Computers, Environment and Urban Systems 104 (2023) 102009

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Predicting mobile users' next location using the semantically enriched geo-embedding model and the multilayer attention mechanism

Yao Yao [a,b,1], Zijin Guo [a,b,1], Chen Dou [c], Minghui Jia [d], Ye Hong [e], Qingfeng Guan [a,*], Peng Luo [f,**]

[a] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, Hubei Province, PR China
[b] Center for Spatial Information Science, The University of Tokyo, Chiba 277-8568, Japan
[c] School of Remote Sensing Information Engineering, Wuhan University, Wuhan 430079, China
[d] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[e] Institute of Cartography and Geoinformation, ETH Zurich, 8093 Zurich, Switzerland
[f] Chair of Cartography, Technical University of Munich, Munich, Germany

ARTICLE INFO

Keywords:
Location prediction
Semantic features
Geo-embedding
Multilayer attention
Human activity pattern

ABSTRACT

Predicting the next location of human mobility and its semantic information can support recommendations for location-based services and trajectory mining, such as human mobility pattern recognition and sequential anomaly detection. Previous studies have ignored the implicit correlation between location and spatiotemporal information thereby limiting the model performance in terms of location prediction accuracy. In this study, we propose a GEMA-BiLSTM (Geographical Embedding and Multilayer Attention -Bidirectional Long Short-Term Memory) model to predict the next-location in users' mobility. The model combines location and spatiotemporal information to extract the semantics of human mobility. The results show that the model can accurately predict the next location with a high accuracy of 87.63%. Compared with BiLSTM-CNN, LSTM, CNN, and Markov, the location prediction accuracy of the model improved by 2.28%, 9.72%, 11.53%, and 17.64%, respectively. In addition, the model has the highest semantic prediction accuracy (75.35%). Compared with the BiLSTM-CNN model, the our method improves the semantic prediction accuracy for residential and industrial function areas by 4.79% and 5.37%, respectively. The accuracy of location prediction for different time periods indicates that the next location of human activity during morning rush and evening rush hours is the most difficult to predict, which corresponds to the increase in human travel demand. Moreover, weekday human activity patterns indicate that the commercial area is still very active at night, which may be linked to nighttime economic policies. This study could improve the accuracy of recommendations for location-based service applications.

1. Introduction

Rapid urbanization has led to complex patterns of human mobility which generates challenges such as traffic congestion and imbalanced regional development (Cobbinah et al., 2022; Wang, Zhang, and Li, 2022). Geographic big data, such as geotagged social media data and mobile phone signaling data, are commonly used to analyze and monitor human mobility (Liu et al., 2020). Human trajectories consist of temporally connected locations where people engage in activities such as shopping, eating, and working. The time interval between the locations is usually in minutes, hours or days (Qian et al., 2021; Tu et al., 2017). Capturing the trends of human movement in a short period can assist in urban planning and management thus providing theoretical support for reallocating public resources (Liu et al., 2020; Tian et al., 2021).

Next-location prediction refers to the prediction of one step ahead location based on the existing historical trajectory. Although location prediction techniques have increasing applications in urban planning

* Corresponding author.
** Corresponding author at: Technical University of Munich, Munich, Germany.
E-mail addresses: yaoy@cug.edu.cn (Y. Yao), gzj2017@cug.edu.cn (Z. Guo), whu_dc@whu.edu.cn (C. Dou), minghui_jia@whu.edu.cn (M. Jia), guanqf@cug.edu.cn (Q. Guan), peng.luo@tum.de (P. Luo).
[1] These authors have equal contributions.

https://doi.org/10.1016/j.compenvurbsys.2023.102009
Received 10 December 2022; Received in revised form 6 July 2023; Accepted 8 July 2023
Available online 14 July 2023
0198-9715/

and management, their adoption is still constrained by predictive performance issues. Current trajectory modeling leverages only time and location data due to the lack of specific activity information in geographical big data (Xia, Hu, and Chen, 2023). However, human trajectories carry information regarding activity, purpose and preferences (Liu et al., 2015; Zhang et al., 2021), and their analysis enables the inference of the purposes of activities (Zhai et al., 2019). Fully exploiting the temporal and spatial characteristics of trajectories will potentially improve location prediction.

Studies have identified temporal features of the trajectory, such as the daily patterns and longer-term weekly or seasonal patterns in human activity (Gonzalez, Hidalgo, and Barabasi, 2008; Liu et al., 2020). Zhao, Chen, Gao, et al., (2023) used workers to demonstrate the daily periodicity of human mobility between home and office. Deschaintres, Morency, and Trépanier (2022) observed the weekly cycle of human mobility in 108-days origin-destination (OD) survey data. The repetitive and periodic activity patterns of historical trajectories can be an important source for location prediction (Liu et al., 2020). However, historical trajectories for individuals can be difficult to obtain as availability is often limited. Learning from human mobility data collections can also support human mobility prediction by relying on common human activity patterns to complement the individual's activity information (Solomon et al., 2021).

Researchers have made efforts at location prediction using the temporal features of trajectories by employing three main location prediction techniques: pattern-based, model-based, and neural network-based. In the pattern-based method, the frequent mobility patterns are extracted from the trajectory set, and then the pattern tree is built to predict the next location (Guo et al., 2010). However, this method is sensitive to the selection of parameters and requires a long time to find frequent patterns.

Model-based methods, such as the probability model based on statistics, have been developed for location prediction (Bernecker, Cheng, Cheung, et al., 2013; Gambs, Killijian, and Del Prado Cortez, 2012; Qiao et al., 2014; Song, Kotz, Jain, et al., 2003). The model-based methods calculate transition probabilities among all locations and use dynamic programming to identify the sequence with the greatest probability. Gambs et al. (2012) used Markov chains to compute the probability of moving between locations based on previous visits. Qiao et al. (2014) improved the hidden Markov model (HMM) to capture the relevant parameters for real-world scenarios. However, the model-based methods assume trajectory independence, meaning that the current location is only related to previous one. This independence assumption limits the model's ability to predict accurately, leading to unsatisfactory performance (Jensen et al., 2006).

Neural network-based methods compose overrepresentation vectors of locations in the trajectory for classification (Alahi et al., 2016; Bao et al., 2021; Solomon et al., 2021). One-hot embedding is a common approach to represent the location (Bao et al., 2021; Erdelić et al., 2021). However, it creates sparse, high-dimensional vectors that hinder neural-network-based methods from learning meaningful and helpful information (Mai et al., 2022). Word embedding uses neural networks to learn to generate word representation vectors from contextual information in the sentences of a corpus (Selva Birunda and Kanniga, 2021). To further consider the semantic information of spatial data, the word embedding method has been introduced into spatial data mining tasks such as urban function mining and geographic semantic representation to extract semantic information (Zhai et al., 2019; Zhang et al., 2021). In a trajectory, semantic information refers to the spatiotemporal transition relationship indicating the user's intent (Yao, Zhang, Huang, et al., 2017; Zhang et al., 2021). Zhang et al. (2021) proposed a Traj2Vec model that embeds the context of locations in trajectories to extract semantic information about how people use urban space. The word-embedded representation vectors reflect human activity purposes, which are helpful to improve the performance of land-use classification. The word embedding method is expected to improve the performance of location prediction models.

Neural network-based methods such as the recurrent neural network (RNN) and long short-term memory (LSTM) have been used to mine long- and short-term dependencies from trajectories. RNN models are able to consider correlations in trajectories but do not extract features in parallel (Alahi et al., 2016). LSTM is a derived RNN model containing long- and short-term memory units. LSTM can efficiently solve gradient boosting and explosion problems, but it can only exploit historical contextual information (Solomon et al., 2021). The Bi-LSTM model combines the forward and the backward hidden layers that can process historical and future contextual information (Bao et al., 2021; Graves and Schmidhuber, 2005). However, due to the uncertainty of human activities and the instability of recording devices (Xu et al., 2015), it is still difficult to eliminate the interference due to meaningless information in prediction results by using Bi-LSTM only.

Focusing on the important information in the trajectory can help neural-network-based methods to produce better results (Li et al., 2020). Useful information for location prediction focuses on important locations and corresponding spatiotemporal features, such as transportation centers in the morning and evening peaks and workplace locations during working hours. The attention mechanism assigns weights to highlight important contextual information for sequence problems (Chen et al., 2019). It has been successfully applied in text classification (Liu and Guo, 2019), image recognition (Li, Jin, Zhou, et al., 2020), and other tasks. A feedforward neural network parameterizes the attention mechanism and can be integrated into the training of deep neural networks (Miao, Luo, Zeng, et al., 2020). Since LSTM models can capture contextual information in trajectories, combining an attention mechanism and Bi-LSTM models further improves the performance of location prediction.

Functional semantics refer to human activities in urban areas, i.e., eating, traveling, shopping, and sleeping (Cai, Xu, Liu, et al., 2019). Studies have shown that the functional semantics of location can help explain or analyze human activity intentions, habits, and activity preferences thus improving location prediction performance. For example, Meng et al. (2017) combined location and point of interest (POI) information within the activity unit to improve the accuracy of inferring the travel purpose. Temporal information about the m'ovements in the trajectory is also important for detecting spatiotemporal trends of human mobility (Zhang, Liu and Wang, 2019; Gao, 2015). For example, people usually move toward their residence at night and toward their office during the morning peak. However, the influence of different features on the prediction location is often ignored. For example, when inferring the trend of human movement at night, it is more likely to consider time information. This is because people's activities at night are fixed and have little relation to the place in the previous time period.

In summary, both the spatiotemporal features of location and the long- and short-term dependencies between locations are essential for location prediction (Yao et al., 2017). However, most studies simply combine external characteristics (POI and time). Ignoring the difference between the spatiotemporal semantics and the previous location of trajectory will introduce bias into location prediction. Based on word embedding for geographical data and an attention mechanism, this study proposes a GEMA-BiLSTM model to fuse the spatiotemporal semantics and context information of location to capture the active intent to enable location prediction. The remainder of this paper is organized as follows: in the methodology section, the composition of the GEMA-BiLSTM model and the corresponding working principle are explained in detail. The third section presents specific location prediction cases, and the last two sections discuss and summarize the model effects. This study can help implement urban management policy and ensure urban traffic stability.

