Knowledge-Based Systems 235 (2022) 107623

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Enabling the interpretability of pretrained venue representations using semantic categories

Ning An [a], Meng Chen [a,*], Li Lian [a], Peng Li [b], Kai Zhang [b], Xiaohui Yu [c], Yilong Yin [a]

[a] School of Software, Shandong University, Jinan, Shandong, China
[b] Shandong Liju Robot Technology Co., Ltd, Jinan, Shandong, China
[c] School of Information Technology, York University, Toronto, Ontario, Canada

ARTICLE INFO

Article history:
Received 23 May 2021
Received in revised form 18 October 2021
Accepted 19 October 2021
Available online 21 October 2021

Keywords:
Venue semantic representation
Interpretable
Embedding learning
Semantic mapping
Check-ins

ABSTRACT

The growing popularity of location-based social networks gives rise to a tremendous amount of social check-ins data, which are broadly used in previous studies to produce dense venue representations for various trajectory mining tasks. In this work, we focus on the interpretability of venue representations, an essential property that existing methods fail to provide. We propose two novel models to generate interpretable and easy-to-understand venue representations. The first model, CEM, is a category-aware (a category may be a restaurant, a mall, etc.) check-in embedding model and generates venue and category representations by capturing the sequential patterns of check-in records. With the second model, XEM, each dimension of the venue representation corresponds to a semantic anchor (i.e., a category) and can be interpreted as a coherent topic. We conduct extensive experiments using real-world check-in datasets for venue similarity computation and venue semantic annotation, and empirically show that introducing interpretability to the venue representations improves the performance of various downstream tasks.

1. Introduction

An increasing number of Location-Based Social Network (LBSN) users nowadays share their locations (including the venue IDs and the corresponding categories) in the form of social check-in. A user's check-in records during a period of time constitute her or his trajectory. Since real-world trajectory data is usually unstructured, in order to leverage the records for various tasks such as check-in data mining, it is essential to transform these unstructured check-ins into numerical vectors, which is referred to as **check-in embedding learning**. Check-in embedding learning is the heart of various check-in data mining tasks such as point-of-interest recommendation [1–6], next location prediction [7,8] and venue category retrieval [9–11].

Recent studies have made significant progress in representation learning of check-ins, which mainly model check-in sequences to learn the embedding vectors of venues based on the Word2vec framework [12–15]. They follow the distributional hypothesis that venues appearing in similar contexts tend to have similar semantic properties and thus ought to be projected into closer embedding vectors in the latent space. These methods can only produce dense venue representations whose coordinates have no meaningful interpretation. However, explaining each dimension of the representations as a coherent and understandable topic is vital to many downstream tasks. For example, consider a case when a user queries with a venue whose category is *Restuarant*, with meaningful interpretations of venue representations, we can suitably arrange the query results and navigate the user to her/his desired results.

In this paper, we aim to learn an interpretable venue representation from users' check-ins whose coordinates have clear meanings to humans. We propose two models for the task:

(1) We propose a Check-in Embedding Model (CEM) which captures the sequential check-in patterns together with the semantic categories of venues, to simultaneously generate venue and category embedding vectors.

(2) We propose an eXplainable Embedding Model (XEM) which utilizes the categories (e.g., *Japanese Restaurant, Museum*) of venues to provide interpretability to the pretrained venue embeddings. More specifically, we take these categories as semantic anchors and compute the similarities between a venue and these anchors based on the embedding vectors learned using CEM. Then, we take these similarity scores as the representation of a venue, each dimension of which is associated with a coherent and easy-to-understand topic (i.e., a semantic category).

To evaluate the effectiveness of the proposed venue representations methods, we perform quantitative experiments on

* Corresponding author.
E-mail addresses: ningan@mail.sdu.edu.cn (N. An), mchen@sdu.edu.cn (M. Chen), lianli@sdu.edu.cn (L. Lian), wonderlp@126.com (P. Li), zbhkkk@126.com (K. Zhang), xhyu@yorku.ca (X. Yu), ylyin@sdu.edu.cn (Y. Yin).

https://doi.org/10.1016/j.knosys.2021.107623
0950-7051/

real-world check-in datasets collected from Foursquare and Yelp. We show that by making venue representations interpretable, the proposed models outperform the baselines across various tasks.

The contributions of this paper are summarized as follows:

- We provide interpretability to venue representations by explaining each dimension of these representations as a coherent and understandable topic. To the best of our knowledge, this is the first attempt to explain venue representations and fills the gap between venues and semantics.
- We devise a check-in embedding model CEM, integrating both the venue context (which embodies the sequential relations between individual venues) and the category context (which is formed by the categories of adjacent venues in the check-in sequence and thus reflects aggregated semantics) for learning representations of venues and categories.
- We design a method XEM for enabling the interpretability of pretrained venue embeddings. XEM explains each dimension of the venue representations as a semantic category.
- We conduct experiments on real-world datasets to demonstrate the superiority of the proposed models over applicable baselines, and empirically prove that interpretable venue representations with the assistance of categories greatly helps to capture semantics.

The rest of this paper is organized as follows. Section 2 reviews the studies on venue representation learning and interpretability of the embedding space. Section 3 presents the research objective of our study. Section 4 presents a new check-in embedding model. Section 5 describes the interpretable venue representations. The experimental results are discussed in Section 6. Section 7 concludes this paper and describes future work.

2. Related work

Our work is related to check-in data mining (including venue embedding learning and sequential learning) and interpretability of the embedding space, and representative works in both areas are summarized in this section.

2.1. Check-in data mining

Check-in data mining, which aims at extracting valuable information from large-scale location-based social networks, locates at pattern & knowledge discovery according to the four dimensions of social network analysis proposed by Camacho et al. [16]. Many methods have been proposed to model multiple patterns (e.g., sequential, temporal, and semantic patterns) from check-in data and can be used for a variety of applications, e.g., venue recommendation, next venue prediction, and user profiling. For example, Ying et al. [17] consider two properties (namely the asymmetric property of transitions between consecutive points-of-interest (POIs) in check-in sequences, and dynamic user preference at different times) of check-ins and propose a time-aware metric embedding approach with asymmetric projection for successive POI recommendation. Yang and Eickhoff [18] transform social check-ins into dense feature vectors using an embedding model, with which they model sequential, temporal, and geographic patterns of venues. The embedding model is employed in many tasks, such as venue recommendation, urban functional zone study, and crime prediction.

