthermore, in the PCA-reduced 2D positional encoding heatmap, adjacent spatial positions exhibit consistent color patterns, suggesting that the encoding maintains local spatial coherence and preserves spatial proximity in the high-dimensional space. Thus, the proposed $2DPE$ provides a structured and absolute positional representation for two-dimensional space. By ensuring that adjacent spatial locations exhibit similar encoding values, $2DPE$ facilitates the modeling of spatial dependencies.

Using the $2DPE$ encoding method, each trajectory point (x_k, y_k) in a user's trajectory sequence $TS(U_i)$ is encoded and mapped to its corresponding $2DPE$ value, $2DPE(x, y)$. For masked trajectory points, their $2DPE$ encoding values are set to -1, enabling the model to identify and handle missing data. To further enhance spatial representation, we introduce learnable embedding vectors $E_x(x)$ and $E_y(y)$ for all 201 possible coordinate positions (including masked positions). These embeddings allow the model to learn adaptive spatial relationships, rather than relying solely on fixed positional encodings. Finally, the Proximity Encoding (PE) for a user U_i is obtained by adding the $2DPE$ and the learnable embeddings:

$$PE(U_i) = 2DPE(x_k, y_k) + E_x(x_k) + E_y(y_k), k \in [1, T] \tag{8}$$

where $PE(U_i) \in R^{T\times h}$ represents the final proximity encoding for user U_i.

2) Tendency Encoding: While Proximity Encoding ensures that spatially adjacent locations maintain similar feature representations, it does not explicitly account for how frequently an individual visits specific locations. Understanding the variation in visit intensity across different locations in a trajectory sequence is crucial for predicting human mobility behavior (Feng *et al*. 2018). For example, when an individual leaves their residence, they may either head to another high-frequency location (e.g., workplace) or engage in exploratory activities nearby (e.g., walking).

To address this, we propose Tendency Encoding, which embeds the location preference information from the SPM into the model, as depicted in Figure 3. This allows the model to recognize and leverage individual spatial preferences, enhancing its understanding of how location visit intensity influences the trajectory sequence and improving prediction accuracy. Specifically, we generate the visit tendency information $F(U_i)$ for each trajectory point (x_k, y_k) in the trajectory sequence $TS(U_i)$ by looking up the visit frequency of the corresponding position in SPM:

$$F(U_i) = (f(x_1, y_1), f(x_2, y_2), \ldots, f(x_T, y_T)) \tag{9}$$

where $f(x_k, y_k)$ represents the visit frequency of the point (x_k, y_k) in SPM. For masked trajectory points, their corresponding tendency values are set to -1 to enable the model to recognize and handle missing data. Next, the visit tendency sequence $F(U_i)$ is input into a multi-layer perceptron (MLP) to generate the tendency encoding $TEN(U_i)$ with a hidden dimension h:

$$TEN(U_i) = MLP(F(U_i)), TEN(U_i) \in R^{T \times h} \tag{10}$$

Finally, the Spatial Encoding (SE) is obtained by adding the proximity encoding $PE(U_i)$ and tendency encoding $TEN(U_i)$, as follows:

$$SE(U_i) = PE(U_i) + TEN(U_i), SE(U_i) \in R^{T \times h} \tag{11}$$

4.3.2. Temporal encoding

In Temporal Encoding, we use three types of embeddings to incorporate temporal information into the model: Day Encoding, Time Encoding, and Time Delta Encoding, as shown in Figure 3. By encoding both the collection time and time gaps, the model can more accurately capture the temporal dependencies within the trajectory sequence, helping the model understand the temporal relationships and trends between trajectory points.

1) Day Encoding: In Day Encoding, we process the date information d_k of each trajectory point to capture the temporal relevance with respect to days. Specifically, for the trajectory sequence $TS(U_i)$, we generate an embedding for each date d_k and map it to a high-dimensional vector space to represent the specific day's information:

$$DE(U_i) = E(d_k), k \in [1, T] \tag{12}$$

This embedding captures the cyclical nature and patterns in the trajectory data, such as weekly or seasonal variations in human mobility patterns. For instance, different days of the week (e.g., weekdays vs. weekends) might exhibit different mobility patterns, which are effectively encoded in this step.

2) Time Encoding: In Time Encoding, similar to Day Encoding, we embed the collection time t_k of each trajectory point to capture the temporal features within a day. Specifically, for the trajectory sequence $TS(U_i)$, we apply an embedding to each time t_k as follows:

$$TIE(U_i) = E(t_k), k \in [1, T] \tag{13}$$

This time encoding introduces a time context into the trajectory sequence, allowing the model to distinguish activity patterns during different time periods, such as work hours, rest periods, or peak commuting times. By incorporating this temporal encoding, the model can capture finer-grained behavioral changes within the daily cycle, thereby improving the prediction of temporal dependencies in the trajectory data.

3) Time Delta Encoding: Due to the sparsity of trajectory data, significant time gaps may exist between consecutive trajectory points. To incorporate this interval information into the model, we utilize Time Delta Encoding, inspired by Terashima *et al.* (2023), which calculates and embeds the time difference between adjacent trajectory points to help the model understand temporal irregularities. Specifically, for the trajectory sequence $TS(U_i)$, we compute the time difference $\Delta t_k = t_{k+1} - t_k$ for each pair of consecutive points and represent it as a high-dimensional vector:

$$TDE(U_i) = E(\Delta t_k), k \in [1, T] \tag{14}$$

Time Delta Encoding provides the model with information about the duration between trajectory points, enabling it to better understand the irregularities caused by data sparsity. By embedding this time interval information, the model's ability to recognize discontinuities and gaps in the data is enhanced, improving the accuracy and robustness of trajectory predictions. Finally, Temporal Encoding (TE) is represented as the combination of Day Encoding, Time Encoding, and Time Delta Encoding, as shown below:

$$TE(U_i) = DE(U_i) + TIE(U_i) + TDE(U_i), TE(U_i) \in R^{T \times h} \tag{15}$$

The Spatial-Temporal Trajectory Sequence Encoding (TSE) is then generated by combining spatial encoding and temporal encoding to capture the comprehensive spatial-temporal characteristics of the trajectory sequence:

$$TSE(U_i) = SE(U_i) + TE(U_i), TSE(U_i) \in R^{T \times h} \tag{16}$$

4.4. Individual mobility feature embedding

This section introduces the Individual Mobility Feature Embedding (IME), which aims to extract and abstract users' overall mobility characteristics from the SPM to enable the model to more accurately differentiate the heterogeneity in individual mobility behavior. Unlike the Spatial-Temporal Trajectory Sequence Encoding, which focuses on trajectory sequence features, IME represents individual-level features and incorporates them into the model as global influence factors, as shown in Figure 3. This allows the model to better account for individual differences during prediction.

Specifically, IME quantifies a user's activity range and mobility diversity using four classic mobility metrics from human mobility research (Xu *et al.* 2021). These metrics

are selected to capture both spatial extent (radius of gyration) and behavioral diversity (entropy and visit count), enabling the model to distinguish between different movement tendencies, such as frequent explorers and routine-based returners. By integrating these features as global conditioning factors, the model can more effectively represent the unique aspects of individual mobility behavior.