2. Methodology

2.1. Overall framework

This study proposes a location prediction framework called GEMA-BiLSTM that couples geographical embedding, a multilayer attention mechanism, and bidirectional long short-term memory. As shown in Fig. 1, GEMA-BiLSTM consists of four parts: the trajectory model, spatiotemporal feature model, attention model, and prediction model. The trajectory model consists of the word embedding component to obtain the matrix representation of the trajectory and the Bi-LSTM component to extract the contextual information from the trajectory matrix. Similarly, the spatiotemporal feature module includes a spatiotemporal feature embedding component to build the feature matrix and a Bi-LSTM component to extract contextual information from the feature matrix. Next, the output of the Bi-LSTM component in the trajectory model and the spatiotemporal feature model are merged and used as the input to the attention model. The attention model includes a local attention layer and a global attention layer. Finally, a fully connected neural network is used in the prediction model to obtain the next location, and the accuracy of the model results is verified.

2.2. Embedding of geographic information

Before location prediction, spatiotemporal data such as trajectories, time, and land function semantics must first be represented as vectors

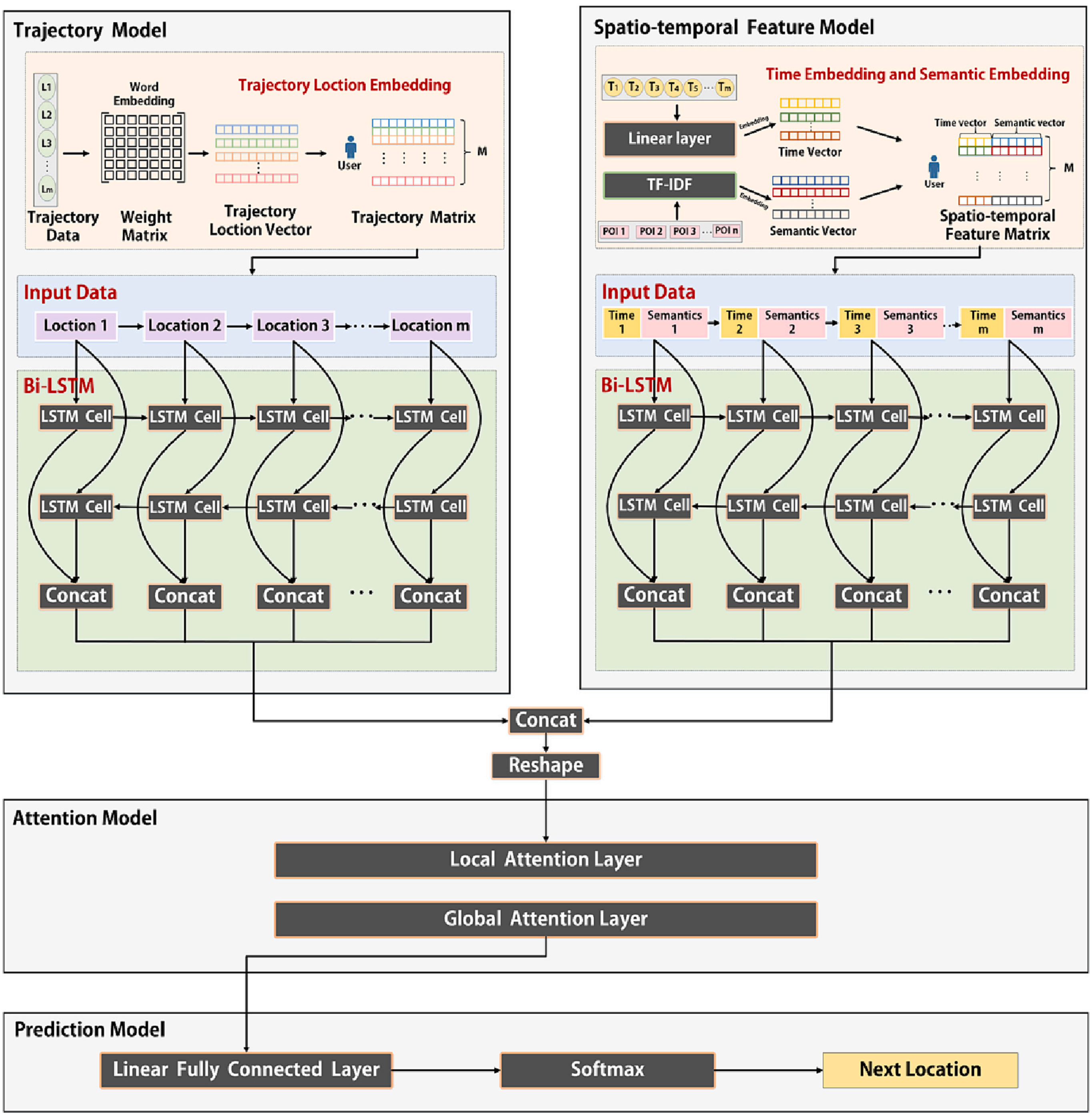

Fig. 1. GEMA-BiLSTM model framework.

that computers can learn and process. Specifically, the trajectory here is about the human activity trajectory, where each location in the trajectory represents an activity anchor point. This anchor point indicates that the human has been in the location for some time and has performed some activities (Zhang et al., 2021). In this study, anchor points are selected on the criteria that the stay is longer than one hour.

2.2.1. Extraction of contextual information in trajectories based on the word embedding model

Human mobility is regular in urban areas (Gonzalez et al., 2008). For example, the next step for a weekday worker at home in the morning is the office, and the next location at work in the evening is the residence. Contextual information from the location helps to infer the intention. The Word2Vec model is used because the representational vector of locations can capture contextual semantic information. Word2Vec is one of the most popular models for learning word embeddings using shallow neural networks (Mikolov, Chen, Corrado, and Dean, 2013). The neural network used in Word2Vec learns the distribution of words in the corpus to obtain vectors representing words. When the representational vectors are similar in distance, the context information of the locations is similar.

The CBOW (continuous bag-of-words) is the Word2Vec architecture used in this study to fully characterize the contextual information of each location in the trajectory (Mikolov et al., 2013). It infers the target word from the given context words (Fig. 2). When using the CBOW model to generate a word embedding vector, we must first build a training dataset. The sequence of trajectories of the *k*-th user input to the CBOW can be expressed as $U^k = \{X_1,\ldots, X_i,\ldots, X_M\}$, where X_i is the *i*-th step and *M* denotes the number of samples. All the users' trajectories form a training dataset for the CBOW model. Subsequently, we must build the set of locations where people have moved from the complete set of trajectories, with each location having a corresponding index. The locations are encoded using one-hot encoding, such as $X_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{iV})$, where *V* is the total number of location sets.

Next, we need to determine the size of the slide window *c*, a parameter that determines how much context information is embedded. The context of X_i in the trajectory includes $\{X_{i-c}, X_{i-c+1}, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{i+c-1}, X_{i+c}\}$.

Then, the CBOW model is trained to obtain the weight matrix W_1, where $W_1 \in R^{V\times N}$, *N* is the vector dimension after embedding. By taking the average of the intermediate vectors v_i in the window, as shown in Formula (2), the output vector h_i can be obtained. Multiplying the hidden layer output h_i by the output matrix W_2 in Formula (3), the vector u_i

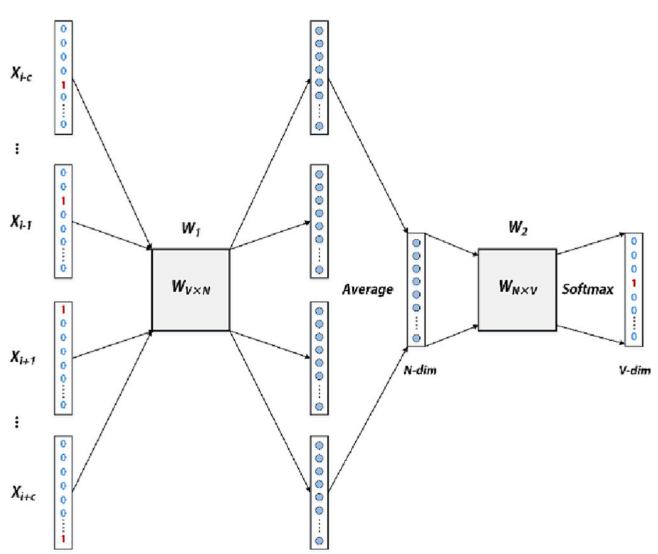

Fig. 2. The CBOW framework of the word2vec model.

can be obtained, where $u_i \in R^{1\times V}$.

$$v_i = W_1 \bullet X_i \tag{1}$$

$$h_i = \frac{1}{2c}(v_{i-c}, v_{i-c+1}, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{i+c-1}, v_{i+c}) \tag{2}$$

$$u_i = W_2 \bullet h_i \tag{3}$$

Finally, the probability of each characteristic dimension is calculated through the softmax layer in Formula (4) and Formula (5). The deviation between the real data X_i and u_i' is calculated through the loss function in Formula (6). The optimal weight matrix W_1 is obtained by minimizing loss. Each location-representing vector encoded with one-hot in the trajectory is transformed into an N-dimensional word vector using the weight matrix W_1, and the trajectory is represented as an $M*N$ matrix.

$$u_i^{'} = \mathrm{softmax}(u_i) \tag{4}$$

$$u_{ij}^{'} = \frac{e^{u_{ij}}}{\sum_{j=1}^{V} e^{u_{ij}}} \tag{5}$$

$$\mathrm{loss}\left(X_i, u_i^{'}\right) = -\sum_{j=1}^{V} X_{ij} \bullet ln\left(u_{ij}^{'}\right) \tag{6}$$

where u_{ij} is the predicted value of the *j*-th location in the *i*-th step, and $u_{ij}^{'}$ is the predicted probability of the *j*-th location in the *i*-th step. X_{ij} is the true probability of the *j*-th location in the *i*-th step.