The common practice to learn representations of venues from check-in data is to use the Word2Vec framework [19,20]. Liu et al. [12] model the check-in sequences and capture the influence of linear context to learn the venue representations, which are used for personalized venue recommendations. In addition to the sequential patterns, Zhao et al. [14] model the dynamic user preference and temporal factor to learn venue representations. Similarly, Zhao et al. [13] propose a temporal location embedding model, which discriminates unvisited venues according to geographical information and incorporates geographical influence into the pairwise preference ranking method. Zhou et al. [15] propose a general multi-context trajectory embedding model which projects users, trajectories, venues, category labels, and the temporal factor in the same latent space. Furthermore, some studies leverage external information (e.g., text content) to learn venue representations. Chang et al. [21] propose a content-aware venue embedding model to utilize the text content of a venue to boost prediction performance. Yao et al. [22] propose a semantics-enriched recurrent model, which jointly learns the representations of multiple factors (e.g., venue, keyword) and the transition parameters of a recurrent neural network. Yu et al. [23] investigate the problem of fine-grained venue discovery by learning the cross-modal correlation between photos and the text descriptions of venues.

In addition, some methods leverage Recurrent Neural Networks (RNNs) to model the sequential patterns of check-ins and learn the venue representations as byproducts. For example, Kong and Wu [24] propose a hierarchical spatiotemporal LSTM model, leveraging historical visit information and spatiotemporal factors for venue prediction. Liu et al. [25] propose spatial–temporal RNNs, which model the local temporal and spatial contexts for mining mobility patterns. Yang et al. [26] employ RNNs to capture the sequential correlations in mobile trajectories. Wang et al. [27] propose a novel light location recommender system based on a lightweight yet effective gated RNN to make successive POI recommendations locally on resource-constrained mobile devices to mitigate the problem that POI recommendations always have to be made on the server-side. Sun et al. [28] propose a novel method, which consists of a nonlocal network for long-term preference modeling and a geo-dilated RNN for short-term preference learning for the next POI recommendation. However, these RNN-based methods focus on mining long-term transitions in a sequence instead of the quality of venue representations.

Our work differ from the aforementioned methods in the following two aspects. First, our focus is to make venue representations interpretable using semantic categories. To the best of our knowledge, this is the first attempt to investigate the interpretability of venue representations. In addition, we consider both the venue context (that embodies the sequential relationships between individual venues) and the category context (which is formed by the categories of adjacent venues in the check-in sequence) to learn venue representations.

2.2. Interpretability of the embedding space

Many studies adopt the idea of topic models and utilize the latent topics to explain the dimensions of the embedding space. For example, Panigrahi et al. [29] propose an LDA-based generative model to learn a distribution over senses for each word and utilize these senses to explain each dimension of the embedding space. Fyshe et al. [30] create an interpretable vector space model that respects the notion of semantic composition via non-negative matrix factorization, in which they select the top scoring words in each dimension as the semantics of the latent dimensions. On the other hand, some methods employ word embedding models to add interpretability to the pretrained embeddings. For instance, Zhao and Mao [31] compute the semantic correlation among words according to the cosine similarities between word embeddings, and explain the dimensions using a group of similar words. Mathew et al. [32] adopt semantic differentials to map pretrained word embeddings into a new interpretable polar space, in which they choose the most discriminative dimensions from many polar dimensions provided by external sources.

Table 1
Several examples of check-ins in Foursquare.

User ID	Venue ID	Category name	Time
1001	4ba90a95f964a52009063ae3	Seafood Restaurant	Apr 30 13:52:32
1001	4fbd7718e4b06a0de5a8fd1a	Mall	May 02 19:06:32
1001	4e737e04d16472c0372f3999	Falafel Restaurant	May 03 13:49:23
1001	510012a9e4b09b94f0d250f7	Coffee Shop	May 03 14:28:06
1001	4c6d7fb406ed6dcb750da422	Asian Restaurant	May 06 14:03:35

Table 2
Notations and descriptions.

Notations	Descriptions
u, w, t, c, s	user, venue, time, category and check-in sequence
S_u	check-in sequence of a user u
N_u	length of S_u
$\mathcal{S}$	set of check-in sequences
$\mathcal{C}$	set of categories
$\mathbf{v}_w, \mathbf{v}_c$	vectors of a context venue w and a context category c
$\mathbf{v}'_w, \mathbf{v}'_c$	vectors of a target venue w and a target category c
M	size of the category set
d	embedding size

Furthermore, it is also a widely-adopted strategy to combine probabilistic topic models and word embedding techniques to add interpretability to word embeddings. Li et al. [33] propose a topic modeling and sparse autoencoder which explicitly capture the mutual influence of global topics and local contexts to improve topic discovery and word embedding simultaneously. Xun et al. [34] consider both the global and local context information in documents and learn latent topics and word embeddings collaboratively based on matrix factorization techniques. Potapenko et al. [35] learn word embeddings by decomposing the word co-occurrence matrix, and use topic models to provide interpretable components. However, these methods cannot be applied to learning interpretable venue representations directly, as venues (which are different from words) do not contain semantics and cannot be understood by humans.

3. Research objective

A check-in can be defined as a tuple $ch = \langle u, w, c, t\rangle$ which depicts that user u visits venue w (whose functional category is c) at time t. Several examples of check-ins in Foursquare are shown in Table 1. The main goal of our study is to learn interpretable venue representations whose coordinates have clear and understandable meanings from check-in data. To achieve this goal, we first propose a check-in embedding model, which constructs both the venue and category contexts of a target venue from users' check-in sequences and projects both venues and categories into a latent embedding space. Next, we devise an explainable embedding model to learn interpretable venue representations. As the categories of venues contain coherent and understandable semantics, XEM takes them as the meaningful interpretation of each dimension of venue representations. Specifically, XEM computes the similarities between a venue and these categories based on the embedding vectors learned from CEM and takes these similarity scores as the numerical values of a venue's representation. The notations used in the paper are listed in Table 2.

4. Our check-in embedding model

Recent works have shown that the latent representation model Word2vec can effectively capture the semantic relationships in word spaces based on the distributed semantic assumption [19]. A user's check-in sequence is naturally analogous to a sentence, where a venue in the sequence corresponds to a word in a sentence. Furthermore, the analysis of Foursquare check-in records from Tokyo and New York reveals that, similar to the word frequencies, venue frequencies also follow a power law distribution, as shown in Fig. 1. The observations naturally lead to the use of the word embedding model and its underlying distributional semantic assumption to the study of venues and categories in the latent embedding space.

The framework of the proposed Check-in Embedding Model (CEM) is shown in Fig. 2. In the upper left plot, we show a sampled check-in sequence, which is generated by sorting the check-ins of a user over a (configurable) period of time in chronological order. The sequence is described by two parallel subsequences: a venue sequence and the corresponding category sequence of the venues. With these check-in sequences, we model the relations between the target venue and the two corresponding types of contexts (including the venue context and the category context), as shown in the diagram at right. Finally, as shown in the bottom left plot, semantically related venues and categories are projected close to each other in the latent space.