To assess an individual's activity range, we employ two well-established metrics: the ordinary radius of gyration and the *k*-th radius of gyration (Xu *et al*. 2018). Leveraging the SPM, these metrics can be efficiently computed to assess mobility characteristics. The radius of gyration r_g measures the average distance of all visited locations relative to the centroid, reflecting the overall activity range of an individual. In the SPM, this is calculated by determining the distance from each coordinate point to the centroid, weighted by the visit frequency as follows:

$$r_g = \sqrt{\frac{1}{LN}\sum_{i=1}^{LN} f(x_k, y_k)\cdot((x_k - x_{cm})^2 + (y_k - y_{cm})^2)} \tag{17}$$

where LN is the total number of visited locations in SPM, $f(x_k, y_k)$ represents the normalized visit frequency at position (x_k, y_k), and (x_{cm}, y_{cm}) is the coordinate of the centroid of SPM. The ordinary radius of gyration provides insight into the scale of an individual's overall activity space, aiding in understanding the extent of their mobility range.

On the other hand, the *k*-th radius of gyration ${r_g}^{(k)}$ focuses on the top k locations with the highest visitation frequencies, measuring the activity distribution within these high-frequency areas. This metric is particularly useful for capturing the core regions of an individual's daily activities, highlighting the size of their main activity areas:

$${r_g}^{(k)} = \sqrt{\frac{1}{LN_k}\sum_{i=1}^{k} f(x_k, y_k)\cdot\left(\left(x_i - x_{cm}^{(k)}\right)^2 + \left(y_i - y_{cm}^{(k)}\right)^2\right)} \tag{18}$$

where LN_k is the total number of visits to the top k most frequently visited locations in SPM, $f(x_k, y_k)$ represents the normalized visit frequency at position (x_k, y_k), and $(x_{cm}^{(k)}, y_{cm}^{(k)})$ are the centroid coordinates based on these k positions. The ordinary radius of gyration r_g describes the overall activity range across all visited locations, while the *k*-th radius of gyration ${r_g}^{(k)}$ focuses on the activity distribution within core areas. Together, these metrics quantify and model an individual's activity range from the SPM, enhancing the model's understanding of individual spatial activity characteristics and heterogeneity.

To assess an individual's mobility diversity, we also use two well-established metrics: the total number of visited locations and activity entropy, both of which can be efficiently derived from the SPM. Total number of visited locations (*LN*) serves as an intuitive indicator of an individual's exploration range. It is determined by counting the number of positions in SPM that have non-zero visit frequencies, denoted as *LN*. In general, a greater number of visited locations indicates a higher potential for future exploratory behavior. On the other hand, activity entropy (*AE*) measures the diversity of an individual's activities across different locations. It quantifies the distribution of activities based on the visit frequency at each location in the SPM. First, we calculate the visit proportion for each location $p(x_k, y_k)$:

$$p(x_k, y_k) = \frac{f(x_k, y_k)}{\sum_{i=1}^{LN} f(x_k, y_k)} \tag{19}$$

where $f(x_k, y_k)$ represents the visit frequency at location (x_k, y_k), and $\sum_{i=1}^{LN} f(x_k, y_k)$ is the total visit frequency across all locations. Using the visit proportion, we can calculate the overall activity entropy:

$$AE = -\sum_{i=1}^{LN} p(x_i, y_i)\log(p(x_i, y_i)) \tag{20}$$

Activity Entropy reflects an individual's behavioral diversity and exploratory inclination. A higher entropy value indicates more dispersed activity, suggesting a stronger exploration tendency. To incorporate this information into the model, we combine these four metrics into a vector termed the Individual Mobility Metric (IMM), defined as: $IMM = [r_g, r_g{}^{(k)}, LN, AE]$, where r_g is the ordinary radius of gyration, $r_g{}^{(k)}$ is the k-th radius of gyration, LN is the total number of visited locations, and AE is the activity entropy. These features are then mapped into the embedding space through a Multi-Layer Perceptron (MLP), resulting in the Individual Mobility Embedding (IME), which is integrated into the model:

$$IME = MLP(IMM), IME \in R^h \tag{21}$$

In subsequent attention mechanism computations, since attention mechanisms can flexibly capture relationships between any points in a sequence, we position the IME at the start of the trajectory sequence encoding $TSE(U_i)$. This allows it to serve as a global user feature that interacts with each position, providing individual-specific guidance for subsequent predictions. IME ensures that the model references an individual's overall characteristics at each prediction step, enabling personalized predictions that account for individual heterogeneity.

4.5. Mobility attention module and output layer

1) Mobility Attention Module: As shown in Figure 3, the mobility attention module is built on the classic multi-head attention mechanism to learn spatial-temporal dependencies within the trajectory sequence. Given that the model has already embedded rich spatial proximity and individual heterogeneity information into the sequence, the attention mechanism can effectively capture correlations between trajectory points. The detailed computation process of the mobility attention module is as follows: First, the initial input to the mobility attention module is the concatenation of the overall user mobility features IME and the trajectory sequence information $TSE(U_i)$:

$$H_0 = Concat(IME, TSE(U_i)), H_0 \in R^{(T+1)\times h} \tag{22}$$

Next, the multi-head attention mechanism projects H_l into multiple subspaces to generate representations of queries (Q), keys (K), and values (V), capturing spatial-temporal associations. The calculations are as follows:

$$Q_l^i = H_l W_l^{(Q^i)}, K_l^i = H_l W_l^{(K^i)}, V_l^i = H_l W_l^{(V^i)}, where\ i \in [1, h] \tag{23}$$

$$A_l^i = Softmax\left(\frac{Q_l^i\left(K_l^i\right)^T}{\sqrt{h'}}\right) \tag{24}$$

$$MultiHeadT(H_l) = FC(concat\left(A_l^1V_l^1, A_l^2V_l^2, \ldots, A_l^iV_l^i\right)) \tag{25}$$

where $l \in [0, L-1]$ represents the index of the current layer, Q_l^i, K_l^i and V_l^i are the query, key, and value matrices for the i-th head, A_l^i is the attention weight, and $MultiHeadT(H_l)$ is the output of the multi-head attention mechanism. The output of each attention layer is passed through layer normalization and a residual connection before being fed to the next layer:

$$H_{l+1} = Layermorm(MultiHeadT(H_l)) + H_l, H_{l+1} \in R^{(T+1)\times h} \tag{26}$$

After multiple layers of mobility attention module, the final spatial-temporal feature representation H_L is obtained, where $H_L \in R^{(T+1)\times h}$ and L is the total number of layers in the mobility attention module.

