Learning Hierarchy-Enhanced POI Category Representations Using Disentangled Mobility Sequences

Hongwei Jia[1], **Meng Chen**[1*], **Weiming Huang**[2], **Kai Zhao**[3], **Yongshun Gong**[1]
[1]School of Software, Shandong University
[2]School of Computer Science and Engineering, Nanyang Technological University
[3] Robinson College of Business, Georgia State University
jhw@mail.sdu.edu.cn, mchen@sdu.edu.cn, weiming.huang@ntu.edu.sg, kzhao4@gsu.edu, yongshun2512@hotmail.com

Abstract

Points of interest (POIs) carry a wealth of semantic information of varying locations in cities and thus have been widely used to enable various location-based services. To understand POI semantics, existing methods usually model contextual correlations of POI categories in users' check-in sequences and embed categories into a latent space based on the word2vec framework. However, such an approach does not fully capture the underlying hierarchical relationship between POI categories and can hardly integrate the category hierarchy into various deep sequential models. To overcome this shortcoming, we propose a Semantically Disentangled POI Category Embedding Model (SD-CEM) to generate hierarchy-enhanced category representations using disentangled mobility sequences. Specifically, first, we construct disentangled mobility sequences using human mobility data based on the semantics of POIs. Then we utilize the POI category hierarchy to initialize a hierarchy-enhanced representation for each category in the disentangled sequences, employing an attention mechanism. Finally, we optimize these category representations by incorporating both the masked category prediction task and the next category prediction task. To evaluate the effectiveness of SD-CEM, we conduct comprehensive experiments using two check-in datasets covering three tasks. Experimental results demonstrate that SD-CEM outperforms several competitive baselines, highlighting its substantial improvement in performance as well as the understanding of learned category representations.

1 Introduction

Points of interest (POIs) are specific locations that individuals may find useful or interesting such as restaurants or shopping malls, and they are often associated with category labels, such as *Bar* or *Nightclub*. Such categories offer rich information to understand the functions and activities afforded by POIs, and also the semantic difference among them. For instance, *Bar* implies the possibility of getting a drink at night, and it shares several characteristics with *Nightclub*. But its semantics are largely divergent from the categories such as *Gas Station* and *Museum*. To quantitatively represent POI categories, a recent popular method is casting them into distributed vectors in a latent embedding space, which helps better understand and reason about POI semantics. In addition, such representations have been demonstrated to be effective in addressing various urban challenges such as land use classification [Huang *et al.*, 2022; Bing *et al.*, 2023; Bai *et al.*, 2023; Xu *et al.*, 2023b], human mobility prediction [Chen *et al.*, 2019; Zhang *et al.*, 2022a; Chen *et al.*, 2020], and POI recommendation [Zhang *et al.*, 2023; Wang *et al.*, 2022; Zhang *et al.*, 2022b; Xu *et al.*, 2023a].

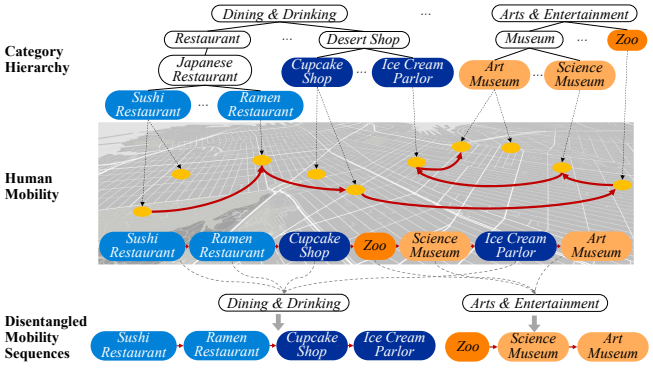


Figure 1: Illustration of disentangled mobility sequences.

Recently, the widespread usage of location-based services (e.g., Foursquare and Twitter) has resulted in the increasing availability of vast amounts of human mobility records attached to POIs (i.e., check-ins). Such mobility data have been commonly used to learn the representations of POI categories. Previous methods mainly model contextual correlations of POI categories in users' check-in sequences using a framework similar to word2vec [Cao *et al.*, 2019; Bing *et al.*, 2023]. However, such methods only capture the co-occurrence patterns between POI categories from the mobility view, neglecting the semantic relatedness inherently entailed by the hierarchical structure of POI categories (as shown in Figure 1) and hidden human travel patterns.

*Corresponding author.

To leverage the hierarchical structure of POI categories, Chen *et al.* [2021] propose Hier-CEM which models both the linear context derived from mobility data and the hierarchical context provided by the category hierarchy. However, this method is tied to the word2vec framework that predicts the target category given multiple contexts to generate the representations of POI categories. This implies that Hier-CEM can hardly be generalized to other learning architectures other than word2vec. With the prevalence of deep sequential models like LSTM and Transformer, it is necessary to explore alternative methods to mine hierarchical POI categories, which can be seamlessly integrated with deep sequential models.

