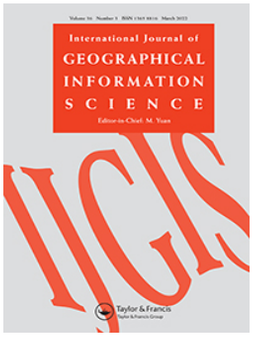

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Estimating urban functional distributions with semantics preserved POI embedding

Weiming Huang, Lizhen Cui, Meng Chen, Daokun Zhang & Yao Yao

To cite this article: Weiming Huang, Lizhen Cui, Meng Chen, Daokun Zhang & Yao Yao (2022): Estimating urban functional distributions with semantics preserved POI embedding, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2022.2040510

To link to this article: https://doi.org/10.1080/13658816.2022.2040510

Published online: 08 Mar 2022.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2022.2040510

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Estimating urban functional distributions with semantics preserved POI embedding

Weiming Huang[a,b], Lizhen Cui[a,b], Meng Chen[a,c], Daokun Zhang[d,e] and Yao Yao[f]

[a]School of Software, Shandong University, Jinan, PR China; [b]C-FAIR, Shandong University, Jinan, PR China; [c]Key Laboratory of Urban Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen, PR China; [d]Department of Data Science and AI, Faculty of Information Technology, Monash University, Melbourne, Australia; [e]Monash Suzhou Research Institute, Suzhou, PR China; [f]School of Geography and Information Engineering, China University of Geosciences, Wuhan, PR China

ABSTRACT

We present a novel approach for estimating the proportional distributions of function types (i.e. functional distributions) in an urban area through learning semantics preserved embeddings of points-of-interest (POIs). Specifically, we represent POIs as low-dimensional vectors to capture (1) the spatial co-occurrence patterns of POIs and (2) the semantics conveyed by the POI hierarchical categories (i.e. categorical semantics). The proposed approach utilizes spatially explicit random walks in a POI network to learn spatial co-occurrence patterns, and a manifold learning algorithm to capture categorical semantics. The learned POI vector embeddings are then aggregated to generate regional embeddings with long short-term memory (LSTM) and attention mechanisms, to take account of the different levels of importance among the POIs in a region. Finally, a multilayer perceptron (MLP) maps regional embeddings to functional distributions. A case study in Xiamen Island, China implements and evaluates the proposed approach. The results indicate that our approach outperforms several competitive baseline models in all evaluation measures, and yields a relatively high consistency between the estimation and ground truth. In addition, a comprehensive error analysis unveils several intrinsic limitations of POI data for this task, e.g. ambiguous linkage between POIs and functions.

ARTICLE HISTORY
Received 4 September 2021
Revised 4 February 2022
Accepted 6 February 2022

KEYWORDS
Urban function; point-of-interest; POI-region embedding; spatial co-occurrence; categorical semantics

1. Introduction

The latest demographic projections forecast the tendency of an increasing urban population in the forthcoming decades; thus, we are confronting the challenge of making cities more fit for human habitation (United Nations 2019). Among numerous perspectives, studying the functional distributions of urban spaces is pivotal for promoting the formation of sustainable and livable cities (Rodrigue *et al*. 2013). Specifically, our cities are composed of many regions that bear various functions, such as *residential*, *commercial* and

CONTACT Lizhen Cui clz@sdu.edu.cn

industrial. In reality, the function of each region is not unitary, but is normally a composition of several functional types. Moderately blended functional regions are advocated, which can make cities more compact, promote urban vibrancy and yield socioeconomic benefits, e.g. reducing the need for long-distance commuting (Burton *et al*. 2003, Koster and Rouwendal 2012, Yue *et al*. 2017). Therefore, studying the proportional distributions of function types (i.e. the proportions of different functions that each region bears) is of paramount relevance for the understanding, planning and management of our cities.

Recently, data mining approaches using a variety of crowdsourcing data sources have become increasingly popular for studying urban functional distributions, in view of the shortcomings of remote sensing data in delineating the socioeconomic perspectives of regions (Yao *et al*. 2017). Such crowdsourcing data sources include, among others, POIs (e.g. Gao *et al*. 2017, Barlacchi *et al*. 2021), social media data (e.g. Zhou and Zhang 2016, Tu *et al*. 2017), and human mobility data (e.g. Zhang *et al*. 2021). A recent survey by Andrade *et al*. (2020) revealed that, POIs are one of the most commonly used crowdsourcing data sources in this regard, owing to their intrinsic connections with human behavior and the socioeconomic perspectives of cities (Janowicz 2012). In addition, POIs can usually be easily obtained compared to other data sources, such as human mobility data, of which the availability is often restricted to a few particular areas and certain user groups. Therefore, although increasing types of crowdsourcing data are emerging, POIs remain a valuable and readily available data source for estimating urban functional distributions.

Early studies that utilized POIs for mining urban functional regions mainly relied on feature engineering methods, with POI frequencies being the most employed feature (Tian and Shen 2011, Jiang *et al*. 2015). Even if one discounts tedious efforts to construct features, such methods suffer from the loss of many types of latent information, e.g. the spatial distribution patterns of POIs. To overcome such limitations, representation learning approaches have been utilized. Such approaches learn low-dimensional latent vector embeddings for POI categories (e.g. *hotel* and *park*) based on spatial co-occurrence information with a certain sampling strategy. A pioneering work was inspired by the seminal idea of Word2Vec in natural language processing (Mikolov *et al*. 2013), in which a string of nearby POIs is constructed in each region, so that spatial co-occurrence information can be captured according to closeness in the strings (Yao *et al*. 2017). A subsequent study modified the sampling strategy to the Place2Vec approach (Yan *et al*. 2017; essentially K-nearest-neighbors [KNN]), which captures spatial co-occurrence information based on POI adjacency relations (Zhai *et al*. 2019). The rationale behind these studies is that there is a linkage between region functions and POI spatial co-occurrence patterns (e.g. *mall* and *parking lot* usually co-occur, and they both imply certain functions). Subsequently, the learned POI (category) embeddings are aggregated with an average operation to generate regional embeddings, which are then mapped to a type of urban function through a supervised classifier, e.g. random forest or fed into clustering methods in an unsupervised setting.

Despite the remarkable performance of such intuitive yet powerful models, previous studies have several limitations that impede the utilization of the rich information in POIs:

1. Incapability to estimate multiple functions: Previous studies usually only assign a single functional label to a certain region, neglecting the general presence of

mixed and multiple functions embodied in a single region. Although there are a few studies dealing with the mixed-function problem with human mobility data (Zhang *et al*. 2021) and social media data (Wu *et al*. 2020), there has not been a supervised approach that could fully utilize the available ground truth data, e.g. Zhang *et al*. (2021) only used regions with a single function for training.

2. Shortage in capturing long-range spatial co-occurrence: The methods to capture spatial co-occurrence relations between POIs mainly concentrate on adjacency relations, i.e. two POIs would only co-occur if they are spatially close to each other. We can readily learn local patterns using such a sampling strategy, e.g. the co-occurrences between *malls* and *parking lots*. However, this strategy falls short in capturing long-range complementary dependencies (Du *et al*. 2019). For example, *schools* tend to distribute evenly in cities, and the dependency between *universities* and *university science parks* could arise outside of close proximity (Mai *et al*. 2020).
3. Omission of categorical semantics: Previous studies only utilized spatial co-occurrence information to learn POI embeddings, overlooking the intrinsic semantic relations in their hierarchical categories (Jin *et al*. 2019). For instance, *Chinese restaurant* and *western-style restaurant*, despite their differences, have substantial similarities in their functional affordance, and they both belong to an upper-level category: *food service* (according to Baidu Map).[1] In fact, hierarchical categories are common for POI data, such as the POIs from Baidu Map, Foursquare and Yelp. POI embeddings should ideally take into account such categorical semantics.
4. Naïve aggregation scheme to obtain region embeddings: Previous studies mainly used an element-wise average operation to aggregate POI vector embeddings and generate regional embeddings, neglecting the different influence levels of POIs on the functions of a region. For example, a *railway station* may only appear once in a region, and *restaurant* appears many times. In this case, the *railway station* should be more definitive for the region. Simply averaging all POI embeddings in a region would water down such key information. Although several works utilized POI frequencies or popularity for determining the semantics of regions (e.g. Gao *et al*. 2017, Yan *et al*. 2017, Liu *et al*. 2020), how to consider different importance levels of POIs to generate region embeddings is largely unexplored.

To overcome the aforementioned limitations, in this article, we formulate the problem of estimating urban functional distributions with POIs as a supervised label distribution learning problem, and propose an approach that learns semantics preserved POI embeddings. The approach comprises four components: 1) spatially explicit random walks in a POI network to capture both local and long-range spatial co-occurrence information; (2) the incorporation of categorical semantics into POI embeddings with a semantic smoothing assumption and a manifold learning method; (3) an aggregation function coupling LSTM (Hochreiter and Schmidhuber 1997) and attention mechanisms (Vaswani *et al*. 2017) to aggregate POI embeddings and generate regional embeddings, to account for the differences of each POI's importance in defining the functions of a region; and (4) an MLP that maps the generated regional embeddings

to functional distributions. The proposed approach is evaluated in a study area in Xiamen, China, against several competitive baseline models.