2) Output Layer: In the output layer, to generate the prediction results, we first remove the global user feature portion from H_L and retain only the trajectory sequence portion, resulting in $H_{Output} \in R^{T\times h}$. To output the value of trajectory coordinates, H_{Output} is then fed into two separate multi-layer perceptrons (MLPs), producing two $T \times 200$ tensors that represent the probability distributions of the x and y coordinates at each time step:

$$X_{Output} = MLP_x(H_{Output}), Y_{Output} = MLP_y(H_{Output}), X_{Output}, Y_{Output} \in R^{T\times 200} \tag{27}$$

Finally, we concatenate X_{Output} and Y_{Output} along a new dimension to obtain the output tensor $Output \in R^{T\times 200\times 2}$, representing the probability distribution of trajectory sequence over the two-dimensional space at each time step. This tensor serves as the final prediction result of the SPM-Former model.

4.6. Optimization

In this section, we present a multi-loss optimization method for SPM-Former that assesses the alignment between predicted and actual sequences from both time-series similarity and spatial distribution perspectives. Specifically, for temporal similarity, we employ the multi-class cross-entropy loss L_{MC}. For spatial distribution similarity, we propose a novel SPM-based loss function, L_{SPM}, as shown in Figure 3. The cross-entropy loss measures the prediction accuracy of each trajectory point in the time series. First, the predicted results at the masked positions are extracted from the SPM-Former's output to obtain the relevant temporal sequence. For a user U_i, the prediction results are represented as $Output(U_i, mask) \in R^{mask\times 200\times 200}$. The sequence loss is then calculated by comparing the predicted results with the true trajectory sequence $TS(U_i, mask)$, defined as (28):

$$L_{MC} = -\sum_{j=1}^{mask} TS\left(U_i, mask_j\right)\log\left(Output(U_i, mask_j)\right) \tag{28}$$

where j denotes each specific time step in the sequence being compared. This loss quantifies the divergence between the predicted and true probability distributions at

each time step, with a smaller loss value indicating higher prediction accuracy for the temporal data.

Given that the SPM can effectively represent users' spatial distribution characteristics and the heterogeneity of individual mobility behavior, we introduce the SPM-Loss L_{SPM} to evaluate the similarity between the model's predicted and true spatial distributions. The closer the predicted SPM is to the true SPM, the higher the model's predictive performance. Specifically, using the SPM generation method described in Section 4.2, we compute the SPM for both the true trajectory sequence $TS(U_i, mask)$ and the predicted output $Output(U_i, mask)$. These are denoted as $\text{SPM}_{mask} \in R^{200\times200}$ and $\text{SPM}_{output} \in R^{200\times200}$, respectively. We then use the mean squared error (MSE) loss function to measure their similarity:

$$L_{SPM} = \frac{1}{N}\sum_{i=1}^{N}(SPM^i_{mask} - SPM^i_{output})^2 \tag{29}$$

where i represents the visit frequency at each location in the SPM, and $N = 200 \times 200$ is the total number of spatial units in the SPM. By minimizing L_{SPM}, the model can better match the true spatial distribution characteristics. To effectively combine the cross-entropy loss L_{MC} and the SPM-based spatial loss L_{SPM}, we design an adaptive weighting method to ensure balanced influence between the two loss components during training. The final total loss function is defined as:

$$L_{total} = \frac{L_{MC}}{|L_{MC}|} + \frac{L_{SPM}}{|L_{SPM}|} \tag{30}$$

where $|L_{MC}|$ and $|L_{SPM}|$ are the detached (non-gradient) values of the respective losses, used as scaling factors to maintain stability in relative weights during gradient updates. This proportional normalization approach allows the model to adaptively adjust the relative weight of the two loss components at different training stages, achieving an optimal balance between sequence similarity and spatial distribution similarity, ultimately enhancing overall prediction performance.

5. Analysis results

5.1. Experimental settings

The SPM-Former model is implemented using the PyTorch library with CUDA version 10.1. All experiments are conducted on an Intel(R) Xeon(R) Gold 5317 CPU @ 3.00 GHz and NVIDIA A40 GPUs. As illustrated in Figure 2(C), the training set consists of the complete trajectories of the first 80,000 users, along with the first 60 days of trajectories for an additional 20,000 users. In contrast, the model's prediction target is the trajectories of these last 20,000 users from days 61 to 75. To generate predictions for this target period, the model is provided with the first 60 days of these users' historical trajectories as input. The model is optimized using the Adam optimizer with an initial learning rate of 10^{-4}, which is adjusted using a cosine annealing scheduler over 200 epochs to improve convergence and avoid suboptimal local minima. The SPM-Former architecture consists of 4 mobility attention modules, each with 4 attention heads and

an embedding dimension of 64. In the Individual Mobility Feature Embedding module, we set the k-th radius of gyration ($r_g^{(k)}$) with k values of 2, 4, and 8.

5.2. Baseline methods

In this study, we conduct a comparative analysis between SPM-Former and eight baseline models, as follows:

1. Periodic Attenuation and Local Feature Match (PALFM) (Guo *et al*. 2023): PALFM is a statistics-based prediction method that operates without training. By modeling human mobility patterns with temporal attenuation, PALFM provides an efficient, generalized approach for real-time trajectory prediction.
2. Mobility Markov Chain (MMC) (Gambs *et al*. 2012): MMC is a traditional statistics-based prediction method. It models transitions between POIs, aiming to capture the movements among them. By learning the transition probabilities between consecutive locations, MMCs can predict the next visited location with relative accuracy.
3. LSTM-Regression (LSTM-R) (Hochreiter and Schmidhuber 1997): LSTM-R frames trajectory sequence prediction as a regression task. It utilizes the standard Long Short-Term Memory (LSTM) architecture to generate predictions, optimized using the Mean Squared Error (MSE) loss function.
4. Transformer-Regression (Transformer-R) (Vaswani 2017): Transformer-R frames trajectory sequence prediction as a regression task, utilizing the vanilla Transformer architecture to generate predictions, optimized through the mean squared error (MSE) loss function.
5. LSTM-Classification (LSTM-C) (Hochreiter and Schmidhuber 1997): In contrast to LSTM-R, LSTM-C frames trajectory sequence prediction as a classification task. It utilizes the standard LSTM architecture to generate predictions, which are optimized using the cross-entropy loss function.
6. Transformer-Classification (Transformer-C) (Vaswani 2017): Transformer-C frames trajectory sequence prediction as a classification task, utilizing the vanilla Transformer architecture to generate predictions, optimized through the cross-entropy loss function.
7. Location Prediction-BERT (LP-BERT) (Terashima *et al*. 2023): LP-BERT uses a unique masking strategy to enable long-term sequential predictions. It frames trajectory sequence prediction as a classification task. Notably, LP-BERT was the top-performing model in the HuMob Challenge 2023. Compared to LP-BERT, SPM-Former aims to enhance prediction by explicitly incorporating spatial proximity and individual heterogeneity.
8. Spatio-Temporal Seq2Seq Encoder-Decoder (ST-seq2seq) (Kim *et al*. 2023): ST-seq2seq integrates spatial and temporal information through a spatiotemporal embedding layer, creating embeddings compatible with a sequence-to-sequence encoder-decoder structure. Like LP-BERT, ST-seq2seq frames trajectory prediction as a classification task.