To summarize, current POI category embedding methods encounter the following two challenges:

1) POI categories are usually organized in a hierarchical manner[1], and the interdependent relations between POI categories at various layers of the hierarchy can assist in establishing semantic relationships from lower to higher levels. It is reasonable to assume that the representation of each category should possess overlapped semantics with its ancestor category nodes, e.g., the representation of *Restaurant* has some common semantics with that of *Dining & Drinking*, and the representation of *Japanese Restaurant* shares some similar semantics with those of *Dining & Drinking* and *Restaurant*. Therefore, the first challenge is: *how to encode the pre-defined hierarchical structure of POI categories into embedding generation for each category?*

2) In principle, human mobility (check-in) sequences exhibit individuals' personal habits and travel intentions. For example, the schedule for a sports enthusiast on a particular day is to play basketball, and then go to a stadium to watch a football match. In this case, two POIs are particularly indicative of the intention of the person, i.e., *Basketball Court* and *Stadium*, which are also two highly relevant POI categories in terms of their semantics. However, in real-world check-in sequences, such prominent POIs are usually buried into numerous trivial ones, as it is barely the case the person only visits these two places. In reality, (a part of) the person's check-in sequence is *Basketball Court* $\rightarrow$ *Subway Station* $\rightarrow$ *Grocery Store* $\rightarrow$ *Sushi Restaurant* $\rightarrow$ *Stadium*. We can observe that it is challenging for current sequential models to identify the real travel intentions of individuals, and to disentangle the visited places into varying types, e.g., some visits indicate real travel intentions, while others are for human habits, daily needs, or merely transport transitions. In this context, the second challenge is: *how to disentangle and streamline the semantically mixed mobility sequences to help unveil the latent semantic relatedness between POIs?*

To solve these challenges, we propose a Semantically Disentangled POI Category Embedding Model (SD-CEM), which learns hierarchy-enhanced category representations using disentangled mobility sequences. SD-CEM consists of three major components: disentangled mobility sequences, hierarchy-enhanced embedding, and pretext tasks for training the category representations. Taking the category hierarchy and users' mobility sequences as input, SD-CEM first groups different types of categories from a mobility sequence according to the top-level nodes in the category hierarchy (most generic categories) and generates multiple semantically disentangled category sequences, as shown in Figure 1. Next, in the component of hierarchy-enhanced embedding, SD-CEM leverages the prior knowledge of the category hierarchy to generate the latent representation for each category in the disentangled mobility sequences, where the representation of a category is the combination of itself and its upper-level ancestors in the hierarchy via an attention mechanism. Finally, the category representations are optimized via two pretext tasks: the masked category prediction task encodes the deep feature interactions of semantically correlated categories based on the disentangled mobility (category) sequences, and the next category prediction task encodes the transition patterns of human mobility based on the original mobility sequences (before the disentanglement).

Our main contributions are as follows:

- We study the POI category embedding problem by exploring category correlations from multiple views, considering both the human mobility and the POI category hierarchy. To the best of our knowledge, we are the first to disentangle the semantically mixed mobility sequences following the category hierarchy to produce effective POI category representations.
- We mine the hierarchical structure of POI categories in two different ways, with the first being semantically disentangled mobility sequences based on the hierarchy, and the second being a hierarchy-enhanced (attention-based) embedding component, which is a detached component that encodes the hierarchy into embeddings. Overall, we present a more flexible way to integrate the category hierarchy into various deep sequential models.
- We experiment with the proposed SD-CEM on two public check-in datasets and evaluate its performance on three tasks. Experimental results show that SD-CEM demonstrates significant performance gains over the baseline methods. Data and source codes are at https://github.com/2837790380/SD-CEM.

2 Problem Formulation

Definition 1 (Mobility Sequence)**.** *The mobility sequence $s_i = (q_1, q_2, \cdots, q_{N_i})$ of a user u_i is a time-ordered sequence composed of N_i track points visited by u_i. A track point can be represented as a tuple of $q = (p, c, t)$, where p is the visited POI, c is the semantic category label of p, and t is the time when the user visits p.*

Remark. Given a mobility sequence, we fetch the POI category from every track point and constitute a category sequence. The category sequence hereinafter is also referred to as the mobility sequence when the context makes it clear.

Definition 2 (POI Category Hierarchy)**.** *POI categories are usually organized with a tree-liked hierarchical structure, where categories on a path (e.g., Dining & Drinking - Restaurant - Japanese Restaurant - Sushi Restaurant) from the root (i.e., Dining & Drinking) to the leaf node (i.e., Sushi Restaurant) have similar semantic properties.*