2.2.2. Calculating the spatiotemporal semantic vectors of locations

Extracting temporal information and functional semantics of land is essential for collecting human activity patterns (Humtsoe, 2022). Human trajectory records contain temporal information, represented by timestamps, which are integers. To obtain the spatiotemporal semantic vectors corresponding to the locations, we first need to compute vectors of length *N* (as mentioned in 2.2.1) to represent the temporal information. This study performs entity embedding for time to generate time feature vectors. Entity embedding refers to the one-hot embedding of timestamps into low-dimensional vectors using a fully connected neural network. The next step is to compute a functional semantic vector of length *Z*, where *Z* is the number of POI categories. Since human activities are associated with land function semantics, POIs express the land function semantics in cities (Zhai et al., 2019). For embedded land function semantics, using only a single category of POI to represent the function of mixed land use can produce a significant bias (Sarkar and Chunchu, 2016). The attractiveness of each POI category is obtained based on the term frequency-inverse document frequency (TF-IDF) algorithm. The formula for the TF-IDF is as follows:

$$W_d = f_{w,d} \bullet log\frac{|D|}{1 + f_{w,D}} \tag{7}$$

where *w* represents a type of POI, *d* represents a location, and $f_{w,\,d}$ represents the frequency of w in *d*. *D* represents the set of locations in human trajectory data. $f_{w,\,D}$ represents the frequency of *w* in *D*. The dimensionality of the semantic vectors is the same as the number of POI categories, namely, *Z*.

Finally, the functional semantic vector is appended to the time information vector to form the spatiotemporal feature vector of the location that has a vector dimension of *(N + Z)*. Corresponding to each step location, the spatiotemporal feature vector is parallel-connected to construct the *M*(N + Z)* spatiotemporal semantic matrix.

2.3. Long space-time mobile mode learning

In location prediction problems, the current location is related to the past state and may be related to the future state (Chen et al., 2017).

However, the single-direction LSTM can only transmit information in one direction. Bi-LSTM is a combination of forward and reverse LSTM (Graves and Schmidhuber, 2005). Bi-LSTM can handle the long- and short-term dependencies of sequences effectively (Bao et al., 2021).

Each location in the trajectory has previous and next locations, except for the starting and the ending locations. To extract the contextual information from trajectory and spatiotemporal feature sequences, we apply the Bi-LSTM model in corresponding modules, comprising feature input and training layers. In the feature input layer before training, the vector of trajectory representations and the vector of spatiotemporal feature sequences need to be used as the input of the Bi-LSTM model. In the trajectory module, the representation vector of each location in the trajectory (detailed in Section 2.2.1) is used as input to the Bi-LSTM model. In the spatiotemporal feature module, the spatiotemporal feature vector corresponding to each location in the trajectory is used as input to the Bi-LSTM model (detailed in Section 2.2.2). In the training layer, each sequence needs to be cut into some fixed-length segments in the neural network. The length of the sequence is set to M. We then use the previous (M-1) location information and the previous (M-1) spatiotemporal feature information to predict the final location information. The last location is used to evaluate the performance of the model.

2.4. Multilayer attention mechanism

The attention mechanism is a mimetic way of observing objects in humans and can select key features from a large amount of information (Luong, Pham, and Manning, 2015). In the location prediction task, the attentional mechanism can assign weights to location and spatiotemporal features and to different human movements in the trajectory. This study employs a multilayer attention model consisting of local and global attention layers. The local attention layer fuses the output of the trajectory and spatiotemporal semantic models, while the global attention layer highlights the influence of location on the prediction results.

2.4.1. Local feature fusion of spatiotemporal semantics

Mobility transitions are governed by multiple factors (Feng, Li, & Zhang, 2018), such as time of day and location preferences. Considering the implicit correlations between these different factors, we need to consider their weights. This study incorporates a local attention layer to fuse location and spatiotemporal features. First, the hidden state vectors in the trajectory model are merged with the hidden state vectors in the spatiotemporal model. After merging, the new hidden state vector h_f is obtained, where $h_f \in R^{L\times 2\times 2N}$, L is the sequence length, and N is the number of neurons at the hidden layer. Since Bi-LSTM is used in this study, the output hidden state vector is twice as long as the input vector. The length of the hidden state vector in the location model and in the temporal semantics are both 2 N. Then, the local attention layer is adopted to calculate the weights of the location and spatiotemporal semantic features. As shown in Formula (8), the initialized coefficient matrix W_l and bias term b_l are used to perform linear transformation on the input h_f, and then nonlinear transformation is realized through the tanh function.

$$\mathrm{u_{li}} = \mathrm{V_l^T}\, tanh\left(\mathrm{W_l} \bullet h_f + \mathrm{b_l}\right) \tag{8}$$

During the training process, the model is iterated continuously to optimize and update the parameters of W_l, b_l and V_l^T. The softmax function is used to realize normalization, and the weight α_{li} of local features is calculated, as shown in (9).

$$\alpha_{\mathrm{li}} = \mathrm{softmax}(\mathrm{u_{li}}) = \frac{exp(\mathrm{u_{li}})}{\sum_{\mathrm{i=1}}^{2} exp(\mathrm{u_{li}})} \tag{9}$$

Finally, the fused features $h_{Atten} \in R^{L\times 2N}$ at each location is obtained by summing all the feature vectors according to the feature weights, as shown in (10).

$$\mathrm{h_{Atten}} = \sum_{\mathrm{i=1}}^{2} \alpha_{\mathrm{li}} \bullet h_{fi} \tag{10}$$

where α_{li} is the influence weight at the i-th feature and h_{fi} is the i-th feature vector.

2.4.2. Global attention allocation of sequences

Human mobility in cities has irregular transition features (Feng et al. 2018). For example, people may visit breakfast or convenience stores in addition to transportation places such as bus and subway stops during their commute. However, these places are generally less important than transportation places in predicting where people will visit next. Therefore, we need to focus on the importance of each location in predicting the next location. This study uses the global attention layer to explore each location information's impact on prediction to give the model a broader vision. The input of the global attention layer is the comprehensive feature h_{Atten_j} obtained in Section 2.4.1, and the calculation method is shown in (11) to (13).

$$\mathrm{u_{gj}} = \mathrm{V_g^T}\, tanh\left(\mathrm{W_g} \bullet \mathrm{h_{Atten_j}} + \mathrm{b_g}\right) \tag{11}$$

$$\alpha_{\mathrm{gj}} = \mathrm{softmax}\left(\mathrm{u_{gj}}\right) = \frac{exp\left(\mathrm{u_{gj}}\right)}{\sum_{\mathrm{j=1}}^{\mathrm{L}} exp\left(\mathrm{u_{gj}}\right)} \tag{12}$$

$$\mathrm{H} = \sum_{\mathrm{j=1}}^{\mathrm{T}} \alpha_{\mathrm{gj}} \bullet \mathrm{h_{Atten_j}} \tag{13}$$

where W_g and b_g are the coefficient matrix and bias term in the attention mechanism, respectively. T represents the total number of moments, α_{gj} is the influence weight of each location at the j-th moment, L is the length of the trajectory sequence, and H is the result.

In the final prediction module, based on the feature vector output from the global attention layer, the fully connected layer and softmax are used to obtain the probabilities of each candidate location. The candidate location with the highest probability is regarded as the prediction result.

2.5. Model prediction and validation

In addition to Markov, we chose to compare our method with the BiLSTM-CNN, CNN, and LSTM models, open-sourced by Bao et al., to prove the validity of the proposed model. The BiLSTM-CNN model combines LSTM with CNN, where the output of LSTM is used as the input of CNN, and 1D convolution with kernel sizes of 2, 3, and 5 is used in CNN. These models were fed with the same trajectory data after word embedding. The study also tested the effectiveness of multilayer attention models for location prediction. Experiments were conducted on GEMA-BiLSTM with no attention, only a local attention layer, and only a global attention layer. The original attention module's input was used directly in GEMA-BiLSTM without an attention layer. GEMA-BiLSTM had one local attention layer, and GEMA-BiLSTM had one global attention layer.

The prediction performance evaluation was conducted from two aspects: location prediction and travel semantic prediction. To evaluate the location prediction performance of the model, we used the mean absolute error (MAE), root mean square error (RMSE), mean relative error (MRE) and accuracy as evaluation indicators, as shown in formulas (14) to (18):

$$\mathrm{MAE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} |\Delta \mathrm{Dis}(\mathrm{y_i}, \widehat{\mathrm{y}}_{\mathrm{i}})| \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\Delta\mathrm{Dis}(y_i, \hat{y}_i)\right)^2} \tag{15}$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\Delta\mathrm{Dis}(y_i, \hat{y}_i)|}{\mathrm{Distance}_i} \tag{16}$$

$$\mathrm{Hit}(y_i, \hat{y}_i) = \begin{cases} 1, |\Delta\mathrm{Dis}(y_i, \hat{y}_i)| \le \gamma \\ 0, |\Delta\mathrm{Dis}(y_i, \hat{y}_i)| \rangle \gamma \end{cases} \tag{17}$$

$$\mathrm{Accuracy} = \frac{\sum_{i=1}^{n}\mathrm{Hit}(y_i, \hat{y}_i)}{n} \tag{18}$$

where n is the number of trajectories, y_i is the actual location of user i, $\hat{y}_i$ is the predicted location of user i, and $\Delta Dis(y_i, \hat{y}_i)$ is the spherical distance between y_i and $\hat{y}_i$. $Distance_i$ is the sum of the moving distances of user i, and γ is the error threshold. Considering the range of human activities, γ is set to 500 m.

Travel semantic prediction performance measures whether the predicted location matches the travel semantics of the actual location. Travel semantics are a category of POI that people are most likely to visit in the neighborhood of the travel destination (Yue, Wang, Hu, et al., 2012, Liu, Wu, & Peng, 2022). To evaluate travel semantic prediction performance, we first identified the functional semantics of each land in the city using POI data. Then, the POI category with the highest attractiveness in the neighborhood of the travel destination was used as the travel semantic. Finally, we used the confusion matrix to assess the travel semantic overall accuracy. It is the proportion of correctly identified samples to the total number of samples in the matrix.