5.3. Evaluation metrics

To comprehensively assess the performance of the proposed model, we employ four evaluation metrics categorized into two main types: sequence similarity metrics and global feature similarity metrics. The sequence similarity metrics include Dynamic Time Warping (DTW) and GEO-BLEU, both of which were used as evaluation criteria in the HuMob Challenge 2023. The global feature similarity metrics compare the radius of gyration (r_g) and activity entropy (ae) between the predicted and true trajectories. Detailed descriptions of each metric are as follows:

1) Dynamic Time Warping (DTW): DTW measures the similarity between two temporal sequences by allowing non-linear alignment to find the optimal path that minimizes the cumulative distance between them. Formally, DTW is defined as the minimum cumulative matching distance over all possible alignment paths π:

$$DTW(TS_{targ}, TS_{pred}) = \min_{\pi} \left(dist(TS_{targ}, TS_{pred}) \right) \tag{31}$$

where TS_{targ} represents the true target sequence, and TS_{pred} is the predicted sequence. A lower DTW value indicates greater similarity between the sequences. To assess the overall prediction performance of the model, we calculate the average DTW across all prediction targets as the final measure of accuracy.

2) GEO-BLEU: GEO-BLEU is a metric for assessing the similarity of geospatial sequences, extending the BLEU metric by incorporating geographical proximity in n-gram comparisons. This allows sequences that are geographically close to show similarity even if their n-grams do not match exactly. GEO-BLEU is computed as follows:

$$GEO - BLEU = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log q_n \right) \tag{32}$$

where BP is the brevity penalty, q_n is the similarity score for geographic n-grams, and w_n denotes the weight of each n-gram. Higher GEO-BLEU values suggest greater similarity between sequences. To evaluate the overall predictive performance of the model, we compute the average GEO-BLEU across all prediction targets as the final accuracy measure.

3) Radius of Gyration Difference (D_{r_g}): The radius of gyration, as detailed in Section 4, measures the average distance of all trajectory points from the centroid, reflecting the overall spatial range of an individual's movement. To assess the accuracy of the model's predictions in capturing this spatial extent, we compute the mean absolute error (MAE) between the r_g values for each user in the true sequence TS_{targ} and the predicted sequence TS_{pred}:

$$D_{r_g} = \frac{1}{N} \sum_{n=1}^{N} \left| r_g\left(TS^i_{targ}\right) - r_g\left(TS^i_{pred}\right) \right| \tag{33}$$

where N is the total number of users, and $r_g\left(TS^i_{targ}\right)$ and $r_g\left(TS^i_{pred}\right)$ represent the radius of gyration of user i in the true and predicted sequences, respectively. A lower D_{r_g} indicates more accurate predictions of spatial activity ranges.

4) Activity Entropy Difference (D_{ae}): Activity entropy, as defined in Section 4, quantifies the diversity of an individual's mobility pattern. To evaluate the model's ability to

capture this behavioral diversity, we also calculate the MAE between the activity entropy values of the true sequence TS_{targ} and the predicted sequence TS_{pred} for each user:

$$D_{ae} = \frac{1}{N}\sum_{n=1}^{N}\left|ae\left(TS^{i}_{targ}\right) - ae\left(TS^{i}_{pred}\right)\right| \tag{34}$$

where $ae\left(TS^{i}_{targ}\right)$ and $ae\left(TS^{i}_{pred}\right)$ represent the activity entropy of user i in the true and predicted sequences. A lower D_{ae} indicates better alignment in predicting the diversity of user behavior.

5.4. Comparison

The predictive performance of SPM-Former is evaluated by comparing it against eight baseline models. All deep learning models are trained for 200 epochs. For the attention-based models, including Transformer-R/C, LP-BERT, and SPM-Former, we consistently set the hidden dimension to 64, the number of attention layers to 4, and the number of attention heads to 4. Table 3 presents the prediction results of all models, evaluated using sequence-based metrics such as DTW and GEO-BLEU, as well as global mobility feature metrics, including D_{r_g} and D_{ae}.

As shown in Table 3, SPM-Former demonstrates superior performance compared to most baseline models. It achieves the best results in sequence-based metrics, outperforming all other models, and shows a slight advantage in D_{r_g} among the global metrics. The only exception is D_{ae}, where it is outperformed by the statistical model PALFM. Specifically, compared to the top-performing model in the HuMob Challenge 2023, LP-BERT, SPM-Former improves by 3%, 3%, 18%, and 24% in DTW, GEO-BLEU, D_{r_g}, and D_{ae}, respectively. These results indicate that SPM-Former not only enhances trajectory sequence similarity but also makes significant advancements in capturing global features. By integrating spatial proximity and individual heterogeneity, SPM-Former improves the representation of human mobility within deep learning frameworks.

A comprehensive analysis reveals that deep learning approaches generally surpass statistical models in terms of trajectory sequence similarity. However, statistical models

Table 3. Performance comparison of different models.

		Sequence Metric		Global Feature Metric	
Model Type	Model	DTW	GEO-BLEU	D_{r_g}	D_{ae}
Statistical model	PALFM	51.023	0.188	7.068	**0.455**
	MMC	55.769	0.153	17.073	3.098
Regression based DL model	LSTM-R	48.964	0.106	12.754	0.701
	Transformer-R	51.243	0.081	12.350	0.653
Classification based DL model	LSTM-C	36.031	0.185	17.079	2.854
	Transformer-C	33.768	0.203	17.567	3.117
	LP-BERT	33.145	0.302	8.604	1.988
	ST-seq2seq	33.582	0.192	10.613	2.067
	SPM-Former	**32.113**	**0.312**	**7.005**	1.496

Note: The metrics DTW, D_{r_g}, and D_{ae} represent differences, with lower values indicating better performance. Conversely, GEO-BLEU measures similarity, where higher values reflect better performance. In the table, bolded numbers indicate the best performance among all models for a given metric, while underlined numbers represent the second-best performance.

often yield predictions that better capture overall global characteristics. This difference likely stems from statistical models being designed to encapsulate global properties. For instance, PALFM models the decaying periodic patterns of human mobility over time, leading to more accurate global predictions. In contrast, deep learning models focus on learning spatial-temporal dependencies within trajectory sequences, resulting in stronger performance for sequence similarity but potentially lagging in global feature representation.

Furthermore, we compare the performance of two LSTM models and two Transformer models that share identical architectures but differ in their task formulations: LSTM-R and Transformer-R, which are regression-based and optimized with a mean squared error loss, against LSTM-C and Transformer-C, which are classification-based and optimized with a cross-entropy loss. As shown in Table 3, the classification-based model significantly outperforms the regression-based model in sequence similarity. This is because the classification approach treats each location as a distinct category, making it more adept at recognizing frequently visited locations such as home or work. Consequently, the classification-based model more accurately captures basic travel patterns like the classic home-to-office-to-home trajectory, thus aligning more closely with real-world sequences. In contrast, the regression model is inclined to predict locations near or between frequently visited places, as this often minimizes the MSE loss.