[1]https://location.foursquare.com/places/docs/categories

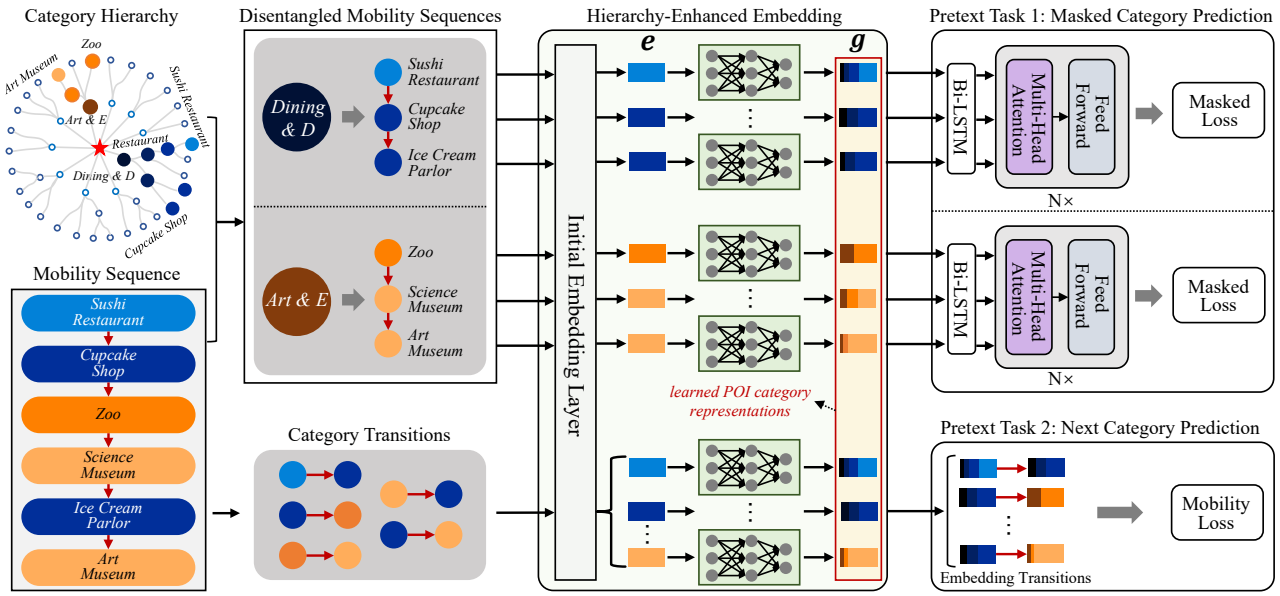


Figure 2: Framework of the proposed SD-CEM.

Definition 3 (Disentangled Mobility Sequence)**.** *Given a mobility (category) sequence as well as the POI category hierarchy, each group of track points sharing the same top-level (most generic) category composes a subsequence (e.g., Zoo - Science Museum - Art Museum in Figure 1). In this way, each mobility sequence is decomposed into several subsequences, which are semantically disentangled based on the top-level POI categories. Such subsequences are thereof denoted disentangled mobility sequences.*

Problem 1 (POI Category Embedding)**.** *Given users' mobility sequences and the POI category hierarchy, the goal is to learn a hierarchy-enhanced representation for each category.*

3 Methodology

Figure 2 presents the overview of the proposed SD-CEM, where it considers both the category hierarchy and users' mobility sequences as input and takes the learned POI category representations as output. SD-CEM consists of three components. The first component generates disentangled mobility sequences from the original user mobility sequential data. In the second component, for each category in the disentangled mobility sequences, it leverages the pre-defined category hierarchy to learn hierarchy-enhanced category representations via an attention mechanism. Finally, our third component optimizes the category representations via two pretext tasks: the masked category prediction task encodes the deep feature interactions of semantically correlated categories in the disentangled category sequences; the next category prediction task on the original mobility sequences (before the disentanglement) encodes the transition patterns of human mobility into the embedding generation process.

3.1 Disentangled Mobility Sequences

As prominent POIs are usually buried into numerous trivial ones in a mobility sequence, it is desirable to streamline the places with mixed intentions to unveil the latent semantic relatedness between POIs. Meanwhile, POI categories are not independent but organized in a tree-like hierarchy and those on the path from the root node to the leaf node contain similar semantics. Thus, we propose to disentangle users' mobility sequences according to the POI semantics.

We group different types of categories in a mobility sequence according to the top-layer (most generic) nodes in the category hierarchy and generate multiple category subsequences. That is, each group of categories sharing the same ancestor category comprises a single semantically disentangled category subsequence. As shown in Figure 2, we disentangle the mobility sequence (*Sushi Restaurant* → *Cupcake Shop* → *Zoo* → *Science Museum* → *Ice Cream Parlor* → *Art Museum*) into two disentangled category subsequences (*Sushi Restaurant* → *Cupcake Shop* → *Ice Cream Parlor* and *Zoo* → *Science Museum* → *Art Museum*) according to the top-layer categories (i.e., *Dining & Drinking* and *Arts & Entertainment*). As a result, categories in the disentangled subsequence contain aggregated semantics.

3.2 Hierarchy-Enhanced Embedding

We now introduce the embedding method to transform each category in the disentangled mobility sequences into a finite-dimensional real-valued vector. Traditionally, the embedding layers are weighted parameter matrices that provide mappings between the original category and the real-valued vector, whose parameters are jointly optimized with the entire deep network. Such an embedding method ignores the semantic relationships among categories predefined by the category hierarchy. Specifically, categories on a path from the root node to the leaf node have similar semantics, and the low-layer categories contain more fine-grained semantics than the top ones. For example, considering *Dining & Drinking-Restaurant-Japanese Restaurant-Sushi Restaurant* as an

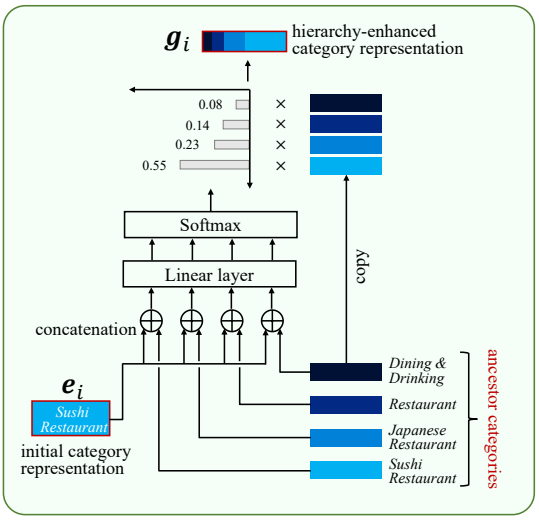


Figure 3: Structure of hierarchy-enhanced embedding.

example, the ancestor nodes (i.e., *Dining & Drinking*, *Restaurant*, and *Japanese Restaurant*) of *Sushi Restaurant* all contain the semantic information of *Sushi Restaurant* to different extents. To solve this problem, we present a hierarchy-enhanced category embedding component to generate the latent representations of POI categories, in which the representation of a POI category is a combination of itself and its upper-level ancestors in the hierarchy via an attention mechanism (see Figure 3).

