International Journal of Geographical Information Science

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

The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types

Jinbao Zhang, Xia Li, Yao Yao, Ye Hong, Jialyu He, Zhangwei Jiang & Jianchao Sun

To cite this article: Jinbao Zhang, Xia Li, Yao Yao, Ye Hong, Jialyu He, Zhangwei Jiang & Jianchao Sun (2020): The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types, International Journal of Geographical Information Science

To link to this article: https://doi.org/10.1080/13658816.2020.1726923

Published online: 11 Feb 2020.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2020.1726923

RESEARCH ARTICLE

The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types

Jinbao Zhang [a], Xia Li [b], Yao Yao [c], Ye Hong [d], Jialyu He [a], Zhangwei Jiang [e] and Jianchao Sun [e]

[a]Guangdong Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou, China; [b]KeyLabof Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai, P.R. China; [c]School of Geography and Information Engineering, China University of Geosciences, Wuhan, China; [d]Department of Civil, Environmental and Geomatic Engineering, ETH Zürich, Zürich, Switzerland; [e]Department of Data Technology and Products, Alibaba Group, Hangzhou, China

ABSTRACT

The formulation of mixed urban land uses is not only intended to find the ideal scenario of land use but also regarded as a way toward sustainable urban development. We propose a geo-semantic mining approach Traj2Vec to quantify the trajectories of residents as high-dimensional semantic vectors. Then, a random forest (RF) method is used to model the relationship between the semantic vectors and mixed urban land uses. The proposed Traj2Vec approach can obtain the highest accuracy (OA = 0.7733, kappa = 0.7245) in urban land-use classification and a high average proportion accuracy (64.0%) in capturing the proportions of urban land-use types. Diversity analysis indicates that Shenzhen has a high degree of mixed urban land use at the scale of a street block. By analyzing the mixing index and the travel distance, we find a weak but significant negative correlation between them ($r = -0.107,\ p < 0.001$), which not only confirms the conclusion that an increase in the degree of mixing will reduce the travel distances of residents but also verifies the mixing index. This suggests that urban planning should focus on mixed urban land uses, which can reduce the travel distances of residents, reduce energy consumption, and make cities more compact.

ARTICLE HISTORY
Received 18 June 2019
Accepted 4 February 2020

KEYWORDS
Mixed urban land-use patterns; mobile phone positioning data; residents' spatial trajectories; traj2vec; geo-semantic mining

1. Introduction

The spatial distribution of urban land use is an important indicator for urban planning and urban studies (Regan *et al*. 2015, Chen *et al*. 2017, Tu *et al*. 2017, Yao *et al*. 2017). In particular, it is argued that moderately mixed urban land-uses should be promoted so that cities can become more compact (Burton *et al*. 2003). Studies have shown that mixing residential and commercial land uses with recreational and public service facilities can provide more work within a shorter commuting distance, potentially reducing the demands for transportation and energy and strengthening connections among individuals to help form a strong community culture (Burton *et al*. 2003, Holden and Norland 2005). With rapid urbanization in recent years, urban land uses

CONTACT Xia Li lixia@geo.ecnu.edu.cn; Yao Yao yaoy@cug.edu.cn

and urban spatial structures have become increasingly diverse and sophisticated (Gao *et al*. 2017). Thus, obtaining qualitative and quantitative data on mixed urban land uses quickly and accurately is very important for understanding and managing cities (Wu *et al*. 2018). There is a general lack of such studies, because mixed urban land use is difficult to estimate using conventional methods. Traditionally, spectral unmixing methods have been applied to remote sensing images for mixed land-use classification or mixed scene recognition (Zhang and Du 2015, Li *et al*. 2017). However, remote sensing images can only reflect the natural attributes of the ground surface (Yao *et al*. 2017). Urban land uses often have strong relationships with social characteristics, which are difficult to obtain directly from remote sensing images alone (Tu *et al*. 2017, Yao *et al*. 2017).

Human activity data can provide more detailed and accurate information for analyzing mixed urban land-uses since urban land use is defined as the use of the urban space by residents and the activities within that area (Siła-Nowicka *et al*. 2016, Tu *et al*. 2017, Wu *et al*. 2018). The rapid development of location-based services (LBSs) provides us with a large amount of human activity information that can be used to measure urban spatial structures and land uses (Gonzalez *et al*. 2008, Wang *et al*. 2014, Wu *et al*. 2018). In particular, trajectory information is generated by residents in their daily lives and can represent the resident's behavioral purposes (Pei *et al*. 2014, Tranos and Nijkamp 2015). Scholars have incorporated trajectory data into urban studies since the data contain valuable information on how people utilize urban spaces (Tu *et al*. 2017). The trajectory information may include mobile phone positioning data (Pei *et al*. 2014, Tu *et al*. 2017), taxi GPS trajectories (Yuan *et al*. 2015, Li *et al*. 2016), social media check-in data (Cranshaw *et al*. 2012, Shen and Karimi 2016), etc. However, previous studies (Pei *et al*. 2014, Lenormand *et al*. 2015, Long and Shen 2015) have only taken some simple features from human activity data, such as the frequency and volume. These methods may waste the majority of the spatial information and the inner spatial correlations in human activity data (Zhai *et al*. 2019).

To address the above research gap, geo-semantic mining techniques have been applied to spatial data to explore the spatial semantic features of geospatial data (Yuan *et al*. 2015, Chen *et al*. 2016, Liu *et al*. 2017). In natural language processing, semantic mining refers to the transformation of words, phrases, signs, and symbols into forms that computers can recognize and understand the relationships between them (Miller 1995). Semantic information is a digital high-dimensional feature vector that can fully characterize these relationships. Therefore, geo-semantic mining refers to mining potential relationships in geographical data (Yan *et al*. 2017). By exploiting the potential relationships, we can fully extract the information inside geographical data and apply it in various geographic applications (Yao *et al*. 2017). Existing research on geo-semantic mining show that the semantic model can well discover the potential semantic information of the geospatial data, and the obtained information can be used to quantify the relationship between the urban land uses and geospatial data (Zhao *et al*. 2013, Zhong *et al*. 2015). In particular, there are some differences between the semantics here and the traditional geographical semantics. The semantics here is an abstract concept that uses feature vectors to represent the potential relationship within geographic data. Traditional geographic semantics refers to describing the meaning of spatial data and the relationship between them, and making the semantics of geographic information explicit (Battle and Kolas 2012).

To further consider the spatial context in spatial data, Yao *et al*. (2017), Yan *et al*. (2017) and Zhai *et al*. (2019) introduced the Word2Vec (Mikolov *et al*. 2013) model to measure the potential contextual relationships between points of interest (POIs) and obtained satisfactory results in the classification of detailed urban land uses. Word2Vec is a deep learning language model that can represent words as high-dimensional semantic vectors that can be recognized by computers (Yao *et al*. 2017). However, POIs are spatially discrete, so some methods were also developed to construct a continuous dataset from POIs (Yao *et al*. 2017, Zhai *et al*. 2019). The method used to construct a continuous POI dataset largely affects the spatial contextual relationships, thus affecting the result of urban land-use identification. A trajectory is continuous in space, and a person's travel information can reflect the use of urban space. Therefore, it is expected that the use of the Word2Vec model to explore the potential semantic information in a trajectory can help us to better understand structures and land uses.