4.1. Modeling the venue context

We model the sequential patterns using the venue sequences to learn the venue embedding vectors. More specifically, given a user, we fetch the venues in her or his check-in tuples to construct a chronologically ordered sequence of venues, $S_u = (w_1, \ldots, w_i, \ldots)$. Given a venue sequence S_u, the context of venue $w_i \in S_u$ contains the venues visited before and after w_i within a predefined window size k. The objective of venue representation learning is to maximize the log probability of the venue context for given w_i:

$$\ell^s_{w_i} = \sum_{i-k<=j<=i+k, j!=i} \log p(w_j|w_i). \tag{1}$$

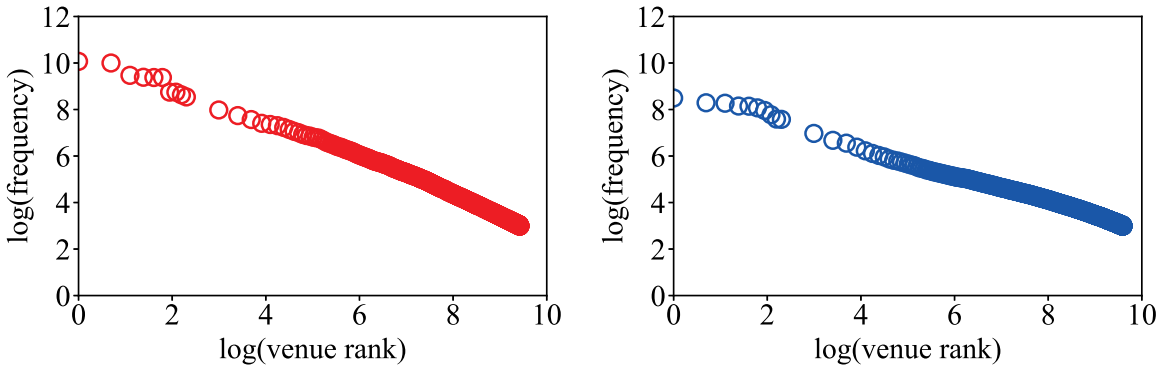
The probability of $p(w_j|w_i)$ is defined with a softmax function:

$$p(w_j|w_i) = \frac{\exp(\mathbf{v}^T_{w_j} \cdot \mathbf{v}'_{w_i})}{\sum_{n=1}^{N} \exp(\mathbf{v}^T_{w_n} \cdot \mathbf{v}'_{w_i})}, \tag{2}$$

where $\mathbf{v}'_{w_i}$ and $\mathbf{v}_{w_j}$ are the latent embedding vectors of the target and context venues, respectively, and N is the number of unique venues in these check-ins.

4.2. Modeling the category context

Each venue is usually associated with a category label, and a category sequence can be generated by considering the category labels of each venue in the sequence, as shown in Fig. 2. In practice, it is not rare that two very different venue sequences correspond to the same or similar category sequence. as many venues share category labels (e.g., *Noodle House* → *Bus Station* → *Gym/Fitness Center*). Such category sequence information imposes additional constraints on the representation of venues and, if properly utilized, could lead to improved representation quality. Therefore, we also model the category context of venues in addition to venue context to enhance the representations. More



(a) Tokyo (b) New York

Fig. 1. Log–log venue rank-frequency plot.

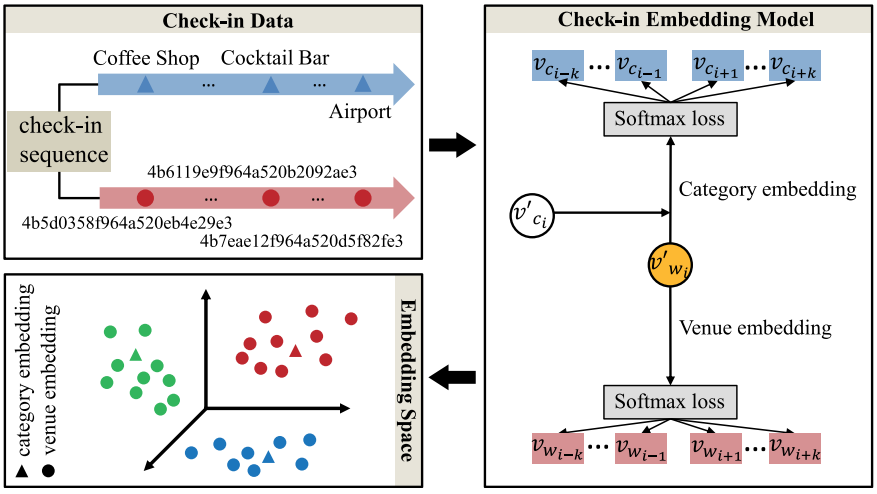


Fig. 2. Graphical representation of the proposed CEM.

specifically, given a venue w_i in sequence S_u, we consider the categories of venues in the linear context as the category context, and the combination of the venue w_i and its category c_i as the target. The objective can be defined as follows:

$$\ell^c_{w_i} = \sum_{i-k<=j<=i+k, j!=i} \log p(c_j|w_i, c_i), \tag{3}$$

where c_i is the category of w_i, and c_j is the categories of venues in the linear context. The probability $p(c_j|w_i, c_i)$ is estimated using the softmax function as follows:

$$p(c_j|w_i, c_i) = \frac{\exp(\mathbf{v}^T_{c_j} \cdot \mathbf{v}'_{wc})}{\sum_{m=1}^{M} \exp(\mathbf{v}^T_{c_m} \cdot \mathbf{v}'_{wc})}, \tag{4}$$

where $\mathbf{v}'_{c_i}$ and $\mathbf{v}_{c_i}$ are the embeddings of the target category and the context category, respectively, and $\mathbf{v}'_{wc}$ is the mean of vectors $\mathbf{v}'_{w_i}$ and $\mathbf{v}'_{c_i}$. M is the number of categories in these check-ins.

The unified objective of our check-in embedding model CEM Considering both the venue and category contexts is defined below:

$$\ell = \sum_{S_u \in \mathcal{S}} \sum_{i=1}^{N_u} \sum_{i-k<=j<=i+k, j!=i} [\log p(w_j|w_i) + \log p(c_j|w_i, c_i)], \tag{5}$$

where $\mathcal{S}$ is the set of check-in sequences and N_u is the length of sequence S_u. To learn the embedding vectors of venues and categories, we adopt the negative sampling method [7,19], which is known to be effective and efficient for the task. To summarize, CEM minimizes the objective function in Eq. (5), whereby it simultaneously learns the sequential and semantic patterns of venues in check-ins.

5. Interpretable venue representations

While the proposed CEM learns embedding vectors of venues and categories and maps semantically similar venues and categories in proximity in the vector space, these representations still lack of clear semantics. In this section, we propose a new model named XEM to learn how to explain each dimension of venue representations using semantic information.