Formally, we start by randomly initializing a category c_i (e.g., *Sushi Restaurant* in Figure 3) as a raw embedding $\mathbf{e}_i \in \mathbb{R}^d$, where d denotes the embedding dimension. Then, we find the ancestor categories of c_i and construct the ancestor set $\mathcal{A}(c_i)$. Here we also add the category c_i into $\mathcal{A}(c_i)$. For each category $c_j \in \mathcal{A}(c_i)$, we compute the semantic correlation score $s(c_i, c_j)$ between c_i and c_j via a linear layer, which takes the concatenation of the two raw embeddings ($\mathbf{e}_i$ and $\mathbf{e}_j$) as input,

$$s(c_i, c_j) = \mathbf{u}_a^{\mathsf{T}} \tanh(\mathbf{W}_a \begin{bmatrix} \mathbf{e}_i \\ \mathbf{e}_j \end{bmatrix} + \mathbf{b}_a), \tag{1}$$

where $\mathbf{W}_a \in \mathbb{R}^{r \times 2d}$ is the weight matrix for the concatenation of $\mathbf{e}_i$ and $\mathbf{e}_j$. $\mathbf{b}_a \in \mathbb{R}^r$ and $\mathbf{u}_a \in \mathbb{R}^r$ represent the bias and the weight. Next, we apply a softmax function on all $s(c_i, c_j), c_j \in \mathcal{A}(c_i)$ to calculate the attention weights,

$$\alpha_{ij} = \frac{\exp(s(c_i, c_j))}{\sum_{c_k \in \mathcal{A}(c_i)} \exp(s(c_i, c_k))}. \tag{2}$$

Finally, based on the attention weights and the raw embeddings of categories from $\mathcal{A}(c_i)$, we compute the hierarchy-enhanced representation for category c_i as

$$\mathbf{g}_i = \sum_{c_j \in \mathcal{A}(c_i)} \alpha_{ij} \mathbf{e}_j. \tag{3}$$

After the hierarchy-enhanced embedding process, we obtain the latent representation for each category in the mobility sequences, which is further optimized with the pretext tasks.

3.3 Pretext Tasks

Masked Category Prediction

Taking the latent representations $\mathbf{g}$ of categories as input, we further model the deep interactions of categories in the semantically disentangled subsequences using a Bi-LSTM layer composed of a forward and a backward LSTM layer. Specifically, given a subsequence $\{c_1, c_2, \cdots, c_n\}$ containing n categories, we take $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_n)$ as input, and generate the hidden feature vector $\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_n)$ as follows,

$$\mathbf{h} = \text{Bi-LSTM}(\mathbf{g}). \tag{4}$$

After generating the feature vector $\mathbf{h}$ via a Bi-LSTM, we adopt masked category prediction (MCP) as our first pretext task, similar to the Bert model, originally designed for language modeling [Kenton and Toutanova, 2019]. We conduct this task on the disentangled category subsequences. For each subsequence, we randomly select 15% of the categories to be replaced. For a selected category c_i^m, we replace its token with the actual [MASK] token with 80% probability, another random category token with 10% probability, and the unchanged one with 10% probability. The masked subsequence is inputted into a module consisting of stacked transformer layers (where parameters across these layers are shared inspired by ALBERT [Lan *et al.*, 2019]), where each layer contains a multi-head self-attention layer and a feedforward layer [Vaswani *et al.*, 2017]. We use the feature vector $\mathbf{h}$ generated via the Bi-LSTM layer as the token embeddings and the rank information of each category in the subsequence as position embeddings. After the stacked transformer layers, the output $\mathbf{o}_i^m$ corresponding to the [MASK] token is used to predict the masked category via a softmax function,

$$\hat{\mathbf{p}}_i = \text{softmax}(\mathbf{W}_{\text{MCP}} \mathbf{o}_i^m + \mathbf{b}_{\text{MCP}}), \tag{5}$$

where $\mathbf{W}_{\text{MCP}} \in \mathbb{R}^{|\mathcal{C}| \times d^o}$ and $\mathbf{b}_{\text{MCP}} \in \mathbb{R}^{|\mathcal{C}|}$ are the parameterized weight matrix and the bias weight vector, respectively. $|\mathcal{C}|$ represents the number of unique categories, d^o is the dimension of the output $\mathbf{o}_i^m$, and $\hat{\mathbf{p}}_i$ is the predicted distribution of the masked category c_i^m over all categories.

Note that we use Bi-LSTM to capture temporal dependencies between categories in the disentangled mobility sequences and the stacked transformer layers to capture the local and global relationship between a single category and an entire sequence by inferring the masked categories based on the context information. By combining them, it could achieve a comprehensive understanding of the sequential patterns of mobility sequences.

For training, we use the cross-entropy loss between the one-hot label $\mathbf{y}_i$ of the true masked category c_i^m and the predicted distribution $\hat{\mathbf{p}}_i$. Given a mobility sequence s, we define the objective function as

$$\mathcal{L}^{MCP} = -\frac{1}{|\mathcal{M}_s|} \sum_{c_i^m \in \mathcal{M}_s} \sum_{j}^{|\mathcal{C}|} \mathbf{y}_{i,j} \log \hat{\mathbf{p}}_{i,j}, \tag{6}$$

where $\mathbf{y}_{i,j}$ and $\hat{\mathbf{p}}_{i,j}$ are the j-th components of $\mathbf{y}_i$ and $\hat{\mathbf{p}}_i$, respectively. $\mathcal{M}_s$ is the set of masked categories in the disentangled category subsequences generated from s, and $|\mathcal{M}_s|$ is the size of $\mathcal{M}_s$.

