GeoInformatica (2023) 27:137–157
https://doi.org/10.1007/s10707-022-00467-0

# ConvGCN-RF: A hybrid learning model for commuting flow prediction considering geographical semantics and neighborhood effects

**Ganmin Yin[1,2] · Zhou Huang[1,2] · Yi Bao[1,2] · Han Wang[1,2] · Linna Li[3] · Xiaolei Ma[4] · Yi Zhang[1,2]**

Received: 25 January 2022 / Revised: 5 April 2022 / Accepted: 27 April 2022 /
Published online: 27 May 2022


**Abstract**

Commuting flow prediction is a crucial issue for transport optimization and urban planning. However, the two existing types of solutions have inherent flaws. One is traditional models, such as the gravity model and radiation model. These models rely on fixed and simple mathematical formulas derived from physics, and ignore rich geographic semantics, which makes them difficult to model complex human mobility patterns. The other is the machine learning models, most of which simply leverage the features of Origin-Destination (OD), ignoring the topological nature of the interaction network and the spatial correlation brought by the nearby areas. In this paper, we propose a 'preprocessing-encoder-decoder' hybrid learning model, which can make full use of geographic semantic information and spatial neighborhood effects, thereby significantly improving the prediction performance. Specifically, in the preprocessing part, we divide the study area into grids, and then incorporates features such as location, population, and land use types. The second step of the encoder designs a convolutional neural network (CNN) to achieve the fusion of neighborhood features, constructs a spatial interaction network with the grids as nodes and the flows as edges, and then uses the graph convolutional network (GCN) to extract the embeddings of the nodes. In the last step of the decoder, a random forest regressor is trained to predict the commuting flow based on the learned embedding vectors. An empirical study on a commuter dataset in Beijing shows that our proposed model is approximately 20% better than XGBoost (state-of-the-art), thus proving its effectiveness.

**Keywords** Human mobility · Commuting flow · Hybrid learning · Graph neural network · Geographical semantics · Neighborhood effects

✉ Zhou Huang
huangzhou@pku.edu.cn

1 Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China

2 Beijing Key Lab of Spatial Information Integration & Its Applications, Peking University, Beijing 100871, China

3 Department of Geography, California State University, Long Beach, California, USA

4 School of Transportation Science and Engineering, Beihang University, Beijing 100191, China

## 1 Introduction

Traveling is the movement behavior of human beings from an origin to a destination in geographic space, and it is also a way of spatial interaction between regions. People often travel for certain purposes, such as work, school, shopping, entertainment, etc., and these activities are undertaken by places with specific functions, which is a manifestation of the coupling relationship between people and the environment [1, 2]. Considering this people-place interaction, [3] gives a definition of geographical semantics, paying particular attention to the functional attributes of places that carry human activities. Commuting is an important type of travel, which refers to the behavior of residents going back and forth between home and workplace, and people's living and working are typical activities that occur in the places, therefore, commuting is supposed to be related to the geographical semantics of the regions (such as land use types, etc.) [4–9].

Accurately predicting the number of commuters between two places not only helps us understand the internal geographical mechanism of human mobility, but also guides policymakers to optimize the transportation network and to configure the urban infrastructure. Commuting flow prediction is different from the traffic forecasting problem in the following aspects [10–13]: (1) *The scale of the spatial unit*. The analysis units of traffic forecasting are mostly road nodes, bus stops, or monitoring sensors, attached to the road network for real-time traffic management, while the units of commuting flow prediction are mainly grids, traffic analysis zones (TAZs), streets, or census tracts, to better configure the urban infrastructure to meet the travel demand; (2) *Time independence*. Traffic forecasting is usually a dynamic time-series-based prediction that requires historical flow information, while commuting is a relatively repetitive and static behavior, which focuses more on spatial correlation and is not closely related to time; (3) *Geographical semantic relevance*. Commuters' choice of location for home and work is closely related to the land use types of the places. For example, homes are generally in residential areas, workplaces are generally in commercial and industrial areas, and availability of public facilities also plays a role in the choice of home and workplace.

Existing commuting flow prediction models can be roughly divided into two categories. One is the traditional *spatial interaction models* [14, 15], such as the gravity model and the radiation model, which are mainly derived from the principles of physics. The former [16, 17] assumes that the intensity of interaction between two places is proportional to their populations and decays with distance, while the latter [18, 19] combines the diffusion process with the theory of intervening opportunities and states that in addition to population, the interaction flow is also related to the number of opportunities between them. These models have relatively fixed inputs and simple expressions, unable to capture nonlinear and irregular patterns. They also ignore rich geographic semantic information which is supposed to be relevant to commuting [3, 4, 6], so it is difficult to model complex human mobility patterns. The other category is the *machine learning models* [20], such as artificial neural networks and tree-based models [21–23]. Although these data-driven and black-box models have poor interpretability, they are good at modeling complex patterns. One limitation of the existing models is that they simply use the features of the Origin-Destination (OD), rarely considering the topological nature of the spatial interaction network and the spatial correlations caused by the geographical and topological neighboring regions [24, 25].

To address these limitations, this paper proposes a hybrid learning (*preprocessing-encoder-decoder*) model based on the graph convolutional network, which can make full use of geographic semantic information and spatial proximity effects, thereby significantly improving the performance of commuting flow prediction. In the first *preprocessing* part, we divide the research area into grids, which have features like location, population, and geographical semantics (such as landuse types). Then the *encoder* achieves the fusion of neighborhood features through convolutional neural network (CNN), uses the grids as nodes and the flows as edges to construct a spatial interaction network, and then trains a graph convolutional network (GCN) to extract the representation vectors of the nodes, effectively utilizing the topological structure properties of the spatial interaction network. In the last step of the *decoder*, the learned representation vectors of the nodes, concatenated with the distance between them, are fed to a random forest regressor to achieve a more accurate prediction of commuting flow.

The main contributions of this research are the following three points:

- We propose a preprocessing-encoder-decoder hybrid learning model, which significantly improves the prediction of commuting flow, especially in the absence of historical flow information;
- We combine CNN and GCN to fuse both geographical and topological neighbor features, thereby effectively modeling the spatial correlations. To the best of our knowledge, we are the first to consider both neighborhood effects in this task;
- We find that commuting is closely related to geographic semantics. The experimental results show that when the geographic semantics such as land use types are considered as input, higher prediction accuracy can be achieved.