The primary goal of this study is to estimate urban functional distributions, which can be applied to the scenarios, such as (1) when the ground truth of functional use in a city is only partially available (e.g. through field survey), our approach can be used to estimate the functional distributions in the unknown areas, as the discovered pattern would most likely hold within the same city; (2) when the urban functional use data is partially obsolete, e.g. due to urban renewal, our approach can be leveraged to discover such updates, as POIs are usually up-to-date; (3) when the urban function map is not completely accurate, one could use our approach to discover errors in such data. In addition, from a methodological viewpoint, our approach can also be used as a general POI embedding method for other downstream tasks, such as POI classification (e.g. Mai *et al*. 2020) and recommendation (e.g. Yu *et al*. 2020).

Following this introduction, we formulate the problem in Section 2. In Section 3, we provide the details and intuitions of our approach. In Section 4, we demonstrate the results of our experiment, including a comparison with several baseline models, an ablation study, a parameter sensitivity analysis, and an error analysis. The article ends with a discussion in Section 5, and the conclusions and outlook in Section 6.

2. Problem formulation

The notations used in this study are as follows. Let $\mathcal{F} = \{\mathrm{f}_1, \ \ldots, \mathrm{f}_m\}$ be the set of m labels indicating the function types of an urban region (e.g. *residential*, *commercial* and *public service and education*). Let $\mathcal{R} = \{\mathrm{r}_1, \ \ldots, \mathrm{r}_n \}$ be a set of n spatially disjoint urban regions, and $y_i^{f_k}$ represents the proportion of the kth function that the region r_i bears, which satisfies the constraints $y_i^{f_k} \in [0, 1]$ and $\sum_k y_i^{f_k} = 1$. Let $\mathcal{P} = \{p_1, \ \ldots, \mathrm{p}_t\}$ be a set of t POIs, and the ith POI p_i (e.g. the national museum) is associated with a two-dimensional geographic location $\boldsymbol{x}_i$ and a category set c_i, where $c_i = \{c_i^1, \ldots, c_i^{h_i}\}$, with c_i^j indicating the category of POI p_i in the j th hierarchical category level and h_i denoting the depth of the categorical hierarchy ($h_i = 2{,}3$ for most POI providers). For example, a POI's category set is {*food service* (first-level category), *Chinese restaurant* (second-level category)}, where each first-level category conceptually contains a number of second-level categories.

The problem of estimating urban functional distribution can be formulated as a label distribution learning problem (Geng 2016): For a region r_i, given the POI set $\mathcal{P}_i \subseteq \mathcal{P}$ that spatially resides in r_i, the goal is to learn a conditional probability distribution $P(f_k|\mathcal{P}_i)$ for each function label $f_k \in \mathcal{F}$, i.e. obtaining an estimated functional distribution $y_i = \{P(f_1|\mathcal{P}_i), \ P(f_2|\mathcal{P}_i), \ \cdots, P(f_m|\mathcal{P}_i)\}$ (proportions of function types that region r_i affords). In this process, each POI p_i is represented as a vector embedding $\boldsymbol{p}_i$ in the latent space $\mathbb{R}^s$, and such a vector captures the information of spatial co-occurrence and the hierarchical structures of categories, which implies that the embedding of a POI is determined by its category. Therefore, in this study, we use the terms *POI embedding* and *POI category embedding* interchangeably.

3. Methodology

The overarching architecture of the proposed approach is illustrated in Figure 1. This approach comprises four major components. First, we construct a POI network and design a spatially explicit random walk strategy to sample POI co-occurrence information (Figure 1(a); Section 3.1). We then capture the categorical hierarchy of POIs (Figure 1(b)). The captured co-occurrence information and the categorical semantic information are fed into a POI encoder ϕ that generates POI embeddings by simultaneously optimizing the objectives of skip-gram (Mikolov *et al.* 2013) and a manifold learning algorithm Laplacian Eigenmaps (LE; Belkin and Niyogi 2001) (Section 3.2). Subsequently, the embeddings of the POIs in a single region are stacked to generate an embedding matrix for each region, and each matrix is passed through an aggregation function Γ coupling LSTM and attention mechanisms to obtain an embedding for each region (Figure 1(c); Section 3.3). Finally, an MLP ψ maps the regional embeddings to functional distributions (Figure 1(d); Section 3.4). In this process, the POI encoder ϕ is trained in an unsupervised manner, whereas the aggregation function Γ and MLP ψ are trained with the supervision of the urban functional distribution ground truth data.

3.1. Capturing spatial co-occurrence information in a POI network

Network, which essentially models entities as nodes and their connections as edges, is a natural data model for modeling and linking discrete spatial vector data

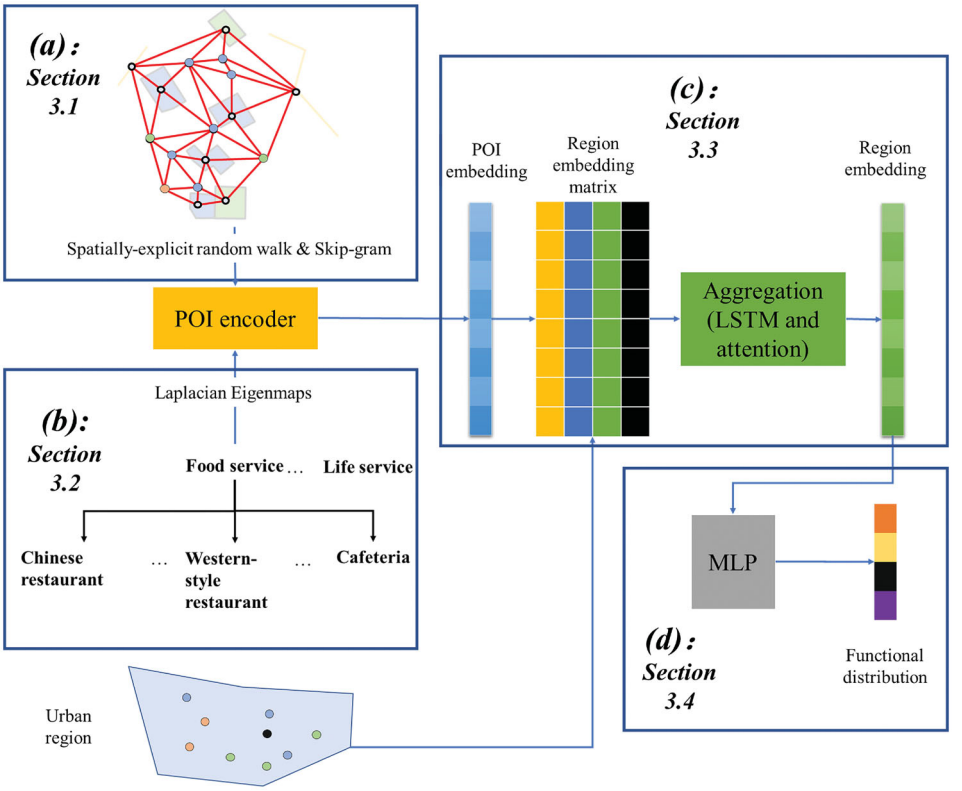

Figure 1. The overarching architecture of the proposed approach.

(e.g. POIs), in virtue of its flexibility which is not tied to a rigid raster (e.g. for remote sensing imageries) or sequence (e.g. for trajectories) notion (Yan *et al*. 2019, 2021). In particular, we can leverage a network representation learning approach (Zhang *et al*. 2020) to flexibly capture the co-occurrence patterns of POI categories.

POIs that scatter around a city do not naturally form a network. Nevertheless, each POI can be viewed as a network node, and the edges between the nodes are expected to foster connectivity. In principle, there are various means for generating a spatial network with a set of points. In this article, we employ the Delaunay triangulation (DT) network, in virtue of its several favorable properties which make the generated network both informative and compact. In addition, previous studies have demonstrated its fitness for learning embeddings of spatial vector data (e.g. Yan *et al*. 2019).

In this article, we propose a spatially explicit random walk sampling strategy to capture the co-occurrence patterns of POI categories. The core of our method lies in the transition probability from point *a* to point *b* along network edges, and such a biased probabilistic transition between nodes (POIs) incorporates spatial distance decay, the balance between local and long-range co-occurrence patterns, and the differentiation between intra- and cross-region co-occurrences. To this end, we define three types of transition bias in the random walk process. In Figure 2, a DT network for the POIs is shown. Assuming that we are going through a random walk process, and have just completed the transition $p_1 \rightarrow p_2$, and for the next step of transition, we have several candidates: the seven neighbors of p_2 : $\{p_1, p_3, \ldots, p_8\}$ (it is allowed to traverse back to the previous node). Each candidate node has a transition probability that depends on three transition biases.

