Point-of-Interest Demand Discovery Using Semantic Trajectories

Ying Jin[1], Guojie Ma[1(✉)], Shiyu Yang[1], and Long Yuan[2]

[1] East China Normal University, Shanghai, China
yjin@stu.ecnu.edu.cn, {xmlin,gjma,syyang}@sei.ecnu.edu.cn
[2] Nanjing University of Science and Technology, Nanjing, China
longyuan@njust.edu.cn

Abstract. Semantic trajectories have become unprecedentedly available because of the rapidly growing popularities of location-sharing services. People's lifestyles and Point-of-Interest demands are hidden in such data. Extracting people's POI needs for different regions from semantic trajectories plays an important role in site selection, which can be widely used in city planning, facility location and other applications. However, most of existing works either use traditional trajectories which need to infer semantic with external information and lead to inaccuracy, or just focus on specific category. Semantic trajectory mining provides us a new way to address the challenges. Based on above motivation, we study the *regional POI demand discovery* problem using semantic trajectories. In this paper, we carefully analyze the features of semantic trajectory data and people' mobility patterns. Then, we propose an effective POI demand modeling method. Furthermore, we propose two efficient algorithms to identify the regional POI demands. The proposed algorithms extract regional patterns and compute the regional POIs demand according to POI demand model. Finally, the ranked POIs demands for regions are obtained. We evaluate the proposed modeling method and algorithms in terms of efficiency and effectiveness on two real data sets. The results show that our proposed methods outperform the competitor for both efficiency and effectiveness.

Keywords: POI discovery · Semantic trajectory · Mobility pattern mining

1 Introduction

With the rapid urbanization process, modern cities have developed urban regions with diverse functionality which naturally have a variety of demands for different Point-of-Interests (POIs) [28,29]. Discovery of regional POI demand, which can help governments allocate resources efficiently and provide suggestions for business investors, is becoming more and more important. Considering a cafe franchise want to open a new store in an urban region, with the help of regional

Y. Jin and G. Ma—The joint first authors.

Y. Nah et al. (Eds.): DASFAA 2020, LNCS 12113, pp. 280–296, 2020.
https://doi.org/10.1007/978-3-030-59416-9_17

POI demand, the investor can identify the potential demand for cafe in this region which is important for his investment decision.

Nowadays, POI demand analysis is still largely dependent on manual survey which is time-consuming. Besides, the markets and land resources are easy to change, if the analysis is conducted on the outdated data, the result inevitably leads to failure. Recently, with the advance of location positioning technique and high prevalence of location-sharing applications such as Foursquare, an increasing volume of semantic trajectories is extracted by combining traditional trajectory with semantic data [1], where each point in a semantic trajectory not only contains a particular address and a corresponding timestamp when person visited this site but also is enriched with a semantic category, such as cafe, gym or office. The appearance of semantic trajectories provides a new way for efficient and effective POI demand analysis which reflects the detailed underlying dynamics of residents in the city.

There are several approaches to recommend proper site location using human mobility data such as traditional or semantic trajectories using angle, velocity and other attributes. Unfortunately, most of existing works are proposed for specific demand [11,14,15,19,32], for instants, cultural planning or gas station. Specific focus leads to the ignorance of other significant categories and loss of generality of common demand discovery. Meanwhile another part of existing works focuses on traditional trajectory data [15], which means that they do not consider the semantic information and are not able to completely infer the POI category information.

In order to address these challenges, we study general POI demand discovery problem with an important economic logic *Foot Voting* [25]. Charles Tiebout points out that people have the ability to choose what they want by traveling. In other words, if people from one region frequently travel to other regions for specific category such as coffee shop, it is much likely that people need fresh or better coffee shop in their origin region. Therefore we can conclude that coffee shop is one of the POI demands of people's origin region. Based on the above observation and the help of people's frequent mobility patterns, the regional POI demands can be identified. Furthermore, governments and companies can understand POI demand better for future planning.

In this paper, we first propose a regional POI demand modeling which takes several relevant features into consideration and come up with a well designed regional POI demand modeling method. Then we develop efficient algorithms for regional POI demand discovery which consist of pre-processing of the raw semantic trajectories, cross-regional pattern mining and POI demand mining. Furthermore, we introduce optimization techniques which are based on some interesting observations to enrich the mining results. Finally, we evaluate our proposed method on two real datasets and show two illustrating cases in London and New York respectively. The experimental results show the effectiveness and efficiency of our proposed methods.

We conclude the main contributions of this paper as follows:

- We propose a POI demand modeling method based on the observation of *Foot Voting*.

– We design two efficient algorithms to compute the regional POI demand using semantic trajectories.
– We evaluate our proposed model and algorithms using real data sets and provide detailed analysis.
– Two case studies are introduced to show the effectiveness of the proposed methods.

The rest of this paper is organized as follows. We briefly review the related work in Sect. 2. In Sect. 3, we first define a few terms, then introduce our POI demand modeling and the formal problem definition. Our proposed mining algorithms are presented in Sect. 4 with some interesting observations. The experimental evaluation and illustrating case study are provided in Sect. 5. Finally, we conclude this paper in Sect. 6.