3. Case study

3.1. Study area

This study selected Shenzhen city to investigate next-location prediction. As shown in Fig. 3, Shenzhen has 10 administrative districts, of which the Nanshan, Futian, Luohu, and Yantian Districts are within the special economic zone. According to the 2012 Shenzhen Statistical Yearbook, the year-end permanent population of Shenzhen was 10.46 million. In terms of urbanization, Shenzhen ranks among the top cities in China. Generally, a developed city takes 20 years of a development planning cycle and does not change much in urban function and structure (Liu, Li, and Yang, 2018). The human activity patterns identified in Shenzhen are representative.

3.2. Study data

The dynamic data from mobile phone users can accurately reflect human movement patterns (Li et al., 2019). This study collected 16.3 million mobile phone signaling data (MPSD) tracks from a large communications operator, recorded regularly in time. The number of users in the MPSD matched the year-end permanent population data from the 2012 Shenzhen Statistical Yearbook (10.46 million), indicating that MPSD is representative of the city's population. Notably, 82.6% of Chinese used mobile phones in 2012 (China Digital Divide Team, 2013). Table 1 displays that the MPSD utilized in this study covers crucial information, including user IDs, time records, and geographical locations. To preserve privacy, user IDs in the dataset were assigned unknown numbers, and the location recorded was not the users' real-time location but rather the mobile base station's location. We counted the time interval of continuous recording of all users and found that 70.47% of the

Table 1
Example of original mobile phone signaling data.

User ID	Record times	Location	Next record times	…
634ea*******9f8d	20,120,322 23:06:14	113.81** 22.70**	20,120,323 00:06:15	…
f92ce*******96d4	20,120,322 23:17:00	113.92** 22.49**	20,120,323 00:17:01	…
f3fdd*******55d0	20,120,322 23:06:14	113.82** 22.70**	20,120,323 00:06:17	…
5790f*******c970	20,120,323 10:55:40	114.35** 22.70**	20,120,323 11:26:35	…
8880f*******2bd9	20,120,322 23:40:58	113.82** 22.70**	20,120,323 00:40:54	…
…	…	…	…	…
8ac5b*******8593	20,120,323 07:27:42	113.98** 22.57**	20,120,323 07:59:45	…

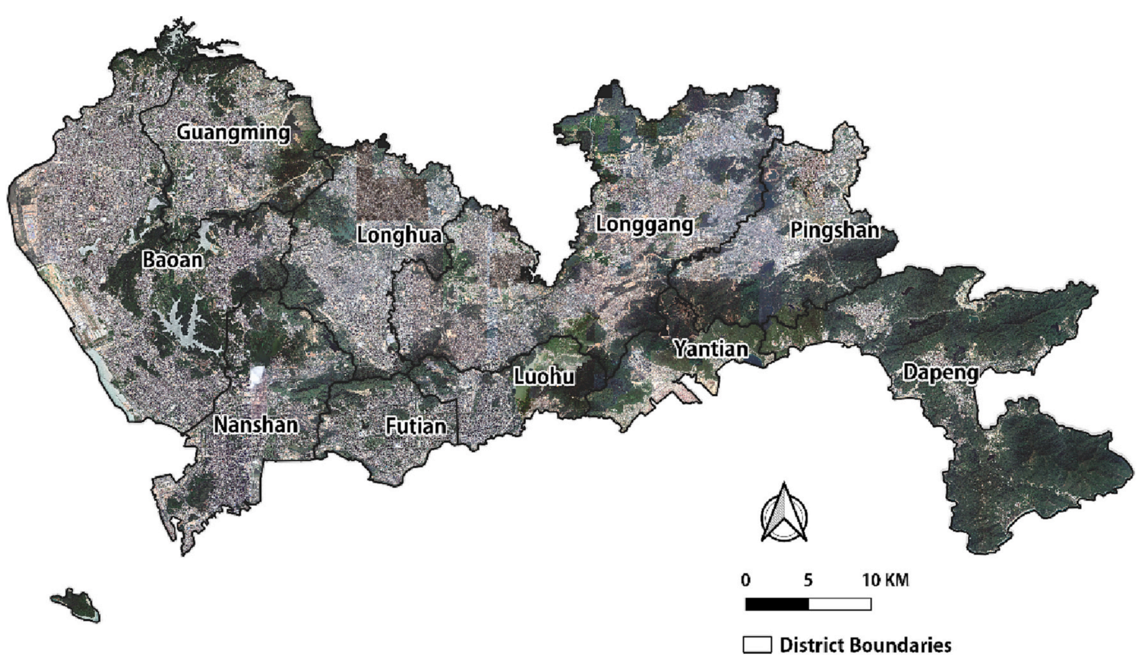

Fig. 3. Remote sensing image of Shenzhen.

time, the interval was between 30 and 60 min.

According to the statistical analysis, there are 5943 base stations in Shenzhen (Fig. 4). We drew Voronoi diagrams based on the base station to obtain the Thiessen polygon and divided and counted the service area of Shenzhen's mobile phone base station. In addition, the performance of the proposed model, such as the distance error level could be compared by calculating the nearest neighbor distance of all Voronoi polygons. Approximately 92% of the base stations had a nearest neighbor at a distance of less than 500 m (Fig. 5).

This study used Shenzhen's Autonavi POI data from 2012 to calculate spatial semantic features. Each row of POI data contained five basic attributes, latitude, longitude, POI category, name, and address. In this study, there were 15 original POI categories, including catering services, factories, government agencies, transportation facilities, education and culture institutions, residential communities, shopping stores, automobile services, hotels, financial institutions, business offices, entertainment venues, medical institutions, tourist attractions, and administrative landmarks.

3.3. Data preprocessing and training details

The MPSD was preprocessed in four steps: (1) Trajectories were selected to form datasets. We selected trajectories where there was at least one location in each hour. When there were multiple locations during a time interval, only one location was retained as the anchor point based on the longest dwell time. (2) Trajectory data was selected for word embedding. We performed location prediction at one-hour intervals. For effective embedding using the CBOW model, the trajectories in the dataset had to correspond with each hour and contain a total of 24 locations. In the CBOW model, the embedding vector size was set to 300, and the window size c was set to 5. (3) The input data was then organized. The record data were organized as follows: $\{l_1, t_1, s_1; l_2, t_2, s_2; \ldots; l_i, t_i, s_i; \ldots; l_M, t_M, s_M\}$. l_i is the Thiessen polygon index in Voronoi diagrams for the i-th location in the users' mobile records. t_i denotes the i-th discretization moment. s_i denotes the geo-semantic vector of the i-th location, consisting of the attractiveness of different POI categories within the Thiessen polygon. In the comparison model, the form of the training dataset was reorganized. The input data were of the form: $\{l_1, t_1; l_2, t_2; \ldots; l_i, t_i; \ldots; l_M, t_M\}$. In general, the last tuple of the input data was used as the predicted label.

To determine the travel semantics of the user, we reclassified the POI categories as suggested by Li et al. (2021). We calculated the proportion of different POI categories within Thiessen polygons using the TF-IDF algorithm. The POI category with the highest attractiveness was used as the functional semantics of the base station. Table 2 shows the classification rules and the proportion of corresponding functional semantics. Similar to the results of the proportion of urban function classification (Li et al., 2021), the commercial services category is more distributed, followed by public services and residential areas (Fig. 6). In addition, we counted the proportion of visits to each functional semantic (Fig. 7), which may help to support the experimental results.

3.4. Results

3.4.1. Next-location prediction results and location prediction accuracy

By preprocessing the original data, we obtained approximately 6 million trajectory data points and divided the human trajectory dataset into training sets and test sets in a 7:3 ratio for the experiments. During the training process for each model, accuracy was the primary metric for evaluating the model's performance.

In Table 3, compared to the traditional Markov model, better results are indicated in all metrics for LSTM, CNN, BiLSTM-CNN, and GEMA-BiLSTM. This shows the superiority of deep learning models in location prediction. The proposed GEMA-BiLSTM model's performance was the best in all metrics, with an improvement in prediction accuracy of 6.21% and 2.28% compared to the LSTM and BiLSTM-CNN models, respectively. Furthermore, we observed a gradual improvement in the performance of deep neural network models by incorporating an

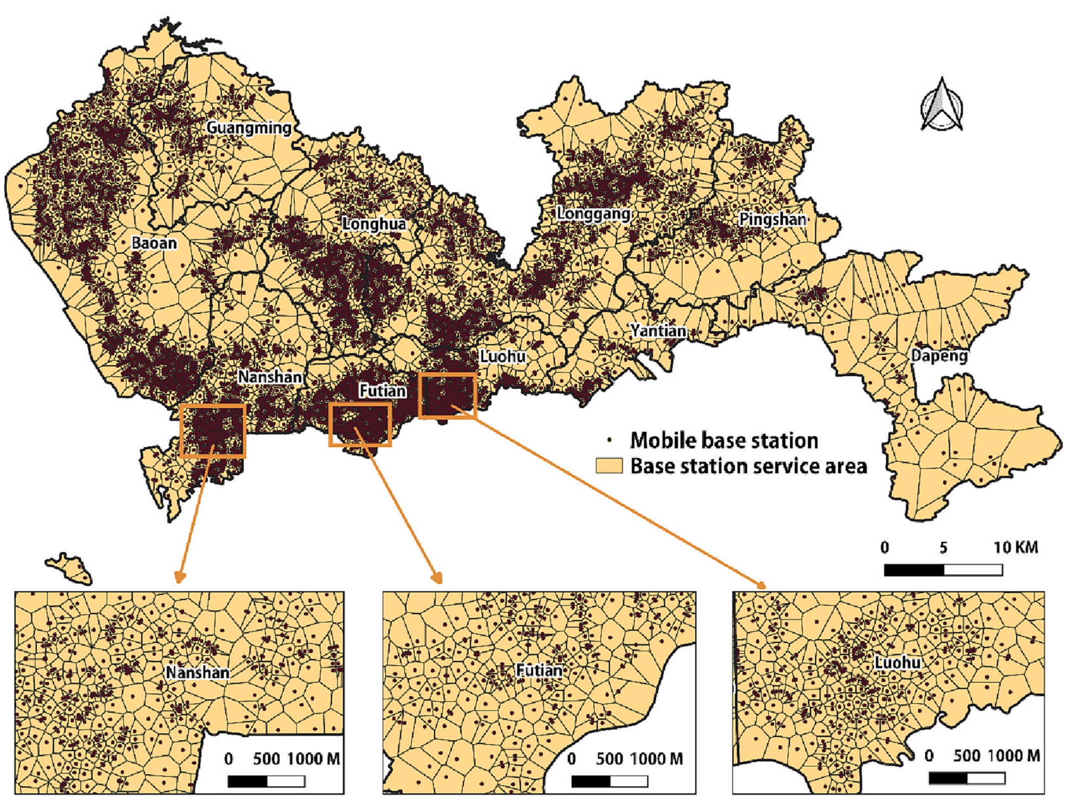

Fig. 4. Distribution of mobile base stations in Shenzhen.