The rest of the paper is organized as follows. Section 2 reviews the literature on commuting flow prediction and graph neural networks in spatial interaction and summarizes the existing problems. Section 3 provides a formal statement of the research problem. Section 4 introduces the implementation principles and details of our proposed hybrid learning model (*preprocessing-encoder-decoder*). Section 5 presents the results of our model compared with other baselines on a Beijing dataset and evaluates its performance. Section 6 discusses the impact of hyperparameters, flow intensity/distance, and landuse data on the model. Section 7 concludes with major findings and future directions of the study.

## 2 Related work

### 2.1 Commuting flow prediction

This research mainly focuses on the prediction of commuting flow, which has attracted wide attention recently, because it is critical for traffic optimization and urban planning. The gravity and radiation models are the most common spatial interaction models. The gravity model assumes that the intensity of the spatial interaction between two places is directly proportional to their populations and inversely proportional to the power of distance [17], while the radiation model assumes that the main factor affecting interaction is not distance, but the number of intervening opportunities [26]. McArthur et al. [27] research the transferability of the parameters of the gravity model and propose a statistical test method based on bootstrapping. Simini et al. [19] believe that the gravity model has shortcomings such as lack of theoretical derivation and the need for known data, so they propose a parameter-free radiation model that can predict commuting and mobility fluxes.

Lenormand et al. [15] systematically compare the gravity and radiation models on different datasets at different scales and find that the gravity model is generally better for predicting commuter flow. Although these models have few parameters and are easy to understand, their abilities to model complex patterns are limited.

With better abilities to capture nonlinear and irregular distributions, machine learning models have been gradually used in commuting flow prediction. Black [28] introduces the artificial neural network into spatial interaction modeling, using the same input as the gravity model. As a result, the prediction error is greatly reduced, so he believes that his model effectively detects patterns in the data. However, the study by [22] presents different results. When fitting the commuting trip distribution, they find that the multi-layer perceptron is much worse than that of the traditional models. Simini et al. [6] design a deep neural network called Deep Gravity, which integrates multisource features such as point of interests (POIs), land use, and transportation, to achieve more accurate mobility flow generation. In addition to neural networks, tree-based models have also been applied to commuting flow prediction. Morton et al. [21] compare the XGBoost model with the gravity and radiation models and find that the former performs better than the latter two models on multiple indicators. Based on Twitter data, [23] compare the gravity model, neural networks, and random forest on the task of fitting commuting trip distribution, and find that the random forest model has the best performance. Spadon et al. [29] select 22 urban socioeconomic indicators and find that the XGBoost model performs best on binary classification and regression tasks of commuter network reconstruction. These machine learning models can model complex patterns to a certain extent but do not consider the network topological properties and spatial proximity effects, which have been proven important in existing works [25, 30, 31].

### 2.2 Graph neural networks in spatial interaction

There are connections of different strengths between things at different locations in geographic space, including matter, energy, and information, which is called *spatial interaction* [32]. From the perspective of the network, as a spatial interaction behavior, human mobility forms a series of human activity networks, such as commuting networks. The rise of graph neural networks (GNNs) provides a unified knowledge extraction framework for these network data [33–36]. The main idea of GNNs is that the attributes of nodes are related to the local topology of the network, which is obtained through the propagation and aggregation of neighbor nodes' attributes. The existing research of GNNs in spatial interaction mostly focuses on traffic forecasting, especially on the modeling of spatial and temporal correlation. Li et al. [10] innovatively propose a diffusion convolutional recurrent neural network (DCRNN) model, which models the traffic flow as a diffusion process on a directed graph, and extracts time series features through the gated recurrent unit (GRU), which effectively improves the prediction ability. Yu et al. [11] design multiple spatiotemporal convolution blocks, which effectively model the spatial and temporal dependence to achieve the state of the art. Zhao et al. [13] propose a T-GCN model, which uses GCN and GRU to extract spatial and temporal features to make up for the shortcomings of long-term traffic flow prediction. Yao et al. [31] focus on the spatial flow imputation and propose a spatial interaction graph convolutional network model, and prove its effectiveness by comparing it with the gravity and radiation models. However, as mentioned in Section 1, there are differences between traffic forecasting and commuting flow prediction. The former models require historical flows between locations to support the prediction, which are often not available in the latter scenario. Therefore, few studies have tried to apply GNNs to extract the features of commuting networks without knowing any historical information [30], and they also lack the grasp of the geographic semantic dependence of commuting behavior.

### 2.3 Summary

In summary, the existing models for predicting commuting flow have inherent flaws. In the spatial interaction models represented by the gravity and radiation models, few parameters make it unable to take geographic semantics as input, and simple formulas limit the ability to model the complex patterns. The machine learning models represented by neural networks and tree-based models simply use the attributes of Origin-Destination, ignoring the topological properties of the spatial interaction network. Graph neural networks can effectively extract information from network data, but they are currently mostly used in traffic forecasting and rarely used in commuting flow prediction.

This research aims to apply graph neural networks to commuting flow prediction. Considering the geographic semantic relevance of commuting, we decide to take land use types as part of the input. Besides, we also believe that the following two proximity effects need to be considered in the prediction of commuting flow, as shown in Fig. 1:

- Topological proximity in the network. The features of nodes in the network are affected by the local structure and neighbor properties. In a spatial interaction network, if the neighbors of two nodes are similar, the intensity of their flow to the same destination should also be similar. The topological proximity effect can be effectively modeled by graph neural networks [34].
- Geographical proximity between the grids. According to Tobler's first law of geography [24], the correlation between neighboring things in geographic space is more significant. The flow of people generated or attracted by a geographic unit will be generally affected by neighboring areas. This has been proven in existing studies [25, 37], but this proximity effect is generally not considered in the prediction of commuting flow [30]. In this study, we model the spatial correlation to a certain extent through convolution at the beginning of the *encoder*.

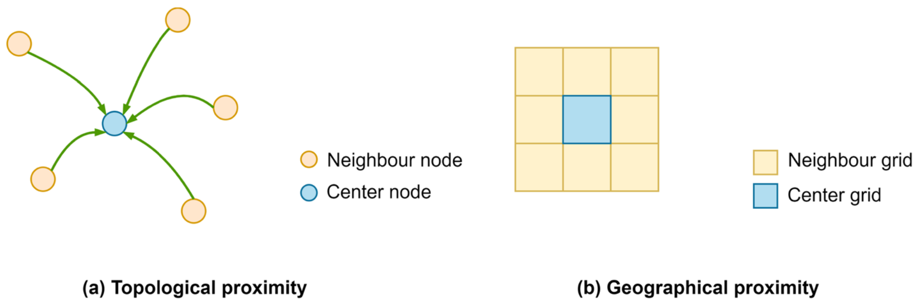



**Fig. 1** Two types of proximity effects: (**a**) Topological proximity; (**b**) Geographical proximity

## 3 Problem statement

In this section, we formalize the definitions and problem to clarify the research goal.