Moreover, most previous studies have assumed that the land uses in each urban land patch are homogeneous and can be labelled with a single category, ignoring the phenomenon of mixed urban land uses in the urban structure (Niu *et al*. 2017, Wu *et al*. 2018). Based on the Word2Vec model and residents' trajectories, this study proposes a Traj2Vec model to retrieve the potential semantic information of locations and trajectories to capture the features of mixed urban land uses. The Traj2Vec model extends the Word2Vec model into a spatial trajectory context to extract semantic information on how residents utilize urban space. Then, these semantic features are processed using a random forest (RF) algorithm to obtain the proportions of each urban land-use type. Our proposed method is a generic method for estimating the proportions of urban land-use types, and we verify it in Shenzhen. This study can help understand the characteristics of cities by measuring quantitative mixed urban land use.

2. Methodology

Figure 1 shows the workflow of the proposed Traj2Vec model that integrates mobile phone data and the Word2Vec approach to obtain the proportions of each urban land-use type within a street block. In general, the procedure contains five steps: 1) Traj2Vec: extracting residents' activity locations to build a dataset and input it into the Word2Vec model to obtain the potential semantic information for each cell tower; 2) generating the semantic vectors of street blocks: obtain the potential semantic vectors of street blocks by using a weighted summation; 3) extracting pure street block samples of urban land uses: statistics by POI types; 4) estimating the proportions of each urban land-use type: train an RF model based on the pure land-use samples, and obtain the proportions of each urban land-use type for all street blocks in the study area; and 5) results validation: verification of model results by accuracy and proportion accuracy.

2.1. Extraction of location semantics based on trajectory data from mobile phones

Urban residents often travel with certain purposes, such as a person first stays at home, then goes to the place he works and returns home at night. In addition, some studies have shown that people's travel trajectories have a certain regularity (Gonzalez *et al*. 2008).

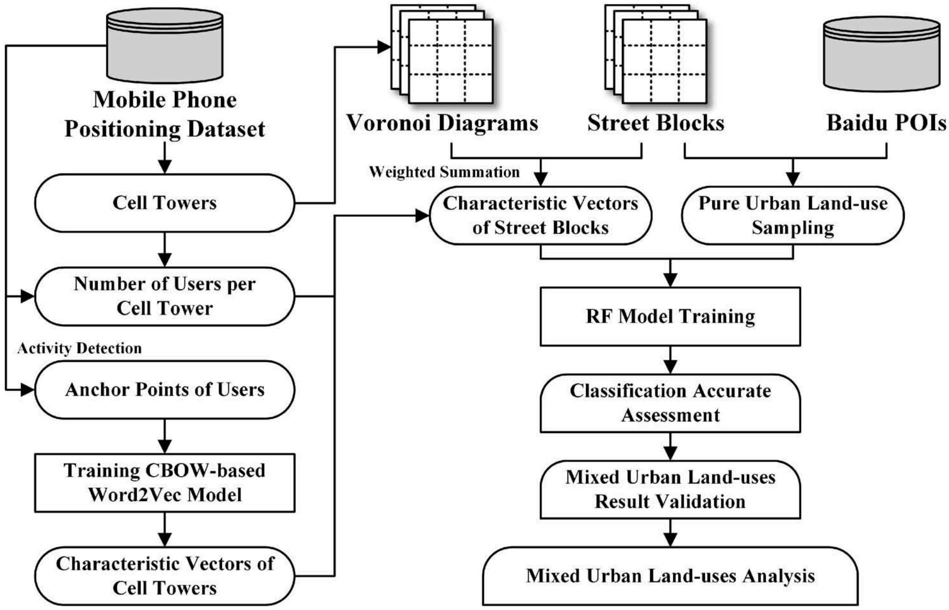

Figure 1. The workflow of the proposed Traj2Vec framework for quantifying the proportions of each urban land-use type.

Therefore, we regard a resident's activity location as a word, and the activity trajectory as a sentence; then, through a natural language model for exploring the activity context, we can obtain feature vectors that can represent the potential relationships between various places. The potential relationship mentioned here is similar to the latent relationship between words in sentences in natural language. In natural language processing, the latent relationship refers to the relationship between word usage, while the potential relationship here refers to the relationship between place uses.

We present the Traj2Vec model, which is based on the Word2Vec model and trajectory data from mobile phones. Since the Traj2Vec model is based on the Word2Vec model for retrieving trajectory information, it is necessary to build a training dataset. In this study, cell towers are regarded as words, and each mobile phone user that has multiple activity points (anchor points) is treated as a sentence; then, all the mobile phone users form a training dataset for the Word2Vec model. The mobile phone user's activity sequence represents the context of the sentence.

In particular, an activity anchor is a point that represents the cell tower where a mobile phone user stays for a certain time; such anchors may indicate the user's potential activity locations (Xu *et al*. 2016). The specific method for extracting the anchor points can be found in Tu *et al*. (2017) and Xu *et al*. (2016). In this study, the anchor points are the same as defining trip stops in trajectories, and we set each location where a user stays for more than an hour as an anchor point. Therefore, the user's activity trajectory is the sequence of anchor points. The potential semantic information corresponding to each cell tower can be obtained by importing the user behavior dataset into the Word2Vec model for training.

By using a user's activity trajectory as an input, Word2Vec maximizes the probability of occurrence of the trajectory by adjusting the semantic vector of each activity point. During the training of the Word2Vec model, the likelihood function of a sentence is shown in Equation (1):

$$\mathrm{l}(\theta) = \mathrm{logL}(\theta) = \frac{1}{T}\sum\nolimits_{i=1}^{T} logp(w_i|w_{i-c}^{i+c}) \quad (1)$$

where T is the number of activity anchor points of a mobile phone user, w_{i-c}^{i+c} is the context of the $i-th$ anchor point, and the context size is c. Finally, by adopting a stochastic gradient descent (SGD) algorithm to maximize the likelihood function $\mathrm{l}(\theta)$, the Word2Vec model optimizes the cell tower vectors during the iterative training process.

After we input the sentence (user's trajectory) into the Word2Vec model, each word (cell tower) will have a vector representing it. Then, the Word2Vec model uses these vectors to predict the probabilities of words appearing in the sentence, maximizing the probabilities of the words (maximum likelihood) and adjusting the vector of each word by backpropagation. By iterating and adjusting the vectors of each word, we can finally learn the relationships between the words in the input data set. In general, we give each tower a unique tag, and a user's trajectory (anchor point sequence) is a tag sentence (such as 345 213 568 234 987 238, which is a user's trajectory). Then, we input the trajectories of all users into Word2Vec, and we can obtain the semantic vector of each tower (such as, the vector of cell tower 1 is [0.3, 0.5, 0.2, ...], and the vector of cell tower 2 is [0.4, 0.2, 0.1, ...]). These vectors contain deep semantic information about the relationships between cell towers.