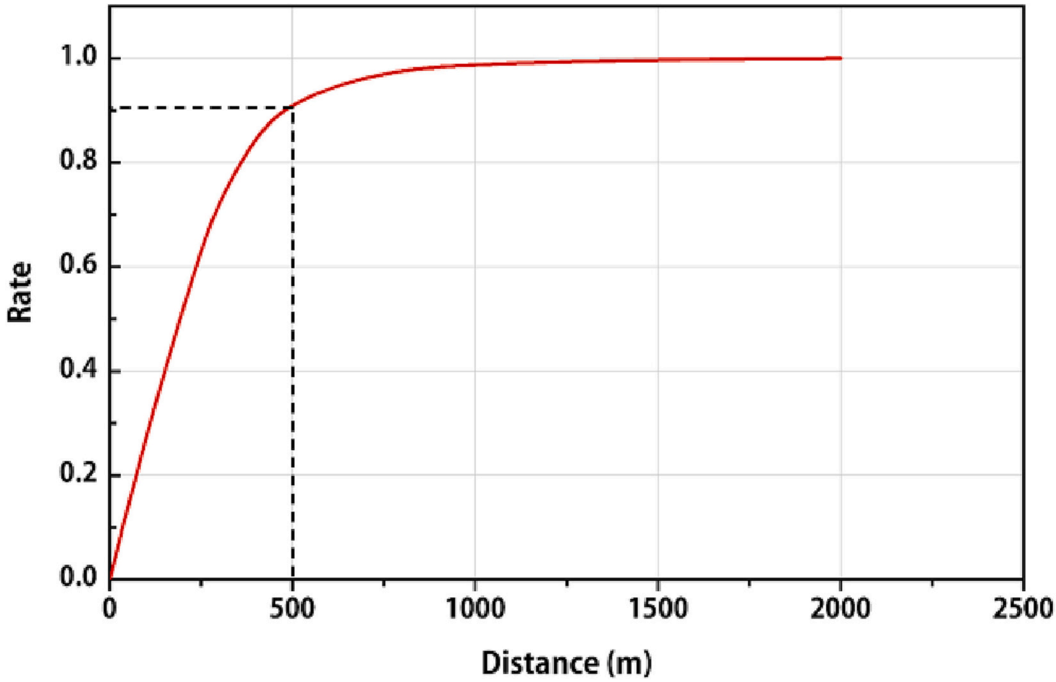

Fig. 5. Cumulative distribution function. The rate is the cumulative frequency, which is calculated from the nearest neighbor distance of all Thiessen polygons in Shenzhen.

Table 2
POI and urban land use type mapping relationship.

POI types	Urban land use types	Percentage (%)
Residential Communities	Residential	5.33%
Medical Institutions	Public Services	26.64%
Education and Culture		
Administrative Landmarks		
Government Agencies		
Catering Services	Commercial Services	49.81%
Shopping Stores		
Automobile Services		
Hotels		
Financial Institutions		
Business Buildings		
Entertainment Venues		
Transportation facilities	Transportation	2.54%
Factories	Industrial	14.40%
Tourist Attractions	Scenery	1.28%

attention module. The addition of a local attention layer and a global attention layer improved performance by 4.58% and 5.69%, respectively, compared to the LSTM model. This improvement may be because the attention module can extract the most relevant contextual information to enhance the model's performance.

The spatial autocorrelation analysis of the prediction errors was performed for each polygon defined by the Voronoi diagram. Statistical analysis revealed that all the models exhibited significant error aggregation at the 1% confidence level, which may be caused by inaccurate information about the user's location. Compared to other models, such as BiLSTM-CNN (Moran's $I = 0.0245$), LSTM (Moran's $I = 0.0707$), CNN (Moran's $I = 0.0095$), and Markov (Moran's $I = 0.7106$), GEMA-BiLSTM (Moran's $I = 0.0076$) showed a weaker spatial autocorrelation of errors, indicating its greater robustness.

Considering the temporal characteristics of human activities, this study divided a day into four parts: morning rush hour (6:00–9:59), daytime slow hour (10:00–15.59), evening rush hour (16:00–20.59), and night slow hour (21:00–5:59 the next day). Based on the accuracy performance of each model during the four periods (Fig. 8), it is evident that there are significant differences in their performance. In the morning rush hour, the accuracy of the proposed GEMA-BiLSTM model (84.66%) is the highest, while the Markov model (36.25%) is the lowest. Meanwhile, the GEMA-BiLSTM model accuracy is significantly higher during all other periods. Moreover, the prediction model's performance is associated with the time period (Fig. 8). All models are less accurate in predicting the location during the evening rush hour than in other periods, indicating the diverse purposes of human activity during the evening rush hour.

3.4.2. Travel semantic prediction accuracy

To further analyze the proposed algorithm's performance, we evaluated its travel semantic prediction accuracy. The results show that GEMA-BiLSTM can effectively tap into users' travel semantics by combining the functional semantics of locations (Fig. 9). Overall, GEMA-BiLSTM outperforms the LSTM (63.27%), Markov (67.04%), CNN (71.70%), and BiLSTM-CNN (72.96%) models with the highest overall accuracy (75.35%) for identifying travel semantics. The model performs well in identifying various functional areas, including residential, public services, scenery, industrial, commercial services, and transportation areas. Notably, the proposed model significantly reduces the percentage of misclassification into diverse functional lands. For example, for inferring the travel semantics to residential areas, GEMA-BiLSTM (19.61%) has fewer errors compared to LSTM (29.96%), Markov (27.40%), BiLSTM-CNNN (24.20%), and CNN (22.96%) for identifying commercial areas. Moreover, the error rate of GEMA-BiLSTM misclassification into commercial is also the lowest in all the inferences of other semantics.

The study found that residential activities are frequently classified as commercial services, with GEMA-BiLSTM misclassifying nearly 20% of residential activities. This suggests that the travel characteristics of commercial activities are complex and have characteristics similar to other travel semantics. The complexity may be due to the diverse functions of urban land use, where commercial land is frequently combined with other functions, such as entertainment, tourism, or cultural education (Cheng et al., 2018).

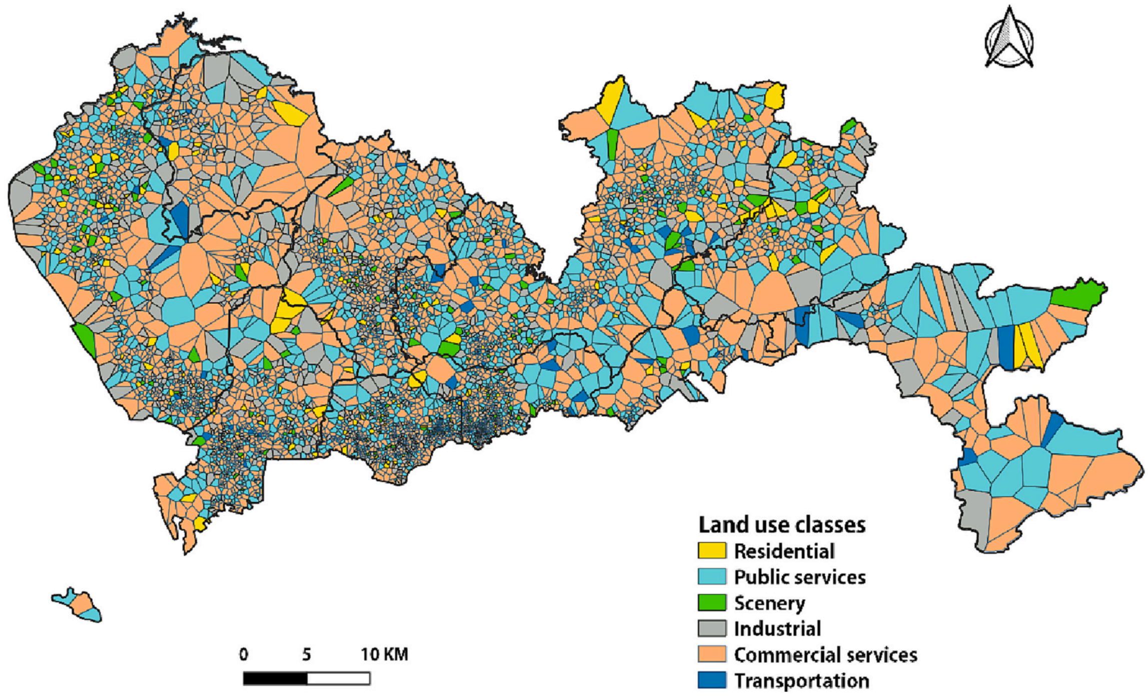

Fig. 6. Land use distribution based on base station service area.