***Definition 1*** **Geographic unit set** $V = \{v_1,v_2,...,v_n\}$. For spatial analysis, the geographic unit can be a grid, traffic analysis zone, street, census tract, etc. In this study, we divide the study area into 500m grids as the geographic units. Each grid has the features of location, population, land use types, etc., represented as a vector.

***Definition 2*** **Commuter flow set** $F = \{(v_i,v_j,f_{ij})\}$. It is a set of triplets in which $v_i$ represents the origin (home), $v_j$ represents the destination (workplace), and $f_{ij}$ is the flow from $v_i$ to $v_j$, that is, the number of commuters.

***Definition 3*** **Commuter network** $G = (V,F)$. The network is a weighted directed graph with the geographic unit set $V$ as the nodes and the commuter flow set $F$ as the edges.

**Problem** When given a commuter network $G = (V,F)$, learn a model $M$ that is effective enough to predict the flow $f_{ij}$ between any two nodes $i$ and $j$, that is, $f_{ij} = M(G,i,j)$, with minimal deviation from the true value.

## 4 Methodology

In this section, we introduce the proposed model framework (Fig. 2) to address the problem mentioned in Section 3. In the preprocessing part, we divide the study area into grids and incorporate multisource features like location, population, and landuse types. Then the encoder achieves the fusion of neighboring features through CNN, thus rebuilding the commuting network and learning the representation vectors of the nodes with GCN as the backbone. Lastly in the decoder, the learned vectors of the origin/destination, concatenated with the distance between them, are used to train a random forest regressor to predict the commuting flow.

### 4.1 Preprocessing

This part focuses on the processing of multisource geo-data as input to subsequent encoder. The first step is rasterization, where we select the study area, get its minimum bounding rectangle, and divide it into 500m grids as geographic analysis units. Then is the feature summary, which integrates the location (normalized coordinates), population (living and working), and land use data as input attributes on the grid-scale. Location and population are to be consistent with the input of gravity model and radiation model, and the land use data is to enrich the geographical semantic information, which characterizes the close relationship between commuting and urban functional attributes.

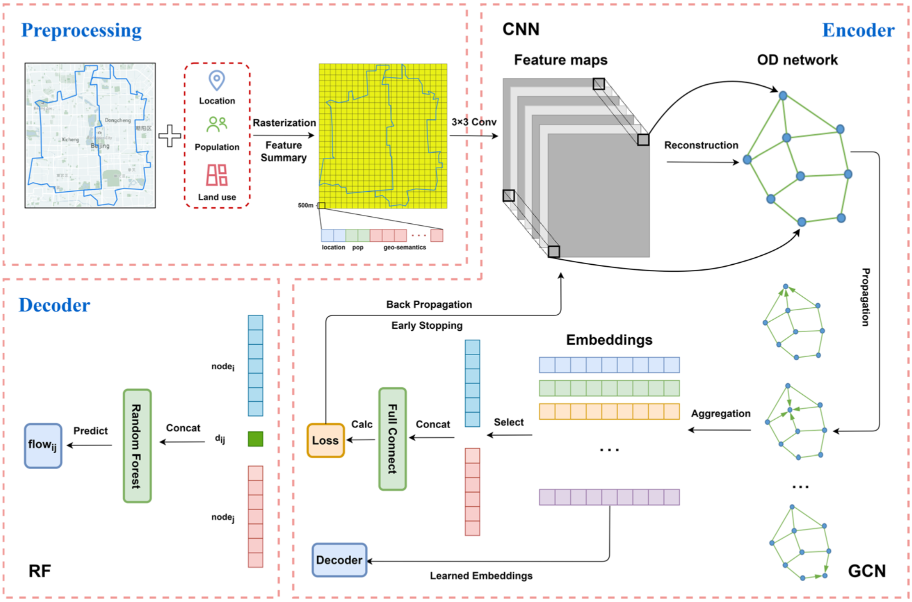



**Fig. 2** The framework of this study: (1) Preprocessing: rasterize the study area and integrate multisource geo-data, including location, population, and land use; (2) Encoder: combine CNN and GCN to extract the embedding vectors of the nodes, the two modules are organically connected and learned at the same time; (3) Decoder: the extracted embeddings of the nodes are fed into a random forest regressor to predict the commuting flow

## 4.2 Encoder

The encoder is an embedding extractor that consists of two modules, CNN and GCN, to model the geographical and topological proximity effects mentioned in Section 2.3, respectively. It should be noted that both modules belong to the same pipeline and the weights of them are updated at the same time during the loss backward process.

First, we design a single-layer CNN to blend features. Convolution is an important operation in functional analysis, which is mainly used in the signal analysis [38]. After [39] propose CNN, it has been widely applied in the field of image processing. CNNs draw on the idea of the receptive field of visual neurons [40] to output multi-dimensional feature maps by sliding different convolutional kernels on the input image and performing an element-wise multiplication and summation (similar to Sliding Average), thus realizing the extraction of local features. The formula for the convolutional layer is as follows.

$$O^{(l)}(u,v) = \sigma\left(\sum_{i=-m}^{m}\sum_{j=-m}^{m} I^{(l)}(u+i,v+j)\cdot K^{(l)}(i,j) + b^{(l)}\right), m = \frac{n-1}{2} \tag{1}$$

where $O^{(l)}$ denotes the output image of the $l$-th layer, which can be input into the next layer after pooling, $I^{(l)}$ is the input image of the $l$-th layer, $(u,v)$ means the coordinates of the pixel, $K^{(l)}$ is the kernel matrix and $b^{(l)}$ is the bias, $n$ is the convolutional kernel size, and $\sigma$ means the activation function. From the above formula, we can find that the value of a pixel in the output image actually comes from the fusion of the values of the $n \times n$ neighborhood pixels in the input image.

In this study, we consider the study area as an "image", the grids as "pixels", and the multi-dimensional attributes of the grid as "channels", then the receptive field can be regarded as the "buffer" in geography, so convolution allows the attributes of the grid neighborhood to blend into itself, thus modeling the geographical proximity effect to some extent. For example, for $Grid_i$ in Fig. 3, after the convolution (the grey window) from input (the red image) to output (the blue image), the features of its neighbors and itself are both taken into account. After the convolution is completed, multiple feature maps are generated, with pixels (i.e., grids, represented as multi-dimensional hidden vectors) on the feature maps as nodes, and the commuter flows between grids as edges to reconstruct the commuter network.