2 Related Work

2.1 Human Mobility Analysis

Understanding human mobility is crucial to location-based services and many related researches with mobility data, such as mobile phone data and transportation data [2,16,23]. [24] explores the urban Region-of-Interest to study agglomeration economies using online map search queries. [26] recommends a region with reliable POIs to a user with deep metric learning.

As we all know, each region has different specific needs. A good location has an effective impact on business and city planning. So correct POI demand analysis is key to site selection. With the help of mobility data, people can improve the accuracy and efficiency of site selection with comparison to traditional manual surveys or analysis models based on census [21,22]. Researchers mainly study specific POI category demand [11,13,14,19], for example, [32] designs a method to properly allocate cultural resources in urban area. However, [15] develops a systematic framework integrating POI and demographic data to identify various demands for developed and underdeveloped regions using traditional trajectories.

Human lifestyle [8] is another research field improved by the popularity of mobility data [20]. Some studies try to find the relation between multiple types of human lifestyles [31] and [10] uses shopping records to extract shopping patterns for divergent urban regions incorporating mobility patterns.

Existing POI demand identification works mainly focus on traditional human mobility pattern which may lead to mis-inference. And researches on assisting decision making with semantic data is few. Different from the above works, we design a more general framework for revealing people's lifestyles in cross-regional behaviors extracted from semantic trajectories to identify their life demands of each region. And recommend POI category from both region and semantic category perspectives in the framework.

2.2 Trajectory Pattern Mining

Trajectory pattern mining is a hot topic in spatial-temporal data mining. [18] explores propagation patterns and influential patterns in traffic and weather data with LSTM. According to the *Foot Voting* principle [25], people's movement can reflect POI demands to some extent. Therefore, frequent pattern mining using semantic trajectories [3,6] can be adopted to discover the POI demand. [9] proposes spatio-temporal containment pattern which requires similar transition time on visiting same sequences of places. [30] utilizes collaborative group of similar POIs rather than independent POIs to mine fine-grained frequent patterns. [12] proposes a probabilistic model to capture movement between semantic regions with coherent topic. They all focus on extracting globally frequent patterns in the entire data space. [7] proposes a new density scheme to quantify frequency of locally significant sequential patterns based on clustering. [4,5] study co-movement patterns problem which is closely related to frequent pattern mining.

However, there still exist challenges on mining POI demands due to the sparsity of trajectory data, especially for the regional POI demands discovery problem. People's different destinations usually distribute in various regions and they often start from diverse origins. This problem becomes even more challenging if we consider the semantic information for each visits in people's trajectories.

3 Preliminaries

In this section, we first define some important terms. Then we introduce the POI demand modeling using semantic trajectories. Finally we formally define the POI demand discovery problem.

3.1 Semantic Trajectories

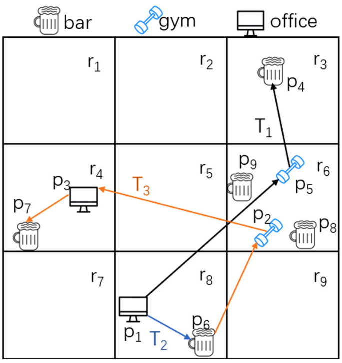
Let $\mathcal{C} = \{c_1,\ c_2,\ \cdots,\ c_n\}$ is a set of semantic categories. Let $\mathcal{P} = \{p_1,\ p_2,\ \cdots,\ p_m\}$ be a set of places where each $p \in \mathcal{P}$ is defined as a tuple $(p.lon, p.lat, p.cat)$ where $p.lon$ and $p.lat$ denote p's latitude and longitude respectively and $p.cat \in \mathcal{C}$ is p's category. Following the existing definitions [7,30], we define *semantic trajectory* as follows:

Definition 1 *(Semantic Trajectory)***.** *A semantic trajectory T is defined as a sequence of pairs of place and a corresponding timestamp $\langle(p_1, t_1),\ (p_2, t_2),\ \cdots,\ (p_l, t_l)\rangle$, where $t_i < t_j$ if $i < j$ and l is the length of the trajectory.*

3.2 Region Partition

There are several methods to partition the urban area into regions, such as grid-based [17], road network-based and neighborhood-based [15] method. For the ease of presentation, we adopt grid-based method for region partition. Please note our proposed method can be used for other partition methods directly, as mapping POIs to regions does not rely on the partition methods.

Given a urban map, we divide the urban area into a set of $g \times g$ grid cells. Then we get a set of regions $\mathcal{R} = \{r_1,\ r_2,\ \cdots,\ r_{|\mathcal{R}|}\}$, where $|\mathcal{R}|$ equals to $g \times g$. According to POIs' location in semantic trajectory T, we can assign each of them into a region grid, as shown in Fig. 1.



Trajectory	Sequence of POIS	Cross-Regional Patterns
T_1	<p_1, p_5, p_4>	<(o, r_8), (g, r_6)> <(g, r_6), (b, r_3)>
T_2	<p_1, p_6>	
T_3	<p_6, p_2, p_3, p_7>	<(b, r_8), (g, r_6)> <(g, r_6), (o, r_4)>

Fig. 1. An example of semantic trajectories (in different colors), T_1, T_2, T_3 and Regional Pattern, where $p_1, p_2, \cdots, p_9$ are places of interest (POIs), $r_1, r_2, \cdots, r_9$ are spatial regions and o, g, b are categories of POIs denoting office, gym and bar respectively.