The framework of the proposed XEM is shown in Fig. 3. With XEM, we fetch the venue embedding vectors and category embedding vectors learned from CEM, take the categories appearing in the check-in records as the semantic anchors, and compute the similarities between the given venue and these anchors based on these embedding vectors. Finally, we utilize these similarity scores as the numerical values of the new venue representations, in which each dimension corresponds to a category, as shown in the upper diagram of Fig. 3.

Let $\mathcal{C} = \{c_1, \ldots, c_M\}$ be the set of all categories in the check-in records. To represent a venue with semantics, a traditional Bag-of-Words (BoW) model selects these categories as the semantic anchors and represents a venue with an M-dimensional vector via exact matching. That is, if the category of a given venue w_i matches a semantic anchor, the output of the corresponding

Table 3
Check-in data statistics.

	♯user	♯venue	♯category	♯category-2	♯category-3	♯check-in
Foursquare (TKY)	9,548	10,321	103	73	27	1,270,977
Foursquare (NYC)	11,097	13,923	138	94	39	799,825
Yelp	41,572	33,884	109	76	11	2,436,826

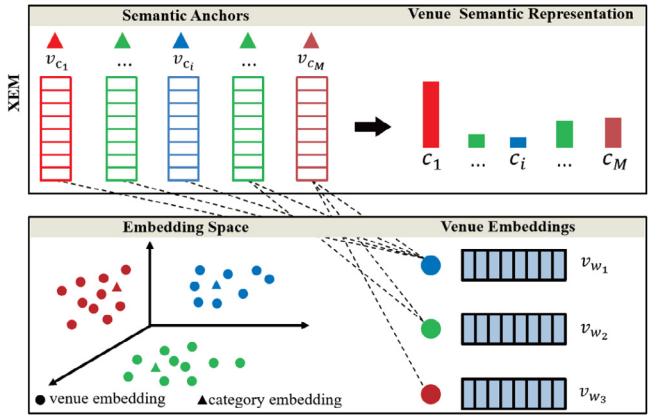


Fig. 3. Graphical representation of the proposed XEM.

mapping function will be 1, and 0 otherwise. Exact matching, equivalent to hard or crisp mapping, suffer from some obvious drawbacks: (1) the generated venue representation is pretty sparse, and (2) the representation does not encode sufficient semantic information.

To tackle these issues incurred by exact matching in BoW model, we propose to use semantic matching which evaluates the similarity between a venue and a semantic anchor based on the cosine similarity between the corresponding embedding vectors. Given a venue w_i and M semantic anchors (i.e., the categories), the new venue semantic representation is denoted as $\mathbf{v}^s_{w_i} = [Sim(w_i, c_1), \ldots, Sim(w_i, c_i), \ldots, Sim(w_i, c_M)]$. Specifically, the similarity score $Sim(w_i, c_i)$ between a venue w_i and a specific semantic anchor c_i is defined as follows:

$$Sim(w_i, c_i) = \begin{cases} \cos(w_i, c_i), & \text{if } \cos(w_i, c_i) > \lambda \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where $\cos(w_i, c_i)$ is the cosine similarity between w_i and c_i, and λ is a threshold. We evaluate the influence of the parameter λ on performance in the experiments. $\cos(w_i, c_i)$ can be calculated using the learned representations of w_i and c_i in the CEM:

$$cos(w_i, c_i) = \frac{\mathbf{v}_{w_i} \cdot \mathbf{v}_{c_i}}{\|\mathbf{v}_{w_i}\| \|\mathbf{v}_{c_i}\|}, \tag{7}$$

where $\mathbf{v}_{w_i}$ and $\mathbf{v}_{c_i}$ denote the embedding vectors of a venue w_i and a category c_i, respectively. The cosine similarity is positive when the venue w_i is semantically similar to the category c_i.

The BoW model employs exact venue matching to semantic anchors (i.e., categories in our study), which only counts the anchor identical to the category of a given venue. In contrast to the BoW model, the proposed XEM allows "vagueness" in the matching venues with semantic anchors, which is able to reduce sparsity, improve robustness and encode more semantic information. Specifically, XEM activates any anchor semantically similar to the venue. For a venue, the value in its representation corresponding to a semantic anchor is calculated according to the semantic similarity between this venue and the anchor. As the representations of venues and categories encode the semantics into vectors, the score of semantic similarity between a venue and a semantic anchor can be calculated using the cosine similarity between the corresponding vectors, as shown in Eq. (7). The cosine similarity could be interpreted as the degree of one venue semantically matching a semantic anchor. XEM would produce similar vectors for semantically similar venues (e.g., *Burger Joint* and *Food Court*).

Table 4
Parameters of CEM and XEM.

Parameters	Tested settings
Embedding size (d)	50, 100, 150, 200, 250, 300
Threshold (λ)	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

6. Experiments

In this section, we conduct experiments on real-world check-in datasets from Foursquare and Yelp to validate the effectiveness of the proposed models. We begin with introducing the datasets and baselines, and then show the evaluation results on two tasks and the qualitative analysis of venue representations.

6.1. Datasets and settings

Foursquare data. We carry out experiments on two real-world check-in datasets (publicly available at Foursquare), collected from New York (NYC) and Tokyo (TKY) during April 2012 to September 2013 [36,37]. Each check-in record in the datasets contains four properties: user ID, venue ID, category name, and check-in time. As data cleaning, we filter out users with number of check-ins fewer than 20 and venues with number of check-ins fewer than 20. The statistical properties of the two datasets are shown in Table 3, where ♯user, ♯venue, ♯category, ♯check-in are the number of users, venues, categories, and check-ins, respectively. Foursquare organizes venue categories using a five-layer tree structure,[1] and we use ♯category-2 and ♯category-3 to represent the number of categories from the second and third layers of the category tree in our datasets.

Yelp data. We get another dataset from Yelp.[2] We regard a user's review record as a check-in record, and randomly choose one of the non-leaf categories (Yelp organizes these categories using a four-layer tree structure) of a venue as its category. We filter those users whose number of reviews is fewer than 20 and those venues whose number of reviews is fewer than 20. The statistical properties of the Yelp data is shown in Table 3.

Baselines. We compare the performance of the proposed models with the state-of-the-art methods listed below:

- **GTM** [38]: a generative model based on LDA for discovering the geographic topics in which a venue is viewed as a word and a user's check-in sequence is modeled as a document. GTM learns a topic-based distribution for each venue.
- **STES** [18]: a spatiotemporal embedding similarity algorithm. As we do not include check-in timestamps in the embedding model, we adapt STES by considering the venue as the feature word and train a vector representation for each venue with the venue sequences.

[1] https://developer.foursquare.com/docs/build-with-foursquare/categories/.
[2] https://www.yelp.com/dataset/.