The first is an inverse-distance transition bias α_d:

$$\alpha_d\ (p_2, x) = \log\left(1 + D^{1.5}/1 + d_{p_2x}^{1.5}\right) \tag{1}$$

α_d ensures that spatially closer candidate nodes are assigned with higher probability, where D denotes the diagonal length of the minimum bounding rectangle of all the

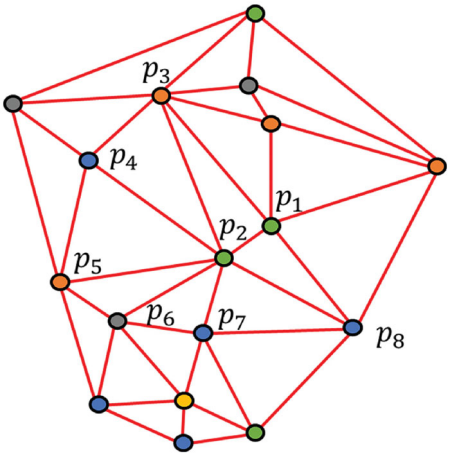

Figure 2. A DT network for POIs. Each node color represents a second-level category.

POIs in the study area, and d represents the spatial distance between two nodes p_2 and x (the candidate node). The rationale behind the distance decay of $\alpha_d \sim d^{-1.5}$ is in view of the previous studies in Calafiore *et al*. (2021) and Chen *et al*. (2015) that revealed the exponent of -1.5 can well capture spatial network structures.

The second transition bias balances local and long-range co-occurrences between POIs, which is inspired by Node2Vec (Grover and Leskovec 2016). Recall our example in Figure 2, the second transition bias α_b is defined as:

$$\alpha_b\ (p_2, x) = \begin{cases} \alpha_b^{loc} \ if\ hop_{p_1 x} = 0 \\ 1 \ if\ hop_{p_1 x} = 1 \\ \alpha_b^{lr} \ if\ hop_{p_1 x} = 2 \end{cases} \tag{2}$$

where $hop_{p_1 x}$ denotes the required minimum number of hops from p_1(the previous node) to the candidate node x. For example, we have $\alpha_b\ (p_2, x) = \alpha_b^{loc}$ for going back ($x = p_1$), $\alpha_b\ (p_2, x) = 1$ for $x = p_3$, and $\alpha_b\ (p_2, x) = \alpha_b^{lr}$ for traversing further to explore long-range information, e.g. $x = p_5$. The searching strategy of α_b^{loc} and α_b^{lr} can reference to Grover and Leskovec (2016). Note that the notion of long-range is within the random walk framework, and is not strictly bonded to spatial distance incrementation. Nevertheless, in reality, making more hops further could reach rather distant places.

The third transition bias balances intra- and cross-region co-occurrences. In previous studies, Yao *et al*. (2017) only sampled intra-region co-occurrences, while Zhai *et al*. (2019) sampled both of them without any differentiation. We argue that neither of the sampling strategies fully respects the influence of region boundaries. Indeed, the co-occurrence between two POIs in different regions entails a certain level of correlation, but is weaker than the co-occurrences in the same region. Therefore, we define the third transition bias $\alpha_r\ (p_2, x)$ as:

$$\alpha_r\ (p_2, x) = \begin{cases} 1 \ if\ \{p_2, x\} \subseteq \mathcal{P}_i \\ \alpha_r^{\text{inter}-\text{region}} \ if\ p_2 \in \mathcal{P}_i,\ x \in \mathcal{P}_k,\ and\ r_i \neq r_k \end{cases} \tag{3}$$

In principle, the value of $\alpha_r^{\text{inter}-\text{region}}$ should be smaller than 1, which implies that the intra-region co-occurrence information should be more likely to be sampled.

Finally, the unnormalized transition probability in our spatially explicit random walk from the current node p_2 to each candidate node x is:

$$tp(p_2, x) = \alpha_d\ (p_2, x) \times \alpha_b\ (p_2, x) \times \alpha_r\ (p_2, x) \tag{4}$$

With the proposed sampling strategy, we perform several walks starting from each node, and can then obtain a number of POI sequences. Subsequently, each POI in the sequence is represented by its second-level category c_i^2. The reason for using the second-level category (e.g. *Chinese restaurant*) is that it is neither too generic (e.g. *food service* in the first level) nor plethorically detailed (*Sichuan restaurant* in the third level) (Zhai *et al*. 2019). For each sampled sequence, the first is regarded as the *target category*, and each of the rest is a *context category*.

The embedding of each POI second-level category is then obtained by optimizing a skip-gram neural network with a negative sampling process, which essentially entails minimizing the objective function:

$$\mathcal{L}_{\text{co-occurrence}} = \sum_{c \in \mathcal{C}^2} \sum_{c_q \in N_{R(c)}} -\log\left(\frac{\exp\left(\boldsymbol{c}^{\mathsf{T}}\boldsymbol{c}'_{\mathrm{q}}\right)}{\sum_{c_n \in \mathcal{C}^2}\exp\left(\boldsymbol{c}^{\mathsf{T}}\boldsymbol{c}'_{n}\right)}\right)$$

$$\approx \sum_{c \in \mathcal{C}^2} \sum_{c_q \in N_{R(c)}} -\left(\log\left(\sigma\left(\boldsymbol{c}^{\mathsf{T}}\boldsymbol{c}'_{\mathrm{q}}\right)\right) - \sum_{i=1}^{k}\log\left(\sigma\left(\boldsymbol{c}^{\mathsf{T}}\boldsymbol{c}'_{n_i}\right)\right)\right) \tag{5}$$

where $\boldsymbol{c}$ denotes the vector embedding of the second-level category c, and $N_{R(c)}$ represents the set of context categories of c captured in the random walks. $\boldsymbol{c}$ denotes the *target* embedding, while $\boldsymbol{c}'$ denotes the *context* embedding of category c. σ denotes the sigmoid function. The first line is the original form of skip-gram's objective function, which is computationally expensive; thus, we use its approximation in the second line, in which c_{n_i} means the categories obtained by the negative sampling process (i.e. c_{n_i} does not co-occur with c).

3.2. Incorporating categorical semantics into POI embeddings

The POI (category) embeddings in previous studies generally encode only their spatial co-occurrence patterns, overlooking the intrinsic semantic relations between them. As we intend to learn the embeddings for second-level POI categories, there is abundant semantic information explicitly defined in the POI category hierarchy. For example, in Baidu POIs, the categories *supermarket*, *mall* and *grocery store* all belong to a first-level category *shopping*, implicating their substantial resemblance in functional affordance (categorical semantics). Such resemblance can only be marginally captured using spatial co-occurrence information under the assumption that semantically similar categories have similar spatial distribution patterns.

In this article, we impose a semantic smoothness assumption on the embeddings of POI categories: if two second-level POI categories c_i and c_j belong to the same first-level category, they should have embeddings $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ close to each other in space $\mathbb{R}^s$. This assumption forces the second-level categories that share the same first-level category to be adjacent in the embedding space, which is akin to the local invariance assumption utilized in manifold learning theory (Guo *et al*. 2015). Therefore, we propose to realize this assumption with a manifold learning algorithm, which is capable of enforcing the learning model to be smooth in terms of the geometric structure of the data (Belkin *et al*. 2006).

Specifically, we leverage the manifold learning algorithm LE, which preserves the local invariance between each data point pair (Belkin and Niyogi 2001). With LE, the semantic smoothness assumption for POI category embeddings can be realized by minimizing the objective function:

$$\mathcal{L}_{\text{categorical semantics}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \| \boldsymbol{c}_i - \boldsymbol{c}_j \|_2^2 \, w_{ij} \tag{6}$$

where $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ are the embeddings of the second-level POI categories c_i and c_j, respectively; w_{ij} is used to measure the semantic smoothness of the embedding space, for which $w_{ij} = 1$ if c_i and c_j belong to the same first-level category, and otherwise $w_{ij} = 0$, and $\| \cdot \|_2^2$ denotes the operation of taking the squared $\mathcal{Q}_2$ norm of the encapsulated vector.

Thus far, we have proposed two objective functions for optimizing POI category embeddings, which encode spatial co-occurrence information (Equation (5)) and categorical semantics (Equation (6)). Our proposed POI encoder ϕ then combines them to foster embeddings entailing information from both perspectives with the overall objective function:

$$\mathcal{L}_{\phi} = \mathcal{L}_{\text{co-occurrence}} + \lambda \mathcal{L}_{\text{categorical semantics}} \tag{7}$$

where λ is a hyperparameter that should be tuned during training to balance the two perspectives. The optimization of the overall objective function $\mathcal{L}$ can be carried out with stochastic gradient descent.

3.3. Generating regional embeddings through aggregating POI embeddings

With the POI encoder ϕ the embedding for each POI (category) is obtained. As we aim to estimate functional distributions of urban regions, the embedding for each urban region should be generated. Specifically, each urban region $r_i \in \mathcal{R}$ is often a traffic analysis zone (TAZ; Yao *et al*. 2017, Zhang *et al*. 2021) or a grid cell (Barlacchi *et al*. 2021); in this article, we opt to use TAZ, although our approach applies to both of the means of region partition.

For each region, usually tens or hundreds of POIs spatially reside in it, and the set of corresponding POI embeddings is fundamentally a region embedding matrix if one stacks them together. As the number of POIs in each region varies, we need an aggregation function to generate an embedding $\boldsymbol{r}_i$ for each region. Such an aggregation function should be *permutation invariant*, which means that the embedding $\boldsymbol{r}_\mathbf{i}$ needs to remain the same regardless of the order of the POIs fed into the function (Zaheer *et al*. 2017). To this end, previous studies generally applied (element-wise) average pooling to generate an region embeddings (e.g. Yao *et al*. 2017, Zhai *et al*. 2019). Nevertheless, using average pooling could let some frequently arisen POI categories (e.g. *restaurant* and *grocery store*) water down the definitive information (e.g. *railway station* and *park*) in the regional embedding, thereby compromising the performance for the task.