After mapping the POIs into grid regions, we can generate regional patterns from a semantic trajectories according to the following definition.

Definition 2 *(Regional Pattern)***.** *Given a semantic trajectory T, a regional Pattern O of T is a sequence of tuples $O = \langle(c_1,\ r_i),\ (c_2,\ r_j), \cdots,\ (c_m,\ r_m)\rangle$, where each $c_i \in C$ and r_i, r_j and $r_m \in R$.*

A regional pattern O may contain several intra-region movements, e.g. movement from p_3 to p_7 in trajectory T_3 in Fig. 1. Such movement does not provide the POI demand information as the user's POI needs can be satisfied in the given region, say r_4 in the example. Therefore we cannot infer any POI demand for this region by considering the intra-region movements. In order to mine the region POI demand, we define the cross-region pattern as follows:

Definition 3 *(Cross-Regional Pattern)***.** *For any adjacent category-region pairs $\langle c_n,\ r_i\rangle$ and $\langle c_m,\ r_j\rangle$ in a given regional pattern, if $i \neq j$, we call such category-region pair as cross-regional pattern.*

A toy example to clarify the above definitions is as following.

Example 1. As shown in Fig. 1, given 3 semantic categories T_1, T_2 and T_3, there are 4 cross-regional pattern $\langle(o,\ r_8),\ (g,\ r_6)\rangle$, $\langle(g, r_6),\ (b, r_3)\rangle$, $\langle(b,\ r_8),\ (g,\ r_6)\rangle$ and $\langle(g,\ r_6),\ (o,\ r_4)\rangle$.

Next we define the regional category number to reflect the POIs appearance in a given region as follows.

Definition 4 *(Regional Category Number). Given a region r_i, we use regional category number $n_{i,j}$ to denote the count of POI with category c_j in region r_i.*

We maintain a list $F_{r_i} = \langle(c_1,\ n_{i,1}),\ (c_2,\ n_{i,2}),\ \cdots,\ (c_m,\ n_{i,m})\rangle$ to count how many POIs of each kind of category c_i in this region r_i, where m is the number of categories in this region. Besides gathering regional category number for each kind of category in each region, we also count the global sum of category c_m and the total number of POIs in region r_j.

Due to the unbalanced development of urban regions, only counting the POI numbers may not reflect the real POI demand accurately. For example, the number of POIs in rural area is usually much smaller than that in city centers. Besides, if there already exists a lot of POIs of same category, we shouldn't suggest the same type of POI for this region because it will lead unbalanced region development and intensify competition which may lead vendor's failure. So it is important to formulate the regional category density.

Next, we define two types of semantic category density, namely, *global category density* and *local category density* which are calculated in the following two equations respectively.

$$d - global_{j,m} = \frac{n_{j,m}}{|c_m|}, \tag{1}$$

where $n_{j,m}$ is the number of category m in region j and $|c_m|$ is total number of category m in the dataset.

$$d - local_{j,m} = \frac{n_{j,m}}{|POI_j|}, \tag{2}$$

where $|POI_j|$ is total number of POI in region j.

Based on the density defined above, we propose regional category density to reveal the category distribution information for a given region as follows.

Definition 5 *(Regional Category Density). A density list of given region r_j, $D_{r_j} = \langle(c_1,\ d_{j,1}),\ (c_2,\ d_{j,2}),\ \cdots,\ (c_m,\ d_{j,m})\rangle$, where m is the total number of semantic categories in this region r_j. A regional category density $d_{j,m} = d - global_{j,m} \times d - local_{j,m}$.*

The key idea of calculating regional category density is that only using local-level density cannot reflect the actual development of region. Assuming this region has very few vendors, any categories in this region can have a high local-level density value, however people need more various vendors. We will omit their demands if we only focus regions with lower local-level density. And high global-level density doesn't mean that they can fulfill each person's need because it may be a very flourishing area which need more shops than other regions.

3.3 Regional POI Demand Modeling

Before introducing the regional POI demand modeling, we first introduce need number which quantifies people's demand for a specific category in the region.

Given a cross-regional pattern E, any length-2 sub-pattern $\langle$(origin.category, origin.region), (destination.category, destination.region)$\rangle$ in E, we update origin region need number of destination's category $\langle origin.region, destination.category\rangle$ by adding $\frac{1}{d}$, where d is regional category density of destination's category in origin's region. The formal definition of need number is given as follow.

Definition 6. *(Need Number) $need_{j,i}$ is the need number representing demand of c_i in r_j. Every time the demand of c_i in r_j emerges in cross-regional pattern E, $need_{j,i}$ adds $\frac{1}{d_{j,i}}$.*

The need number reflects the demand of category in the given region. We use $\frac{1}{d}$ as the need number increase step. According to our analysis in Sect. 3.2, the regional category density d reflects the influence of local density and global density of the category in a given region which is important when considering the unbalance development in urban area. A toy example to show how to compute the need number is provided as follow.