2.2. Calculating the characteristic vector of each street block

Each cell tower has a certain range of services (Pei *et al.* 2014). An example is to use a Voronoi diagram (VD) to represent a cell tower's service area (Pei *et al.* 2014). However, street blocks are used as a spatial unit that is more suitable for urban research due to the similar socioeconomic properties within such a unit (Gao *et al.* 2018). In this study, street block vectors are calculated by a weighted summation of the cell tower vectors and the corresponding Voronoi diagrams that intersect with each street block according to the first law of geography (Tobler 1970), i.e. closer regions have more similar land uses. The $i-th$ street block vector can be calculated by Equation (2):

$$Vec_i^B = \sum\nolimits_j \frac{Pop_j * Pro_j^{VD} * Pro_j^B}{\sum_j Pop_j * Pro_j^{VD} * Pro_j^B} * Vec_j^{VD} \quad (2)$$

In Equation (2), j represents the $j-th$ Voronoi diagram that intersects the $i-th$ street block, Vec_j^{VD} is the characteristic vector of the $j-th$ cell tower, and Pop_j is the number of mobile phone signals accumulated over a day at the $j-th$ cell tower. Pro_j^B can be calculated by Equation (3) as the ratio of the area of the intersection of the $i-th$ street block and $j-th$ Voronoi diagram to the area of the $i-th$ street block. Similarly, Pro_j^B can be calculated by Equation (4) as the ratio of the area of the intersection of $i-th$ street block and $j-th$ Voronoi diagram to the area of the $j-th$ Voronoi diagram.

$$Pro_j^B = \frac{Area_{B_i \cap VD_j}}{Area_{B_i}} \tag{3}$$

$$Pro_j^{VD} = \frac{Area_{B_i \cap VD_j}}{Area_{VD_j}} \tag{4}$$

In Equations (3) and (4), $Area_{B_i \cap^v D_j}$ is the area of the intersection of the $i-th$ street block and the $j-th$ Voronoi diagram, while $Area_{B_i}$ and $Area_{VD_j}$ represent the areas of the $i-th$ street block and the $j-th$ Voronoi diagram, respectively.

2.3. Extracting pure samples of urban land uses in street blocks

Pure samples of urban land use are those street blocks that are relatively pure (including only one type of land use). Extracting these pure street blocks as training samples is a very important step in obtaining mixed urban land use. Generally, the selection methods of pure samples include manual selection, clustering, and endmember extraction (Wu *et al*. 2018). These methods have different shortcomings, such as large workloads and large uncertainties, and are not suitable for our research.

Since urban residents' activities usually take place in POIs, POIs are a way of expressing how residents use a city (Gao *et al*. 2017, Zhai *et al*. 2019). In this context, many studies have used POIs to obtain urban land use (Zhong *et al*. 2014, Liu *et al*. 2015, Gao *et al*. 2017). Therefore, in this study, we use the same method as Wu *et al*. (2018) and Zhang and Du (2015) to select pure samples of urban land use. POIs are divided into urban land-use types according to the POI types, and then the proportion of POIs of each urban land-use type in each street block is calculated. Finally, we select the street blocks with the highest proportions of corresponding urban land-use types as pure samples. This method can help us obtain relatively pure samples of urban land use.

2.4. Estimating the proportions of each urban land-use type by a random forest model

In the above selection process, we obtain pure samples of urban land uses for training. This study further uses an RF algorithm to obtain the quantitative proportions of each urban land-use type. Previous studies have proven that RF algorithms are some of the most powerful machine learning models and can perform well in many situations (Biau 2012, Fern A Ndez-Delgado *et al*. 2014, Yao *et al*. 2018).

RF training involves building multiple decision trees by applying the bootstrap method to the samples that were randomly extracted from the original training samples. Each node in a decision tree is split, selecting only a portion of the variables, and an information gain strategy is adopted to select one variable as the split attribute of each node (Breiman 2001). The out-of-bag (OOB) error represents the generalization error of the decision trees; we can obtain the RF model's error by averaging the OOB errors (Wolpert and Macready 1999). After training, the RF model can be used to obtain the proportion of each urban land-use type in a street block through the voting of internal decision trees.

To measure the overall proportions of different land-use types, the area-weighted type proportions (AWTPs, Equation (5)) are used in this study (Zhang and Du 2015).

$$WeightedP_i = \frac{\sum_{j=1}^{N} Area_j \times P_{ji}}{\sum_{j=1}^{N} Area_j} \tag{5}$$

where $WeightedP_i$ represents the area-weighted proportion of the $i-th$ land-use type, N is the number of street blocks, and P_{ji} refers to the proportion of the $i-$th land-use type in the $j-$th street block.

2.5. Results validation

The proposed Traj2Vec model was validated in two aspects: comparing it with other urban land-use classification methods in pure land-use sample classification and evaluating its accuracy in obtaining the proportions of urban land-use types.

To evaluate the classification accuracy in pure land-use sample classification, the proportions obtained by the RF model were analyzed. If the urban land-use type with the highest proportion in a pure street block is the same as its corresponding label, the street block is a correct sample. Overall accuracy (OA) and the kappa coefficient are used for classification performance assessment.

To evaluate the accuracy of our proposed model in obtaining the proportions of urban land-use types, a reasonable method is to investigate the actual proportions. Because the ground truth was difficult to obtain, we manually interpreted the proportions of some street blocks as references. Like Zhang and Du (2015), the proportion accuracy (PA, Equation (6)) was used to evaluate the accuracy of the proportion results.

$$\mathrm{PA} = 1 - \left(\sum_{i=1}^{K} \left|p_i^r - p_i^d\right|\right) \tag{6}$$

where p_i^r represents the proportion of the $i-$th urban land-use type in the references and p_i^d represents that in the model results.

3. Results

3.1. Study area and data

To validate the proposed Traj2Vec model and assess its capability for quantifying the proportions of urban land-use types, a case study was conducted in Shenzhen, which is one of the four largest cities in China. Shenzhen has a total area of 1,991.64 square kilometers with ten administrative districts, as shown in Figure 2. As a fast-growing international city, Shenzhen had a total GDP exceeding 2.42 trillion yuan (equivalent to $352.75 billion) in 2018.

The street blocks used in this study are also used for land-use planning by the Bureau of Land and Resources of Guangdong Province (Yao *et al*. 2017). A street block is divided by urban roads and is the basic unit of urban structure, urban land use, and urban management (Liu and Long 2016). It is of great significance to carry out the study at the block scale. There are a total of 6,835 street blocks in the study area, with a total area of 797.003 square kilometers (Figure 2). The distribution of the street blocks is highly coincident with the built-up area.

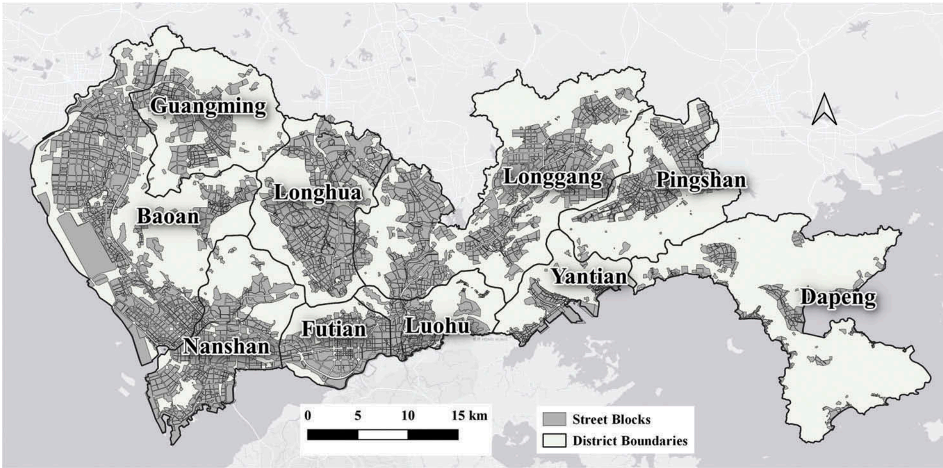

Figure 2. Study Area: Shenzhen city in Guangdong Province. Shenzhen has 6,835 street blocks at the unit of a street block, as determined from the Bureau of Land and Resources of Guangdong Province.