Next Category Prediction

We further model the category transitions (i.e., the original mobility sequences without the disentanglement) based on the latent representations $\mathbf{g}$ of categories and design the second pretext task of the next category prediction (NCP). Specifically, given a category c_i, we model the distribution of the next visited category c_j as

$$\hat{p}(c_j|c_i) = \frac{\exp(\mathbf{g}_i^\mathsf{T}\mathbf{g}_j)}{\sum_{c_k \in \mathcal{C}} \exp(\mathbf{g}_i^\mathsf{T}\mathbf{g}_k)}, \tag{7}$$

where $\mathbf{g}_i$ and $\mathbf{g}_j$ are the hierarchy-enhanced representations of c_i and c_j, respectively.

Given the set $\mathcal{T}_s$ of category transitions generated from a mobility sequence s, we define the objective function as the negative log-likelihood of predicted distributions,

$$\mathcal{L}^{NCP} = -\frac{1}{|\mathcal{T}_s|} \sum_{(c_i,c_j)\in\mathcal{T}_s} \log \hat{p}(c_j|c_i), \tag{8}$$

where $|\mathcal{T}_s|$ is the number of category transitions generated from s.

The final loss is the combination of the two losses. Given training mobility sequences in mini-batches of size B, the parameters are optimized by minimizing the following loss,

$$\mathcal{L} = \sum_{b=1}^{B} \lambda \mathcal{L}^{MCP}(b) + (1-\lambda)\mathcal{L}^{NCP}(b), \tag{9}$$

where $\mathcal{L}^{MCP}(b)$ and $\mathcal{L}^{NCP}(b)$ correspond to the masked category prediction loss and the next category prediction loss of the b-th training mobility sequence in the mini-batch, respectively, and λ is a weight that balances the two losses.

4 Experiments

We start by describing the experimental settings, followed by the evaluation of the learned POI category representations on three tasks. In the end, we visualize the learned category representations to gain an intuitive understanding of the results.

4.1 Experimental Settings

Datasets. We adopt the public check-in data of Foursquare collected from the United States and Japan [Yang *et al.*, 2016], and pre-process these check-ins following [Chen *et al.*, 2021] to generate mobility sequences. The JP dataset consists of 10,336 sequences of POI categories, encompassing 330 distinct categories; the US dataset consists of 21,898 category sequences, with 396 unique categories.

Model parameters. For Bi-LSTM layers, we set the number of layers at 4 and the dimension of the hidden vector at 128. For the stacked transformer layers, we set the number of layers and attention heads at 8, and the hidden size at 128. The weight λ is 0.8. The Adam learning rate is initialized as 0.001 with a linear decay. The dimension of the category representations $\mathbf{e}$ and $\mathbf{g}$ is set at $\{20, 30, 40, 50, 60, 70, 80\}$. The optimal parameters are selected using grid search with a small but adaptive step size.

Baselines. We compare SD-CEM with seven state-of-the-art category embedding methods, i.e., the word2vec-based methods (including MC-TEM [Zhou *et al.*, 2016], STES [Yang and Eickhoff, 2018], SC [Yan *et al.*, 2017], and CatEM [Bing *et al.*, 2023]), the tensor factorization-based method (LBPR [He *et al.*, 2017]), the RNN-based method (CatDM [Yu *et al.*, 2020]), and the hierarchy-based methods (Hier-CEM [Chen *et al.*, 2021]). Details of these methods are introduced in the section of Related Work.

4.2 Performance Comparison

Surface-Level Semantic Task

To assess the effectiveness of different methods in encoding semantic properties, we evaluate the degree of semantic overlap among various categories following [Chen *et al.*, 2021]. Specifically, for each category c_i, we determine its closest neighbor c_j in the embedding space by measuring the cosine similarity between their respective category representations. If c_i and c_j share the top-layer category in the category hierarchy, we consider a match to have occurred between c_i and c_j. The match rate of a given test set is then calculated as the ratio of matched categories to the total number of categories in the set. That is,

$$\textit{match rate} = \frac{\textit{number of matched categories}}{\textit{number of categories in the test set}}. \tag{10}$$

A higher *match rate* indicates a greater extent to which the category representations capture semantic information.

We assess all the methods using the same dataset and conduct 10 runs to derive 10 match rate values. The mean of these 10 values is reported in Table 1. We observe that:

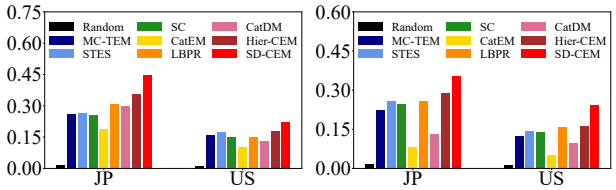
1) The word2vec-based methods achieve satisfactory results, suggesting that explicitly modeling mobility sequences can effectively capture category semantics. Specifically, MC-TEM exhibits relatively inferior performance as it predicts the target POI based on multiple contexts and incidentally learns category representations; STES, SC, and CatEM directly model check-in category sequences to generate representations, consequently yielding superior performance.

2) LBPR exhibits the poorest performance as it solely focuses on modeling users' preference on category transitions, without taking into account the high-order sequential patterns. CatDM demonstrates superior performance compared to LBPR by utilizing LSTM-based techniques to capture long-term transitions in the check-in sequences. However, it does not exhibit significant advantages in comparison to word2vec-based approaches.