Example 2. As shown in Fig. 1, T_1 is a cross-regional pattern, and there exists 2 length-2 sub-pattern $\langle(o,\ r_8),\ (g,\ r_6)\rangle$ and $\langle(g,\ r_6),\ (b,\ r_3)\rangle$. And T_3 contains a length-3 cross-regional pattern, which includes 2 length-2 sub-pattern $\langle(b,\ r_8),\ (g,\ r_6)\rangle$ and $\langle(g,\ r_6),\ (o,\ r_4)\rangle$. We take bar in region r_6 as example. Bar in region r_6's local density is $\frac{2}{4}$ and its global density is $\frac{2}{5}$, so final regional category density of Bar in r_6 is 0.2, and r_6's need number of bar is $\frac{1}{0.2} = 5$.

The need number $need_{i,j}$ quantifies the demand of category c_j in the region r_i. And larger $need_{i,j}$ means a stronger demand for the category. Now we are ready to formalize the problem studied in this paper.

Problem Statement: Given places set $\mathcal{P}$, categories set $\mathcal{C}$ and grid parameter g which partition the space into $g \times g$ regions, the aim of our problem is that, for each region, to compute the need number of categories which are demanded in this region and return the rank of categories according to their need numbers.

4 Regional POI Demand Discovery

In this section, we first introduce the framework of our proposed method for regional demand mining. Then we present our algorithms that efficiently compute the need numbers and return ranked regional demands in detail.

4.1 Framework Overview

Figure 2 shows the framework of proposed regional POI demand discovery method. We take check-ins data, map data and POIs data as the input and each of these data goes through an offline preprocessing step. Specifically, the check-ins data are used to construct semantic trajectories. Then the semantic trajectories are separated into sub-trajectories according to the max time gap.

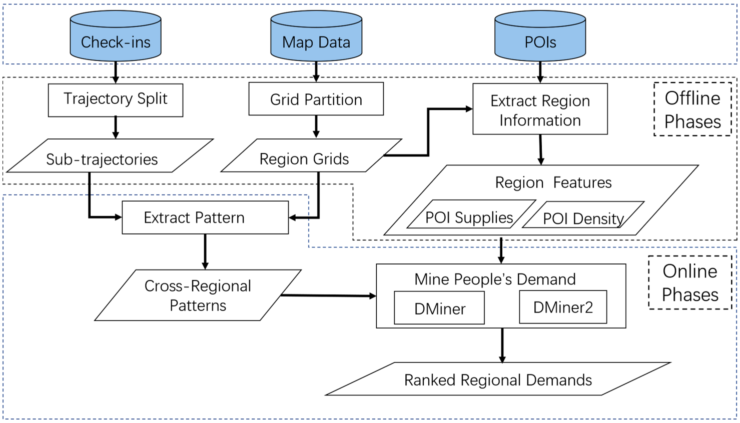


Fig. 2. An overview of framework

The map data are partitioned into regions by the grid size parameter g. And the POIs data together with the region information are used to extract regional POIs information as the region features.

The time difference between two adjacent check-ins is called time gap. In our daily life, the reason for people's movement is the demand for a variety of purposes or just semantic categories, if they need a kind of POI category strongly, they will go to this site as soon as possible. According to actual experience, a strong demand driven movement often happens in a short time period, for example period is from several hours up to a few days. So we introduce a max time gap threshold Δt, and split semantic trajectories into sub-trajectories to make sure that each time gap in one sub-trajectories is smaller than Δt. Another task in offline phase is to convert the semantic trajectories to regional patterns as we defined in Definition 2.

In the online phase, we conduct our proposed mining algorithm to extract cross-regional pattern, compute the need numbers and return ranked regional demands accordingly. Next, we will introduce our proposed algorithms in details.

4.2 Cross-Regional Pattern Extraction

In order to compute the need numbers, we first extract cross-regional patterns from regional patterns set $\mathcal{O}$ which are generated in the offline phase. We filter out pattern which length is less than 2 because it will not produce cross-regional pattern. As shown in Algorithm *patternMine()*, we first scan remaining patterns to check two adjacent category-regional pairs of a pattern whether located in the same region or not without considering category, in order to find a cross-region pattern. We add such category-region pair into the result set and try to extend this length-2 cross-regional pattern by repeatedly checking location information demonstrated as line 7. we finish the extension of this pattern until the next check-in is in the same region and start to find another cross-regional pattern from the remaining patterns.