Mobile phone positioning data were collected for Shenzhen on 23 March 2012, recording the activity locations at half-hour intervals. The record of a user's mobile phone positioning is not the real location but the location of the cell tower that received the signal from the user's mobile phone. As shown in Figure 3, Shenzhen contains 5,818 cell towers. A total of 16.3 million mobile phone users were identified. Each user has 48 records, including the user ID, time, and longitude and latitude of the cell tower.

In addition to the mobile phone data, a POI dataset was used in this study. The POI dataset, containing 269,818 records in 16 categories (Table 1) in the study area, was collected via Baidu Map Services (http://map.baidu.com) in 2013. These POIs are used as auxiliary data to select pure street block samples of urban land-use types.

3.2. Semantic feature extraction based on traj2vec

As mentioned before, we regard a cell tower as a word and each user's travel trajectory as a sentence; then, we input the user's behavior sentence into the Word2Vec model to obtain the semantic feature vector of each cell tower. Each cell tower has a unique identifier, and each user's trajectory is a sentence that is connected by multiple identifiers. Therefore, the input to the model is sequences of these cell tower identifiers, and the output of the model is the feature vectors of each identifier, i.e. the semantic feature vector of each cell tower. After training, we obtain 100-dimensional feature vectors of each cell tower, which can represent the potential relationships with other locations.

To verify the validity of the semantic vectors we obtained, we first calculated the cosine similarity between semantic vectors of all cell towers, and then the average spatial distances between all cell towers and their several most similar cell towers were calculated. The results show that the average distances between all cell towers and their top-1, top-5, and top-10 similar cell towers are 575.521 meters, 776.008 meters, and 967.557 meters, respectively. That is, in the semantic feature vectors of the cell towers we obtained, those cell towers with similar vectors are spatially closer. For the four selected

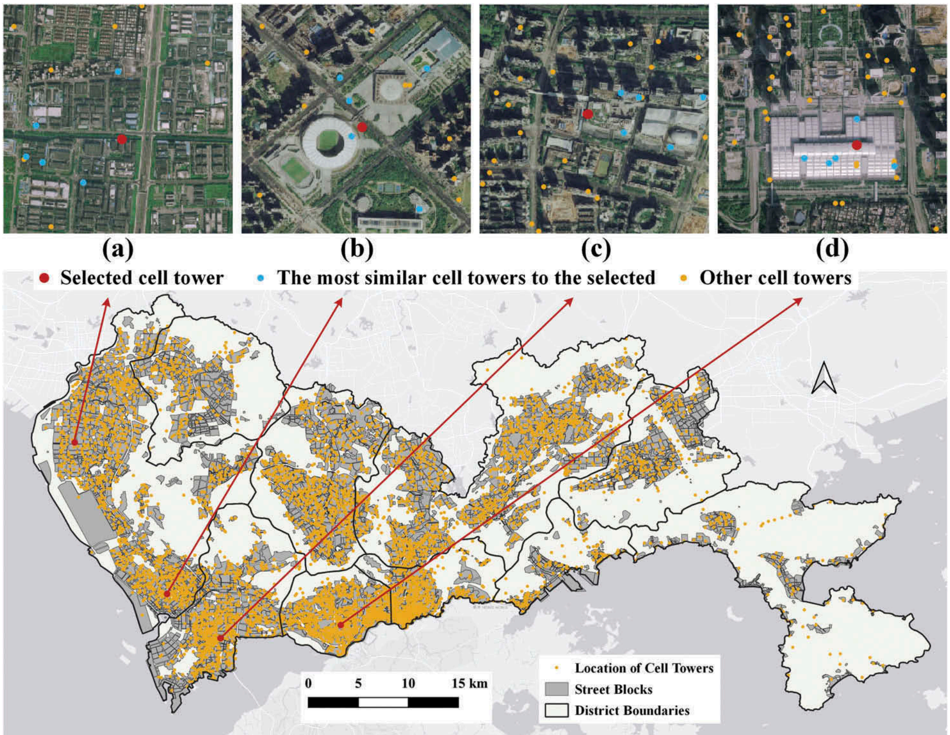

Figure 3. Selected cell towers at four locations: (a) Industrial park in the Baoan district. (b) Baoan Sports Center. (c) Coastal City Shopping Center. (d) Shenzhen Convention and Exhibition Center.

Table 1. Mapping relationships between POI types and urban land-use types.

POI Types	Count	Urban Land Use Types
Life Service	29,344	Commercial
Shopping	39,670	
Automobile Service	9,196	
Restaurant	34,673	
Hotel	5,760	
Financial Industry	7,785	
Business Building	848	
Residential Community	15,222	Housing
Clinic	11,416	Public
Traffic Facility	20,040	
Educational Institution	6,371	
Road	4,397	
Government Agency	11,353	
Recreation	7,461	Recreation
Scenic Spots	40	
Enterprise	66,242	Working

cell towers shown in Figure 3, the five most similar cell towers to them are located near them. This suggests that the semantic vectors of cell towers obtained by fully exploiting residents' trajectories through the Traj2Vec model can represent the potential relationships between cell towers. Additionally, we find that the semantic feature vector of a cell tower agrees with the first law of geography (Tobler 1970).

Since adjacent cell towers have similar semantic feature vectors, it is reasonable to calculate the feature vectors of each street block using the weighted summation presented in Section 2.2. Similarly, by calculating the similarity between the semantic feature vectors of street blocks, we found that adjacent patches have high similarity. The above results indicate that the features of each street block obtained by our method can represent the potential relationships between street blocks and facilitate the subsequent extraction of mixed urban land use.

3.3. *Extracting the pure samples*

Urban land-use pure sample extraction is the most important step and has a great impact on the mixed urban land-use results. Due to the lack of ready-made samples, POIs are used as auxiliary data for pure sample extraction, similar to Wu *et al*. (2018) and Zhang and Du (2015). The activities of urban residents in cities have certain purposes. According to these purposes, we divide urban land use into five types: commercial, recreation, housing, working, and public. As shown in Table 1, POIs are classified into these five types of urban land use according to their own types.

We calculate the proportions of POIs of these five urban land-use types in each street block, sort them, and extract the top-60 street blocks of each type as pure samples. The 300 samples obtained are verified by remote sensing images and electronic maps to confirm that these samples are relatively pure street blocks with a single land-use type.

3.4. *Validation of the proposed traj2vec model*

We first compare the classification accuracies of different methods in pure urban land-use sample classification. The land-use classification methods of Yao *et al*. (2017) and Zhai *et al*. (2019) are compared with the proposed Traj2Vec model.

(A) Proposed Traj2Vec model; Cell towers are taken as words and user trajectories as sentences. The vector of each cell tower is obtained by training Word2Vec, and then the semantic feature vectors of street blocks are obtained by weighted summation.
(B) Word2Vec + POIs (Yao *et al*. 2017); The POI types are treated as words, and the POIs in a street block are connected into a sentence by the shortest path method. Then, the semantic vector of each POI type is obtained by training Word2Vec, and the feature vector of each street block is obtained by averaging POI vectors with weightings.
(C) Place2Vec + POIs (Zhai *et al*. 2019); The overall idea of this method is similar to (B). The difference between (C) and (B) is that a simple augmented method is used to construct a corpus for Word2Vec training. In addition, the Place2Vec model improves the objective function of the Word2Vec model, making the model more suitable for POI data.