After building the commuter network, we use a two-layer GCN as a backbone to extract the structural properties of the network, to model the topological proximity effect. Given a binary adjacency matrix $A$ and a feature matrix $X$, GCN can learn the embedding vectors of nodes through feature propagation and aggregation of local network neighbors (including themselves). These two steps can be described by the following formula [34]:

$$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{2}$$

where $H^{(l)}$ denotes the hidden representation in the $l$-th layer, meaning the features of the $l$-hop neighbors can be leveraged, and $H^{(0)} = X$, representing the original input. $\hat{A} = I + A$, where $I$ is the identity matrix for self-connection. Because $A$ represents the connection between each node and its neighbors, when $A$ is added to $I$, the features of both itself and its neighbors are taken into account. $\hat{D}$ is the degree matrix, $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$, which is used to normalize the matrix $\hat{A}$. $W^{(l)}$ is the weight matrix to be learned of the $l$-th layer. $\sigma$ is the activation function, such as RELU: $\sigma(x) = max(0,x)$. GCN can map nodes to latent

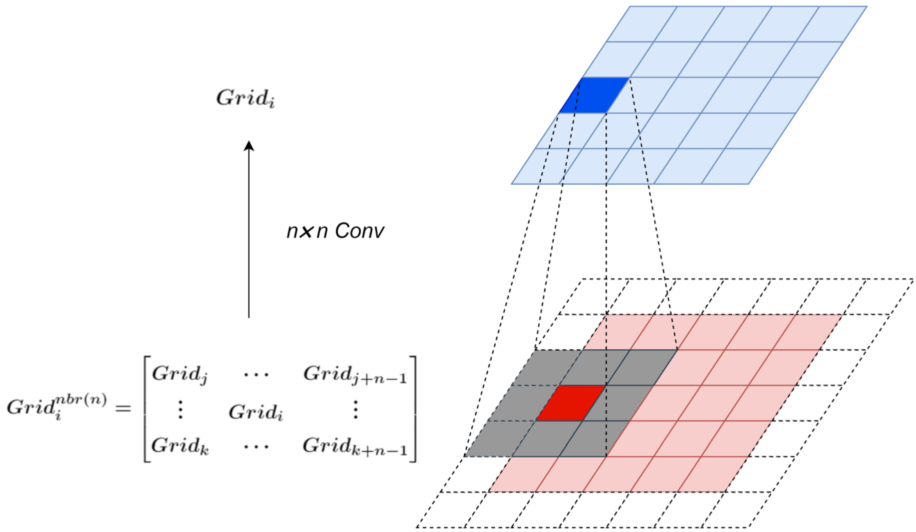



**Fig. 3** The fusion of neighborhood features is achieved through convolution: the red image is the input, the blue image is the output, and the grey window is the convolutional kernel. After convolution, the value of $Grid_i$ includes the features of its $n \times n$ neighbors and itself

representation vectors, and similar nodes are also close to each other in the representation space.

After learning the embedding vectors of nodes, we concatenate the vectors of the origin/destination and design a multi-layer perceptron to map it to the flow intensity between them. Then, based on the predicted value $\hat{f}_{ij}$ and the true value $f_{ij}$ of flow intensity, the loss is calculated and back-propagated, and the gradients are updated to train the parameters of the model. The loss function we select is the Mean Square Error (MSE), as shown in Equation 3. In training, we use the Early Stopping strategy to avoid overfitting, which is to retain the best-performing models and learned embeddings on the validation set every five epochs. Finally, the best embeddings are input into the decoder to predict the flow intensity.

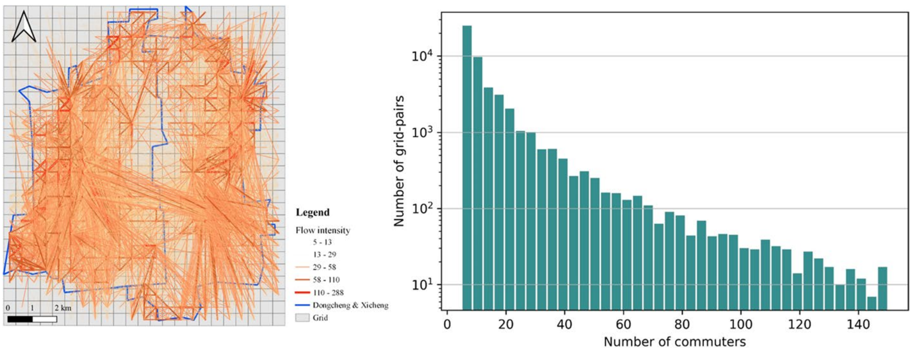
$$MSE = \frac{1}{n} \sum_{i,j} \left(\hat{f}_{ij} - f_{ij}\right)^2 \tag{3}$$

## 4.3 Decoder

Based on the learned embeddings of nodes in the encoder, we choose any two nodes $i$ and $j$ as OD (i.e., home and workplace), concatenate their embeddings with their distance, as input to a random forest regressor, and finally predict the commuter flow between $i$ and $j$ effectively. The decoder is formularized in (4):

$$\hat{f}_{ij} = RF\left(e_i \| d_{ij} \| e_j\right) \tag{4}$$

where $e_i$ and $e_j$ denote the embeddings of nodes $i$ and $j$, $d_{ij}$ is the distance between them, and ‖ represents the concatenation operation.




(a) Study Area

(b) The distribution of commuting flows

**Fig. 4** Study area and the distribution of commuting flows

**Table 1** The field description of commuter flow data

| Field | Type | Description |
|---|---|---|
| Grid id | String | Unique grid id (home/work). |
| Grid coordinates | Float | The coordinate of the grid's center (home/work). |
| Commuter flow | Int | The number of commuters between two grids. |

**Table 2** The category information of land use data

| Category | Code | Description |
|---|---|---|
| Residential | 01 | Buildings where people live, like houses and apartments. |
| Commercial | 02 | Buildings where people work, eat, entertain, etc. |
| Industrial | 03 | Land and buildings for manufacturing, storage, mining, etc. |
| Transportation | 04 | Roads, transportation stations, and airport facilities. |
| Public management and service | 05 | Land for administration, education, medical, sports, greening, etc. |

## 5 Case study

### 5.1 Datasets

In this study, we select Dongcheng District and Xicheng District of Beijing as the study area, as shown in Fig. 4(a). We mainly use two datasets: one is commuter flow data, and the other is land use data; the details of them are further described below.