On the contrary, in terms of global mobility feature similarity, the regression-based model shows better performance in both D_{r_g} and D_{ae}. The advantage of regression-based models in D_{r_g} can be attributed to their capacity to understand spatial continuity, as its loss function minimizes the Euclidean distance between predicted and actual trajectories. This results in its outputs being more aligned with real-world spatial patterns. Conversely, the classification-based model always predicts a limited set of frequently visited locations, which can lead to a smaller radius of gyration and increased deviation from real-world behavior. Moreover, the advantage of regression-based models in D_{r_g} arises because their predictions are naturally more dispersed. Although this dispersion may not perfectly match real-world patterns, it objectively results in higher activity entropy values, making it more aligned with actual conditions compared to the more constrained classification-based model.

5.5. Ablation study

To verify the effectiveness of each proposed module, we conduct an ablation study. Beginning with a basic Transformer model as the foundation, we incrementally add each module to observe its impact on predictive performance. Figure 5 presents the results of the ablation study, where the "Basic Model" refers to the Transformer model without additional modules, consisting only of the Mobility Attention Module and the output layer in SPM-Former. Additionally, we analyze the independent effects of the two submodules within the core innovation—Proximity Encoding and Tendency Encoding from the Spatial Encoding module.

As shown in Figure 5, adding any individual module significantly improves the model's performance in human trajectory prediction, particularly concerning global

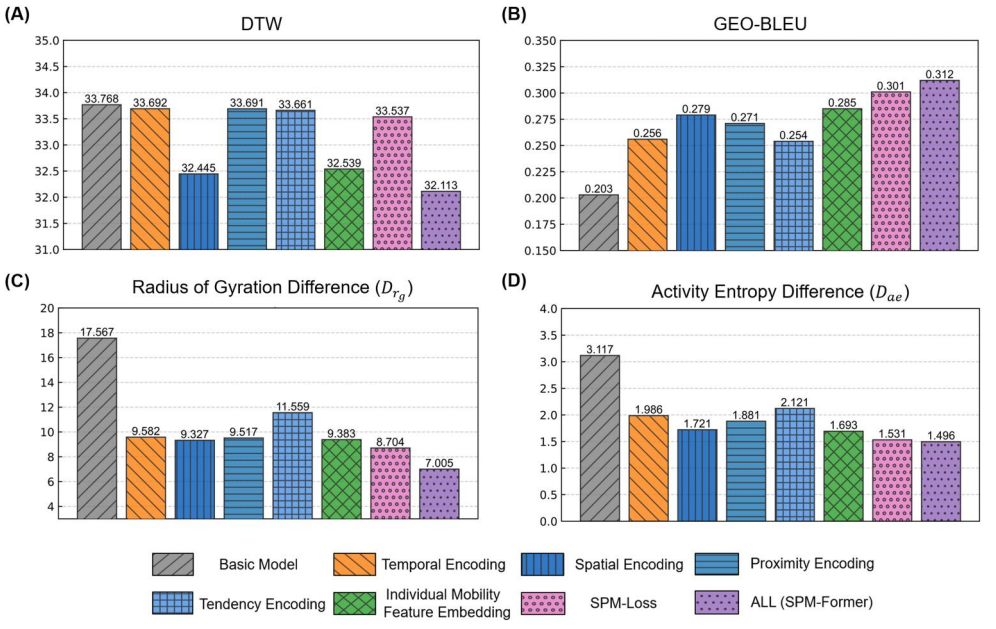

Figure 5. Ablation study of key modules.

mobility characteristics. The inclusion of the Temporal Encoding module, in particular, led to a notable increase in the GEO-BLEU metric, demonstrating that integrating temporal information enables the model to better capture human mobility preferences at different times, thereby enhancing the accuracy of local trajectory predictions. Additionally, the Temporal Encoding module significantly improved overall mobility feature modeling, as evidenced by substantial reductions in D_{r_g} and D_{ae} errors, highlighting the critical role of temporal information in modeling human mobility.

The addition of the Spatial Encoding module significantly boosts the model's performance across all evaluation metrics, with the most pronounced improvement observed in the DTW metric. This finding suggests that the model's overall trajectory sequence similarity is greatly enhanced. By explicitly encoding spatial relationships and user preferences, the Spatial Encoding module enables the model to generate trajectory outputs that more closely resemble real-world conditions. This advantage is particularly evident in the evaluation of global mobility features, showing a 47% improvement in D_{r_g} accuracy and a 45% improvement in D_{ae} accuracy.

When further analyzing the importance of the two submodules within the Spatial Encoding module, we find that Proximity Encoding, which provides explicit spatial relationships, offers more substantial optimization across all metrics compared to Tendency Encoding, which encodes user location preferences. Proximity Encoding enhances the model's spatial awareness, enabling it to simulate the spatial distribution of trajectories more accurately and achieve better results. Although the improvement brought by Tendency Encoding alone is relatively limited, it still aids in capturing user mobility patterns by explicitly integrating visit preferences. The improved performance achieved through Spatial Encoding suggests that Tendency Encoding and Proximity

Encoding complement each other, and their combination leads to more accurate trajectory prediction.

The introduction of the Individual Mobility Feature Embedding (IME) significantly enhances the model's predictive capability. Unlike Spatial Encoding, which focuses on sequence-level information, IME incorporates global individual features into the model, achieving a comparable level of performance improvement. This phenomenon underscores the importance of individual heterogeneity, as IME enables the model to distinguish between individuals and provide personalized predictions.

Finally, the contribution of SPM-Loss is notable, as it significantly improves all metrics except DTW, with particular enhancements in D_{r_g} and D_{ae}. This improvement can be attributed to SPM-Loss evaluating the similarity between the predicted and true trajectory sequences based on overall spatial characteristics, thereby indirectly optimizing consistency in the radius of gyration and activity entropy. Among all the modules, SPM-Loss proves to be the most effective in enhancing the overall mobility feature similarity of the predictions.

In summary, the combined use of all modules in SPM-Former demonstrates significant synergistic effects. The ablation study results indicate that when all modules are integrated, the model achieves the best performance across all metrics. Specifically, compared to the basic model, SPM-Former shows improvements of 4%, 53%, 60%, and 52% in DTW, GEO-BLEU, D_{r_g} and D_{ae}, respectively. This suggests that the interaction among various modules collectively enhances the model's ability to capture spatial-temporal dependencies and individual mobility features. By integrating multi-dimensional information, SPM-Former provides more comprehensive and accurate prediction results, highlighting its potential and advantages in trajectory sequence prediction tasks.

5.6. Improvements of SPM-former over individuals with different mobility characteristics

In this section, we investigate how incorporating SPM improves predictive accuracy at the individual level and identify which user groups benefit the most from this approach. Our analysis consists of two approaches: (1) quantitatively assess the extent to which SPM-Former enhances various performance metrics relative to individual mobility characteristics and (2) visually compare the prediction accuracy of different models for two representative users.