The above three methods can obtain the semantic feature vector of each street block. We set the output vector dimension to 100 and the sample window size to 3 for all methods. The feature vectors of these three methods were used to train three RF models separately, and then the quality of these three methods was determined by comparing the accuracy of each RF model. We used 75% of the 300 pure street block samples for training and 25% for testing. The decision trees of all RF models were set to 100. Specifically, each decision tree was trained with 50% of the training data, and the remaining 50% of the training data were used as the OOB data. To ensure a reliable estimation of the classification accuracy, each RF model was run 100 times. Table 2 shows that our proposed method obtains the highest accuracy in urban land-use classification. By exploring the potential semantics of the activity trajectory information from mobile phone users using the proposed Traj2Vec model, our method quantifies the correlation between urban land uses and the activities of urban residents. In addition, our method consumes only slightly more time and resources in the whole process than the other methods. The differences in time and resource consumption are negligible for the entire process. When we use other parameters for the experiments, we find that the effect on the results is not significant.

In addition, the PA (Equation (6)) was used to evaluate the accuracy of the proportion results. Since the ground truth of mixed urban land use was unavailable, we randomly selected 100 street blocks and obtained the proportions of urban land-use types as subjective references through visual interpretation. The average PA of these 100 street blocks obtained by our method is 64.0%.

Six street blocks are employed to further evaluate the performance in obtaining the proportions of urban land-use types (Figure 4). Figure 4(a–e) show street blocks with high proportions of working, recreation, housing, commercial and public land-use types, respectively, while Figure 4(f) shows a street block with a high degree of mixing. Figure 4(g) shows that the proportions of urban land-use types determined by our model are similar to those from the visual interpretation. The PA values of these six street blocks are 80.0%, 86.0%, 86.0%, 76.0%, 74.0%, and 78.0%, respectively.

For street blocks with a single urban land-use type, our model not only determines the main land-use type but also determines other land-use types with small proportions, which may be affected by the surrounding other land-use types when the semantic feature vector is obtained by weighted summation. For example, street block (e) represents the Baoan District Government, with residential areas, parks, and shopping centers nearby, resulting in a variety of urban land-use types in our model in addition to the public type. However, the proportions of other urban land-use types are very low and have little effect on the overall results, especially in street blocks with a high mixing index.

These results suggest that our proposed method can perform well in obtaining the proportions of urban land-use types, whether in a single land-use street block or a street block with a higher degree of mixing.

Table 2. Accuracy assessment of the urban land-use classification results by various models.

		Training process		Prediction process	
Expr. ID	Method	OOB error	OOB RMSE	OA	Kappa
(A)	**Proposed method**	0.1034 ± 0.0012	0.2515 ± 0.0016	0.7733 ± 0.0033	0.7245 ± 0.0010
(B)	**Word2Vec + POIs**	0.1475 ± 0.0017	0.2833 ± 0.0013	0.7155 ± 0.0025	0.6876 ± 0.0012
(C)	**Place2Vec + POIs**	0.1249 ± 0.0022	0.2721 ± 0.0015	0.7363 ± 0.0036	0.7002 ± 0.0011

Note: OOB error: Relative classification error of the OOB data. **OOB RMSE**: Root mean square error of estimating the posterior probabilities for the OOB data. **OA**: Overall accuracy.

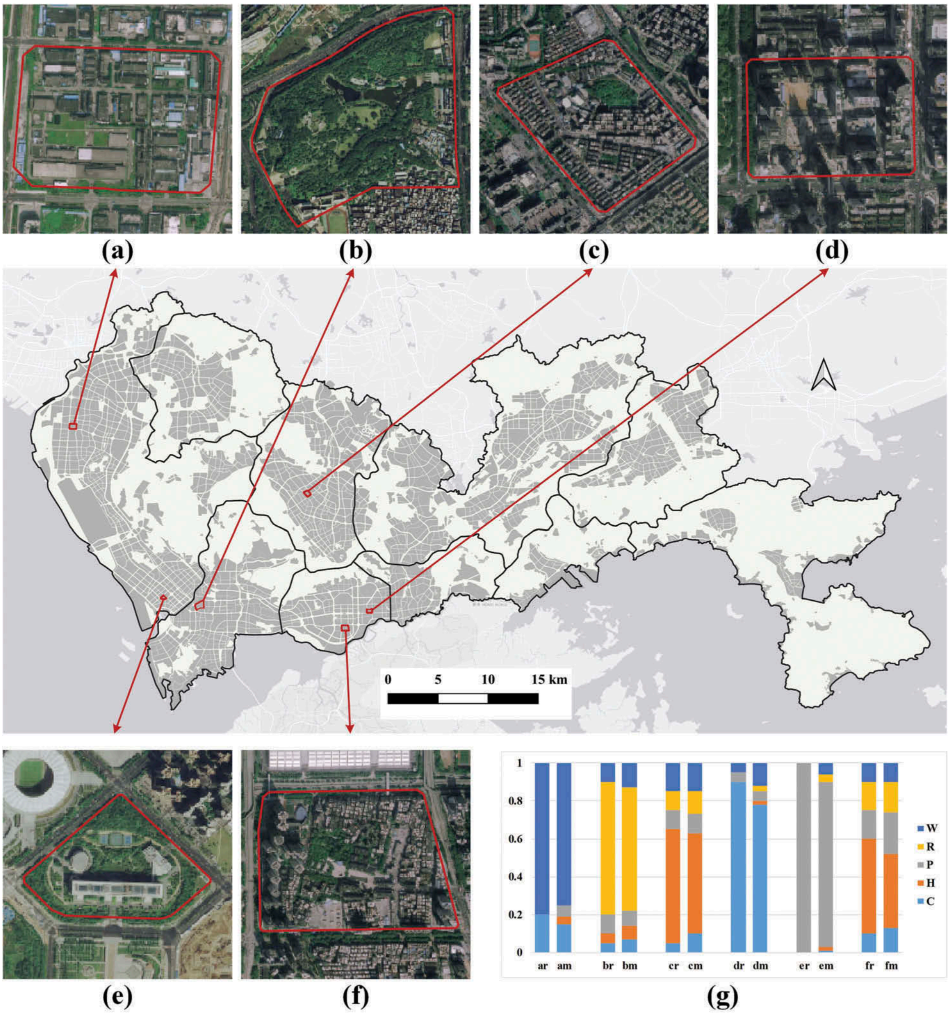

Figure 4. Examples of proportions of urban land-use types. (a) Industrial park in the Baoan district. (b) Zhongshan Park. (c) Residential area in Longhua District. (d) Huaqiang North Business District. (e) Baoan District Government. (f) Areas near Huanggang Village. (g) The proportions of urban land-use types in these six regions. 'xr' represents the reference proportions from visual interpretation of region (x), and 'xm' represents the proportions from the proposed model results for region (x).

3.5. Estimating the proportions of each urban land-use type

In this section, the 300 pure samples are used as training samples to train an RF model. The RF model is used to estimate the proportions of each urban land-use type in all street blocks in Shenzhen. The proportions of each urban land-use type (commercial, housing, public, recreation, and working) are illustrated in Figure 5(a– e), respectively. To describe the land-use diversity in street blocks, we calculated the Shannon index of the five urban land uses in each street block (Figure 5(f)).