The commuter flow data, that runs from January to June 2019, is provided by Amap (https://www.amap.com/), China's largest mobile navigation map company. The home and workplace locations that make up the commuter flows in this dataset are inferred using a machine learning method based on millions of users' trajectory data collected from many smartphone apps including Alipay, Douyin, Weibo, etc. The accuracy of these locations is more than 90% when compared with ground truth locations of registered users, so this data product is highly reliable. Each row of the data represents a pair of commuter flow, including fields such as the OD grids' ID, coordinates, and the corresponding number of commuters, as shown in Table 1. According to statistics, the resident and working population in the study area are 1,903,441 and 3,269,591, respectively. To avoid the effects of random noise, we remove grid pairs with flow intensity less than 5 (Fig. 4(b)) and eventually form a spatial interaction network with 598 nodes and 49,766 edges, which is sparse obviously.

The land use vector data for China 2018 is provided by [41], and open access (http://data.ess.tsinghua.edu.cn/). This dataset is generated based on multi-source geo-big data (such as remote sensing images, POIs, etc.) and a random forest classifier. It provides two levels of results, and we select Level-I, including five categories: residential, commercial, industrial, transportation, and public management and service, as shown in Table 2. We make an intersect operation between the land parcels and the divided grids using QGIS, count the area of each category within each grid, and divide it by the grid area for normalization to form a 5-dimensional vector. This vector, as a part of the grid feature, is used to verify the geographical semantic relevance of commuting mentioned in Section 1.

Here we provide some details of the experiment for reproduction. In the study, we randomly divide the commuter flow data into training set (60%), validation set (20%), and test set (20%) [31, 42]. It is worth mentioning that the encoder is trained and validated on the first two sets, and the decoder is tested on the latter. The three main hyperparameters, namely, the dimension of embedding vectors, the convolution kernel size, and the number of GCN layers, are set to 16, 3, 2. In the encoder, we use Adam as the optimizer and set the learning rate to 0.001, and train the model for 200 epochs. We also adopt the early stopping strategy to preserve the model with the least loss on the validation set every five epochs and the corresponding embeddings. In the decoder, we use 10-fold cross-validation to tune the hyperparameters of the random forest regressor. All parts of our proposed model are implemented based on Sklearn and Pytorch.

## 5.2 Baselines

To prove the validity of the hybrid learning model we propose, we choose the following baselines for comparison: (1) spatial interaction models like the gravity and radiation models; (2) machine learning models like random forest and XGBoost; (3) network-based models with Node2vec or GCN as the encoder and random forest as the decoder, without the convolution part.

- **Gravity model** [17]. It assumes that the commuter flow $f_{ij}$ between two regions $i$ and $j$ is proportional to their populations $p_i$ and $p_j$, and inversely proportional to the power of the distance $d_{ij}$. The formula is as follows, after a log-transformation, the parameters $K$ and $\beta$ can be estimated by linear regression.

$$f_{ij} = K\frac{p_i p_j}{d_{ij}^{\beta}} \tag{5}$$

- **Radiation model** [19]. Inspired by the intervening opportunities law and diffusion process [15, 26], the radiation model states that in addition to population size, flow intensity is influenced by the number of intervening opportunities. The formula is as follows, which is parameter-free and can be directly used to calculate the flow.

$$f_{ij} = T_i\frac{p_i p_j}{\left(p_i + s_{ij}\right)\left(p_i + p_j + s_{ij}\right)} \tag{6}$$

  where $T_i$ represents the total flow originating from $i$, $s_{ij}$ represents the number of opportunities (approximated by the population here) within a circle with $i$ as the center and $d_{ij}$ as the radius (excluding the origin and destination).
- **Random forest** [43]. It is an ensemble learning method based on decision trees, where multiple trees are built independently and combined later to avoid overfitting and to improve model performance. In this study, the features of the OD are concatenated directly as input to predict the commuting flow.
- **XGBoost** [44]. As a gradient boosting tree model, XGBoost takes into account the computational efficiency while improving the model effectiveness. In many classification and regression tasks, XGBoost can be regarded as the state-of-the-art model. In this study, XGBoost has the same input as random forest.
- **Node2vec-RF** [45]. Node2vec is a graph embedding method that learns the feature representations of nodes in an unsupervised manner on graph data. Inspired by Word2vec

[46], Node2vec does random walk on a network, regards the nodes as words, the node sequence after walking as a sentence, to obtain the embeddings of the nodes [47], and the walking strategies include breadth-first sampling and depth-first sampling. In this study, we use Node2vec as an encoder to learn the embeddings of the nodes and then input them into a random forest to make predictions.

- **GCN-RF** [34]. As mentioned earlier, GCN introduces the idea of convolution into the graph structure data, which can fully extract the topology characteristics of the network. Compared with the complete model we propose in Section 4, only the convolution part is removed to analyze the impact of neighborhood feature fusion on the model for comparison.

## 5.3 Evaluation metrics

We adopt three metrics to evaluate the performance of the models: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Common Part of Commuters (CPC). RMSE and MAPE are common metrics of regression tasks, representing a deviation from the true value, the smaller the better. CPC is a measure of the similarity between the predicted value and the true value in commuter flow prediction, which is between 0 and 1, and the greater the better [14]. The formulas of these metrics are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i,j}\left(\hat{f}_{ij} - f_{ij}\right)^2} \tag{7}$$

$$MAPE = \frac{1}{n}\sum_{i,j}\left|\frac{\hat{f}_{ij} - f_{ij}}{f_{ij}}\right| \tag{8}$$

$$CPC = \frac{2\sum_{ij}\min\left(\hat{f}_{ij}, f_{ij}\right)}{\sum_{ij}\hat{f}_{ij} + \sum_{ij}f_{ij}} \tag{9}$$

where $n$ represents the total number of commuter flows, $f_{ij}$ and $\hat{f}_{ij}$ represent the true and predicted values of the flow intensity from grid $i$ to grid $j$, respectively.

## 5.4 Results

The performance of the above models on the test set is shown in Table 3. Obviously, the hybrid learning model we propose (ConvGCN-RF) outperforms XGBoost

**Table 3** The performance comparison of the models on the test set, including RMSE, MAPE, and CPC

| Model | RMSE | MAPE | CPC |
|---|---|---|---|
| Gravity model | 16.064 | 0.751 | 0.707 |
| Radiation model | 19.048 | 0.933 | 0.337 |
| Random forest | 13.541 | 0.617 | 0.758 |
| XGBoost | 12.443 | 0.604 | 0.756 |
| Node2vec-RF (ours) | 11.087 | 0.550 | 0.771 |
| GCN-RF (ours) | 9.871 | 0.452 | 0.810 |
| ConvGCN-RF (ours) | 9.553 | 0.439 | 0.817 |

(state-of-the-art), by 23.2%, 27.3%, and 8.1% in terms of RMSE, MAPE, and CPC, respectively, thus demonstrating the effectiveness of our model. We further analyze and explain the possible causes of each model's performance, as shown below.