Algorithm 1: patternMine ($\mathcal{O}$)

```
Input: the set of regional patterns O
Output: cross-regional sub-pattern set E
for each regional pattern o ∈ O with length ≥ 2 do
    i = 2;
    while i <= o.length do
        if o[i − 1].r! = o[i].r then
            generate a cross-regional pattern e ;
            for idx ∈ [i + 1, o.length + 1) do
                if o[idx − 1].r! = o[idx].r then
                    add o[idx] to e ;
                else
                    i = idx+1;
                    E = E ∪ e;
                    break;
                end
            end
        else
            i = i+1;
        end
    end
end
return E;
```

Example 3. T_3 as demonstrated in Fig. 1 can be translated into regional pattern $\langle(b,\ r_8),\ (g,\ r_6),\ (o,\ r_4),\ (b,\ r_4)\rangle$. We scan first two adjacent (category, region) pairs $(b,\ r_8)$ and $(g,\ r_6)$, and find these check-ins distributed in two regions r_8 and r_6 so we get a length-2 cross-regional pattern. Move to next pair $(o,\ r_4)$ which is placed at r_4 so the cross-regional pattern is extended to length-3. However, $(b,\ r_4)$ is also in region r_4, hence T_3 only includes a length-3 cross-regional pattern.

4.3 Computing Need Numbers for Regional Demand Mining

Next step is mining people's need from cross-regional pattern set $\mathcal{E}$ obtained by *patternmine().* Algorithm **DMiner** shows how to compute the need numbers. According to daily experience, person can travel from one region to another region for a specific demand no matter the category of his(her) current starting point and the destination region of trip. Same category need starting from same region may ended in different regions. So in our algorithm, the idea to simplify the POIs scatter problem is that we only put focus on origin's starting region and destination's semantic category.

DMiner scans cross-regional pattern e in $\mathcal{E}$ from the first beginning and captures any two adjacent items in e to generate new category-region pairs.

Algorithm 2: DMiner ($\mathcal{E}$)

1 **Input:**cross-regional pattern set $\mathcal{E}$
2 **Output:**set of regional demand candidates $\mathcal{M}$
3 **for** *each* e *in* $\mathcal{E}$ **do**
4   **for** i = *1 to e.length-1* **do**
5   origin $= e[i]$ //$e[i]$ is the i-th check-in in e;
6   destination $= e[i+1]$;
7   generate a new need pair $m := ((destination.category, origin.region), 1/d_{x,y})$ //x is origin.region, y is destination.category
8         $\mathcal{M} = \mathcal{M} \cup m$;
9         $need_{x,y} += 1/d_{x,y}$
10   **end**
11 **end**
12 return $\mathcal{M}$;

And the need number is continually updated according to the definitions (Line 9). After proceeding all the cross-regional patterns in $\mathcal{E}$, the regional demand candidates set $\mathcal{M}$ is obtained.

4.4 Optimization

By far, we only split cross-regional pattern e into a set of length-2 sub-patterns in *DMiner*. For example, if there exists a cross-regional pattern ⟨a, b, c⟩ (a, b, c all include category and region information), in *DMiner* we only pay attention to the pattern ⟨a, b⟩, ⟨b, c⟩. However, in real applications, people may not move from one region to another region for a specific category directly, they may stop-by somewhere first and then move to the destination. Trajectories split by max time gap ensure that stationary point in pattern doesn't cost too much time for visiting. So we can take ⟨a, c⟩ into consideration as well. In the optimized algorithm, we add another moving arrow for potential cross-regional category-region pair which may be farther in cross-regional pattern. And we define a new weighting factor σ to reflect the importance of such kind of extended cross-regional patterns.

$$\sigma = \frac{1}{j - i} \tag{3}$$

where i and j stands for origin's and destination's place index in cross-regional pattern e, where $1 \leq i < j \leq e.length$. So σ of need acquired from two adjacent cross-regional pair is 1 as same as the weight in *DMiner*. In the optimized algorithm, the need number is updated by multiplying weight factor σ for the categories which have extended cross-regional pattern.

We call the optimized algorithm **DMiner2** which connects two category-region pairs if they are in different regions. Because cross-regional pattern may contain some behaviors happened in same region, for example cross-regional pattern $\langle$(bar, r_8), (gym, r_2), (office, r_8)$\rangle$, bar and office both locate in r_8. **DMiner2** should examine category distribution again.

4.5 Complexity Analysis

Our proposed framework consists of two online phases, namely cross-regional pattern extraction and need number computation. In the cross-regional pattern extraction phase, **patternMine** scans each regional pattern in regional pattern set $\mathcal{O}$ and records the adjacent movement pairs which are in different regions. Therefore the complexity of **patternMine** is $O(N)$ where N is the total length of the trajectory data set. In the need number computation phase, **DMiner** and **DMiner2** checks the cross-regional patterns generated in **patternMine** and generates regional demand candidates. Regional category density can be easily calculated by pre-stored regional category density list. The time complexity of **DMiner** and **DMiner2** is also $O(N)$, as in the worst case, **DMiner** and **DMiner2** have to check all movement pairs in the trajectory data set. Therefore, the overall time complexity of our framework is $O(N)$ which is linear to the size of data set. The experimental result also confirms the efficiency of our method.

Having the need numbers of the needed categories in a given region, we can return the ranked categories according their need numbers as regional POI demand.

5 Experiment

In this section, we evaluate performance of our proposed algorithms in real data sets. All algorithms are implemented in JAVA conducted on a computer running Linux (CentOS 7.3.1611) with 40 Intel Xeon CPU E5-2630 v4 2.2 GHz and 128 GB memory.