Figure 5 shows that the street blocks with high proportions of commercial land use are areas with frequent economic activities, such as shopping malls and commercial streets ((a) is the

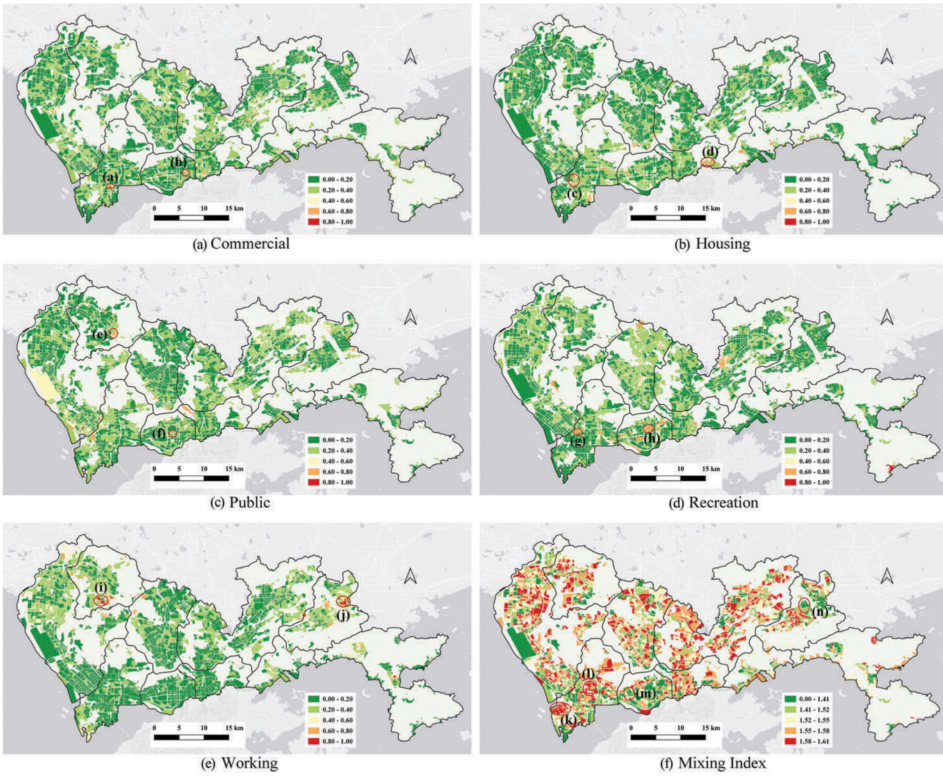

Figure 5. The proportions of each urban land-use type in Shenzhen. (a–e) show the proportions of commercial, housing, public, recreation, and working, respectively. (f) shows the spatial distribution of the mixing index (Shannon index).

Coastal City Shopping Center, and (b) is the Huaqiang North Business District). Similarly, the same is true for other urban land-use types. The areas with high proportions of the housing land-use type are residential areas ((c) is located in the residential area near Daxin, while (d) is a residential area in Luohu District). The internal structures of urban villages are complex; thus, the proportion of the housing land-use type is not as high as that in a residential area. For the public land-use type, the areas with high proportions are government agencies, transportation service facilities, etc. ((e) and (f) are the Guangmingcheng high-speed railway station and Shenzhen Municipal Government, respectively). The places where parks and amusement parks are located have higher proportions of recreation land use ((g) and (h) are Zhongshang Park and Lianhuashan Park, respectively). Most of the areas with high proportions of working land use are industrial areas, which are mainly distributed in the northeast and northwest of Shenzhen ((i) and (j) are two industrial parks).

From Figure 5 and the above analysis, it can be found that there are few street blocks with high proportions of single land-use types, and most of them have certain specific uses. Additionally, we found that the number of street blocks with a proportion of commercial land-use type greater than 0.5 is 156, 121 street blocks for housing land use, 170 street blocks for public land use, 124 street blocks for recreation land use, and 270 street blocks for working land use. The AWTPs (Equation (5)) of commercial, housing, public, recreation, and working land use are 19.45%, 21.55%, 18.20%, 20.20%, and 20.60%, respectively. These results indicate that Shenzhen has a small number of street blocks with high proportions of single land-use types, and the differences in the overall proportions of different land-use types are small.

A street block with a high index value means there is a high degree of mixed land use in the region. This reflects the complexity and diversity of urban land uses. As shown in Figure 5(f), the Shannon index is relatively high (the average Shannon index is 1.41) in most parts of Shenzhen, indicating that the land uses of Shenzhen are highly diverse and have a high degree (80.5% of street blocks have a Shannon index higher than 1.41) of mixing after years of rapid development. Moreover, the aggregation area with a high degree of mixing is complicated, and there are various land uses ((k) is near the Qianhaiwan Garden, and (l) is near the Shenzhen Software Park). In contrast, the aggregation area with a low degree of mixing often has a single land use ((m) is Xiangmi Park, and (n) is near the Guangdong Shenzhen Export Processing Zone).

3.6. Correlation between degree of mixing and trip distances

Some scholars have suggested that mixed urban land uses tend to shorten trip distances (Cervero 1996, Burton *et al*. 2003). On this basis, we take the average taxi and shared bike trip distances of a working day in 2017 as a proxy to validate the mixing index. Although they represent very different modes of transportation and most likely also different user groups in terms of socioeconomic status, they are both ways that people travel in cities. We regard the distance between the origin and destination (OD) of all shared bikes as the bike trip distances of residents and the OD of taxi trajectories as the taxi trip distances. A trip is included only if the origin point is within the street blocks.

The number of shared bikes used in this study is 275,248, with a total of 1,305,573 travel trajectories and an average OD distance of 509.207 meters. The spatial distribution of the average OD travel distance of shared bikes in each street block is shown in Figure 6(a). This figure shows that there is spatial heterogeneity in the travel distances of shared bikes in different street blocks. The Pearson correlation between the mixing index and average travel distances of shared bikes shows a weak negative correlation ($r = -0.093,\ p = 0.008$). This indicates that an increase in the degree of mixing can reduce the travel distance of shared bikes to a certain extent.

The total number of taxis is 14,268, with a total of 377,722 travel trajectories and an average OD distance of 5,706.898 meters. The spatial distribution of the average OD travel distance of taxis is shown in Figure 6(b). We find that most street blocks with higher average OD travel distances are distributed in the north, but the spatial distribution of the degree of mixing is quite different (Figure 6(d)). Therefore, the Pearson correlation between them shows a weaker and nonsignificant negative correlation ($r = -0.069,\ p = 0.159$), which is the same as the finding of Wu *et al*. (2018).

When we considered both shared bikes and taxis (Figure 6(c)), we found a weak but significant negative correlation between the degree of mixing and the average OD travel distance ($r = -0.107,\ p < 0.001$). This shows that an increase in mixing will lead to a decrease in the trip distance, which confirms that moderately mixed urban land uses will promote cities to become more compact (Burton *et al*. 2003, Holden and Norland 2005). As the degree of mixing increases, the daily needs of residents may be satisfied in a shorter distance. The weak relationship may be affected by the bias in the trip data.