3) Hier-CEM demonstrates superior performance compared to the other baselines as it not only models linear contextual information but also incorporates hierarchical relations among categories to acquire category representations. Despite both Hier-CEM and the proposed SD-CEM model the category sequences and category hierarchy, the latter outperforms the former. This can be attributed to the inclusion of a hierarchy-enhanced embedding component and semantically disentangled category sequences in SD-CEM. In comparison to Hier-CEM, SD-CEM achieves an average improvement of 26.2% on the JP data and 12.3% on the US data. Moreover, the results of the superiority paired t-test indicate that the improvement of SD-CEM over the baselines is statistically significant, with a p-value less than 0.01.

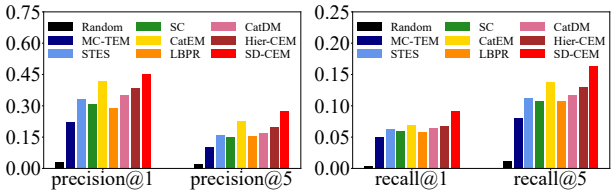
Methods	JP							US						
	Embedding size (d)							Embedding size (d)						
	20	30	40	50	60	70	80	20	30	40	50	60	70	80
MC-TEM	0.298	0.345	0.339	0.329	0.342	0.373	0.379	0.371	0.356	0.351	0.361	0.4	0.387	0.376
STES	0.549	0.564	0.552	0.552	0.552	0.542	0.536	0.634	0.609	0.578	0.558	0.553	0.525	0.508
SC	0.6	0.624	0.621	0.649	0.633	0.624	0.639	0.626	0.588	0.599	0.634	0.611	0.634	0.611
CatEM	0.5	0.558	0.591	0.618	0.615	0.645	0.661	0.588	0.646	0.669	0.659	0.715	0.692	0.725
LBPR	0.294	0.33	0.397	0.324	0.358	0.427	0.421	0.308	0.338	0.333	0.419	0.429	0.379	0.386
CatDM	0.512	0.548	0.545	0.57	0.545	0.603	0.591	0.669	0.611	0.641	0.634	0.604	0.646	0.644
Hier-CEM	0.718*	0.773*	0.749*	0.727*	0.749*	0.755*	0.749*	0.854*	0.851*	0.849*	0.846*	0.851*	0.816*	0.843*
SD-CEM	**0.93**	**0.918**	**0.942**	**0.939**	**0.949**	**0.945**	**0.961**	**0.95**	**0.962**	**0.947**	**0.959**	**0.942**	**0.955**	**0.924**
Improvements	29.5%	18.8%	25.8%	29.2%	26.7%	25.2%	28.3%	11.2%	13.0%	11.5%	13.4%	10.7%	17.0%	9.6%

Table 1: Performance comparison of different methods in terms of *match rate*, where the performance improvements of SD-CEM are compared with the best of these baseline methods, marked by the asterisk.



(a) Accuracy (b) MRR

Figure 4: Performance comparison on the mobility task.



(a) Precision (b) Recall

Figure 5: Performance comparison on POI recommendation.

Surface-Level Mobility Task

To assess the ability of category representations to preserve sequential patterns, we present the task of the next check-in category prediction. The objective is to predict the next check-in categories given the previous category sequence visited by a user. Specifically, we derive the conditional probability $p(c_j|c_i) \propto \mathbf{g}_i \cdot \mathbf{g}_j$, where $\mathbf{g}_i$ and $\mathbf{g}_j$ represent the representations of the current category c_i and the candidate category c_j, respectively. By ranking the candidate categories based on the values of $p(c_j|c_i)$, we generate a list from the most probable to the least probable. The performance is evaluated using the metrics of *accuracy* and Mean Reciprocal Rank (*MRR*). *Accuracy* measures the frequency of the true next category appearing in the top-5 predicted list; *MRR* is defined as $MRR = (\sum_{i=1}^{|\mathcal{C}_{test}|} \frac{1}{rank_i})/|\mathcal{C}_{test}|$, where $\mathcal{C}_{test}$ is the set of test categories, and $rank_i$ is the rank of the true next category for c_i in the ranking list.

Figure 4 shows the comparison results on the mobility task on both datasets. Here the *Random* method pertains to the act of randomly selecting a category from the category set as the anticipated next category, which serves as a benchmark. Evidently, the proposed SD-CEM (red-colored) performs the best in terms of *accuracy* and *MRR*, indicating that the learned POI category representations have also captured mobility patterns.

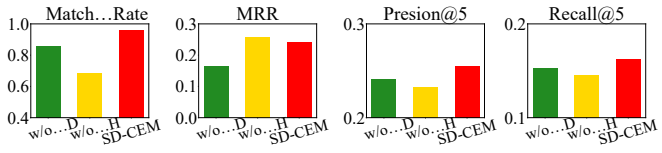
Downstream POI Recommendation Task

In addition to the surface-level tasks, an inquiry into the potential application of POI category representations in downstream tasks is desired. To illustrate this point, we evaluate the category representations with the task of POI recommendation, which aims to forecast a list of top-k POIs that a user is likely to visit in the future. To accomplish this, we employ the POI recommendation algorithm outlined in [Chen *et al.*, 2021] and evaluate model performance using the metrics of *precision* and *recall* [Zhang and Chow, 2015].