5.1 Dataset

In the experiments, we use two real-world data sets which are selected from the world-wide Foursquare check-ins data sets issued by Yang *et al.* [27]. Specifically, BR, US is the part of these numerous data sets where POIs and check-ins located at United Kingdom and United States respectively. The details of these datasets are listed in Table 1.

5.2 Competitors

We compare our proposed methods with state-of-the-art semantic trajectory mining algorithm **RegMiner** [7].

Since **RegMiner** is not designed for regional demand mining, we made the following modifications: First, we adopt the idea of **GridMiner** in [7] to fit **RegMiner** for our grid partition. Second, for the frequent threshold, we set

Table 1. The statistics of data sets

Dataset	BR	US
Number of POIs	54,278	168,625
Number of trajectories	4,893	13,489
Number of categories	414	427

1.5 as default which is much smaller than the value used in [7]. In our grid based partition, large frequent threshold will reduce number of the frequent patterns significantly and results in insufficient need demand, therefore we set the frequent threshold a reasonable small number which can archive balance between efficiency and effectiveness. We call the modified **RegMiner** as **RegMiner-Grid** in the rest of this paper.

5.3 Parameter Setting

We examine the impact of parameter for algorithms on BR dataset. Default parameters are set as follows: max time gap is 24 h and grid number is 100.

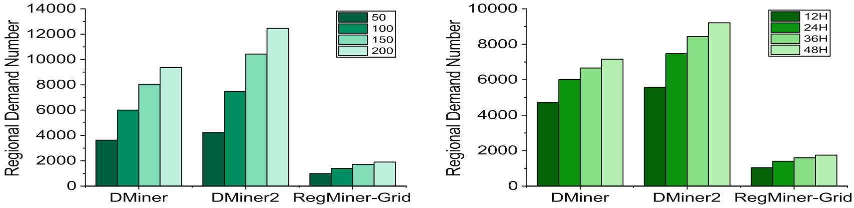
a. Varying grid number
The whole region is split into $g \times g$ size of small grids. Figure 3(a) shows grid number has a large impact on regional demand number. Different region distributions create various cross-regional patterns and result in diverse regional demands and large need number difference. On the other hand, smaller g such as 30 forms larger region area which lets trajectories hard to pass through. In the rest of the experiments, we set g as 100 and the length of each cell is round 6km and 8 km for BR and US which is reasonable for a neighborhood area in real cities.

b. Varying max time gap
We split raw trajectories into small trajectories which time between each two adjacent check-ins is no more than max time gap to filter out not very strong demand meeting. We conduct the experiments with max time gap varying from 12 h to 48 h. In Fig. 3(b), Δt represents max time gap. It shows that larger max time gap can produce large number of demands which confirms our analysis. In the rest of the experiments, we set Δt as 24 h.

5.4 Efficiency Study

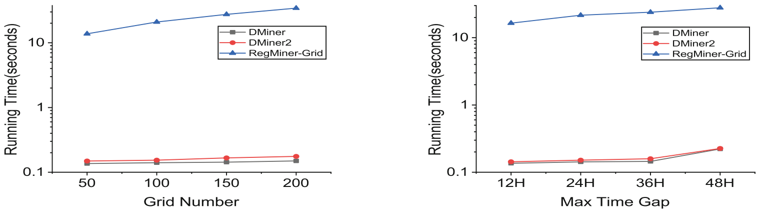
In this section, we study the efficiency of the algorithms on BR with regard to two parameters, grid number and max time gap.

As shown in Fig. 4(a), the running time of all algorithms increases as the grid number increases. This is because the large grid numbers will produce more cross-regional patterns as the region is small. Therefore, it takes more time to compute the numbers for our algorithms and compute the frequent patterns for *RegMiner-Grid*. And our algorithms are always better than *RegMiner-Grid*,



(a) results w.r.t g (b) results w.r.t Δ t

Fig. 3. Regional demand number



(a) Running time w.r.t g (a) Running time w.r.t Δt

Fig. 4. Running time on BR

as *RegMiner-Grid* takes more time when mining the frequent patterns with complicated computation of the support.

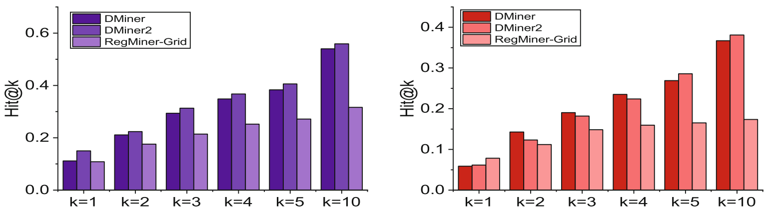
Similar trend can be observed in Fig. 4(b). The reason is that setting large max time gap will produce longer regional patterns and increase the computation for all the algorithms.

5.5 Effectiveness Study

In order to evaluate the effectiveness of the algorithms, we select New York city from US and Greater London from BR as represented. We divide the data sets into training and testing set as follows. We choose the first 80% check-ins to construct the regional patterns for training, and the remaining 20% check-ins are used for testing. First, Given a region, we rank the demands for POI categories and give a top-k ranking list. We use Hit@k as metric. For a region, if top-k predicted demands meet any actual open POI, hit number pluses 1. And hit@k is hit number divided by total number of regions. Figure 5 shows final result. Our algorithms outperforms the competitor in almost all cases.