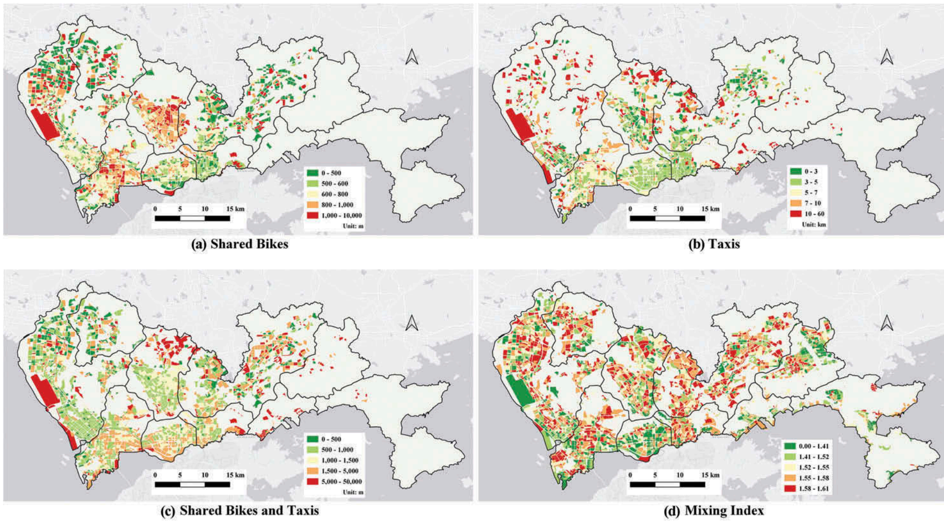

Figure 6. The average OD travel distances of shared bikes and taxis for a working day in 2017. (a) The average OD travel distances of 275,248 shared bikes. (b) The average OD travel distances of 14,268 taxis. (c) The average OD travel distances of shared bikes and taxis. (d) The spatial distribution of the mixing index (Shannon index).

Limited by the availability of data, we have adopted two relatively flexible modes of travel. Other trip modes, such as walking and public transport, are not included, and adding that data may lead to a stronger relationship. Nevertheless, the results not only confirm the conclusion that an increase in the degree of mixing will reduce the travel distances of residents but also verify the mixing index.

4. Conclusions

In this study, we propose the Traj2Vec model to retrieve potential semantic information of urban residents' trajectories based on the Word2Vec model. By obtaining and exploring the semantic information of spatial trajectories, we find that the Traj2Vec model can well establish a relationship between residents' trajectories and mixed urban land uses. We also find that the Traj2Vec model is able to utilize rich semantic information to discover the consistencies and inconsistencies between human activities and urban land uses.

First, the analysis shows that the proposed Traj2Vec approach can obtain the highest accuracy (OA = 0.7733, kappa = 0.7245) in detailed urban land-use classification and obtain a high average PA (64.0%) in capturing mixed land use. With the rapid development of a city, the street blocks in the city often have a mix of land uses, which may provide convenience for residents' daily lives. Due to the limitations of data and methods, there have been few studies on the identification of mixed urban land use (Wu *et al.* 2018). Since trajectory information reflects urban residents' use of urban space, we obtained the proportions of urban land-use types by quantifying residents' activity trajectories as semantic information using our proposed Traj2Vec model. The results show that with rapid development, Shenzhen has a high degree of mixed urban land use. After some improvement of the proposed Traj2Vec model, we can not only use the method for land use but also for the study of residents' travel patterns. For example, we can determine the trajectory semantic vectors of residents so that residents with the same travel patterns can be determined. In addition, our method can also be easily applied to other places. As long as there are mobile phone data and POI data in one place, we can obtain mixed urban land use results in the area through our method.

Second, by exploring the vectors obtained by the Traj2Vec model, we find that the Traj2Vec model can effectively obtain the spatial semantic information contained in residents' trajectories. After treating cell towers as words and travel trajectories as sentences, we obtain the semantic vector of each cell tower through the Word2Vec model. For the semantic feature vectors of cell towers that we obtained, those cell towers with similar vectors are closer in spatial patterns, which agrees with the first law of geography, i.e. closer things have more similarity (Tobler 1970). Additionally, the semantic features of street blocks obtained by weighted summation follow the same law. These results indicate that the semantic vectors obtained by fully exploring residents' trajectories through the proposed model can represent the potential relationships between street blocks.

Third, by analyzing the mixing index and the travel distance, we found that there is a weak but significant negative correlation between them ($r = -0.107,\ p < 0.001$). Related urban theory suggests that mixed urban land uses tend to shorten trip distances (Cervero 1996, Burton *et al.* 2003). In this study, we take the average taxi and shared bike trip distances as proxies to validate this theory. The results are consistent with the urban theory, which means that when a city develops to a certain stage, moderate mixed urban land uses will promote the city to become more compact. The results not only confirm the conclusion that an increase in the degree of mixing will reduce the travel distances of residents but also verify the mixing index obtained in this study. This suggests that urban planning should focus on mixed urban land uses, which can reduce the travel distances of residents, reduce energy consumption, and make cities more compact.

Several issues should be addressed in further studies. The cell phone data used in this study contain data from only one day and may reflect some occasional user movements. In the future, a time series of mobile phone data or other trajectories can be used for more accurate data mining. The selection of pure samples is based on POIs that may be affected by the quality of data. Moreover, a more accurate and objective method should be used to select samples that are crucial to building the model. Despite the above limitations, the information about mixed urban land uses and spatial distributions obtained by this study should be very important for understanding cities and can be used to assist in decision making and policy formulation.

Acknowledgments

We would like to thank Prof. May Yuan, Prof. Urska Demsar and three anonymous reviews for their constructive suggestions that significantly strengthened this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Key National Natural Science Foundation of China [41531176]; National Key R&D Program of China [2017YFA0604402]; National Natural Science Foundation of China [41801306].

Notes on contributors

Jinbao Zhang is a PhD candidate in the School of Geography and Planning at Sun Yat-sen University. His research focuses on spatio-temporal data mining and urban computing.

Xia Li is currently a professor in the School of Geographic Sciences at East China Normal University. His research focuses on geographic information science, cellular automata, global land use simulation, multi-agent system, and remote sensing applications.

Yao Yao is currently an associate professor in the School of Geography and Information Engineering at China University of Geosciences. His research focuses on spatio-temporal data mining, urban computing and social sensing.

Ye Hong is a master candidate in the Department of Civil, Environmental and Geomatic Engineering at ETH Zürich. His research focuses on spatio-temporal data mining and urban computing.

Jialyu He is a PhD candidate in the School of Geography and Planning at Sun Yat-sen University. His research focuses on scene classification, land use simulation and urban computing.

Zhangwei Jiang is currently an algorithm expert at Alibaba Group. His research focuses on data mining.

Jianchao Sun is currently an algorithm expert at Alibaba Group. His research focuses on data mining.

ORCID

Jinbao Zhang http://orcid.org/0000-0001-8510-149X
Xia Li http://orcid.org/0000-0003-3050-8529
Yao Yao http://orcid.org/0000-0002-2830-0377
Ye Hong http://orcid.org/0000-0002-8996-3748
Jialyu He http://orcid.org/0000-0002-0997-558X

Data and codes availability statement

Data and codes supporting the findings of this study are available in 'figshare.com' with the identifier 'doi: 10.6084/m9.figshare.11655864'. Mobile phone data cannot be shared publicly without authorization. Readers interested in this study can find the result in the shared package.