We conduct a comparison between SD-CEM and the baseline methods on the POI recommendation task. The results on the US dataset are presented in Figure 5. Similar results are observed with the JP dataset. Here *Random* refers to the selection of recommended categories at random, rather than using users' preferred categories calculated based on POI category representations. This random selection serves as the benchmark for comparison. Observed from Figure 5, SD-CEM (red-colored) achieves the best performance in terms of *recall* and *precision*. These findings indicate that integrating the category hierarchy into deep sequential models leads to improved POI category representations.

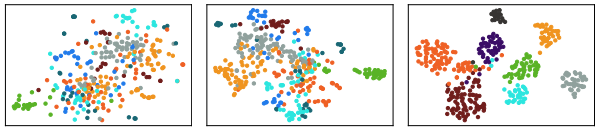
4.3 Ablation Study

We design two variants to further verify the effectiveness of SD-CEM: 1) SD-CEM w/o D: it removes the component of disentangled mobility sequences and directly models the mobility sequences using the pretext masked category prediction task. Note that it also takes the hierarchy-enhanced category representations as input; 2) SD-CEM w/o H: it directly feeds the initial category representations $\mathbf{e}$ into the pretext tasks, without utilizing the hierarchy-enhanced embedding component. The results of SD-CEM and its variants on the US dataset are depicted in Figure 6. We observe that both SD-CEM w/o D and SD-CEM w/o H yield poorer results than SD-CEM in most cases, indicating that modeling the POI cat-



(a) Semantic (b) Mobility (c) POI Recommendation

Figure 6: Performance comparison of different variants.



(a) MC-TEM (b) Hier-CEM (c) SD-CEM

Figure 7: Exhibition of POI category representations in 2D planes.

egory hierarchy explicitly or implicitly could help better encode POI semantics into these representations. Additionally, an interesting observation is that SD-CEM w/o H performs better than SD-CEM in the mobility task, as encoding the hierarchical relations of categories may take the ancestors of the actual next category as the prediction, hindering the sequential prediction performance. Similar results have been observed in the variants of Hier-CEM [Chen *et al.*, 2021].

4.4 Qualitative Analysis of Representations

We visualize the pre-trained category representation by projecting them onto 2D planes using t-Distributed Stochastic Neighbor Embedding (t-SNE) [Maaten and Hinton, 2008]. The obtained results on the US data, including those from SD-CEM as well as the baseline methods (MC-TEM and Hier-CEM), are presented in Figure 7. In these plots, each point represents a category, and categories belonging to the same top-layer category are assigned the same color. Analysis of Figure 7 (a) reveals that the vectors generated by MC-TEM appear to be disordered and intertwined in the space. Conversely, Hier-CEM exhibits a discernible trend where categories with similar semantics tend to be located in close proximity. This observation supports the notion that incorporating hierarchical category information enables the acquisition of more semantic knowledge. Figure 7 (c) demonstrates that SD-CEM successfully embeds the categories into distinct and well-separated clusters. This indicates that the proposed SD-CEM is capable of learning superior category representations compared to both MC-TEM and Hier-CEM.

5 Related Work

Recently, researchers have developed multiple embedding methods to learn effective representations for POI categories by mining human mobility information. Many traditional category embedding methods [Zhou *et al.*, 2016; Yang and Eickhoff, 2018; Yan *et al.*, 2017] mainly adopt the word2vec framework [Mikolov *et al.*, 2013] to construct the category embeddings. In addition, certain approaches [Liu *et al.*, 2013; He *et al.*, 2017] utilize matrix/tensor factorization on a user-category check-in matrix to learn category embeddings.

Furthermore, prominent strides have been made in analyzing check-in data by using deep sequential models. These approaches typically employ sequential modeling techniques to identify patterns within check-in sequences and generate representations of POI categories as a byproduct. For instance, Yu *et al.* [2020] propose a category-aware deep model (CatDM) for predicting the next POI using check-ins. This involves two LSTM encoders to capture the temporal patterns of POI categories and user preferences, as well as considering multiple correlations to determine the probability of each POI within the candidate set. Lian *et al.* [2020] introduce a geography-aware sequential recommender for POI recommendation, which incorporates geography encoders and self-attention networks to capture long-term sequential dependencies within check-in trajectories. However, their focus primarily lies in capturing sequential patterns to make next POI recommendations, rather than comprehending the semantic aspects of POIs.

Moving further from only using check-in data, several studies have sought to enhance the quality of POI category representations by incorporating external information, such as the POI category hierarchy and spatial distribution of POIs. For example, Chen *et al.* [2021] present a pre-trained method (Hier-CEM) based on the word2vec framework for embedding POI categories that considers both the linear context derived from check-in sequences and the hierarchical context provided by the POI category hierarchy. Bing *et al.* [2023] introduce a category embedding method called CatEM, which leverages both the sequential context derived from check-in sequences and the spatial distributions of POI categories within a specific region to learn category representations.

6 Conclusion

In this paper, we propose a Semantically Disentangled POI Category Embedding Model (SD-CEM) that enhances the representations of POI categories by utilizing disentangled mobility sequences. The original human mobility sequences are first processed to generate disentangled sequences based on the semantics of POIs. Subsequently, the pre-defined category hierarchy is modeled to learn a hierarchy-enhanced representation for each category within these disentangled sequences. To further optimize these representations, two pretext tasks, the masked category prediction, and the next category prediction, are utilized. The performance of the proposed SD-CEM is evaluated on two public check-in datasets across three tasks. The experimental results show that SD-CEM outperforms the state-of-the-art baselines, indicating that incorporating the category hierarchy into deep sequential models is effective for POI category embedding.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61906107, 62202270, and 42101421, the Young Scholars Program of Shandong University. W.H. was supported by the Knut and Alice Wallenberg Foundation.