Building on the overall performance of Transformer-C and SPM-Former reported in Table 3, we first collect four prediction metrics for each individual from the 20,000 test subjects. We then compute the differences between these metrics for the two models to quantify each individual's improvement due to SPM-Former, denoted as ΔGEO-BLEU, ΔDTW, ΔD_{r_g}, and ΔD_{ae}. Next, we calculate each individual's radius of gyration and activity entropy over a 75-day period to represent overall mobility characteristics. Finally, as shown in Figure 6, we use scatter plots to visualize the relationship between these improvements and the two mobility features.

From Figure 6, two notable positive correlations emerge: one between the radius of gyration and ΔD_{r_g}, and another between activity entropy and ΔD_{ae}. These findings

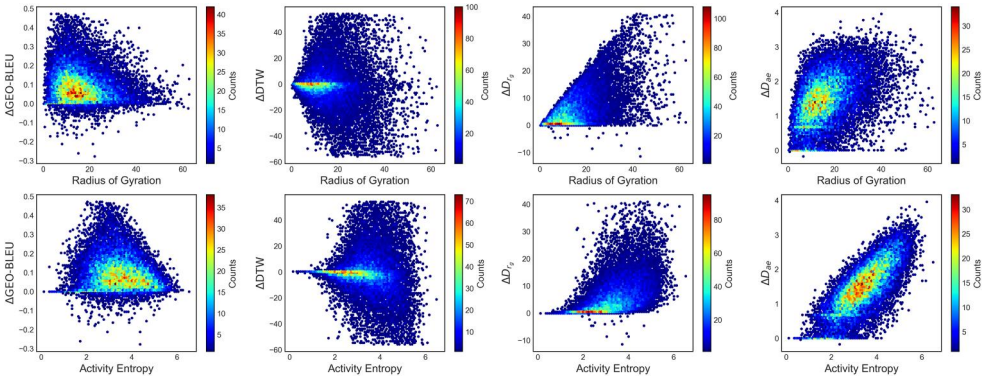

Figure 6. Relationships among individual radius of gyration, activity entropy, and the performance improvements achieved by SPM-Former. For GEO-BLEU, a higher value indicates a higher similarity between an individual's predicted trajectory and the ground truth. Therefore, given ΔGEO-BLEU = GEO-BLEU(SPM-Former) - GEO-BLEU(Transformer-C), a positive value of ΔGEO-BLEU indicates that SPM-Former outperforms Transformer-C in predicting an individual's trajectory. For other metrics including DTW, D_{r_g}, and D_{ae}, a lower value indicates a higher similarity between an individual's predicted trajectory and the ground truth. As a result, given ΔDTW = DTW(Transformer-C) - DTW(SPM-Former), a positive value of ΔDTW indicates that SPM-Former outperforms Transformer-C in predicting an individual's trajectory. Values of ΔD_{r_g} and ΔD_{ae} can be interpreted similarly.

suggest that individuals who travel across larger geographic areas or exhibit greater mobility diversity benefit more from SPM-Former compared to the baseline. Additionally, the radius of gyration shows a positive correlation with ΔD_{ae}, while activity entropy correlates positively with ΔD_{r_g}, underscoring the importance of Individual Mobility Feature Embedding. By explicitly incorporating individual mobility characteristics into the model, SPM-Former more accurately captures overall human mobility patterns. Furthermore, SPM-Former consistently improves GEO-BLEU scores across nearly all individuals, with typical gains around 0.1 and the highest reaching approximately 0.5. Individuals with a radius of gyration below 20 and activity entropy between 2 and 4 experience particularly significant benefits. In contrast, improvements in DTW remain limited, with similar proportions of individuals experiencing gains and declines. Table 3 indicates that SPM-Former outperforms Transformer-C on DTW by only about 5%, likely because DTW emphasizes temporal alignment, where random factors can amplify alignment errors in mobility data. Since SPM-Former primarily emphasizes overall mobility pattern modeling, its impact on DTW is less pronounced.

Through a further visual analysis, we compare the prediction results of a regression-based model (Transformer-R), a classification-based model (Transformer-C), and the proposed SPM-Former against the ground truth. We select two representative users—#80017 and #80547—from the beginning of the test dataset and employ two visualization approaches. First, we compare predicted and actual trajectories to highlight sequence-level similarity, as shown in Figure 7. Second, we analyze spatial preference maps generated from predicted sequences versus those derived from actual trajectories, emphasizing overall spatial distribution similarity, as shown in Figure 8.

Figure 7 shows that classification-based models (e.g., Transformer-C and SPM-Former) generate trajectories that more closely align with actual human mobility. Their

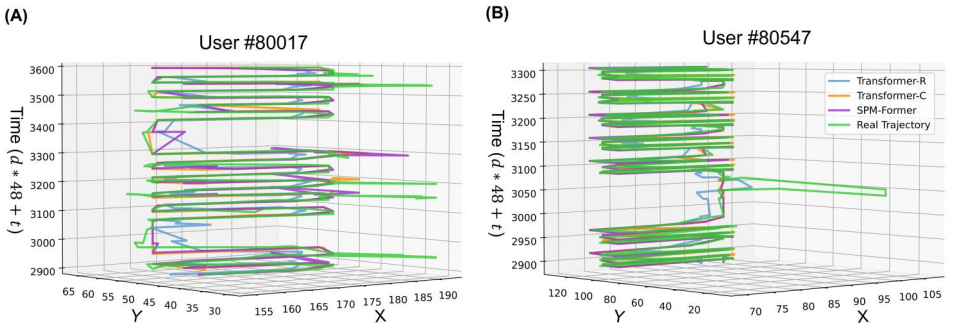

Figure 7. Comparison of Trajectories: Predictions of different models vs. Real Trajectory.

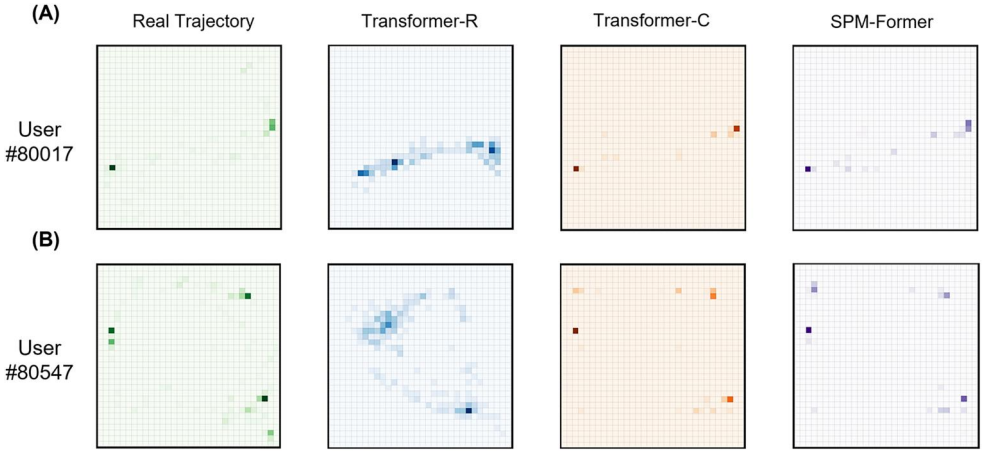

Figure 8. Comparison of Spatial Preference Maps: Predictions of different models vs. Real Trajectory.

predicted points cluster around the ground truth, indicating that they effectively capture frequently visited locations and spatial relationships. In contrast, regression-based models exhibit greater deviations, as reflected in more dispersed spatial distributions, as shown in Figure 8. Although Table 3 indicates that regression-based models can approximate the activity entropy of real data, their reliance on minimizing Euclidean distance may lead to deceptively low errors. Even when predictions fall between actual high-frequency locations, the short distance to each point creates a misleading impression of accurate performance.