Table 5
Illustrative cases of interpretable venue representations.

Venue ID	4b780e35f964a520c0b32ee3	4b5bdb3df964a520131a29e3
Top 5 Dimensions	*Dumpling Restaurant*@0.609 *Food Court*@0.570 *Sporting Goods Shop*@0.522 *Theater*@0.467 *Bike Shop*@0.472	*Dumpling Restaurant*@0.572 *Food Court*@0.499 *Sporting Goods Shop*@0.469 *Theater*@0.449 *Bike Shop*@0.436
Actual Category	Mall	Mall
Venue ID	4bd5266b637ba5939913f670	4e53023d45ddffa8d188a50e
Top 5 Dimensions	*Airport Gate*@0.624 *Scenic Lookout*@0.553 *Airport Lounge*@0.524 *Food Court*@0.536 *Burger Joint*@0.509	*Airport Gate*@0.641 *Scenic Lookout*@0.559 *Airport Lounge*@0.550 *Food Court*@0.533 *Bike Shop*@0.536
Actual Category	Airport	Airport Terminal

- **Geo-Teaser** [13]: a geotemporal sequential embedding rank model that incorporates the personal and temporal information into the Skip-gram model. We omit the check-in time for fair comparison since such information is not considered in our models.
- **MC-TEM** [15]: a multi-context trajectory embedding model in which the check-in venue and its corresponding category are considered as context, and the CBOW model is deployed to learn trajectory attribute embeddings.

The parameters of the proposed models used in the experiments are provided in Table 4. Grid search is employed to identify the optimal parameters with a small but adaptive step size. In addition, for CEM, we set the context window size to 5, the learning rate (η) to 0.01, and the regularization term to 0.001. XEM takes all categories as the semantic anchors.

6.2. Qualitative analysis of interpretable venue representations

A major merit of the proposed XEM is that each dimension of venue representations can be interpreted as a coherent and easy-to-understand topic. Therefore, we use several examples to examine whether XEM could capture semantics with each of its dimensions. Given two venues w_i and w_j, we first fetch their semantic representations $\mathbf{v}^s_{w_i}$ and $\mathbf{v}^s_{w_j}$ learned with XEM, where each dimension of $\mathbf{v}^s_{w_i}$ and $\mathbf{v}^s_{w_j}$ is denoted by a category. Then, for each dimension a, we calculate the value of $(\mathbf{v}^s_{w_i a} - \mathbf{v}^s_{w_j a})/(\mathbf{v}^s_{w_i a} + \mathbf{v}^s_{w_j a})$ and sort them in ascending order. Finally, we retrieve the top 5 dimensions and leverage them to represent the common features between the two venues, which could reveal in which topics the two venues are similar.

Table 5 reports two illustrative cases, where *Dumpling Restaurant*@0.609 represents the topic and the value of the dimension. For a pair of venues with the same category *Mall*, we find that the top 5 dimensions are *Dumpling Restaurant*, *Food Court*, *Sporting Goods Shop*, *Theater*, and *Bike Shop*, which could summarize the features of the venues. We observe similar results for the two venues with categories *Airport* and *Airport Terminal*. Therefore, when querying for venues similar to a given venue, we can explain why the recommended venues are relevant to the query based on the categories corresponding to the top dimensions.

6.3. Evaluation of venue similarity

The proposed models represent venues with embedding vectors. We expect that semantically related venues will be close to each other in the embedding space. To evaluate to what degree these venue representations retain the semantics, we utilize a binary evaluation scheme to measure venue similarities. For this task, we randomly generate 10,000 triplets with each element in the triplet being a venue. In each triplet, two venues share a common category label, and the other has a different category label. The task is to identify the venue in each triplet that has a different category label than the other two.

To evaluate the venue representations generated by the proposed models, for each triplet, we calculate the pairwise similarity score using 2-combination. More specifically, for triplet (w_1, w_2, w_3), we calculate the similarity scores of the three pairs (w_1, w_2), (w_1, w_3) and (w_2, w_3). We identify the one with the highest score and return the other venue as the result for this test. For instance, if (w_1, w_2) has the highest score, then w_3 will be the result. We evaluate the accuracy of different models over all triplets.

Table 6
Performance comparison in terms of accuracy using the Foursquare(TKY) data.

Methods	Embedding size					
	50	100	150	200	250	300
GTM [38]	0.366	0.371	0.373	0.366	0.367	0.361
STES [18]	0.401	0.410	0.408	0.417	0.401	0.405
Geo-Teaser [13]	0.438	0.438	0.440	0.436	0.427	0.438
MC-TEM [15]	0.390	0.395	0.399	0.398	0.392	0.400
CEM	0.528	0.535	0.545	0.536	0.531	0.529
XEM	**0.566**	**0.623**	**0.658**	**0.656**	**0.660**	**0.660**

Table 7
Performance comparison in terms of accuracy using the Foursquare(NYC) data.

Methods	Embedding size					
	50	100	150	200	250	300
GTM [38]	0.400	0.419	0.404	0.416	0.411	0.418
STES [18]	0.432	0.441	0.429	0.430	0.428	0.424
Geo-Teaser [13]	0.436	0.445	0.425	0.439	0.434	0.434
MC-TEM [15]	0.403	0.424	0.408	0.417	0.421	0.421
CEM	0.554	0.581	0.560	0.572	0.564	0.562
XEM	**0.581**	**0.652**	**0.681**	**0.693**	**0.706**	**0.701**

6.3.1. Comparison with baselines

In our experiments, we compare different methods on the same datasets and perform 10 runs to report the average accuracy. Tables 6–8 show the accuracy with different embedding sizes of all the methods on the Foursquare and Yelp datasets correspondingly (where the best scores are highlighted in boldface), from which we can make the following observations.

- GTM performs worst among all methods, as it only models the co-occurrences of venues in a check-in sequence and ignores the order of venues.
- STES, Geo-Teaser and MC-TEM perform better than GTM, as they model the sequential patterns of check-in sequences

Table 8
Performance comparison in terms of accuracy using the Yelp data.

Methods	Embedding size					
	50	100	150	200	250	300
GTM [38]	0.379	0.412	0.418	0.310	0.385	0.402
STES [18]	0.356	0.285	0.356	0.286	0.333	0.370
Geo-Teaser [13]	0.319	0.291	0.332	0.366	0.372	0.368
MC-TEM [15]	0.305	0.408	0.404	0.402	0.394	0.427
CEM	0.470	0.507	0.455	0.509	0.481	0.471
XEM	**0.496**	**0.570**	**0.563**	**0.527**	**0.606**	**0.596**

to learn venue representations. More specifically, STES and Geo-Teaser utilize the preceding and succeeding venues of the target venue in the check-in sequences as context, and MC-TEM considers multiple contexts including the contextual venues and their categories, to predict the target venue and generate venue representations.