The traditional spatial interaction models (i.e., gravity model, radiation model) have achieved poor performance on all metrics, and the radiation model is the worst. In our view, such models have fixed inputs and simple formulas, which may limit their ability to model complex patterns, especially for human mobility.

Machine learning models, represented by random forest and XGBoost, are significantly better than spatial interaction models on all indicators, in line with existing studies [21, 23, 29]. Machine learning models do not require the introduction of prior knowledge and can mine nonlinear and irregular patterns directly from the data, which makes them outstanding in this task.

Node2vec-RF is based on our proposed framework: the encoder is Node2vec while the decoder is random forest. The comparison of this model with XGBoost also proves that our framework is effective. In addition, Node2vec is a shallow model, so it can also achieve high efficiency in terms of running time as an encoder of our framework. Node2vec, based on a random walk, can take advantage of the topological properties of the commuter network, which is lacking in the machine learning models above, but does not take into account the features of itself and its neighbors, which is something that needs to be improved.

GCN-RF is also a hybrid learning model that we designed, with the encoder being converted to GCN, a more efficient model in graph data mining [34]. GCN can learn the feature representations of nodes in a supervised or semi-supervised manner through propagation and aggregation of the attributes of neighbors, including themselves. Supervised learning enables GCN to optimize model parameters by back-propagation theory, and the integration of neighbor characteristics enables GCN to fully capture the features of itself and neighbors, thus better modeling the topological proximity effect mentioned in Section 2.3. These advantages make GCN work better as an encoder than Node2vec, as shown in Table 3.

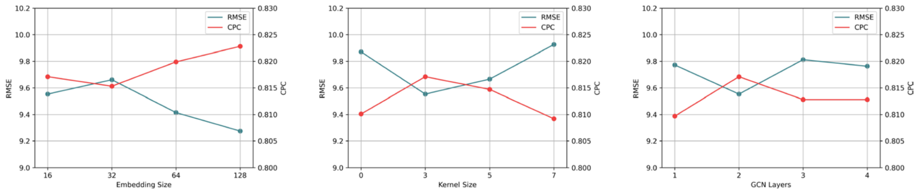
ConvGCN-RF is based entirely on the hybrid learning framework, combining CNN and GCN as the encoder and random forest as the decoder, which performs better than all baselines and proves the effectiveness of our model. Convolution merges the features of the neighborhood grids, and to a certain extent considers the spatial spillover effect, which improves the model performance. The combination of CNN modeling geographical proximity effect and GCN modeling topological proximity effect makes the model perform best.

## 6 Discussions

In this section, we focus on the main factors that may affect model performance, including the model's hyperparameters, the flow intensity and distance, and land use data as geographic semantics.

### 6.1 The effect of hyperparameters

1) *Embedding size*. This parameter determines the dimension of the embedding vectors extracted by our encoder. As you can see from Fig. 5(a), the higher the dimension, the

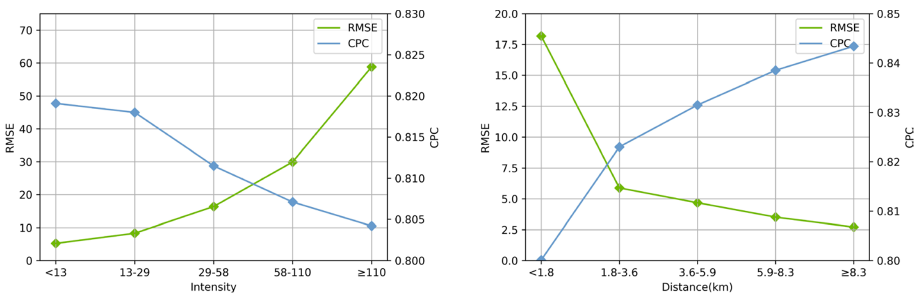



(a) Effect of embedding size (b) Effect of kernel size (c) Effect of GCN layers

**Fig. 5** The effect of hyperparameters on the model

lower the RMSE and the higher the CPC in general, which means the model has better performance. This is because the higher the hidden vector dimension, the more parameters the model needs, and models with larger "capacity" in deep learning tend to have better fitting capabilities [48]. The 16-dimensional embedding is ideal for maintaining the model performance while taking into account the model size.

2) *Kernel size*. As mentioned in Section 4.2, the receptive field of the convolution kernel can be compared to the concept of "buffer" in spatial analysis, so the convolution kernel size determines the scope of neighborhood impacts. As shown in Fig. 5(b), when the kernel size is 0, it indicates that the convolution is not added, which is exactly GCN-RF in baselines, and its performance is not very exciting. When the kernel size is 3, the model performance is the best. Because the research unit is a 500m grid, the 750m buffer is the best distance for neighborhood effect in this task. When the kernel size becomes larger, the considered neighborhood also becomes wider, and the features after fusion tend to be averaged, so the model becomes worse, and even worse than when the kernel size is 0.
3) *GCN layers*. The number of GCN layers determines the depth of the network neighborhood that each node leverages. As shown in Fig. 5(c), the model works best when the number of GCN layers is 2, i.e., when all neighbor nodes within 2-hop have an impact on the central node. When the number of GCN layers is too small, the model fails to fully capture the properties of the network neighborhood, and when it is too large, it takes too many nodes into account, dilutes the difference between nodes, and causes the problem of "over-smoothing" [49]. These factors all make the model perform worse.




(a) Effect of flow intensity (b) Effect of flow distance

**Fig. 6** The effect of flow on the model

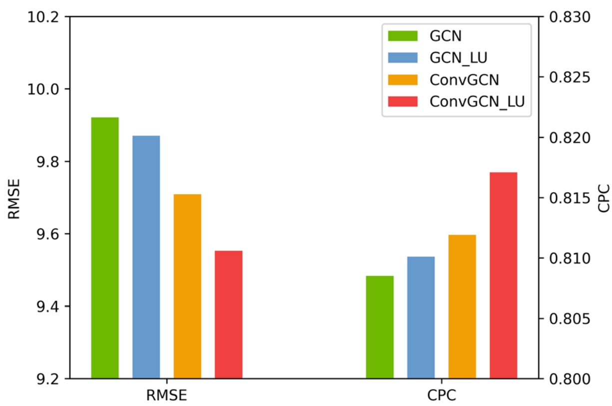
## 6.2 The effect of flow

1) *Flow intensity*. We divide the test set into five groups by flow intensity based on the Jenks natural breaks [50] to compare the model's performances on the flows of different intensities, as shown in Fig. 6(a). We can see that in general, with the increase of flow intensity, RMSE gradually increases and CPC decreases, which indicates that the model has a larger error in the prediction of stronger flows, and the prediction of smaller flows is more accurate.
2) *Flow distance*. Using the same method to group data by the flow distance (keeping a decimal place), we get the result of Fig. 6(b). It can be found that as the flow distance increases, the RMSE decreases and the CPC increases, which indicates that the model predicts long-distance flows better. The reason may be that long-distance flows are usually less intense due to the distance decay effect, and the model's prediction of small-intensity flows is more accurate (as shown in Fig. 6(a)).