A comparison with Transformer-C reveals that SPM-Former's results align more closely with the actual trajectory shapes, which is particularly evident in Figure 8. For example, User #80017's mobility pattern, concentrated around two frequently visited locations, is more accurately represented in the SPM generated by SPM-Former. Moreover, SPM-Former outperforms Transformer-C in terms of prediction diversity. It captures not only routine movements but also exploratory behaviors, identifying transitions between significant locations and activities nearby. This capability illustrates the advantages of integrating Proximity Encoding and Individual Mobility Feature Embedding, which model spatial proximity and individual heterogeneity, thereby

providing a more precise simulation of human mobility patterns. Despite these strengths, discrepancies remain in the case of User #80547, whose trajectory is more complex. As Figure 7(B) indicates, both models adequately capture the periodicity of human mobility, with five distinct sequences corresponding to workdays. However, in capturing irregular exploratory behaviors, both models show potential for improvement. Future research should aim to enhance predictions of non-periodic and exploratory mobility patterns to improve model performance in complex scenarios.

6. Conclusion and future works

In this study, we propose the Spatial Preference Map-based Transformer (SPM-Former) to enhance the accuracy of human trajectory sequence predictions in urban environments. Integrated with the Spatial Preference Map (SPM), SPM-Former achieves significant improvements by accounting for both spatial proximity and individual heterogeneity. SPM-Former is composed of several innovative modules, each of which is based on the SPM. These modules function synergistically to strengthen the model's ability to understand complex spatial-temporal dependencies.

Among these modules, the trajectory sequence encoding module enriches trajectory sequences with detailed spatial-temporal information by decoding proximity relationships and visit tendencies from the SPM. By decomposing the SPM, this module explicitly represents spatial relationships across all locations and effectively captures individual mobility preferences. Concurrently, the individual mobility feature embedding module leverages key mobility metrics (e.g., radius of gyration and activity entropy) to quantify overall travel characteristics derived from the SPM, thereby significantly enhancing the model's capacity to identify individual heterogeneity. Finally, the SPM-based loss function (SPM-Loss) redefines the assessment of prediction accuracy by considering the overall spatial distribution, complementing traditional cross-entropy loss and reducing the gap between predicted and actual trajectories.

Experiments on a large-scale real-world dataset from Japan, demonstrate that SPM-Former achieves improvements of approximately 3% in trajectory sequence similarity and 20% in overall spatial feature similarity, confirming its superior predictive capability in practical applications. Additionally, we compare the accuracy improvements achieved by incorporating SPM across different user categories. The results indicate that individuals with larger radius of gyration or higher activity entropy experience the most significant improvements. This performance enhancement can be attributed to the explicit integration of individual mobility features, allowing SPM-Former to better capture unique mobility behaviors. The progress achieved in this study underscores the significance of explicitly incorporating spatial relationships and individual heterogeneity information into deep learning models. These findings could further be applied to fields such as next-location prediction, trajectory generation, and individual trajectory identification.

While SPM-Former demonstrates strong performance on the current dataset, future research should consider balancing model efficiency and computational complexity to enhance practical applicability. Moreover, further validation across diverse geographic regions is necessary to comprehensively assess the model's generalizability and robustness. Human mobility is not merely a sequence of spatial-temporal trajectories; rather, it

represents a complex behavior shaped by each city's unique socioeconomic and cultural context, alongside its distinct physical structure (Gao 2015). Our future research will address a central challenge: developing a mobility prediction model with robust cross-city generalization capabilities. Addressing this challenge first requires enhancing model adaptability. Consequently, we will pursue more diverse international urban datasets and leverage techniques such as transfer learning. This approach will enable the model to learn from data-rich cities and rapidly adapt for predictions in novel urban environments, thereby augmenting its generalizability. However, true generalization extends beyond adaptation as it fundamentally requires understanding. We contend that an ideal universal model must possess a dual capability: first, it must accurately capture and differentiate the macroscopic mobility patterns of various cities. Second, and more critically, it must encode the city's physical structure, including road network topology and the distribution of functional POIs, into explicit structural representations embedded within the model's architecture (Mai *et al*. 2022). This integration will enable the model to mechanistically explain the variations in these mobility patterns. Therefore, future work will concentrate on developing model architectures capable of deeply integrating urban structural information, ultimately aiming for cross-city mobility prediction that is genuinely context-aware and broadly generalizable.

Acknowledgments

The authors would like to thank the editors and anonymous reviewers for their valuable comments on earlier versions of the manuscript. The authors also wish to express their appreciation to Dr. Takahiro Yabe for organizing the HuMob Challenge 2023 and providing access to the dataset used in this study.

Author contributions

Guangyue Li: Conceptualization, Formal analysis, Software, Visualization, Methodology, Writing - original draft, Writing - review & editing. Yang Xu: Resources, Writing - review & editing, Supervision, Funding acquisition. Zhipeng Gui: Writing - review & editing. Xiaogang Guo: Conceptualization, Methodology. Luliang Tang: Conceptualization, Supervision.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the National Natural Science Foundation of China (Grant No. 42171454) and Hong Kong Polytechnic University Research Grant (Grant No. 4-ZZNC).

Notes on contributors

Guangyue Li is a Ph.D. student in the Department of Land Surveying and Geo-Informatics at the Hong Kong Polytechnic University. His research interests include spatial-temporal data mining, GeoAI, and human mobility.

Yang Xu is an Associate Professor in the Department of Land Surveying and Geo-Informatics at the Hong Kong Polytechnic University. His research interests include GIScience, human mobility, and urban informatics.

Zhipeng Gui is a full professor in the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China. His research interests include spatiotemporal data mining, GeoAI and distributed computing.

Xiaogang Guo is currently pursuing his Eng.D degree at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing. His research interests include human mobility, urban systems, and GIS.

Luliang Tang received the Ph.D. degree from Wuhan University, Wuhan, China, in 2007. He is currently a Professor at Wuhan University. His research interests include space–time GIS, deep learning, GIS for transportation.

ORCID

Yang Xu http://orcid.org/0000-0003-3898-022X
Zhipeng Gui http://orcid.org/0000-0001-9467-9680

Data and codes availability statement

The data and codes that support the findings of this study are available at https://doi.org/10.6084/m9.figshare.27916944.

References

Aggarwal, R., *et al.*, 2021. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digital Medicine*, 4 (1), 65.