Intuitively, there is an intrinsic importance order of the POIs in a region to define its functions. For example, given the POIs that reside in a region {*restaurant_1*, *railway station*, *restaurant_2*, *grocery store*, *restaurant_3*}, its importance order is largely apparent to humans, while unknown *a priori* to machines. The work by Vinyals *et al*. (2015) sheds light on this problem, in which they argued that there is generally an optimal

hidden ordering for a set of entities, e.g. for the task of number sorting, and they proposed an aggregation function coupling LSTM and attention mechanisms to discover such an optimal order. This function is *permutation invariant* and has exceeded average pooling in several applications, e.g. in quantum chemistry (Gilmer *et al*. 2017). In this article, we employ this aggregation function Γ to aggregate POI embeddings. The function can be formally defined as:

$$\boldsymbol{q}_t = LSTM(\boldsymbol{r}_{j,\ t-1}) \tag{8}$$

$$a_{i,t} = \frac{\exp\left(\boldsymbol{p}_i \boldsymbol{q}_t\right)}{\sum_j \exp\left(\boldsymbol{p}_j \boldsymbol{q}_t\right)} \tag{9}$$

$$\boldsymbol{v}_t = \sum_i a_{i,t} \boldsymbol{p}_i \tag{10}$$

$$\boldsymbol{r}_{j,t} = [\boldsymbol{q}_t\ \boldsymbol{v}_t] \tag{11}$$

where $\boldsymbol{p}_i$ is the embedding of the POI p_i in the region r_j (note that the POI embedding is actually the embedding of its second-level category, i.e. $\boldsymbol{p}_i = \boldsymbol{c}_i^2$); $\boldsymbol{q}_t$ is a query vector used to compute the POI weights in the attention mechanism; *LSTM* is an LSTM network that computes a recurrent state with a randomly initialized input in the first step; t is the processing steps of the attention computation; $a_{i,t}$ can be viewed as the weight of the POI p_i at step t; the region embedding $\boldsymbol{r}_{j,t} \in \mathbb{R}^{2s}$ is finally generated by concatenating the query vector $\boldsymbol{q}_t$ and the POI weighted summation ($\boldsymbol{v}_t$) at step t.

The intuition behind the aggregation function Γ is that the query vector $\boldsymbol{q}_t$ can be understood as the regional embedding in its infancy, and it serves as a 'benchmark' to measure the importance levels of the POIs in a region; the 'benchmark' improves to be increasingly expressive and mature during the training of the LSTM, and the regional embedding is finally obtained through combining the 'benchmark' itself and the weighted summation with regard to the 'benchmark'. The generated region embedding $\boldsymbol{r}_{j,t}$ would implicitly account for the different contribution levels of the POIs in defining the region's functional distribution.

3.4. Mapping regional embeddings to functional distributions

With the aggregation function Γ, each region is represented as a vector embedding $\boldsymbol{r}_{j,t}$ The last component of our proposed approach utilizes an MLP ψ to map each regional embedding to its functional distribution, i.e. $\psi(r): \mathbb{R}^{2s} \rightarrow \mathbb{R}^m$, where $2s$ is the dimension of the region embeddings, and m is the number of function types. Specifically, the MLP takes a region embedding $\boldsymbol{r}_{j,t}$ as input, and outputs an m dimensional distribution in which each dimension represents a function type $\mathrm{f}_k \in \mathcal{F}$. Between the input and output layers, there can be one or several hidden layers with nonlinear activation functions. The activation function used for the output layer is a *softmax* function to guarantee that the elements (proportions) sum to 1, i.e. $\sum_k \widehat{y}_i^{f_k} = 1$. The aggregation function Γ and the MLP ψ can be jointly trained by minimizing the Kullback – Leibler divergence (KL divergence) objective function:

$$\mathcal{L}_{\Gamma,\ \psi} = \sum_{i=1}^{n} \sum_{k=1}^{m} \widehat{y}_i^{f_k} \log\left(\frac{\widehat{y}_i^{f_k}}{y_i^{f_k}}\right) \tag{12}$$

where $\widehat{y}_i^{f_k}$ is the estimated proportion of the function type f_k that region i bears, and $y_i^{f_k}$ is the corresponding ground truth proportion. By minimizing this objective function, the aggregation function Γ and MLP ψ can be trained to learn the correlations between the functional distribution of a region and the POIs that it contains, and produce the estimated functional distribution in a supervised manner.

4. Experiment and results

4.1. Study area and data

We demonstrate the effectiveness of the proposed approach in the study area of Xiamen Island, which is the central part of the southeast coastal city of Xiamen, China. Xiamen Island has a population of more than 2 million, and its area is approximately 136.3 km^2. Xiamen Island is widely known as an economically prosperous yet extremely land-scarce area. Thus, sensible urban planning has been a priority in the management of the city, in which obtaining updated and accurate urban functional distributions is pivotal (Song *et al*. 2018). In this study, we utilize three datasets from Xiamen Island:

1. A POI dataset from Baidu Map harvested in June 2020, which contains 45,033 POIs belonging to 22 first-level categories and 184 second-level categories; the first-level categories include, among others, *life service*, *shopping*, *food service*, *medical service*, *governmental agency*, and so on.; several second-level categories can be subsumed under one first-level category.
2. An urban region partition (TAZ) dataset containing 661 regions. The regions are divided by the network of the major roads in the study area, and they are the

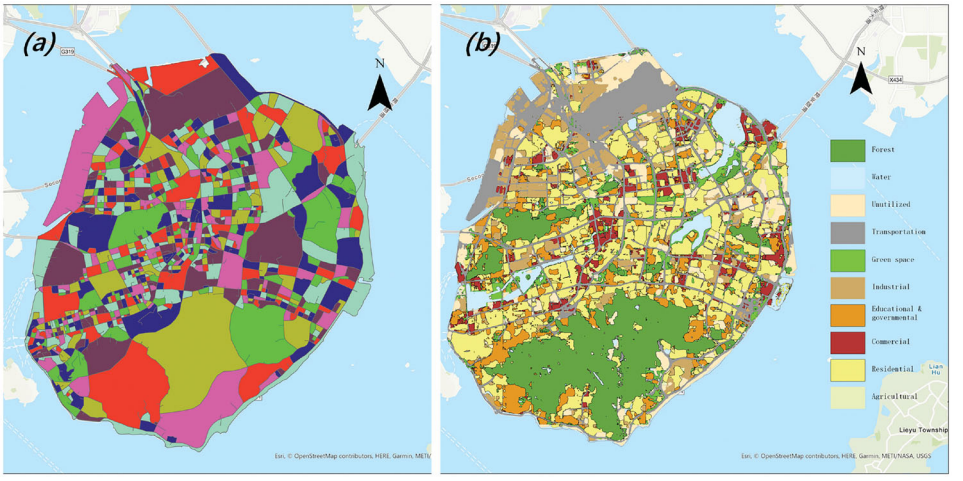

Figure 3. The study area of Xiamen Island. (a) The urban region partition (colors are merely used to differentiate regions, and have no connection with region functions). (b) The spatial distribution of urban functions in the study area. The ground truth proportional distributions of urban functions are derived through spatially overlapping (a) and (b).

basic units of urban structure and land use (Liu and Long 2016); the urban regions are demonstrated in Figure 3(a).

3. An urban function classification dataset from the *Urbanscape Essential Dataset of Peking University* as shown in Figure 3(b), which is produced through information extraction from remote sensing data, POI data and extensive human correction, and modification in 2019 (Zhang *et al*. 2017, 2018, Du *et al*. 2019, 2020); this dataset provides detailed spatial distributions of 10 different urban functions (functional land use): (1) *forest*, (2) *water*, (3) *unutilized*, (4) *transportation*, (5) *green space*, (6) *industrial*, (7) *educational and governmental*, (8) *commercial*, (9) *residential* and (10) *agricultural*. This dataset is spatially overlapped with the urban region partition dataset, to obtain the proportional functional distribution of each region, namely the area proportion of each functional use in the region. For example, after the overlapping operation, a region r_i is assigned with its proportional function distribution y_i (a 10-dimentioal vector distribution), e.g. 0.5 for *commercial*, 0.2 for *education and governmental*, 0.3 for *residential* and 0 for all other function types. The derived proportional distributions are then used as ground truth data.

4.2. Generating POI embeddings

4.2.1. Implementation details

We train the POI encoder ϕ to generate embeddings for POI (second-level) categories. First, we build a DT network based on POIs using Scipy.[2] The constructed POI DT network contains 45,033 nodes and 270,123 edges. During the random walk, five walks with a length of 10 are conducted starting from each node. In total, 225,165 walks are conducted, and each walk yields nine co-occurrence pairs of $\{$*target category*, *context category*$\}$. Negative sampling is performed to generate five negative pairs for each co-occurrence pair.