- CEM yields better results than STES, Geo-Teaser, and MC-TEM, as it takes into consideration the category context (which reflects aggregated semantics) in the check-in sequences to enhance venue representations in addition to the information integrated in other methods.
- XEM outperforms CEM and other baselines on the three datasets, validating the effectiveness of mapping the venue representations into a new semantic space based on these categories. For example, compared with Geo-Teaser, XEM achieves an average improvement of 46.12% for the Foursquare(TKY) data, 53.68% for the Foursquare(NYC) data, and 64.96% for the Yelp data for the various embedding sizes of venue representations. Furthermore, we perform paired t-tests for XEM and these baseline methods and conclude that the improvement of XEM over these baseline methods is of statistical significance with p value < 0.01 [39].

6.3.2. Model analysis

To further study the effectiveness of the proposed models, we design two variants of XEM, XEM_{second} and XEM_{third}, for comparison purpose. As Foursquare organizes venue categories using a tree structure, we take categories from different layers as the semantic anchors. XEM, XEM_{second}, and XEM_{third} use all categories, categories from the second layer and categories from the third layer as the semantic anchors, respectively. We report the comparison results using the Foursquare datasets in Fig. 4. From the results, we can see that XEM_{third} performs poorly because it only includes a small number of fine-grained categories as semantic anchors and thus cannot fully cover the semantic expressions. For example, as shown in Table 3, only 27 categories from the third layer are considered in the TKY dataset and 39 categories are considered in the NYC dataset.

6.3.3. Parameter sensitivity

In this section, we vary parameters of interests, including the embedding size (d) in CEM, and the threshold (λ) in XEM, and study their influences on the performance of the methods using the Foursquare datasets.

Effect of embedding size d. The size of the embedding vectors is an important parameter for latent representations. As reported in Tables 6 and 7, we vary embedding sizes in range 50 to 300 with a step interval of 50. As is clear from the figure, in most cases these models achieve better performance when the embedding size is larger than 200 for the fact that with larger embedding size the representations can capture finer-grained semantics.

Effect of threshold λ. The threshold λ in Eq. (7) would determine the similarity score between a venue and a semantic anchor, and consequently influence the performance of XEM. We vary λ in range 0 to 0.9 with a step of 0.1 and show the accuracy on both datasets in Fig. 5. Evidently, as λ increases, the accuracy improves and reaches a peak at $\lambda = 0.2$. When $\lambda > 0.2$, the accuracy drops dramatically, as the large threshold results in many zeros in the new venue representations. Hence, we set $\lambda = 0.2$ in other experiments.

Table 9
Venue categories and their percentages (x%,y%) in the Foursquare datasets.

Category	TKY(x%)	NYC(y%)
Arts & Entertainment (A&E)	6.6%	7.0%
College & University (C&U)	2.9%	4.1%
Food	13.5%	19.2%
Outdoors & Recreation (O&R)	6.9%	9.4%
Residence (Res)	2.6%	11.9%
Shop & Service (S&S)	23.1%	10.2%
Nightlife Spot (NS)	2.7%	9.7%
Travel & Transport (T&T)	31.5%	12.6%
Professional & Other Places (P&O)	10.2%	15.9%

Table 10
Performance comparison of venue semantic annotation.

Methods	Foursquare (TKY)		Foursquare (NYC)	
	micro-F1	macro-F1	micro-F1	macro-F1
GTM [38]	0.346	0.117	0.272	0.151
STES [18]	0.475	0.268	0.351	0.274
Geo-Teaser [13]	0.528	0.366	0.403	0.343
MC-TEM [15]	0.507	0.342	0.399	0.359
CEM	0.593	0.474	0.524	0.499
XEM	**0.607**	**0.521**	**0.530**	**0.536**

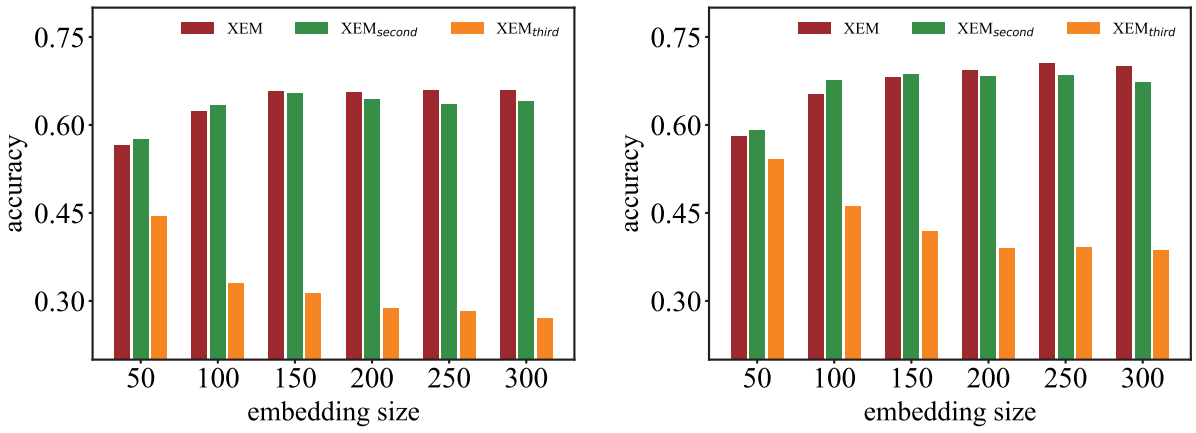
6.4. Evaluation of venue semantic annotation

Venue semantic annotation refers to the process of assigning a meaningful and suitable label to each venue. With check-in data, the semantic labels may be *Mall*, *Bar*, and *Restaurant*, given to places of interest (i.e., venues) in LBSNs. Such semantic labels are important for people to infer activities, explore new places, and conduct recommendation services. Despite their benefits, studies [40,41] analyzing the check-ins collected from Foursquare observe that approximately 30% of venues lack semantic labels. It is thus highly desirable to automatically and precisely annotate a venue using labels from a database of semantic categories.

Venue semantic annotation can be formulated as a multi-class classification problem mapping a representation to one of the many labels. For the task, we first learn venue representations using the proposed models. Note that we randomly replace the categories of 20% of all venues with the "NULL" tag and then learn the venue representations using these data. We then choose the nine top categories in Table 9 to form the evaluation set. We take these venues with category labels as the training set and the rest with "NULL" tags as the test set. Taking the venue representations as features, we apply an SVM [42] to perform the classification, i.e., mapping a representation to a category. To evaluate the performance, we adopt two metrics: the micro-F1 score and the macro-F1 score.