## 6.3 The effect of land use data

We further discuss the impact of land use data on the model, as shown in Fig. 7, where GCN and ConvGCN are models without land use types as inputs. The first thing we can see is that the addition of land use data has improved the effectiveness of both GCN and ConvGCN, as a result of the introduction of more supplementary information. Secondly, land use types have different improvements in the two models. For ConvGCN, the RMSE and CPC have a greater improvement than GCN. The reason may be that when there is abundant semantic information in geographical units, the influence on each other (especially neighborhoods) is more prominent, and convolution captures this neighborhood effect well, thus improving the performance of the model better. In other words, the rich geographical semantic information enhances the improvement on the model caused by the fusion of neighborhood features through convolution.

**Fig. 7** The effect of landuse data on the models



## 7 Conclusions

Accurate prediction of commuter flow between any two places is important for transportation optimization and infrastructure configuration. However, there are inherent shortcomings in the two existing mainstream methods, such as spatial interaction models and machine learning models. In this paper, we propose a *preprocessing-encoder-decoder* hybrid learning model to solve the above problem: 1) in the preprocessing part, we integrate multisource geo-semantic data, including location, population, and land use types; 2) in the encoder, we combine CNN and GCN to extract the representation vectors of the nodes in the commuter network, thus effectively modeling geographical and topological proximity effects, respectively; 3) in the decoder, the learned embeddings of the nodes are fed into a random forest regressor to achieve accurate prediction of commuter flow, which has the best performance compared to all baselines.

We further discuss the main factors affecting model performance. We find that: 1) in terms of model hyperparameters, the higher the embedding dimension, the better the model performance in general, while the convolution kernel size and the number of GCN layers are different, too large and too small will make the prediction effect worse; 2) the model is more accurate in predicting flows with less intensity and greater distance; 3) finally, land use data, as a supplement to geographical semantics, makes the fusion of neighborhood features improve the model effect to a greater extent.

There are still some shortcomings in our research that need to be improved. For example, CNN can only handle regular matrix-like inputs, and when the research unit is a traffic analysis zone, census tract, etc., designing a deep learning-enhanced spatial lag model or introducing the attention mechanism may be a solution. Besides, convolution can only take advantage of local attributes, without considering global information, the attention mechanism can consider the similarity of geographical units both at the global and local scale, to better model spatial relevance. In addition to land use data, we can add more features, such as POIs, social media check-in texts, residents' income, and other information to enrich geographic semantics.

**Acknowledgements** We thank the editors and reviewers for their suggestions to improve the quality of this paper. This research was supported by grants from the National Natural Science Foundation of China (41830645, 41971331).

**Data availability** The commuting flow data is provided by Amap company, and they have not given their permission for researchers to share their data. Data requests can be made to Amap company (https://lbs.amap.com/). The land use data is publicly available online at http://data.ess.tsinghua.edu.cn/.

## Declarations

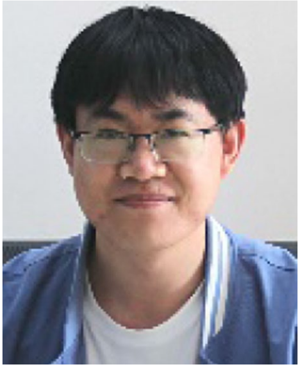
**Conflict of interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Liu Y, Liu X, Gao S, Gong L, Kang C, Zhi Y, Chi G, Shi L (2015) Social sensing: A new approach to understanding our socioeconomic environments. Ann Assoc Am Geogr 105(3):512–530
2. Walmsley DJ, Lewis GJ (2014) People and environment: Behavioural approaches in human geography. Routledge

3. Hu Y (2017) Geospatial semantics. arXiv:1707.03550
4. Huang W, Li S (2016) Understanding human activity patterns based on space-time-semantics. ISPRS J Photogramm Remote Sens 121:1–10
5. Liu X, Kang C, Gong L, Liu Y (2016) Incorporating spatial interaction patterns in classifying and understanding urban land use. Int J Geogr Inf Sci 30(2):334–350
6. Simini F, Barlacchi G, Luca M, Pappalardo L (2021) A deep gravity model for mobility flows generation. Nat Commun 12(1):1–13
7. Tardy C, Falquet G, Moccozet L (2016) Semantic enrichment of places with vgi sources: a knowledge based approach. In: Proceedings of the 10th workshop on geographic information retrieval, pp 1–2
8. Van Acker V, Witlox F (2011) Commuting trips within tours: how is commuting related to land use? Transportation 38(3):465–486
9. Wu L, Yang L, Huang Z, Wang Y, Chai Y, Peng X, Liu Y (2019) Inferring demographics from human trajectories and geographical context. Comput Environ Urban Syst 77:101368
10. Li Y, Yu R, Shahabi C, Liu Y (2018b) Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: International conference on learning representations
11. Yu B, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: Proceedings of the 27th international joint conference on artificial intelligence, pp 3634–3640
12. Zhang Y, Cheng T, Ren Y, Xie K (2020) A novel residual graph convolution deep learning model for short-term network-based traffic forecasting. Int J Geogr Inf Sci 34(5):969–995
13. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2019) T-gcn: A temporal graph convolutional network for traffic prediction. IEEE Trans Intell Transp Syst 21(9):3848–3858
14. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: Models and applications. Phys Rep 734:1–74
15. Lenormand M, Bassolas A, Ramasco JJ (2016) Systematic comparison of trip distribution laws and models. J Transp Geogr 51:158–169
16. Masucci AP, Serras J, Johansson A, Batty M (2013) Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Phys Rev E 88(2):022812
17. Zipf GK (1946) The p 1 p 2/d hypothesis: on the intercity movement of persons. Am Sociol Rev 11(6):677–686
18. Ren Y, Ercsey-Ravasz M, Wang P, González MC, Toroczkai Z (2014) Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. Nat Commun 5(1):1–9
19. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. Nature 484(7392):96–100
20. Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. Science 349(6245):255–260
21. Morton A, Piburn J, Nagle N (2018) Need a boost? a comparison of traditional commuting models with the xgboost model for predicting commuting flows (short paper). In: 10th International Conference on Geographic Information Science (GIScience 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik
22. Mozolin M, Thill J.-C., Usery EL (2000) Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. Transp Res B Methodol 34(1):53–73
23. Pourebrahim N, Sultana S, Niakanlahiji A, Thill J-C (2019) Trip distribution modeling with twitter data. Comput Environ Urban Syst 77:101354
24. Tobler WR (1970) A computer movie simulating urban growth in the detroit region. Econ Geogr 46(sup1):234–240
25. Xing X, Huang Z, Cheng X, Zhu D, Kang C, Zhang F, Liu Y (2020) Mapping human activity volumes through remote sensing imagery. IEEE J Sel Top Appl Earth Obs Remote Sens 13:5652–5668
26. Stouffer SA (1940) Intervening opportunities: a theory relating mobility and distance. Am Sociol Rev 5(6):845–867
27. McArthur DP, Kleppe G, Thorsen I, Ubøe J (2011) The spatial transferability of parameters in a gravity model of commuting flows. J Transp Geogr 19 (4):596–605
28. Black WR (1995) Spatial interaction modeling using artificial neural networks. J Transp Geogr 3(3):159–166
29. Spadon G, de Carvalho AC, Rodrigues-Jr JF, Alves LG (2019) Reconstructing commuters network using machine learning and urban indicators. Sci Rep 9(1):1–13
30. Liu Z, Miranda F, Xiong W, Yang J, Wang Q, Silva C (2020) Learning geo-contextual embeddings for commuting flow prediction. In: Proceedings of the AAAI conference on artificial intelligence, vol 34, pp 808–816