Barbosa, H., *et al.*, 2018. Human mobility: Models and applications. *Physics Reports*, 734, 1–74.

Brébisson, A.D., *et al.*, 2015. Artificial neural networks applied to taxi destination prediction. *Proceedings of the 2015th International Conference on ECML PKDD Discovery Challenge - Volume 1526*. Porto, Portugal: CEUR-WS.org, 40–51.

Brown, T.B., 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Chu, C., *et al.*, 2024. Simulating human mobility with a trajectory generation framework based on diffusion model. *International Journal of Geographical Information Science*, 38 (5), 847–878.

Devlin, J., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Feng, J., *et al.*, 2018. Deepmove: Predicting human mobility with attentional recurrent networks. ed. *Proceedings of the 2018 world wide web conference*, 1459–1468.

Gambs, S., Killijian, M.-O., and Del Prado Cortez, M.N., 2012. Next place prediction using mobility markov chains. ed. *Proceedings of the first workshop on measurement, privacy, and mobility*, 1–6.

Gao, S., 2015. Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spatial Cognition & Computation*, 15 (2), 86–114.

Gui, Z., *et al.*, 2021. LSI-LSTM: An attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points. *Neurocomputing*, 440, 72–88.

Guo, X., *et al.*, 2023. Large-Scale Human Mobility Prediction Based on Periodic Attenuation and Local Feature Match. ed. *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, 16–21.

Hochreiter, S., and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9 (8), 1735–1780.

Hong, Y., *et al.*, 2023. Context-aware multi-head self-attentional neural network model for next location prediction. *Transportation Research Part C: Emerging Technologies*, 156, 104315.
Huang, H., *et al.*, 2021. Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. *Computers, Environment and Urban Systems*, 90, 101712.
Huang, Q., *et al.*, 2013. Using adaptively coupled models and high-performance computing for enabling the computability of dust storm forecasting. *International Journal of Geographical Information Science*, 27 (4), 765–784.
Kim, T., Kim, K.-S., and Matono, A., 2023. Cell-Level Trajectory Prediction Using Time-embedded Encoder-Decoder Network. ed. *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, 37–40.
Kumar, G., Jain, S., and Singh, U.P., 2021. Stock market forecasting using computational intelligence: A survey. *Archives of Computational Methods in Engineering*, 28 (3), 1069–1101.
Law, S., *et al.*, 2020. Street-Frontage-Net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34 (4), 681–707.
Li, F., *et al.*, 2020. A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction. *Neurocomputing*, 403, 153–166.
Li, G., *et al.*, 2024a. Advancing complex urban traffic forecasting: A fully attentional spatial-temporal network enhanced by graph representation. *International Journal of Applied Earth Observation and Geoinformation*, 134, 104237.
Li, G., *et al.*, 2024b. Towards integrated and fine-grained traffic forecasting: A spatio-temporal heterogeneous graph transformer approach. *Information Fusion*, 102, 102063.
Li, M., *et al.*, 2021. Prediction of human activity intensity using the interactions in physical and social spaces through graph convolutional networks. *International Journal of Geographical Information Science*, 35 (12), 2489–2516.
Luca, M., *et al.*, 2023. A survey on deep learning for human mobility. *ACM Computing Surveys*, 55 (1), 1–44.
Lv, J., *et al.*, 2018. T-CONV: A convolutional neural network for multi-scale taxi trajectory prediction. ed. *IEEE international conference on big data and smart computing (bigcomp)*, 2018, 82–89.
Mai, G., *et al.*, 2022. A review of location encoding for GeoAI: methods and applications. *International Journal of Geographical Information Science*, 36 (4), 639–673.
Nilsson, I., and Delmelle, E.C., 2023. An embedding-based text classification approach for understanding micro-geographic housing dynamics. *International Journal of Geographical Information Science*, 37 (12), 2487–2513.
Pappalardo, L., *et al.*, 2019. Human mobility from theory to practice: Data, models and applications. ed. *Companion Proceedings of The 2019 World Wide Web Conference*, 1311–1312.
Pappalardo, L., *et al.*, 2015. Returners and explorers dichotomy in human mobility. *Nature Communications*, 6 (1), 8166.
Perofsky, A.C., *et al.*, 2024. Impacts of human mobility on the citywide transmission dynamics of 18 respiratory viruses in pre-and post-COVID-19 pandemic years. *Nature Communications*, 15 (1), 4164.
Shaw, S.L., and Sui, D., 2021. Understanding the spatial and temporal dynamics of a global pandemic. *In*: S.L. Shaw and D. Sui, eds. *Mapping COVID-19 in space and time. Human dynamics in smart cities*. Cham: Springer, 1–9. https://doi.org/10.1007/978-3-030-72808-3_1
Solatorio, A.V., 2023. GeoFormer: Predicting Human Mobility using Generative Pre-trained Transformer (GPT). ed. *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, 11–15.
Song, C., *et al.*, 2010a. Modelling the scaling properties of human mobility. *Nature Physics*, 6 (10), 818–823.
Song, C., *et al.*, 2010b. Limits of predictability in human mobility. *Science (New York, N.Y.)*, 327 (5968), 1018–1021.
Suzuki, M., Furuta, S., and Fukazawa, Y., 2023. Personalized human mobility prediction for HuMob challenge. ed. *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, 22–25.

Tang, J., *et al.*, 2021. Trip destination prediction based on a deep integration network by fusing multiple features from taxi trajectories. *IET Intelligent Transport Systems*, 15 (9), 1131–1141.

Terashima, H., *et al.*, 2023. Human Mobility Prediction Challenge: Next Location Prediction using Spatiotemporal BERT. ed. *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, 1–6.

Vaswani, A., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Wang, C., and Deng, Z., 2023. Multi-perspective Spatiotemporal Context-aware Neural Networks for Human Mobility Prediction. ed. *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, 32–36.

Xu, Y., *et al.*, 2018. Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems*, 72, 51–67.

Xu, Y., *et al.*, 2021. Towards a multidimensional view of tourist mobility patterns in cities: A mobile phone data perspective. *Computers, Environment and Urban Systems*, 86, 101593.

Yabe, T., *et al.*, 2024a. Enhancing human mobility research with open and standardized datasets. *Nature Computational Science*, 4 (7), 469–472.

Yabe, T., *et al.*, 2024b. YJMob100K: City-scale and longitudinal dataset of anonymized human mobility trajectories. *Scientific Data*, 11 (1), 397.

Yang, D., *et al.*, 2020. Location prediction over sparse user mobility traces using rnns. ed. *Proceedings of the twenty-ninth international joint conference on artificial intelligence*, 2184–2190.

Yang, H., *et al.*, 2025. Exploring human mobility: a time-informed approach to pattern mining and sequence similarity. *International Journal of Geographical Information Science: IJGIS*, 39 (3), 627–651.

Zhou, Z., *et al.*, 2024. Predicting collective human mobility via countering spatiotemporal heterogeneity. *IEEE Transactions on Mobile Computing*, 23 (5), 4723–4738.