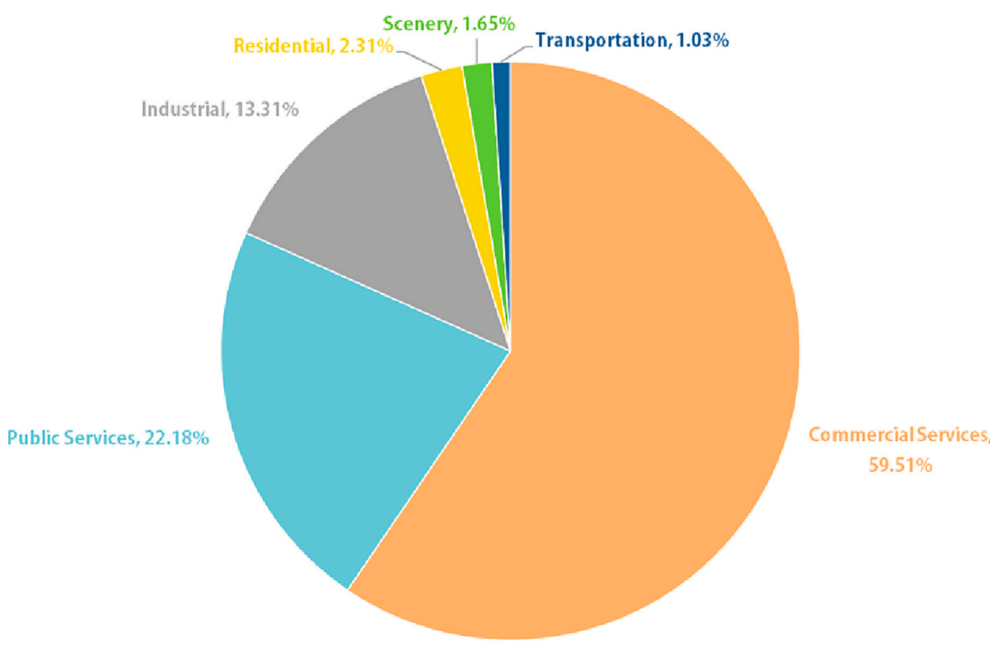

Fig. 7. Proportion of mobile phone users performing different activities.

3.4.3. Human activity pattern analysis

To analyze the pattern of human activity, we counted the proportion of functional areas in population activities during the four periods (Fig. 10). The study identified the "residence-work-residence" activity pattern using location prediction on weekdays. The morning and evening rush hour periods had a higher proportion of transportation, as office workers predominantly use public transport. The proportion of activities in industrial, public service and scenery function areas increased initially until the daytime slow hour and then decreased, whereas the residential areas exhibited an inverse trend. These findings suggest that geospatial big data can reflect activity preferences, with the weekday activity patterns of users aligning with those of residents (Liu et al., 2012).

Nighttime attraction to commercial services is a unique phenomenon, as suggested by the changing trend in the proportion of activities from the daytime slow hour to the evening rush hour. The proportion of residential and transportation activities increased, reflecting people's commuting and rest needs. Furthermore, the proportion of commercial service areas has increased, potentially resulting from government policies promoting the development of the nighttime economy. The land for commercial services has launched night leisure projects to promote economic development (Seijas and Gelders, 2021).

Table 3
Comparison of the predicted performance.

Model	MAE (m)	RMSE (m^2)	MRE	Accuracy (%)
Markov	1344.90	4157.06	0.22	69.99
CNN	967.32	3162.40	0.14	76.10
LSTM	833.98	2819.13	0.11	77.91
GEMA-BiLSTM (no attention model)	758.19	2698.33	0.11	79.09
GEMA-BiLSTM (only local attention layer)	677.52	2478.21	0.08	83.67
GEMA-BiLSTM (only global attention layer)	616.44	2405.71	0.07	84.78
BiLSTM-CNN	579.17	2286.45	0.07	85.35
GEMA-BiLSTM	498.23	2081.89	0.06	87.63

3.4.4. Future trajectory prediction based on next step location prediction

To apply the next-location prediction model, we implemented the prediction of future trajectories. In the multistep forward prediction process, we used an iterative prediction strategy (Guen and Thome, 2020). Specifically, the location prediction at time i and the location prediction at time i + 1 were both based on the information from the most recent past n time steps, and the output at time i served as the input at time i + 1. In the study, we used the most recent twenty-three locations to predict six steps forward.

The accuracy of the future trajectory prediction was high; the average prediction accuracy was 61.20% (Fig. 11). The prediction performance monotonically decreased during the six forward steps, which was reflected by the MAE and *accuracy* metrics. This is due to the accumulation of errors caused by the inaccurate predicted location being implicated in the next step of location prediction. The multistep

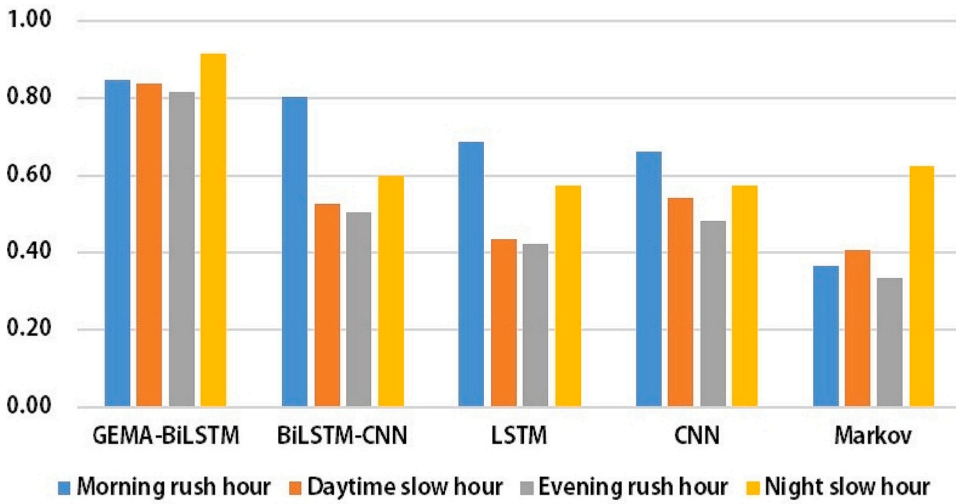

Fig. 8. Comparison of the accuracy of different models in four periods.

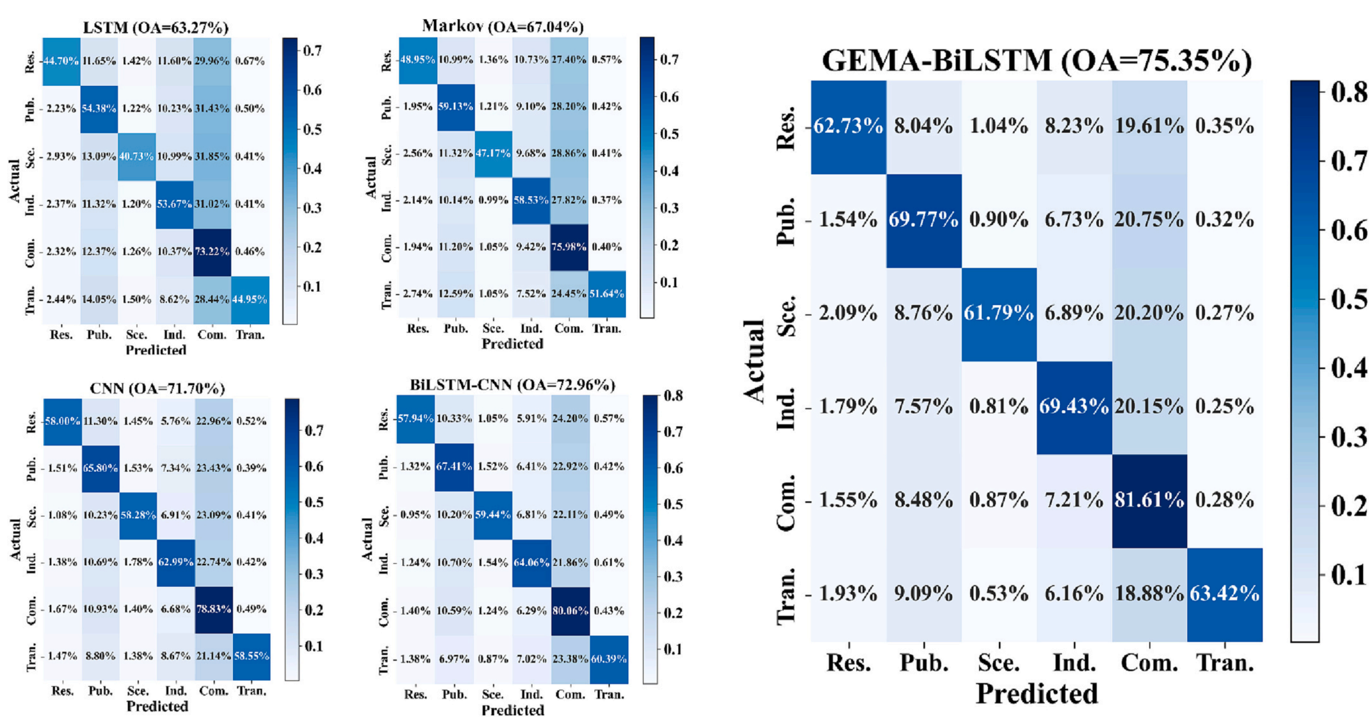

Fig. 9. Confusion matrix for semantic prediction of location by seven models. The semantic categories include: Residential (Res), Public services (Pub), Scenery (Sce), Industrial (Ind), Commercial services (Com), and Transportation (Tran).

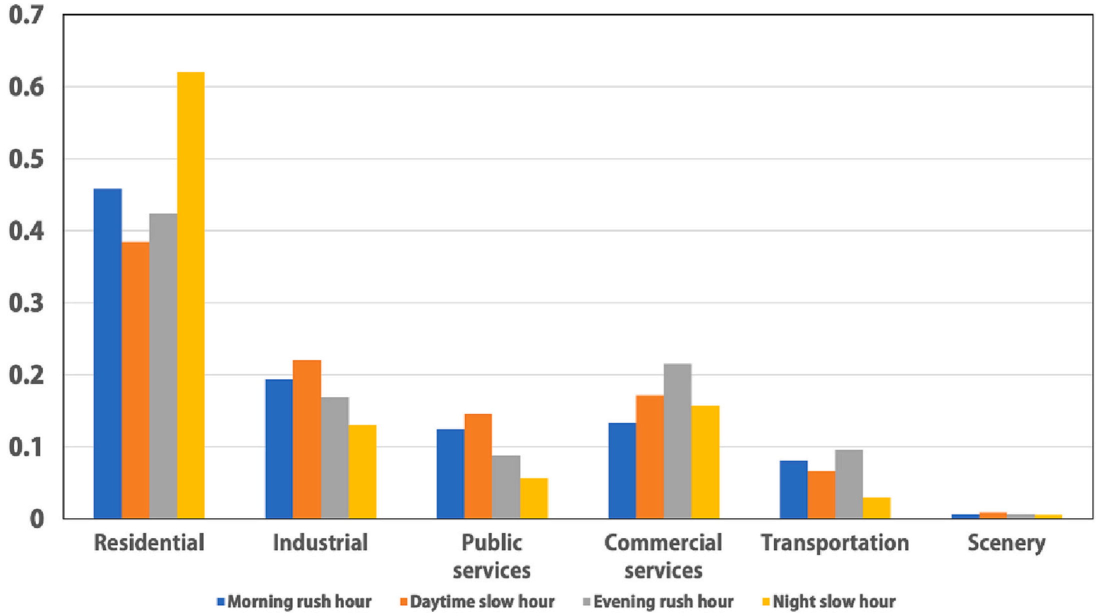

Fig. 10. Proportion trend of different functional land use types in different periods.