31. Yao X, Gao Y, Zhu D, Manley E, Wang J, Liu Y (2020) Spatial origin-destination flow imputation using graph convolutional networks. IEEE Trans Intell Transp Syst
32. Ullman EL, Boyce RR (1980) Geography as spatial interaction. University of Washington Press
33. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 30th international conference on neural information processing systems, pp 3844–3852
34. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR)
35. Wang Y, Zhu D, Yin G, Huang Z, Liu Y (2020) A unified spatial multigraph analysis for public transport performance. Sci Rep 10:9573–9581
36. Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY (2020) A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst 32(1):4–24
37. Miller HJ (2004) Tobler's first law and spatial analysis. Ann Assoc Am Geogr 94(2):284–289
38. Hirschman II, Widder DV (2012) The convolution transform. Courier Corporation
39. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
40. Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160(1):106–154
41. Gong P, Chen B, Li X, Liu H, Wang J, Bai Y, Chen J, Chen X, Fang L, Feng S, et al. (2020) Mapping essential urban land use categories in China (euluc-china): Preliminary results for 2018. Sci Bull 65(3):182–187
42. Zhu D, Zhang F, Wang S, Wang Y, Cheng X, Huang Z, Liu Y (2020) Understanding place characteristics in geographic contexts through graph convolutional neural networks. Ann Am Assoc Geogr 110(2):408–420
43. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
44. Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp 785–794
45. Grover A, Leskovec J (2016) node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pp 855–864
46. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
47. Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 701–710
48. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press
49. Li Q, Han Z, Wu X.-M. (2018a) Deeper insights into graph convolutional networks for semi-supervised learning. In: Proceedings of the AAAI conference on artificial intelligence, vol 32
50. Jenks GF (1967) The data model concept in statistical mapping. Int Yearb Cartogr 7:186–190



**Ganmin Yin** received the B.S. degree from Peking University in 2020. He is currently pursuing the Ph.D. degree in GIScience with the Institute of Remote Sensing and Geographical Information Systems, Peking University. His research interests include human mobility, transportation, urban data mining, social sensing and GeoAI.

**Zhou Huang** is Associate Professor of GIScience at the Institute of Remote Sensing and Geographical Information Systems, Peking University. He received the BSc degree in GIS and the Ph.D. degree in Cartography and GIS from Peking University, China, in 2004 and 2009, respectively. Currently, his main research interests include big geo-data, high-performance geocomputation, distributed geographic information processing, spatial data mining, and spatial database. He has published more than 60 academic papers in international journals or conferences. In addition, Zhou Huang serves as Deputy Director of Institute of Remote Sensing and GIS, Peking University, and Deputy Director of Engineering Research Center of Earth Observation and Navigation, Ministry of Education, China. He was also selected for the Youth Talent Innovation Plan in Remote Sensing Science and Technology in 2015, funded by the Ministry of Science and Technology of China.

**Yi Bao** received the B.S. degree from China University of Geosciences, Wuhan in 2018. He is currently a PhD candidate in GIScience with the Institute of Remote Sensing and Geographical Information Systems, Peking University. His main research interests include spatial data mining and location based services.

**Han Wang** received the B.S. degree from Wuhan University in 2020. He is currently pursuing a Master's degree in the Institute of Remote Sensing and Geographic Information Systems, Peking University. His current research interests include trajectory data mining, representation learning and generative models.

**Linna Li** is an associate professor in Geography at California State University, Long Beach. She got her B.S. degree in Geographic Information Science from Peking University in China in 2004, her M.S. degree in Geography from the University of South Carolina in 2006, and her Ph.D. in Geography from the University of California, Santa Barbara in 2010. Her teaching and research interests include geographic information science theories and applications in public health, transportation, and urban development. Her current research focuses on big geo-data analytics and spatio-temporal data mining using social media.

**Xiaolei Ma** is a Professor in the school of transportation science and engineering at Beihang University, China. He received his Ph.D. degree from University of Washington, Seattle in 2013. His research areas mainly lie in public transit operation and planning, shared mobility modeling and transportation Big Data analytics. To date, he has published over 100 articles in peer-reviewed journal articles and conference papers. Dr. Ma servers as an associate editor of IEEE Transactions on Intelligent Transportation Systems and IET Intelligent Transport Systems while also being an editorial member of Transportation Research Part C. He was also a member of the TRB Artificial Intelligence and Advanced Computing Applications Committee. Dr. Ma received several academic awards including National Ten Thousand Talent Program, Young Elite Scientist Sponsorship Program by the China Association for Science and Technology.

**Yi Zhang** obtained his bachelor's degree from school of computer science, Beijing University of technology in 1994, master's degree and doctor's degree in geographic information science, Peking University in 2001 and 2011 respectively. His research interests include qualitative geographic knowledge representation, geographic information retrieval, text spatiotemporal semantic processing, and tourism big data.