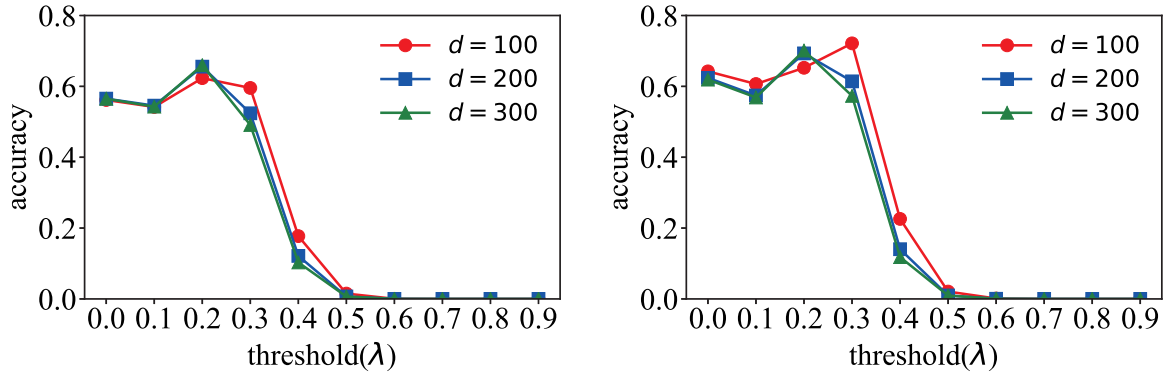
We report the overall micro-F1 scores and macro-F1 scores in Table 10, with the best results highlighted in bold. From this table, we have the following observations.

- GTM performs the worst, as it adopts generative method to capture the global venue distribution without modeling the sequential patterns in the check-in records.



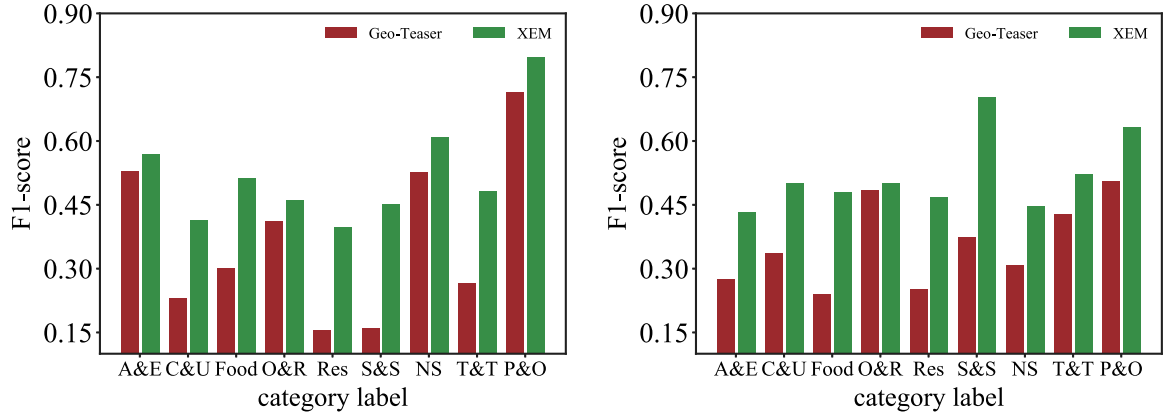
(a) Foursquare(TKY) (b) Foursquare(NYC)

Fig. 4. The performance of XEM and its variants.



(a) Foursquare(TKY) (b) Foursquare(NYC)

Fig. 5. Effect of threshold λ in XEM.



(a) Foursquare(TKY) (b) Foursquare(NYC)

Fig. 6. Performance comparison per venue category on venue semantic labeling based on Geo-Teaser and XEM.

- STES, Geo-Teaser and MC-TEM perform better than GTM, indicating that it is effective to take the latent venue representations as the classification features, as these representations could retain venue semantics. However, they perform worse than CEM, as they do not consider the category context in the check-in sequences that reflect aggregated semantics when learning venue representations.
- CEM obtains good performance because it is able to learn representations of venues and categories that preserve certain semantics. Among all the methods, XEM performs the best. More specifically, compared with Geo-Teaser, XEM achieves an improvement of 42.35% on the Foursquare(TKY) data and 56.27% on the Foursquare(NYC) data in terms of

macro-F1. The results demonstrate that XEM is more effective than the other competitors for venue category classification. Furthermore, based on the results of the paired t-tests, we know that the improvement in XEM over these baselines is of statistical significance (under significance level 0.01) [39].

Furthermore, we report the venue semantic labeling results per venue category of Geo-teaser and XEM in Fig. 6. It is evident that XEM is superior to Geo-Teaser for every venue category in both datasets.

7. Conclusion and future work

We have proposed a framework to learn interpretable venue representations from users' check-ins, where each dimension of the venue representations can be represented as a coherent and easy-to-understand topic. We start with a category-aware check-in embedding model named CEM, which models both the sequential patterns and the categorical information of venues to learn embedding vectors of venues and categories simultaneously. Next, we present XEM, which takes easy-to-understand categories as semantic anchors and designs a mapping scheme based on semantic correlation between venues and anchors to yield interpretable venue representations. Furthermore, we evaluate the performance of the proposed models on real check-in datasets with quantitative and qualitative tasks. Experimental results show that the proposed models outperform the baselines across various settings.

Several interesting research problems exist for further exploration. First, it will be worth investigating different weighting schemes for the semantic anchors in computing the similarity scores. Second, as we adopt a two-step strategy to learn interpretable venue representations, we can consider how to build a joint model to yield interpretable representations from check-ins directly.

CRediT authorship contribution statement

Ning An: Conceptualization, Methodology, Software, Writing – original draft. **Meng Chen:** Methodology, Supervision, Writing – review & editing. **Li Lian:** Data curation. **Peng Li:** Resources. **Kai Zhang:** Writing – original draft. **Xiaohui Yu:** Writing – review & editing. **Yilong Yin:** Supervision.

Declaration of competing interest

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

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61906107, the Natural Science Foundation of Shandong Province of China under Grant No. ZR2019BF010, the Young Scholars Program of Shandong University.