References

Battle, R. and Kolas, D., 2012. Enabling the geospatial semantic web with parliament and geosparql. *Semantic Web*, 3 (4), 355–370. doi:10.3233/SW-2012-0065.

Biau, G.A.V.S., 2012. Analysis of a random forests model. *Journal of Machine Learning Research*, 13 (Apr), 1063–1095.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32. doi:10.1023/A:1010933404324.

Burton, E., Jenks, M., and Williams, K., 2003. *The compact city: a sustainable urban form?*. London, England: Routledge. doi:10.4324/9780203362372.
Cervero, R., 1996. Mixed land-uses and commuting: evidence from the American Housing Survey. *Transportation Research Part A: Policy and Practice*, 30 (5), 361–377.
Chen, S., *et al*., 2016. Discovering urban functional regions using latent semantic information: spatiotemporal data mining of floating cars GPS data of Guangzhou. *Journal of Geographical Sciences*, 71, 471–483.
Chen, Y., *et al*., 2017. Delineating urban functional areas with building-level social media data: a dynamic time warping (DTW) distance based k-medoids method. *Landscape and Urban Planning*, 160, 48–60. doi:10.1016/j.landurbplan.2016.12.001.
Cranshaw, J., *et al*., 2012. The livehoods project: utilizing social media to understand the dynamics of a city. *In: Sixth international AAAI conference on weblogs and social media*. Dublin, Ireland.
Fern A Ndez-Delgado, M., *et al*., 2014. Do we need hundreds of classifiers to solve real world classification problems. *Journal of Machine Learning Research*, 15 (1), 3133–3181.
Gao, Q., *et al*., 2018. Exploring changes in the spatial distribution of the low-to-moderate income group using transit smart card data. *Computers, Environment and Urban Systems*, 72, 68–77. doi:10.1016/j.compenvurbsys.2018.02.006.
Gao, S., Janowicz, K., and Couclelis, H., 2017. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21 (3), 446–467. doi:10.1111/tgis.2017.21.issue-3.
Gonzalez, M.C., Hidalgo, C.A., and Barabasi, A., 2008. Understanding individual human mobility patterns. *NATURE*, 453 (7196), 779. doi:10.1038/nature06958.
Holden, E. and Norland, I.T., 2005. Three challenges for the compact city as a sustainable urban form: household consumption of energy and transport in eight residential areas in the greater Oslo region. *Urban Studies*, 42 (12), 2145–2166. doi:10.1080/00420980500332064.
Lenormand, M., *et al*., 2015. Comparing and modelling land use organization in cities. *Royal Society Open Science*, 2 (12), 150449. doi:10.1098/rsos.150449.
Li, J., *et al*., 2016. Application of GPS trajectory data for investigating the interaction between human activity and landscape pattern: a case study of the Lijiang River Basin, China. *ISPRS International Journal of Geo-Information*, 5 (7), 104. doi:10.3390/ijgi5070104.
Li, X., *et al*., 2017. Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps. *Remote Sensing of Environment*, 196, 293–311. doi:10.1016/j.rse.2017.05.011.
Liu, X., *et al*., 2017. Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science*, 31 (8), 1675–1696. doi:10.1080/13658816.2017.1324976.
Liu, X. and Long, Y., 2016. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design*, 43 (2), 341–360. doi:10.1177/0265813515604767.
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. doi:10.1080/00045608.2015.1018773.
Long, Y. and Shen, Z., 2015. Discovering functional zones using bus smart card data and points of interest in Beijing. *In*: *Geospatial analysis to support urban planning in Beijing*, 193-217. Basel, Switzerland: Springer, Cham. doi:10.1007/978-3-319-19342-7_10.
Mikolov, T., *et al*., 2013. Distributed representations of words and phrases and their compositionality. *In: Advances in neural information processing systems 26*, 3111–3119. Lake Tahoe, Nevada, USA.
Miller, G.A., 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38 (11), 39–41. doi:10.1145/219717.219748.
Niu, N., *et al*., 2017. Integrating multi-source big data to infer building functions. *International Journal of Geographical Information Science*, 31 (9), 1871–1890.

Pei, T., *et al.*, 2014. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28 (9), 1988–2007. doi:10.1080/13658816.2014.913794.

Regan, C.M., *et al.*, 2015. Real options analysis for land use management: methods, application, and implications for policy. *Journal of Environmental Management*, 161, 144–152. doi:10.1016/j.jenvman.2015.07.004.

Shen, Y. and Karimi, K., 2016. Urban function connectivity: characterisation of functional urban streets with social media check-in data. *CITIES*, 55, 9–21. doi:10.1016/j.cities.2016.03.013

Siła-Nowicka, K., *et al.*, 2016. Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30 (5), 881–906. doi:10.1080/13658816.2015.1100731.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (sup1), 234–240. doi:10.2307/143141.

Tranos, E. and Nijkamp, P., 2015. Mobile phone usage in complex urban systems: a space–time, aggregated human activity study. *Journal of Geographical Systems*, 17 (2), 157–185. doi:10.1007/s10109-015-0211-9.

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. doi:10.1080/13658816.2017.1356464.

Wang, J., *et al.*, 2014. Discovering urban spatio-temporal structure from time-evolving traffic networks. *In: Asia-Pacific Web Conference*, 93–104. Changsha, China: Springer, Cham. doi:10.1007/978-3-319-11116-2_9.

Wolpert, D.H. and Macready, W.G., 1999. An efficient method to estimate bagging's generalization error. *Machine Learning*, 35 (1), 41–55. doi:10.1023/A:1007519102914.

Wu, L., *et al.*, 2018. A framework for mixed-use decomposition based on temporal activity signatures extracted from big geo-data. *International Journal of Digital Earth*, 1–19. doi:10.1080/17538947.2018.1556353.

Xu, Y., *et al.*, 2016. Another tale of two cities: understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers*, 106 (2), 489–502.

Yan, B., *et al.*, 2017. From ITDL to Place2Vec: reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. *In: Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems*, 35, 1–10. Redondo Beach, California, USA. doi:10.1145/3139958.3140054.

Yao, Y., *et al.*, 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and google Word2Vec model. *International Journal of Geographical Information Science*, 31 (4), 825–848. doi:10.1080/13658816.2016.1244608.

Yao, Y., *et al.*, 2018. Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Transactions in GIS*, 22 (2), 561–581. doi:10.1111/tgis.2018.22.issue-2.

Yuan, N.J., *et al.*, 2015. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 27 (3), 712–725. doi:10.1109/TKDE.2014.2345405.

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. doi:10.1016/j.compenvurbsys.2018.11.008.

Zhang, X. and Du, S., 2015. A linear dirichlet mixture model for decomposing scenes: application to analyzing urban functional zonings. *Remote Sensing of Environment*, 169, 37–49. doi:10.1016/j.rse.2015.07.017

Zhao, B., Zhong, Y., and Zhang, L., 2013. Scene classification via latent dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery. *Remote Sensing Letters*, 4 (12), 1204–1213. doi:10.1080/2150704X.2013.858843.

Zhong, C., *et al.*, 2014. Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems*, 48, 124–137. doi:10.1016/j.compenvurbsys.2014.07.004.

Zhong, Y., Zhu, Q., and Zhang, L., 2015. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery. *Ieee Transactions on Geoscience and Remote Sensing*, 53 (11), 6207–6222. doi:10.1109/TGRS.2015.2435801.