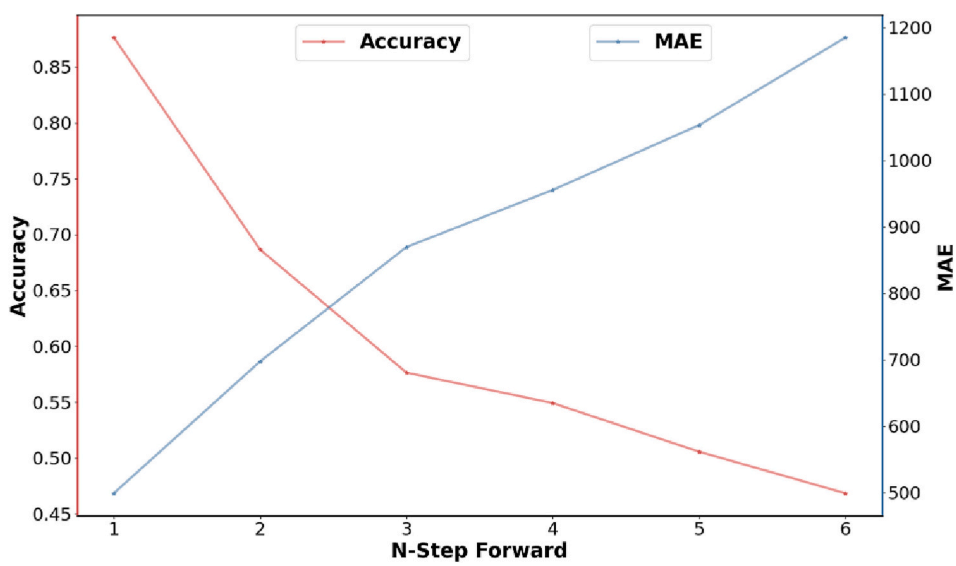

Fig. 11. Multistep forward prediction results at one-hour intervals.

forward prediction converged to one location using the GEMA-BiLSTM model. In Fig. 12, the predicted locations from T14 to T17 in the trajectory of user A do not shift. Similarly, the predicted locations from T10 to T14 obtained in the trajectory of user B do not shift. The period from T10 to T14 is at the night slow hour, when people are resting and the convergence point may be the residence.

4. Discussion and conclusion

The location prediction of human mobility helps uncover complex human behavior and travel patterns. Previous studies have revealed the interaction between human travel patterns and land use within cities (Choi, No, Park, et al., 2022; Lee, Hwang, Park, et al., 2022). However, the correlation between trajectories and spatio-temporal features has not been adequately considered in location prediction studies. This limits the ability of models to extract deep semantic information from trajectories and, consequently, impacts the application of location prediction. In this study, we have proposed the GEMA-BiLSTM model for location prediction. The main priority of the model is on extracting deep semantic information through encoding geolocation information such as MPSD and POI and fusing spatio-temporal feature using attention mechanisms module.

4.1. Semantically enriched geo-embedding is effective for location prediction

Our case study on Shenzhen demonstrates that geo-information embedding and attention mechanisms are effective. The GEMA-BiLSTM (no attention model) model generally outperforms both neural-network-based (LSTM, CNN) prediction and model-based prediction models (Markov). This indicates that spatio-temporal features play a crucial role in revealing human activity patterns (Xu, Li, and Xia, 2023). More, the addition of the attention module leads to a further improvement in the accuracy of location prediction, which may be due

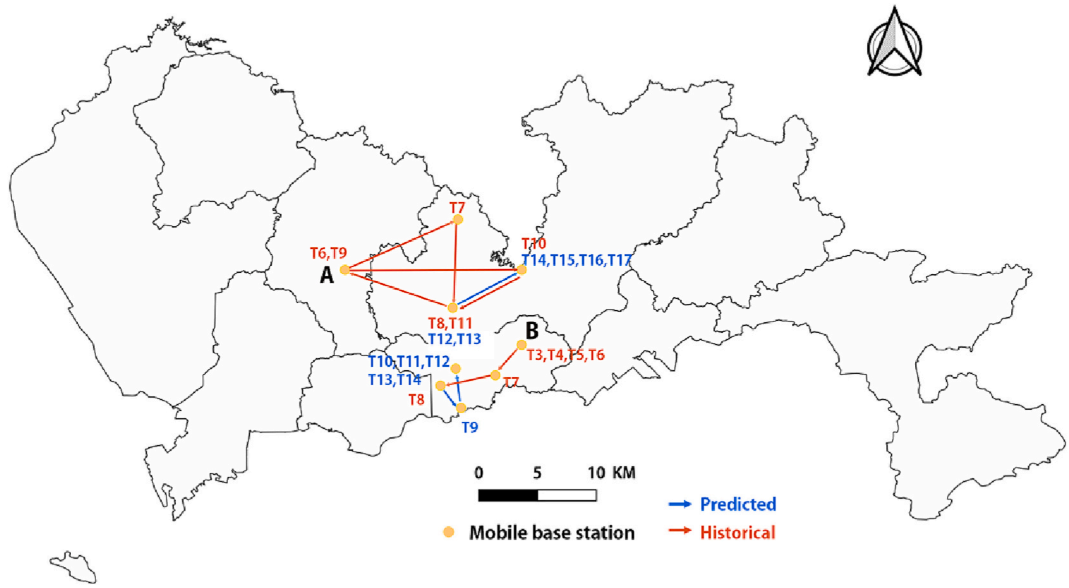

Fig. 12. The figure shows the predicted future trajectory prediction results for user A and user B. To show the trajectory for convenience, we give an example of multistep location prediction using the recent six locations with a two–hour interval. The blue text indicates the period to be predicted, and the orange text indicates the historical period used for the first step of the multistep prediction. T3 denotes the period from 4:00 am–6:00 am, and so on until T17 which denotes the period from 8:00 am–10 am of the next day. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to the fact that the attention mechanism enhances the neural network model for mining complex dependency patterns of trajectories and learning of spatio-temporal features (Huang, Ma, Wang, et al., 2019). In general, the GEMA-BiLSTM model's accuracy improves by 17.64%, 11.53%, 9.72%, and 2.28% compared with Markov, CNN, LSTM, and BiLSTM-CNN, respectively. Accurately predicting the next location in user activities is a major advantage of the proposed GEMA-BiLSTM model.

Utilizing the attention mechanism and embedding spatio-temporal feature information, the GEMA-BiLSTM model effectively captures the travel semantics of users. Typically, commuting usually occurs at a specific time and place, and has a clear continuity and regularity (Hadachi, Pourmoradnasseri, and Khoshkhah, 2020). The GEMA-BiLSTM model shows significantly improved performance in identifying Transportation and Residential, indicating its ability to capture travel features in human activities. The high percentage of misclassification into commercial areas may be attributed to the fact that commercial areas are often mixed with other functional attributes (Zhou, Ming, Lv, et al., 2020). To avoid biases in predicted locations caused by inaccurate spatial features, it is imperative to give special attention to mixed functional lands.

4.2. Human activity pattern revealed by the proposed model

From the perspective of the temporal characteristics of human mobility, the model highlights the complexity of human activity during different periods. Predicting stationary periods such as the night slow hours is considerably simpler compared to other periods. This is primarily because individuals' behavior during these periods follows a more predictable pattern, allowing for more accurate prediction. At night, most people rest at home, and urban residents go out for fewer activities and travel for a single purpose (Dai et al., 2017). These findings reinforce the importance of considering temporal characteristics when modeling human mobility.

Through travel pattern analysis and multi-step forward prediction, governments can develop strategies for effectively managing cities. During the working day, in addition to residential areas, transportation and workplaces, governments also need to focus on the allocation of public transport resources in commercial areas. This is because unusual activity patterns during evening rush hour periods, especially with increased commercial services. The diverse patterns of movement are related to the development of the nighttime economy (Seijas and Gelders, 2021). By collecting multistep forward location data of individual users, we can complete the flow assessments based on the number of individuals within a statistical measurement unit (Huang, Ling, Wang, et al., 2018). Since multi-step forward predictions eventually stop at one place which typically denotes long-stay locations such as residential areas or workplaces, researchers need to be concerned that the reliability of urban travel demand assessments also decreases.

While the performance of our model can be satisfactory, further research is needed to consider additional factors such as weekends and holidays. Human mobility is determined by geographical and socio-economic factors (Gao, Liu, Wang, et al., 2013). Abitbol and Morales (2021) have shown that different income or educational levels affect the behavior of urban residents. We will incorporate this human characteristics data to improve the prediction accuracy in our future work.

Fundings

This work was supported by the National Key Research and Development Program of China (Grant No. 2019YFB2102903), the National Natural Science Foundation of China (Grant No. 42171466 and 41801306), the “CUG Scholar” Scientific Research Funds at China University of Geosciences (Wuhan) (Project No. 2022034), the Alibaba Group through Alibaba Innovation Research Program [20228670] and the China Scholarship Council [202208440090].

Disclosure statement

No potential competing interest was reported by the authors.

Declaration of Competing Interest

No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part.

Data availability

The data and codes that support the findings of the present study are available on Figshare at https://figshare.com/s/b9cd906a5bd8931366cc

References

Abitbol, J. L., & Morales, A. J. (2021). Socioeconomic patterns of Twitter user activity. *Entropy, 23*(6), 780.

Alahi, A., et al. (2016). Social lstm: Human trajectory prediction in crowded spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)* (pp. 961–971). IEEE: Las Vegas, NV, USA.