5.6 Illustrating Cases

In order to better illustrate the ranking results for regions, we pick up several example regions with the top 5 identified demands in New York as shown in Table 2. As shown in the table, our method can effectively discovery the most needed categories and newly opened POIs in the ground-truth confirms that the proposed method is effective.



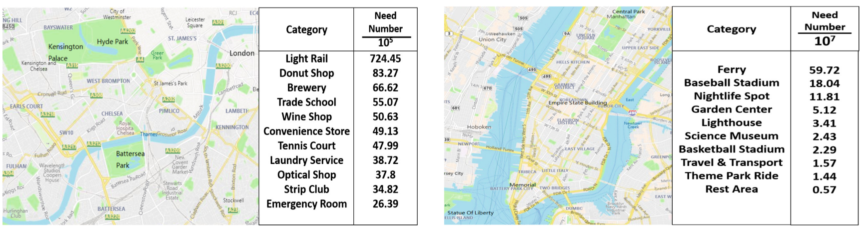
(a) In New York (b) in Greater London

Fig. 5. Hit@k

Table 2. Identified POI demands for regions in New York

Region	Identified demands p@5	Groundtruth
84266	**House**; **Gas Station**; Drugstore; **Fast Food Restaurant**; American Restaurant	Residential Building; House; Gas Station Fast Food Restaurant; Convenience Store
84669	Community; **Department Store**; **Furniture Store**; **Road**; Drugstore	Department Store; Furniture Store; Road; Hotel; Bank
85461	Miscellaneous Shop; Bank; **Doctor'sOffice**; **Bakery**; **Automotive Shop**	Doctor's Office; Chinese Restaurant; Italian Restaurant; Bakery; Automotive Shop

Besides, We choose two hot tourism destinations, e.g. region 5116 in London and region 5367 in New York and list 10 representative regional demand in these regions. In reality, region 5116 is in London, which includes Kensington Palace, Battersea park and other famous spots. And Region 5367 is a part of Manhattan. There exists a lot check-ins scattered in theses regions, so our approaches are more easier to mine their regional category needs as shown in Fig. 6. Need Number is too large so these number are divided by 10^5 and 10^7 respectively in London and New York to show clearly.



(a) Region 5116 in London (b) Region 5367 in New York

Fig. 6. Regional demand in real area

6 Conclusion

In this paper, we focus on inferring diverse POI demands of urban area using semantic trajectories. We carefully analyze the features of semantic trajectory data and people' mobility patterns. Then we propose a general data-driven framework **DMiner** for regional POI demand mining. The framework uses enriched traditional category sequence with region information and **patternMine** to identify cross-regional pattern efficiently. We also introduce some interesting observations to enrich the results and further propose a improved method **DMiner2**. We apply our framework on real data sets to show the effectiveness and efficiency with comparison to state-of-the-art locally frequent pattern mining method. Furthermore, we also present several example mining results for readers' better understanding.

Acknowledgements. We sincerely thank the anonymous reviewers for their feedback which helped improve our work. The research of Shiyu Yang is supported by NSFC61802127 and Shanghai Sailing Program 18YF1406700. The research of Long Yuan is supported by NSFC61902184 and NSF of Jiangsu Province BK20190453

References

1. Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., et al.: A model for enriching trajectories with semantic geographical information. In: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, p. 22. ACM (2007)
2. Batran, M., Mejia, M., Kanasugi, H., Sekimoto, Y., Shibasaki, R.: Inferencing human spatiotemporal mobility in greater Maputo via mobile phone big data mining. ISPRS Int. J. Geo-Inf. **7**(7), 259 (2018)
3. Cao, H., Mamoulis, N., Cheung, D.W.: Mining frequent spatio-temporal sequential patterns. In: Proceedings of the 5th IEEE ICDM, pp. 82–89. IEEE (2005)
4. Chan, H.K.H., Long, C., Yan, D., Wong, R.C.W.: Fraction-score: a new support measure for co-location pattern mining. In: Proceedings of the 35th IEEE ICDE, pp. 1514–1525. IEEE (2019)
5. Chen, L., Gao, Y., Fang, Z., Miao, X., Jensen, C.S., et al.: Real-time distributed co-movement pattern detection on streaming trajectories. Proc. VLDB Endow. **12**(10), 1208–1220 (2019)
6. Chen, Z., Shen, H.T., Zhou, X.: Discovering popular routes from trajectories. In: Proceedings of the 27th IEEE ICDE, pp. 900–911. IEEE (2011)
7. Choi, D.W., Pei, J., Heinis, T.: Efficient mining of regional movement patterns in semantic trajectories. Proc. VLDB Endow. **10**(13), 2073–2084 (2017)
8. Cohen, S.A., Duncan, T., Thulemark, M.: Lifestyle mobilities: the crossroads of travel, leisure and migration. Mobilities **10**(1), 155–172 (2015)
9. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD, pp. 330–339. ACM (2007)
10. Hu, T., Song, R., Wang, Y., Xie, X., Luo, J.: Mining shopping patterns for divergent urban regions by incorporating mobility data. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 569–578. ACM (2016)

11. Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., Mascolo, C.: Geo-spotting: mining online location-based services for optimal retail store placement. In: Proceedings of the 19th ACM SIGKDD, pp. 793–801. ACM (2013)
12. Kim, Y., Han, J., Yuan, C.: TOPTRAC: topical trajectory pattern mining. In: Proceedings of the 21th ACM SIGKDD, pp. 587–596. ACM (2015)
13. Li, Q., Yu, Z., Guo, B., Lu, X.: Inferring housing demand based on express delivery data. In: Proceedings of the 6th IEEE International Conference on Big Data, pp. 1445–1454. IEEE (2018)
14. Li, Y., Zheng, Y., Ji, S., Wang, W., Gong, Z., et al.: Location selection for ambulance stations: a data-driven approach. In: Proceedings of the 23rd SIGSPATIAL, p. 85. ACM (2015)
15. Liu, Y., Liu, C., Lu, X., Teng, M., Zhu, H., et al.: Point-of-interest demand modeling with human mobility patterns. In: Proceedings of the 23rd ACM SIGKDD, pp. 947–955. ACM (2017)
16. Liu, Y., Liu, C., Yuan, N.J., Duan, L., Fu, Y., et al.: Exploiting heterogeneous human mobility patterns for intelligent bus routing. In: Proceedings of the 14th IEEE ICDM, pp. 360–369. IEEE (2014)
17. Ma, S., Zheng, Y., Wolfson, O.: T-share: a large-scale dynamic taxi ridesharing service. In: Proceedings of the 29th IEEE ICDE, pp. 410–421. IEEE (2013)
18. Moosavi, S., Samavatian, M.H., Nandi, A., Parthasarathy, S., Ramnath, R.: Short and long-term pattern discovery over large-scale geo-spatiotemporal data. In: Proceedings of the 25th ACM SIGKDD, pp. 2905–2913. ACM (2019)
19. Niu, H., Liu, J., Fu, Y., Liu, Y., Lang, B.: Exploiting human mobility patterns for gas station site selection. In: Navathe, S.B., Wu, W., Shekhar, S., Du, X., Wang, X.S., Xiong, H. (eds.) DASFAA 2016. LNCS, vol. 9642, pp. 242–257. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32025-0_16
20. Noulas, A., Scellato, S., Mascolo, C., Pontil, M.: An empirical study of geographic user activity patterns in foursquare. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. AAAI (2011)
21. Pilinkienė, V.: Market demand forecasting models and their elements in the context of competitive market. Eng. Econ. **60**(5) (2008)
22. Pilinkienė, V.: Selection of market demand forecast methods: criteria and application. Inžinerinė ekonomika (3) 19–25 (2008)
23. Shi, H., Li, Y., Cao, H., Zhou, X., Zhang, C., et al.: Semantics-aware hidden Markov model for human mobility. IEEE Trans. Knowl. Data Eng. (2019)
24. Sun, Y., Zhu, H., Zhuang, F., Gu, J., He, Q.: Exploring the urban region-of-interest through the analysis of online map search queries. In: Proceedings of the 24th ACM SIGKDD, pp. 2269–2278. ACM (2018)
25. Tiebout, C.M.: A pure theory of local expenditures. J. Polit. Econ. **64**(5), 416–424 (1956)
26. Xu, H., Zhang, Y., Wei, J., Yang, Z., Wang, J.: Spatiotemporal-aware region recommendation with deep metric learning. In: Li, G., Yang, J., Gama, J., Natwichai, J., Tong, Y. (eds.) DASFAA 2019. LNCS, vol. 11448, pp. 491–494. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-18590-9_73
27. Yang, D., Zhang, D., Qu, B.: Participatory cultural mapping based on collective behavior data in location-based social networks. ACM Trans. Intell. Syst. Technol. (TIST) **7**(3) (2016). Article no. 30
28. Yao, Z., Fu, Y., Liu, B., Hu, W., Xiong, H.: Representing urban functions through zone embedding with human mobility patterns. In: IJCAI, pp. 3919–3925. Morgan (2018)

29. Yuan, J., Zheng, Y., Xie, X.: Discovering regions of different functions in a city using human mobility and POIs. In: Proceedings of the 18th ACM SIGKDD, pp. 186–194. ACM (2012)
30. Zhang, C., Han, J., Shou, L., Lu, J., La Porta, T.: Splitter: mining fine-grained sequential patterns in semantic trajectories. Proc. VLDB Endow. **7**(9), 769–780 (2014)
31. Zhong, Y., Yuan, N.J., Zhong, W., Zhang, F., Xie, X.: You are where you go: inferring demographic attributes from location check-ins. In: Proceedings of the 8th ACM International Conference on Web Search and Data Mining, pp. 295–304. ACM (2015)
32. Zhou, X., Noulas, A., Mascolo, C., Zhao, Z.: Discovering latent patterns of urban cultural interactions in WeChat for modern city planning. In: Proceedings of the 24th ACM SIGKDD, pp. 1069–1078. ACM (2018)