With the POI co-occurrences collected, we can then train the POI encoder $\mathcal{L}_\phi$ with the objective function $\mathcal{L}_\phi$ (Equation (7)) using the co-occurrence information and the POI category hierarchy. To this end, we utilize the Adam optimizer built in Pytorch[3] with an embedding size of 64 in view of previous practices (e.g. Yan *et al*. 2017). As the number of co-occurrence pairs is enormous, we perform training in minibatch mode with a batch size of 128 for 100 epochs. In this context, the hyperparameter λ is pivotal in generating POI category embeddings, which is used to balance the spatial co-occurrence information and categorical semantics, and it can be expressed as

$$\lambda = bn \times \lambda' \tag{13}$$

where *bn* denotes the number of batches in a single training epoch. As *bn* is a constant and the absolute value of the LE item in $\mathcal{L}_\phi$ is large, we tune the hyperparameter λ' in $\{10^{-7}, 10^{-8}, 10^{-9}, 10^{-10}\}$.

4.2.2. Analysis

We generate POI category embeddings while tuning the hyperparameter λ', and we also test the scenario where the embeddings are completely learned with spatial co-

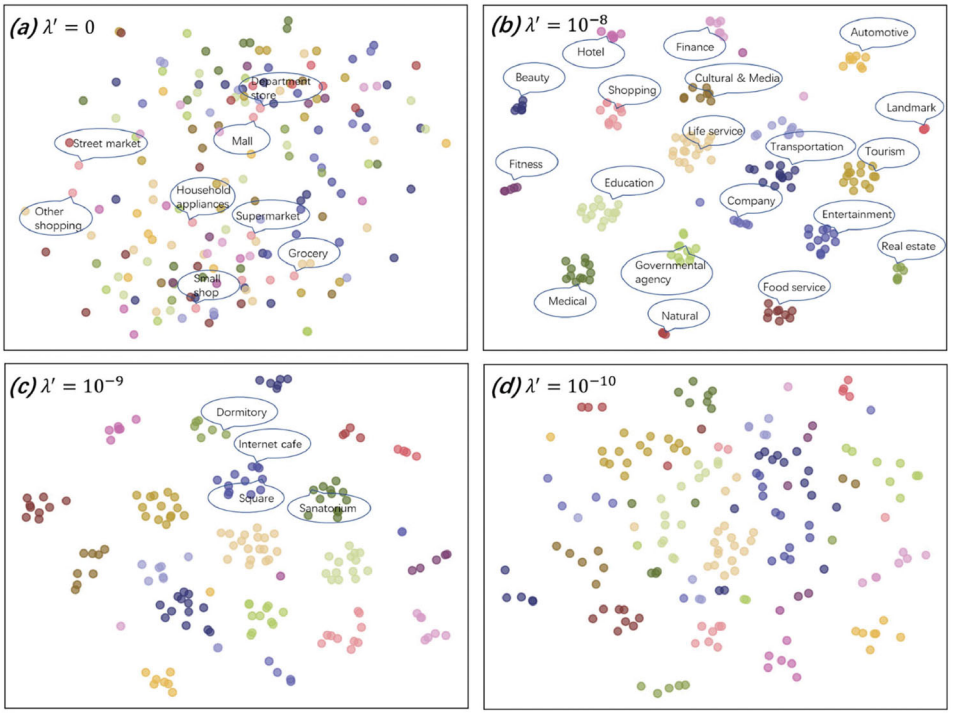

Figure 4. Exhibition of POI category embeddings in 2D planes. The second-level categories belonging to the same first-level category are colored identically. (a)—(d) represent several scenarios with different enforcement strength of categorical semantics, in which (a) corresponds to no enforcement. The annotations in (a) and (c) are second-level categories, while (b) shows first-level categories (clusters).

occurrence information ($\lambda' = 0$). Upon completion of the training, we project the obtained POI category embeddings into 2D planes using the t-SNE algorithm (Van der Maaten and Hinton 2008) for visualizations in Figure 4, where each point represents a second-level category, and those belonging to the same first-level category are colored identically. The rationale of using t-SNE is that it has great capacity in maintaining the relations between high-dimensional points after casting them to a 2D plane.

Figure 4(a) shows the scenario in which the embeddings for second-level categories are trained completely with spatial co-occurrence information ($\lambda' = 0$). For the annotated second-level categories (colored pink) belonging to the first-level category *shopping*, we can observe that they generally scatter around the plane and are mixed with other categories, and no particular pattern arises. We recognize that semantic similarities can only be marginally captured; for example, *mall* and *department store* are similar in their functional affordance, and are also adjacent in the embedding space. Nevertheless, some second-level categories, such as *grocery*, *street market* and *supermarket*, also have substantial semantic resemblance, which can seldom be reflected here.

The embeddings in Figure 4(b) are generated by considering both spatial co-occurrence information and categorical semantics ($\lambda' = 10^{-8}$), in which the embeddings are

semantically preserved. The second-level categories belonging to the same first-level category exhibit an evident clustering phenomenon (each cluster is annotated with its first-level category). Meanwhile, the distribution of the first-level category clusters expresses the spatial co-occurrence patterns. For example, cluster *shopping* is adjacent to *education*, *life service* and *hotel*; cluster *tourism* lies closely to *landmark*, and *transportation*. The two examples both concur with human perception.

In Figure 4(c,d), the embeddings are generated by progressively decreasing the values of λ', thereby weakening the enforcement of categorical semantics. We can observe that with smaller values of λ', the first-level category clustering phenomenon is mitigated. With $\lambda' = 10^{-10}$ the visual boundaries between the clusters have almost vanished, but practically, we can still identify the mitigated clusters. It can also be observed that the spatial co-occurrence patterns for second-level categories can still be expressed after the preservation of categorical semantics. For example, in Figure 4(c), the second-level categories *Internet café* and *square* are both in the first-level category cluster of *life service*, in which the former lies close to the *dormitory* (second-level) in the *real-estate* (first-level) cluster, and the latter is close to the *sanatorium* in the *medical* cluster, which conforms with human experiences.

Based on the above observations, we believe that learning POI embeddings merely using spatial co-occurrence information cannot suffice in capturing the underlying semantic similarities between POI categories, and our approach could resolve this problem by considering both spatial co-occurrence information and categorical semantics.

4.3. Estimating urban functional distributions

4.3.1. Baseline models

We compare our proposed approach with several baseline models, including:

1. Place2Vec (Yan *et al*. 2017): This approach considers spatial co-occurrence information with a KNN sampling strategy and distance decay. In fact, the full version of Place2Vec also incorporates POI popularity, but we ignore such information which is unavailable in our POI dataset. In this article, we utilize the same aggregation function and the MLP architecture in the proposed approach to produce estimated region functional distributions for a fair comparison.
2. One-hot: This approach considers only categorical semantics. Each POI's embedding is the concatenation of the one-hot vectors of its first- and second-level categories (206-dimensional). Subsequently, the aggregation function and MLP architecture remain the same.
3. Random guess (random): Unlike classification problems, it is difficult to gauge the difficulty level of the target problem. Therefore, we provide the results on random guess, i.e. randomly guessing a uniform functional distribution for each region. This shapes the lower bound in the evaluation.

4.3.2. Evaluation measures

Estimating urban functional distributions is essentially a label distribution learning problem. There are generally two types of evaluation measures in this regard: distance

measures and similarity measures. Cha (2007) analyzed 41 measures for this problem, and Geng (2016) selected several representative ones among them, in which each one could reflect a certain perspective of an algorithm. We partially follow these previous works, and pick five measures to obtain a comprehensive understanding of the performance of our approach (↓denotes the smaller the better, and ↑indicates the opposite):

1. L1 distance (L1) ↓: $\sum_{k=1}^{m}\left|\widehat{y_i}^{f_k}-y_i^{f_k}\right|$
2. Canberra distance (Canberra)↓: $\sum_{k=1}^{m}\left|\widehat{y_i}^{f_k}-y_i^{f_k}\right|/\left(\widehat{y_i}^{f_k}+y_i^{f_k}\right)$
3. KL divergence (KL)↓: $\sum_{k=1}^{m}\widehat{y_i}^{f_k}\log\left(\widehat{y_i}^{f_k}/y_i^{f_k}\right)$
4. Chebyshev distance (Chebyshev)↓: $\max_k\left|\widehat{y_i}^{f_k}-y_i^{f_k}\right|$
5. Cosine similarity (Cosine)↑: $\frac{\left(\sum_{k=1}^{m}\widehat{y_i}^{f_k}y_i^{f_k}\right)}{\left(\sqrt{\sum_{k=1}^{m}\widehat{y_i}^{f_k}{}^2}\sqrt{\sum_{k=1}^{m}y_i^{f_k}2}\right)}$

where $\widehat{y_i}^{f_k}$ is the estimated proportion of the function type f_k that region i bears, and $y_i^{f_k}$ is the corresponding ground truth proportion.

4.3.3. Implementation details

The embeddings of the POIs in each region are stacked to compose a region embedding matrix for each region. Each region embedding matrix is then mapped to a functional distribution of the region through the aggregation function Γ and the MLP ψ, which are trained using the objective function $\mathcal{L}_{\Gamma,\ \psi}$ (Equation (12)). We set the processing step number t to 5 in the aggregation function Γ, and set the MLP to have one hidden layer with a size of 64 and the *tanh* activation function. These processes are implemented using Pytorch.