References

[Bai *et al.*, 2023] Lubin Bai, Weiming Huang, Xiuyuan Zhang, Shihong Du, Gao Cong, Haoyu Wang, and Bo Liu. Geographic mapping with unsupervised multi-modal representation learning from vhr images and pois. *ISPRS Journal of Photogrammetry and Remote Sensing*, 201:193–208, 2023.

[Bing *et al.*, 2023] Junxiang Bing, Meng Chen, Min Yang, Weiming Huang, Yongshun Gong, and Liqiang Nie. Pretrained semantic embeddings for poi categories based on multiple contexts. *IEEE Transactions on Knowledge and Data Engineering*, 35(09):8893–8904, 2023.

[Cao *et al.*, 2019] Hancheng Cao, Fengli Xu, Jagan Sankaranarayanan, Yong Li, and Hanan Samet. Habit2vec: Trajectory semantic embedding for living pattern recognition in population. *IEEE Transactions on Mobile Computing*, 19(5):1096–1108, 2019.

[Chen *et al.*, 2019] Meng Chen, Xiaohui Yu, and Yang Liu. Mpe: A mobility pattern embedding model for predicting next locations. *World Wide Web*, 22:2901–2920, 2019.

[Chen *et al.*, 2020] Meng Chen, Yan Zhao, Yang Liu, Xiaohui Yu, and Kai Zheng. Modeling spatial trajectories with attribute representation learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(4):1902–1914, 2020.

[Chen *et al.*, 2021] Meng Chen, Lei Zhu, Ronghui Xu, Yang Liu, Xiaohui Yu, and Yilong Yin. Embedding hierarchical structures for venue category representation. *ACM Transactions on Information Systems*, 40(3):1–29, 2021.

[He *et al.*, 2017] Jing He, Xin Li, and Lejian Liao. Category-aware next point-of-interest) recommendation via listwise bayesian personalized ranking. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, volume 17, pages 1837–1843, 2017.

[Huang *et al.*, 2022] Weiming Huang, Lizhen Cui, Meng Chen, Daokun Zhang, and Yao Yao. Estimating urban functional distributions with semantics preserved poi embedding. *International Journal of Geographical Information Science*, 36(10):1905–1930, 2022.

[Kenton and Toutanova, 2019] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186, 2019.

[Lan *et al.*, 2019] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. In *International Conference on Learning Representations*, 2019.

[Lian *et al.*, 2020] Defu Lian, Yongji Wu, Yong Ge, Xing Xie, and Enhong Chen. Geography-aware sequential location recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2009–2019, 2020.

[Liu *et al.*, 2013] Xin Liu, Yong Liu, Karl Aberer, and Chunyan Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, pages 733–738, 2013.

[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*, pages 3111–3119, 2013.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[Wang *et al.*, 2022] Xiaolin Wang, Guohao Sun, Xiu Fang, Jian Yang, and Shoujin Wang. Modeling spatio-temporal neighbourhood for personalized point-of-interest recommendation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 3530–3536, 2022.

[Xu *et al.*, 2023a] Ronghui Xu, Meng Chen, Yongshun Gong, Yang Liu, Xiaohui Yu, and Liqiang Nie. Tme: Tree-guided multi-task embedding learning towards semantic venue annotation. *ACM Transactions on Information Systems*, 41(4):1–24, 2023.

[Xu *et al.*, 2023b] Ronghui Xu, Weiming Huang, Jun Zhao, Meng Chen, and Liqiang Nie. A spatial and adversarial representation learning approach for land use classification with pois. *ACM Transactions on Intelligent Systems and Technology*, 14(6):1–25, 2023.

[Yan *et al.*, 2017] Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Song Gao. 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*, page 35, 2017.

[Yang and Eickhoff, 2018] Jing Yang and Carsten Eickhoff. Unsupervised learning of parsimonious general-purpose embeddings for user and location modeling. *ACM Transactions on Information Systems*, 36(3):32, 2018.

[Yang *et al.*, 2016] Dingqi Yang, Daqing Zhang, and Bingqing Qu. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology*, 7(3):30, 2016.

[Yu *et al.*, 2020] Fuqiang Yu, Lizhen Cui, Wei Guo, Xudong Lu, Qingzhong Li, and Hua Lu. A category-aware deep model for successive poi recommendation on sparse check-in data. In *Proceedings of the Web Conference 2020*, pages 1264–1274, 2020.

[Zhang and Chow, 2015] Jia-Dong Zhang and Chi-Yin Chow. Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 443–452, 2015.

[Zhang *et al.*, 2022a] Chao Zhang, Kai Zhao, and Meng Chen. Beyond the limits of predictability in human mobility prediction: Context-transition predictability. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4514–4526, 2022.

[Zhang *et al.*, 2022b] Lu Zhang, Zhu Sun, Ziqing Wu, Jie Zhang, Yew Soon Ong, and Xinghua Qu. Next point-of-interest recommendation with inferring multi-step future preferences. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 3751–3757, 2022.

[Zhang *et al.*, 2023] Dabin Zhang, Ronghui Xu, Weiming Huang, Kai Zhao, and Meng Chen. Towards an integrated view of semantic annotation for pois with spatial and textual information. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2441–2449, 2023.

[Zhou *et al.*, 2016] Ningnan Zhou, Wayne Xin Zhao, Xiao Zhang, Ji-Rong Wen, and Shan Wang. A general multi-context embedding model for mining human trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, 28(8):1945–1958, 2016.