Bao, Y., et al. (2021). A BiLSTM-CNN model for predicting users' next locations based on geotagged social media. *International Journal of Geographical Information Science, 35*(4), 639–660.

Bernecker, T., Cheng, R., Cheung, D. W., et al. (2013). Model-based probabilistic frequent itemset mining. *Knowledge and Information Systems, 2013*(37), 181–217.

Cai, L., Xu, J., Liu, J., et al. (2019). Sensing multiple semantics of urban space from crowdsourcing positioning data. *Cities, 93*, 31–42.

Chen, G., et al. (2019). Complete trajectory reconstruction from sparse mobile phone data. *EPJ Data Science, 8*(1), 30.

Chen, T., et al. (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications, 72*, 221–230.

Cheng, Z., et al. (2018). Mix leading to success? Exploring the innovative development model in peri-urban China. *Habitat International, 82*, 1–8.

China Digital Divide Team. (2013). National Information Center, report on China's digital divide [Chinese]. *Information Research, 9*.

Choi, J., No, W., Park, M., et al. (2022). Inferring land use from spatialtemporal taxi ride data. *Applied Geography, 142*, Article 102688.

Cobbinah, P. B., et al. (2022). Contested urban spaces in unplanned urbanization: Wetlands under siege. *Cities, 121*, Article 103489.

Dai, L., et al. (2017). Spatiotemporal structure features of network check-in activities of urban residents and their impacting factors: A case study in six urban districts of Beijing. *Journal of Asian Architecture and Building Engineering, 16*(1), 131–138.

Deschaintres, E., Morency, C., & Trépanier, M. (2022). Cross-analysis of the variability of travel behaviors using one-day trip diaries and longitudinal data. *Transportation Research Part A: Policy and Practice, 163*, 228–246.

Erdelić, T., et al. (2021). Estimating congestion zones and travel time indexes based on the floating car data. *Computers, Environment and Urban Systems, 87*, Article 101604.

Feng J., Li Y., Zhang C., et al. (2018). Deepmove: Predicting human mobility with attentional recurrent networks. Proceedings of the 2018 world wide web conference, 1459-1468.

Gambs, S., Killijian, M., & Del Prado Cortez, M. N. (2012). Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility* (pp. 1–6). Bern, Switzerland: Association for Computing Machinery.

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.

Gao, S., Liu, Y., Wang, Y., et al. (2013). Discovering spatial interaction communities from mobile phone data. *Transactions in GIS, 17*(3), 463–481.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*(5–6), 602–610.

Guen, V. L., & Thome, N. (2020). Disentangling physical dynamics from unknown factors for unsupervised video prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11474–11484.

Guo, L., et al. (2010). Uncertain path prediction of moving objects on road networks. *Journal of Computer Research and Development, 47*(1), 104.

Hadachi, A., Pourmoradnasseri, M., & Khoshkhah, K. (2020). Unveiling large-scale commuting patterns based on mobile phone cellular network data. *Journal of Transport Geography, 89*, Article 102871.

Huang, L., Ma, Y., Wang, S., et al. (2019). An attention-based spatiotemporal lstm network for next poi recommendation. *IEEE Transactions on Services Computing, 14*(6), 1585–1597.

Huang, Z., Ling, X., Wang, P., et al. (2018). Modeling real-time human mobility based on mobile phone and transportation data fusion. *Transportation Research Part C: Emerging Technologies, 96*, 251–269.

Humtsoe, T. Y. (2022). Travel mode choice in the north-eastern Indian City of Kohima: Lessons from empirical study. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 1–15.

Jensen, C. S., et al. (2006). Effective density queries on continuouslymoving objects. In *In: 22nd International Conference on Data Engineering (ICDE'06)* (p. 71). Atlanta, GA, USA: IEEE.

Lee, M., Hwang, S., Park, Y., et al. (2022). Factors affecting bike-sharing system demand by inferred trip purpose: Integration of clustering of travel patterns and geospatial data analysis. *International Journal of Sustainable Transportation, 16*(9), 847–860.

Li, F., et al. (2020). A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction. *Neurocomputing, 403*, 153–166.

Li, J., Jin, K., Zhou, D., et al. (2020). Attention mechanism-based CNN for facial expression recognition. *Neurocomputing, 411*, 340–350.

Li, M., et al. (2019). Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data. *Computers, Environment and Urban Systems, 77*, Article 101346.

Li, Z., et al. (2021). Understanding the pattern and mechanism of spatial concentration of urban land use, population and economic activities: A case study in Wuhan, China. *Geo-Spatial Information Science*, 1–17.

Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing, 337*, 325–338.

Liu, W., et al. (2020). The geography of human activity and land use: A big data approach. *Cities, 97*, Article 102523.

Liu, Y., Li, J., & Yang, Y. (2018). Strategic adjustment of land use policy under the economic transformation. *Land Use Policy, 74*, 5–14.

Liu, X., Wu, M., Peng, B., et al. (2022). Graph-based representation for identifying individual travel activities with spatiotemporal trajectories and POI data. *Scientific Reports, 12*(1), 15769.

Liu, Y., et al. (2012). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning, 106*(1), 73–87.

Liu, Y., et al. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers, 105*(3), 512–530.

Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In *arXiv Preprint arXiv, 1508.04025*.

Mai, G., et al. (2022). A review of location encoding for GeoAI: Methods and applications. *International Journal of Geographical Information Science*, 1–35.

Meng, C., et al. (2017). Travel purpose inference with GPS trajectories, POIs, and geotagged social media data. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 1319–1324). IEEE: Boston, MA, USA.

Miao, C., Luo, Z., Zeng, F., et al. (2020). Predicting human mobility via attentive convolutional network. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (pp. 438–446).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *arXiv Preprint arXiv, 1301.3781*.

Qian, J., et al. (2021). Quantify city-level dynamic functions across China using social media and POIs data. *Computers, Environment and Urban Systems, 85*, Article 101552.

Qiao, S., et al. (2014). A self-adaptive parameter selection trajectory prediction approach via hidden Markov models. *IEEE Transactions on Intelligent Transportation Systems, 16*(1), 284–296.

Sarkar, P. P., & Chunchu, M. (2016). Quantification and analysis of land-use effects on travel behavior in smaller Indian cities: Case study of Agartala. *Journal of Urban Planning and Development, 142*(4), 4016009.

Seijas, A., & Gelders, M. M. (2021). Governing the nighttime city: The rise of night mayors as a new form of urban governance after dark. *Urban Studies, 58*(2), 316–334.

Selva Birunda, S., & Kanniga, D. R. (2021). A review on word embedding techniques for text classification. *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA, 2020*, 267–281.

Solomon, A., et al. (2021). Analyzing movement predictability using human attributes and behavioral patterns. *Computers, Environment and Urban Systems, 87*, Article 101596.

Song, L., Kotz, D., Jain, R., et al. (2003). Evaluating location predictors with extensive Wi-Fi mobility data. *ACM SIGMOBILE Mobile Computing and Communications Review, 2003, 7*(4), 64–65.

Tian, Y., et al. (2021). An individual-based spatio-temporal travel demand mining method and its application in improving rebalancing for free-floating bike-sharing system. *Advanced Engineering Informatics, 50*, Article 101365.

Tu, W., et al. (2017). Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science, 31*(12), 2331–2358.

Wang, R., Zhang, X., & Li, N. (2022). Zooming into mobility to understand cities: A review of mobility-driven urban studies. *CITIES, 130*, Article 103939.

Xia, C., Hu, Y., & Chen, J. (2023). Community time-activity trajectory modeling based on Markov chain simulation and Dirichlet regression. *Computers, Environment and Urban Systems, 100*, Article 101933.

Xu, C., Li, F., & Xia, J. (2023). Fusing high-resolution multispectral image with trajectory for user next travel location prediction. *International Journal of Applied Earth Observation and Geoinformation, 116*, Article 103135.

Xu, J., et al. (2015). Trajectory big data: Data, applications and techniques. *Journal on Communications, 36*(12), 97.

Yao, D., Zhang, C., Huang, J., et al. (2017). Serm: A recurrent model for next location prediction in semantic trajectories. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 2411–2414).

Yue, Y., Wang, H., Hu, B., et al. (2012). Exploratory calibration of a spatial interaction model using taxi GPS trajectories. *Computers, Environment and Urban Systems, 36*(2), 140–153.

Zhai, W., et al. (2019). Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment and Urban Systems, 74*, 1–12.

Zhang, J., et al. (2021). The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types. *International Journal of Geographical Information Science, 35*(1), 193–211.

Zhang, Y., Liu, L., & Wang, H. (2019). A new perspective on the temporal pattern of human activities in cities: The case of Shanghai. *Cities, 87*, 196–204.
Zhao, Y., Chen, B. Y., Gao, F., et al. (2023). Dynamic community detection considering daily rhythms of human mobility. *Travel Behaviour and Society, 31*, 209–222.
Zhou, W., Ming, D., Lv, X., et al. (2020). SO–CNN based urban functional zone fine division with VHR remote sensing image. *Remote Sensing of Environment, 236*, Article 111458.

Yao Yao is a Professor at China University of Geosciences (Wuhan), a researcher at the University of Tokyo and a senior algorithm engineer at Alibaba Group. His research interests are geospatial big data mining, analysis, and computational urban science.

Zijin Guo is a graduate student at China University of Geosciences (Wuhan) and a visiting graduate student at the University of Tokyo. His research interests are trajectory data mining and complex network analysis.

Chen Dou is a graduate student at Wuhan University. His research interests are trajectory big data mining and geographic visualization.

Minghui Jia is a graduate student at Wuhan University. His research interests are nighttime light remote sensing and geospatial data mining.

Ye Hong is a Doctoral Candidate in the Institute of Cartography and Geoinformation at the Swiss Federal Institute of Technology (ETH) Zurich. His research interests include spatiotemporal data mining and analysis, simulation & prediction of human mobility, and sustainability mobility assessments.

Qingfeng Guan is a Professor at China University of Geosciences (Wuhan). His research interests are high-performance spatial intelligence computation and urban computing.

Peng Luo is a PhD candidate at Technical University of Munich. His research interests are spatiotemporal big data mining and uncertainty analysis.