To train the models in the proposed approach, we have 661 regions that can serve as input data, each of which has a ground truth distribution. We split the dataset into a training set (80%) and a test set (20%). Within the training set, 20% is used as the validation set. The models are first trained on those instances left in the training set and tested on the validation set to select the best parameters. Then, the models are trained on the entire training set and tested on the test set. The training is also performed in minibatch mode with a batch size of 64 for 100 epochs. For reliability, the entire dataset (661 regions) are randomly shuffled 10 times to repeat the abovementioned training, validation and testing processes.

Table 1. Performance of our approach and baseline models.

	Evaluation measures					
Approach	L1↓	Canberra↓	KL↓	Chebyshev↓	Cosine↑	Avg. rank
Ours	**0.696 ± 0.024**	**7.467 ± 0.106**	**0.058 ± 0.002**	**0.290 ± 0.012**	**0.808 ± 0.013**	**1.0**
Place2Vec	0.784 ± 0.034	7.537 ± 0.096	0.069 ± 0.003	0.328 ± 0.009	0.764 ± 0.014	3.0
One-hot	0.729 ± 0.023	7.473 ± 0.108	0.063 ± 0.003	0.302 ± 0.009	0.786 ± 0.015	2.0
Random	1.412	8.002	0.146	0.555	0.458	4.0

The best value with regard to each evaluation measure is presented in bold.

4.3.4. Performance

The performances of our approach and the baseline models are presented in Table 1, where the results are represented by 'mean ± standard deviation'. We can observe that our proposed approach prevails in all evaluation measures. The L1 distance indicates that the average absolute value of the estimation error for each function is approximately 0.0696 (as we have 10 function types). The Canberra distance is sensitive to small changes near zero, and thus its values indicate that estimating small functional proportions is generally challenging. The KL divergence performance indicates that the relative entropy between the estimated functional distributions and the real distributions is the smallest with our approach. The Chebyshev distance only cares about the worst match between the estimated and real functional distributions, and the results reveal that the average largest discrepancy for a single function is less than 0.3. In terms of the cosine similarity, the result of our approach is larger than 0.8, indicating a rather high level of consistency between the estimated and real distributions.

With regard to the baselines, Place2Vec (simplified) only encodes spatial co-occurrence information, whereas one-hot only embodies categorical semantics. Surprisingly, we observe that one-hot outperforms Place2Vec in all evaluation measures, indicating that categorical semantics, which was usually neglected in previous studies, is more informative and relevant than spatial co-occurrence information for estimating urban functional distributions. Our approach incorporates both perspectives and thus outperforms the baselines. In addition, the performance of random guess indicates that this task is generally difficult, and the results of our approach are significant.

4.3.5. Ablation study

As our proposed approach comprises four major components, we perform an ablation study to verify the necessity of each component. The results of the ablation study are presented in Table 2. The first row is the performance of the proposed approach with all the components. In the second row, we drop the semantic smoothing technique in the process of learning POI embeddings, thereby considering only spatial co-occurrence information, and the performance declines. Nevertheless, the performance in such a setting is better than Place2Vec (see Table 1), which implies that the network representation learning method outperforms Place2Vec from the perspective of capturing and expressing spatial co-occurrence information. In the third row, we replace the aggregation function with the average pooling method, which degrades the performance. Finally, we replace the distribution learning component (mapping regional embeddings to functional distributions) with a support vector machine (SVM) and random forest. In these settings, we cannot train the aggregation function coupling LSTM and attention mechanisms; therefore, we use the average pooling method. It turns out that SVM and random forest produce unsatisfactory performance, and induce serious performance decline. In the ablation study, we observe that the performance is ranked as follows: our approach > replacing aggregation function > dropping categorical semantics > replacing MLP with random forest or SVM. The results clearly indicate that all the components in our approach are necessary, and the incorporation of categorical semantics in POI embeddings seems to be a pivotal factor in underpinning the superiority of our approach.

Table 2. Results of the ablation study.

POI embedding				Evaluation measures					
Spatial Co-occurrence	Categorical semantics	aggregation	Distribution learning	L1↓	Canberra↓	KL↓	Chebyshev↓	Cosine↑	Avg. rank
Network	LE	LSTM + attention	MLP	**0.696 ± 0.024**	**7.467 ± 0.106**	**0.058 ± 0.002**	**0.290 ± 0.012**	**0.808 ± 0.013**	**1.0**
Network		LSTM + attention	MLP	0.763 ± 0.015	7.558 ± 0.115	0.066 ± 0.002	0.318 ± 0.007	0.775 ± 0.007	3.0
Network	LE	Average pooling	MLP	0.747 ± 0.032	7.470 ± 0.097	0.061 ± 0.004	0.311 ± 0.014	0.794 ± 0.020	2.0
Network	LE	Average pooling	SVM	1.159 ± 0.003	7.820 ± 0.003	0.100 ± 0.0004	0.484 ± 0.001	0.641 ± 0.002	5.0
Network	LE	Average pooling	Random Forest	0.873 ± 0.006	7.500 ± 0.015	0.070 ± 0.003	0.369 ± 0.004	0.769 ± 0.004	4.0

The components that are dropped or replaced are shaded, and the best value with regard to each evaluation measure is presented in bold.

4.3.6. Parameter sensitivity analysis

As revealed in the ablation study, the incorporation of categorical semantics in POI embeddings is pivotal, and the hyperparameter λ' controls the strength of the enforcement of categorical semantics in the process of learning POI embeddings. Therefore, we tune λ' to find the best balance point. The results of the parameter sensitivity analysis are presented in Table 3. Specifically, we find that the value of 10^{-8} leads to the best performance, and the performance decreases as λ' moves away from 10^{-8}. Such results imply that a relatively strong enforcement of categorical semantics is necessary for estimating urban functional distributions (cf. Figure 4), and a subtle balance point exists for our certain task (but we speculate that such a balance point would shift in other tasks).

In addition, we also search the parameters α_b^{loc}, $\alpha_b^{lr} \in \{0.25, 0.50, 1, 2, 4\}$, and $\alpha_r^{inter-region} \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$, which are the transition biases in random walks. We find out that the parameters yield the optimal performance are $\alpha_b^{loc} = 2$, $\alpha_b^{lr} = 2$, and $\alpha_r^{\text{inter}-\text{region}} = 0.4$, which indicates that both local and long-range spatial co-occurrence information play important roles, and there is a subtle balance between them. However, we observe that varying these parameters only leads to small shifts of the final performance, which strengthens our argument that categorical semantics plays a more important role.

4.3.7. Error analysis

We perform a thorough investigation and analysis of estimation errors against ground truth data through visualizations and manual inspections. In Figure 5, the results of the error analysis are presented. In Figure 5(a,b), the L1 distance and cosine similarity between the real and estimated functional distributions are visualized. We then dig into the regions where large errors arise, and select four representative regions to illustrate the underlying reasons that lead to large discrepancies; see Figure 5(c–j); in Figure 5(g–j), the solid blue lines represent the ground truth, and the orange dashed lines are the estimated functional distributions.

The presence of few predominating POIs. In Figure 5(c), a region is shown where a lake (*water*: 0.59) dominates, and around the lake there are also some *green space*, *commercial* and *residential* areas, and so on. However, our approach falls short in sensing such a large proportion of *water*, and only comes to an estimation of 0.10 for *water*. Instead, the estimated functional distribution is relatively flat, with large proportions of *residential*, *transportation* and *water*. In this region, there are 349 POIs, mainly with first-level categories, such as *life service*, *food service*, *real estate* and

Table 3. Performances of our approach with different values for the parameter λ'.

Parameter	Evaluation measures					
λ'	L1↓	Canberra↓	KL↓	Chebyshev↓	Cosine↑	Avg. rank
10^{-7}	0.729 ± 0.023	7.510 ± 0.065	0.062 ± 0.004	0.304 ± 0.011	0.790 ± 0.016	3.2
10^{-8}	**0.696 ± 0.024**	7.467 ± 0.106	**0.058 ± 0.002**	**0.290 ± 0.012**	**0.808 ± 0.013**	**1.2**
10^{-9}	0.726 ± 0.023	**7.456 ± 0.058**	0.060 ± 0.002	0.302 ± 0.012	0.794 ± 0.011	1.8
10^{-10}	0.742 ± 0.035	7.551 ± 0.101	0.061 ± 0.005	0.309 ± 0.016	0.790 ± 0.022	3.8

The best value with regard to each evaluation measure is presented in bold.

transportation. Only eight POIs fall under the first-level category *tourism*, e.g. *park* and *scenic*. The large discrepancy can be ascribed to the low presence of the POIs related to the function of *water*, even with the employed importance-aware aggregation function. Nevertheless, we believe that our approach is somehow competent in view of its estimation for *water* (0.1), given that there is nearly no POI that is directly related to *water*, e.g. *park* is only implicitly linked to *water*.