References

[1] S. Feng, L.V. Tran, G. Cong, L. Chen, J. Li, F. Li, HME: A hyperbolic metric embedding approach for next-POI recommendation, in: SIGIR, 2020, pp. 1429–1438.
[2] Y. Si, F. Zhang, W. Liu, An adaptive point-of-interest recommendation method for location-based social networks based on user activity and spatial features, KBS 163 (2019) 267–282.
[3] J. Manotumruksa, C. Macdonald, I. Ounis, A contextual recurrent collaborative filtering framework for modelling sequences of venue checkins, IPM 57 (6) (2020) 102092.
[4] M. Aliannejadi, F. Crestani, Personalized context-aware point-of-interest recommendation, TOIS 36 (4) (2018) 1–28.
[5] H. Yin, W. Wang, H. Wang, L. Chen, X. Zhou, Spatial-aware hierarchical collaborative deep learning for POI recommendation, TKDE 29 (11) (2017) 2537–2551.
[6] M. Shi, D. Shen, Y. Kou, T. Nie, G. Yu, Attentional memory network with correlation-based embedding for time-aware POI recommendation, KBS 214 (2021) 106747.
[7] M. Chen, X. Yu, Y. Liu, MPE: A mobility pattern embedding model for predicting next locations, WWWJ 22 (6) (2019) 2901–2920.
[8] W. Liang, W. Zhang, X. Wang, Deep sequential multi-task modeling for next check-in time and location prediction, in: DASFAA, 2019, pp. 353–357.
[9] Y. Duan, W. Lu, W. Xing, P. Bao, X. Wei, PBEM: A pattern-based embedding model for user location category prediction, in: ICMU, 2019, pp. 1–6.
[10] B. Yan, K. Janowicz, G. Mai, S. Gao, From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts, in: SPATIAL, 2017, pp. 1–10.
[11] D.Y. Zhang, D. Wang, H. Zheng, X. Mu, Q. Li, Y. Zhang, Large-scale point-of-interest category prediction using natural language processing models, in: BigData, 2017, pp. 1027–1032.
[12] X. Liu, Y. Liu, X. Li, Exploring the context of locations for personalized location recommendations, in: IJCAI, 2016, pp. 1188–1194.
[13] S. Zhao, T. Zhao, I. King, M.R. Lyu, Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation, in: WWW, 2017, pp. 153–162.
[14] W.X. Zhao, N. Zhou, A. Sun, J.-R. Wen, J. Han, E.Y. Chang, A time-aware trajectory embedding model for next-location recommendation, KAIS 56 (3) (2018) 559–579.
[15] N. Zhou, W.X. Zhao, X. Zhang, J.-R. Wen, S. Wang, A general multi-context embedding model for mining human trajectory data, TKDE 28 (8) (2016) 1945–1958.
[16] D. Camacho, A. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, E. Cambria, The four dimensions of social network analysis: An overview of research methods, applications, and software tools, Inf. Fusion 63 (2020) 88–120.
[17] H. Ying, J. Wu, G. Xu, Y. Liu, T. Liang, X. Zhang, H. Xiong, Time-aware metric embedding with asymmetric projection for successive POI recommendation, WWWJ 22 (5) (2019) 2209–2224.
[18] J. Yang, C. Eickhoff, Unsupervised learning of parsimonious general-purpose embeddings for user and location modeling, TOIS 36 (3) (2018) 1–33.
[19] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: NIPS, 2013, pp. 3111–3119.
[20] O. Levy, Y. Goldberg, Neural word embedding as implicit matrix factorization, in: NIPS, 2014, pp. 2177–2185.
[21] B. Chang, Y. Park, D. Park, S. Kim, J. Kang, Content-aware hierarchical point-of-interest embedding model for successive POI recommendation, in: IJCAI, 2018, pp. 3301–3307.
[22] D. Yao, C. Zhang, J. Huang, J. Bi, SERM: A recurrent model for next location prediction in semantic trajectories, in: CIKM, 2017, pp. 2411–2414.
[23] Y. Yu, S. Tang, K. Aizawa, A. Aizawa, Category-based deep CCA for fine-grained venue discovery from multimodal data, TNNLS 30 (4) (2018) 1250–1258.
[24] D. Kong, F. Wu, HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction, in: IJCAI, 2018, pp. 2341–2347.
[25] Q. Liu, S. Wu, L. Wang, T. Tan, Predicting the next location: A recurrent model with spatial and temporal contexts, in: AAAI, 2016, pp. 194–200.
[26] C. Yang, M. Sun, W.X. Zhao, Z. Liu, E.Y. Chang, A neural network approach to jointly modeling social networks and mobile trajectories, TOIS 35 (4) (2017) 1–28.
[27] Q. Wang, H. Yin, T. Chen, Z. Huang, H. Wang, Y. Zhao, N.Q. Viet Hung, Next point-of-interest recommendation on resource-constrained mobile devices, in: WWW, 2020, pp. 906–916.
[28] K. Sun, T. Qian, T. Chen, Y. Liang, Q.V.H. Nguyen, H. Yin, Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation, in: AAAI, 2020, pp. 214–221.
[29] A. Panigrahi, H.V. Simhadri, C. Bhattacharyya, Word2Sense: sparse interpretable word embeddings, in: ACL, 2019, pp. 5692–5705.
[30] A. Fyshe, L. Wehbe, P. Talukdar, B. Murphy, T. Mitchell, A compositional and interpretable semantic space, in: ACL, 2015, pp. 32–41.
[31] R. Zhao, K. Mao, Fuzzy bag-of-words model for document representation, TFS 26 (2) (2017) 794–804.
[32] B. Mathew, S. Sikdar, F. Lemmerich, M. Strohmaier, The polar framework: Polar opposites enable interpretability of pre-trained word embeddings, in: WWW, 2020, pp. 1548–1558.
[33] D. Li, J. Zhang, P. Li, Tmsa: a mutual learning model for topic discovery and word embedding, in: SDM, SIAM, 2019, pp. 684–692.

[34] G. Xun, Y. Li, J. Gao, A. Zhang, Collaboratively improving topic discovery and word embeddings by coordinating global and local contexts, in: KDD, 2017, pp. 535–543.
[35] A. Potapenko, A. Popov, K. Vorontsov, Interpretable probabilistic embeddings: bridging the gap between topic models and neural networks, in: AINL, 2017, pp. 167–180.
[36] D. Yang, D. Zhang, L. Chen, B. Qu, Nationtelescope: Monitoring and visualizing large-scale collective behavior in LBSNs, JNCA 55 (2015) 170–180.
[37] D. Yang, D. Zhang, B. Qu, Participatory cultural mapping based on collective behavior data in location-based social networks, TIST 7 (3) (2016) 1–23.
[38] X. Long, L. Jin, J. Joshi, Exploring trajectory-driven local geographic topics in foursquare, in: UbiComp, 2012, pp. 927–934.
[39] D. Hull, Using statistical testing in the evaluation of retrieval experiments, in: SIGIR, 1993, pp. 329–338.
[40] M. Ye, D. Shou, W.-C. Lee, P. Yin, K. Janowicz, On the semantic annotation of places in location-based social networks, in: KDD, 2011, pp. 520–528.
[41] T. He, H. Yin, Z. Chen, X. Zhou, S. Sadiq, B. Luo, A spatial-temporal topic model for the semantic annotation of POIs in LBSNs, TIST 8 (1) (2016) 1–24.
[42] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, TIST 2 (3) (2011) 1–27.