Ambiguous linkage between POIs and functions. The region shown in Figure 5(d) is a lakeside industrial area, and thus majorly contains the functions of *industrial*, *green space* and *water*. However, our approach assigns the highest proportion to *transportation*, and less to *industrial* and *commercial*. Through inspection, we found that there are many *transportation*-related POIs in this area, e.g. *car wash*, *parking lot*, etc. There are also many POIs that fall in *company*, and we speculate that our approach has limited capability to discriminate whether such POIs indicate *industrial* or *commercial*, and in the end certain proportions are assigned to both of the functions.

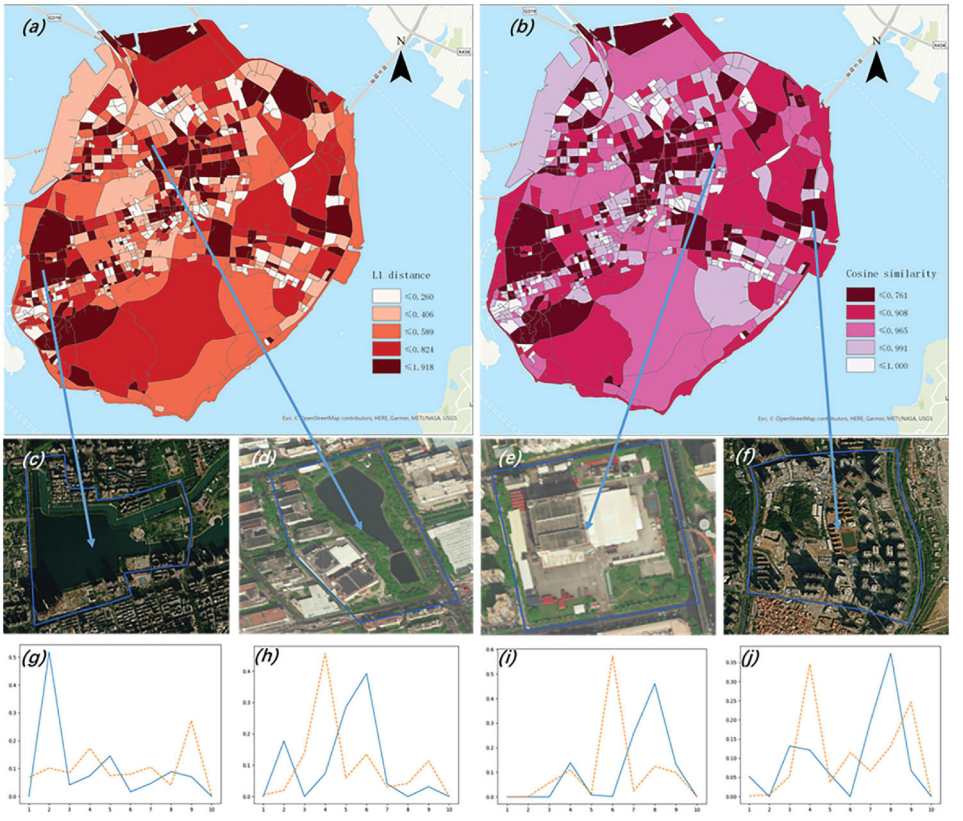

Figure 5. Error analysis. (a) is the visualization for L1 distance, while (b) is for cosine similarity; (c)–(f) are the visualizations of the selected representative regions; in (g)–(j), solid blue lines represent the ground truth-functional distributions of the selected regions, while the orange dashed lines are the estimated functional distributions.

Error in ground truth. The region shown in Figure 5(e) is an industrial park. In fact, our approach performs well in this case, as it assigns the highest proportion to *industrial*; this discrepancy is induced by the error in the ground truth data. It is common that the ground truth data are not completely correct, and we believe this case strengthens the superiority of our approach, as it has a strong tolerance to errors in ground truth data, which also demonstrates the utility of our approach in discovering such problems.

Diversity and granularity of POIs. The region shown in Figure 5(f) is a busy region and contains 593 POIs covering a diversity of categories, e.g. *governmental agency*, *transportation*, *automotive*, *medical*, *finance*, *education*, etc. Such many and diverse POIs, we suspect, lead to the difficulties for our model to gauge what functions this region really has, and thus some relatively close proportions are assigned to several functions, e.g. *residential*, *commercial* and *transportation*. This discrepancy is also linked to the granularity difference between POIs and ground truth data. Yue *et al*. (2017) argued that POIs represent a finer-grained picture than traditional land use maps produced from remote sensing and surveying. This implies that the discrepancies are not necessarily due to the limited expressiveness of our model, but can be attributed to the granularity difference between the POIs and the ground truth.

5. Discussion

Through a thorough experiment, we find that POIs are indeed a competent proxy for sensing urban functions. The results clearly demonstrate that the potential and rich information of POIs are incompletely mined in previous studies. In particular, the incorporation of categorical semantics leads to substantial performance gains. We observe that simply encoding POI categorical semantics (one-hot) with nearly no computation cost could already excel encoding only spatial co-occurrence information. The power of categorical semantics has also been revealed in studies, such as Liu *et al*. (2018) and Jin *et al*. (2019), where only categorical information was utilized in POI embeddings to search for similar urban regions. To this end, this work is the first attempt to encode both categorical semantics and spatial co-occurrence information in POI embeddings.

Spatial co-occurrence information of POIs still matters, as our approach that encodes information from both perspectives leads to the best performance. To this end, we essentially propose a new POI co-occurrence sampling strategy, i.e. the spatially explicit random walk in a network structure. In fact, previous studies mainly concentrated on designing various sampling strategies, e.g. Yao *et al*. (2017) and Zhai *et al*. (2019). In the experiment, we demonstrate that the proposed sampling strategy in a network structure outperforms the others. We believe that the rationale behind this is that our approach could capture both local and long-range spatial co-occurrence patterns, i.e. several ten-length walks see further than a ten-neighbor KNN. To this end, one might argue that we could increase the searching radius in KNN to peep at long-range dependencies. However, such a strategy would capture a plethora of co-occurrence information, which compromises the model's efficiency, and the capability to

discriminate POI categories (likely each category could co-occur with all other categories).

It has also been clearly shown that the aggregation function utilized in our approach with LSTM and attention mechanisms lift the performance compared to average pooling. The aggregation function underlays a more expressive model with inevitably more parameters that have to be trained (mainly those in LSTM). The aggregation function has a certain capacity to gauge the different importance levels of the POIs in a region, but not perfectly. For example, in Figure 5(c), although our model could sense that the POI with the second-level category *park* might be indicative of the region, while such a large proportion of *water* (0.59) is still difficult to estimate. We believe that this should be ascribed to the intrinsic limitation of POI data that models each entity as a point, and we speculate that this problem could be further alleviated by incorporating complex geometries, e.g. from OpenStreetMap.

To the best of our knowledge, this work is a first attempt to estimate urban functional distributions using POIs in a supervised manner with full utilization of the available ground truth data (such ground truth data can be partially available in a city, in which case full utilization is even more pivotal). To this end, the problem formulation and evaluation measures provided in this work could form a solid basis for further studies. We also believe that estimating proportional distributions is more meaningful than classification from an application perspective, as the users of the results (e.g. urban planners) would have a comprehensive understanding of the naturally mixed functions embodied in each region. In addition, if the ground truth is completely unavailable in a city, one could also use our approach in a fully unsupervised setting, but in such a case the aggregation function should be replaced with an average pooling. Subsequently, the region embeddings can be fed into certain clustering methods to discover the functional structure in a city.

In our experiment, we have once again verified the power of POIs as a proxy for urban functions, and we demonstrate a real case where POI data have great potential to rectify erroneous ground truth data and sense the changes of region functions that are yet to be updated. In addition, several limitations of POI data have been unveiled. Apart from the limitation of modeling all entities as points, we also find that the expressiveness of POIs is limited, e.g. whether a *company* POI indicates *industrial* or *commercial* is ambiguous. The problem could be mitigated with multiple POIs in a region, but in many such cases, expressiveness still remains an issue. We believe that estimating distributions with fewer functions could be a cure at the cost of reducing the granularity of the functional distributions, which should be weighed depending on the applications.

6. Conclusions and outlook

In this article, we present a framework for estimating the functional distributions of urban regions (proportions of urban function types in each urban region) based on POIs. In this framework, each POI is represented as a low-dimensional vector embedding that embodies the information of spatial co-occurrence and categorical semantics. The embeddings of the POIs are then aggregated to generate region embedding using an aggregation function coupling LSTM and attention mechanisms, which is

aware of the different importance levels of the POIs in a region. Finally, the regional embeddings are mapped to functional distributions using an MLP. A comprehensive experiment and thorough result analyses are performed in the study area of Xiamen Island, China. The results reveal that the proposed approach substantially outperforms the baseline models in all evaluation measures.

In the error analysis, we exhibit the reasons behind estimation errors and find that POIs are a competent proxy for urban functions, which can sometimes help rectify erroneous ground truth data, and provide a picture of urban functional distributions at a finer granularity than traditional means of remote sensing data and land surveying. At the same time, this study also reveals several intrinsic limitations of POIs for this task, such as (1) all entities are modeled as points, which makes it difficult to sense the large area functions with only few POIs (e.g. a lake) and (2) the linkages between POI categories and urban function types are sometimes ambiguous, e.g. it is unclear whether the POI category *company* implies *industrial* or *commercial*.

Future studies can be conducted in three directions. The first is to further improve the approach for estimating urban functional distributions by considering the uniqueness of individual POIs, as thus far all the POIs belonging to the same categories have the same embeddings, which leads to the loss of the uniqueness of each POI. Second, POIs can be integrated with a data source with complex geometries (e.g. OpenStreetMap) to improve the aggregation mechanism. Third, the proposed approach can be adapted and applied in other downstream tasks to explore its fitness and superiority, such as in POI recommendation, housing price estimation and geographic risk analysis in the insurance industry.

Notes

1. See https://map.baidu.com/.
2. See https://www.scipy.org/.
3. See https://pytorch.org/.

Acknowledgements

We would like to acknowledge the comments and insights from the editors and anonymous reviewers that helped lift the quality of the article, and also the input from Prof. Shihong Du and Dr. Jinchao Song at Peking University.

Disclosure statement

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

Data and codes availability statement

The data and codes that support this work are available in GitHub at https://github.com/RightBank/Semantics-preserved-POI-embedding. The repository contains mocked ground truth data, as the real ground truth data can only be requested at http://geoscape.pku.edu.cn/en.html for copyright reasons.

Funding

This work was supported in part by the National Natural Science Foundation of China (No. 42101421, 91846205, 61906107, and 41801306); National Key R&D Program of China (No. 2021YFF0900800, and 2019YFB2102903); the SDNSFC (No. ZR2021QD007, and ZR2019LZH008); Shandong Provincial Key R&D Program (Major Scientific and Technological Innovation Project) (No. 2021CXGC010108); the China Postdoctoral Science Foundation (No. 2020M682160); and the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources.

Notes on contributors

Weiming Huang obtained his PhD in Geographical Information Science at Lund University, Sweden in 2020, and was a visiting researcher at the Center for Spatial Studies, University of California, Santa Barbara. He is the recipient of the EuroSDR best PhD thesis award in Geoinformation in 2021. His research interests include spatial data mining and management, and knowledge graphs.

Lizhen Cui received his PhD MSc, and BSc degrees from Shandong University in 2005, 2002 and 1999 respectively. During 2013, he was a visiting scholar in Georgia Tech. He is a professor in the School of Software and C-FAIR at Shandong University, and also a visiting professor at Nanyang Technological University Singapore. He has published over 100 articles in journals and refereed conference proceedings. His research interests include big data management and analysis.

Meng Chen received his Ph.D. degree in computer science and technology in 2016 from Shandong University, China, and was a Postdoctoral fellow at the School of Information Technology, York University, Canada. He is currently an associate professor in the School of Software, Shandong University, China. His research interests are in trajectory data mining and traffic management.

Daokun Zhang is currently an Adjunct Lecturer with Department of Data Science and AI, Faculty of Information Technology, Monash University, Australia, and a Research Fellow with Monash Suzhou Research Institute, China. Prior to that, he was a Postdoc at The University of Sydney Business School, Australia, and received his Ph.D degree in data science from the University of Technology Sydney, Australia. His research interests include graph neural networks, knowledge graphs and weakly supervised learning.

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

ORCID

Weiming Huang http://orcid.org/0000-0002-3208-4208
Lizhen Cui http://orcid.org/0000-0002-8262-8883
Meng Chen http://orcid.org/0000-0002-6633-9205
Daokun Zhang http://orcid.org/0000-0002-1803-5768
Yao Yao http://orcid.org/0000-0002-2830-0377

References

Andrade, R., Alves, A., and Bento, C., 2020. POI mining for land use classification: a case study. *ISPRS International Journal of Geo-Information*, 9 (9), 493.

Barlacchi, G., Lepri, B., and Moschitti, A., 2021. Land use classification with point of interests and structural patterns. *IEEE Transactions on Knowledge and Data Engineering*, 33 (9), 3258–3269.

Belkin, M. and Niyogi, P., 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing System*, 14, 585–591.

Belkin, M., Niyogi, P., and Sindhwani, V., 2006. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7 (11), 2399–2434.

Burton, E., Jenks, M., and Williams, K., 2003. *The compact city: a sustainable urban form?* London: Routledge.

Calafiore, A., *et al.*, 2021. A geographic data science framework for the functional and contextual analysis of human dynamics within global cities. *Computers, Environment and Urban Systems*, 85, 101539.

Cha, S.H., 2007. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1 (2), 1.

Chen, Y., Xu, J., and Xu, M., 2015. Finding community structure in spatially constrained complex networks. *International Journal of Geographical Information Science*, 29 (6), 889–911.

Du, S., *et al.*, 2019. Context-enabled extraction of large-scale urban functional zones from very-high-resolution images: a multiscale segmentation approach. *Remote Sensing*, 11 (16), 1902.

Du, S., *et al.*, 2020. Large-scale urban functional zone mapping by integrating remote sensing images and open social data. *GIScience & Remote Sensing*, 57 (3), 411–430.

Du, J., *et al.*, 2019. Beyond geo-first law: learning spatial representations *via* integrated autocorrelations and complementarity. *IEEE International Conference on Data Mining (ICDM)*. New York: IEEE, 160–169.

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.

Geng, X., 2016. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28 (7), 1734–1748.

Gilmer, J., *et al.*, 2017. Neural message passing for quantum chemistry. *In: International conference on machine learning*. Sydney: PMLR, 1263–1272.

Grover, A. and Leskovec, J., 2016. node2vec: scalable feature learning for networks. *In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 855–864.

Guo, S., *et al.*, 2015. Semantically smooth knowledge graph embedding. *In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing*. The Association for Computer Linguistics, 84–94.

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

Janowicz, K., 2012. Observation-driven geo-ontology engineering. *Transactions in GIS*, 16 (3), 351–374.

Jiang, S., *et al.*, 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53, 36–46.

Jin, X., *et al.*, 2019. Learning region similarity over spatial knowledge graphs with hierarchical types and semantic relations. *In: Proceedings of the 28th ACM international conference on information and knowledge management*. ACM, 669–678.

Koster, H. R. and Rouwendal, J., 2012. The impact of mixed land use on residential property values. *Journal of Regional Science*, 52 (5), 733–761.

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.

Liu, K., *et al.*, 2020. Investigating urban metro stations as cognitive places in cities using points of interest. *Cities*, 97, 102561.

Liu, Y., Zhao, K., and Cong, G., 2018. Efficient similar region search with deep metric learning. *In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data Mining*. ACM, 1850–1859.

Mai, G., *et al.*, 2020. Multi-scale representation learning for spatial feature distributions using grid cells. In: *Eighth international conference on learning representations (ICLR 2020), 26–30 April Addis Ababa, Ethiopia*.

Mikolov, T., *et al.*, 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 2, 3111–3119.

Rodrigue, J. P., Comtois, C., and Slack, B., 2013. *The geography of transport systems*. London: Routledge.

Song, J., *et al.*, 2018. Mapping urban functional zones by integrating very high spatial resolution remote sensing imagery and points of interest: a case study of Xiamen, China. *Remote Sensing*, 10 (11), 1737.

Tian, L. and Shen, T., 2011. Evaluation of plan implementation in the transitional China: a case of Guangzhou city master plan. *Cities*, 28 (1), 11–27.

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.

United Nations. 2019. *World urbanization prospects 2018: highlights*. New York, NY: Department of Economic and Social Affairs, Population Division.

Van der Maaten, L. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9 (11), 2579–2605.

Vaswani, A., 2017. Attention is all you need. *The Thirty-first Annual Conference on Neural Information Processing Systems (NeurIPS 2017),* December 4-9, 2017 Log Beach, California, USA, 5998–6008.

Vinyals, O., Bengio, S., and Kudlur, M., 2015. Order matters: sequence to sequence for sets. In: *International conference on learning representations (ICLR 2016)*, 2–4 May San Juan, Puerto Rico.

Wu, L., *et al.*, 2020. A framework for mixed-use decomposition based on temporal activity signatures extracted from big geo-data. *International Journal of Digital Earth*, 13 (6), 708–726.

Yan, X., *et al.*, 2021. Graph convolutional autoencoder model for the shape coding and cognition of buildings in maps. *International Journal of Geographical Information Science*, 35 (3), 490–512.

Yan, X., *et al.*, 2019. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 259–273.

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*. ACM, 1–10.

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.

Yu, F., *et al.*, 2020. A category-aware deep model for successive POI recommendation on sparse check-in data. *In: Proceedings of the web conference 2020*. ACM, 1264–1274.

Yue, Y., *et al.*, 2017. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*, 31 (4), 658–675.

Zaheer, M., *et al.*, 2017. Deep sets. *In: Proceedings of the 31st conference on neural information processing systems*, 3391–3401.

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

Zhang, X., Du, S., and Wang, Q., 2017. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132, 170–184.

Zhang, X., Du, S., and Wang, Q., 2018. Integrating bottom-up classification and top-down feedback for improving urban land-cover and functional-zone mapping. *Remote Sensing of Environment*, 212, 231–248.

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

Zhang, D., *et al.*, 2020. Network representation learning: a survey. *IEEE Transactions on Big Data*, 6 (1), 3–28.

Zhou, X. and Zhang, L., 2016. Crowdsourcing functions of the living city from Twitter and Foursquare data. *Cartography and Geographic Information Science*, 43 (5), 393–404.