Mobile Networks and Applications (2020) 25:459–473
https://doi.org/10.1007/s11036-019-01309-4

Mining Regional Mobility Patterns for Urban Dynamic Analytics

Jing Lian[1] · Yang Li[1] · Weixi Gu[2] · Shao-Lun Huang[1] · Lin Zhang[1]

Published online: 24 July 2019

Abstract

City management plays an important role in the era of urbanization. Understanding city regions and urban mobility patterns are two vital aspects of city management. Numerous studies have been conducted on these two aspects respectively. However, few work has considered combining city region partition and mobility pattern mining together while these two problems are closely related. In this paper, we propose region-aware mobility pattern mining framework, which jointly finds the precise origin and destination region partitions while extracting mobility patterns. We formulate it as an optimization problem of maximizing OD's correlations with spatial constraints. Kernelized ACE, is proposed to solve the problem by learning feature representations that guarantee both objectives. Evaluation results using Beijing's taxi data show that the extracted features are appropriate for this problem and our approach outperforms all the other methods with $\sim 0.3\%$ spatial overlap and 86.43% OD correlation. Our case studies on New York City's urban dynamics and Beijing's three-year consecutive analysis also yield insightful findings that reveal city-scale mobility patterns and propose potential improvement for city management.

Keywords Urban dynamics · Mobility pattern · Region partition · Feature extraction · Co-clustering

1 Introduction

Urban population is growing rapidly worldwide. According to United Nations World Urbanization Prospects [28], 66% of world's population will be residing in urban areas by 2050. Rapid urbanization is exerting severe pressure on city management, which has two vital aspects: the spatial distribution of an urban population and its internal migration within the city [28]. The spatial distribution of population is formed by people residing in and interacting with different regions of the city. Understanding how cities are partitioned into regions and how these regions evolve, *a.k.a. city region partition*, is therefore an important problem. Internal migration refers to how people move from one region to another in the city. An effective way to study internal migration is through *mobility pattern mining*, i.e. using mobility data to infer where people are most likely to go. Numerous studies have been conducted on these two aspects separately.

For city region partition, previous work focused on identifying city regions with semantic meanings were based on similar functionality [30], social interactions [22] and traffic information [9]. However, these work were mainly from a socio-economic perspective but none of them analyzed city structures considering people's mobility movements. Thus these regions with certain semantic meanings cannot provide insights for mobility pattern mining applications. While people's mobility movements reflect functionalities of different city regions. For example, many people from region A go to region B in the morning may indicate a commute route from residential to work. Partitioning city structures based on mobility similarities could identify regions reflecting how people interact with the city, thus benefit urban traffic management. Specifically, this type of partition can be utilized for newly surfaced applications like customized bus service area design, which provide

✉ Lin Zhang
linzhang@tsinghua.edu.cn

Jing Lian
lianj16@mails.tsinghua.edu.cn

Yang Li
yangli@sz.tsinghua.edu.cn

Weixi Gu
guweixigavin@gmail.com

Shao-Lun Huang
shaolun.huang@sz.tsinghua.edu.cn

[1] Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen, China

[2] University of California, Berkeley, Berkeley, CA, USA

on-demand transit services for passengers with similar origins and destinations to help reduce peak-hour congestion [13].

For mobility pattern mining, previous work tried to extract salient mobility patterns while lacked of focus on city management perspective. Traditional ways of mobility pattern extraction were based on single-view methods by clustering OD trips as points in high dimensional space [1, 32]. This approach has a limitation that, when clusters are viewed by the spatial distribution of their origins (O) and destinations (D), regions corresponding to different clusters tend to overlap. Such overlaps produce spatial ambiguities, which lead to limited guidance for administrative region-based city management. Some methods addressed the overlapping problem by a two-stage process. They first partitioned the continuous space of OD locations into a finite discrete set of non-overlapping regions (e.g. grid cells [12, 31] or regions-of-interest (ROI) [25, 27]); then extracted mobility patterns among those regions or collection of regions. However, not only does two-stage process lose fine-grained information and the region shapes cannot be arbitrary; more importantly, there is no guarantee that the partition would be appropriate in terms of mobility pattern extraction.

While city region partition and mobility pattern mining are closely related to each other, few work has considered combining these two problems together. In this paper, we propose the **region-aware mobility pattern mining** framework, which jointly finds the origin and destination region partitions, while extracting mobility patterns directly in their own space. Specifically, our work partitions a city into different regions based on spatial proximity and mobility similarity for mobility pattern mining. These partitions contains trip information appropriate for mobility pattern mining applications, and the partitions reflect people's interactions with the city benefiting urban traffic management. Informally, we formulate the problem as learning feature representations of trip origins and destinations that guarantee both objectives. For mobility pattern mining, we adopt the *HGR maximal correlation* [23], a well-defined dependence measure to represent the mobility similarities among different OD trips. For region partition, we add spatial constraints as spatial proximity to minimize the amount of overlaps between O/D regions of different clusters. Therefore, we obtain OD features containing both information of mobility patterns and region partitions. Using the extracted features, we are able to partition the city into non-overlapping OD regions, which also provide salient mobility patterns between regions.

To solve the **region-aware mobility pattern mining** problem efficiently, we propose the *Kernelized Alternating Conditional Expectation (KACE)* algorithm based on a statistical technique. The original ACE algorithm [16] is an efficient way to compute HGR maximal correlation for discrete random variables. However, it does not impose any constraints on the extracted features, such as the spatial constraints of regions in our case. Moreover, it cannot be used on continuous data. Therefore, we applied kernel smoothing technique [2] on ACE algorithm with an optimal kernel to handle continuous data with spatial constraints.

We evaluated our algorithm on real taxi trip data for both feature extraction and clustering results. For feature extraction, we compared with both CCA and KCCA, which are typical methods to achieve maximal correlation. Features extracted by KACE have smaller kurtosis thus could better differentiate between similar data points. For clustering results, we obtained mobility regions with only 0.33% and 0.22% overlap for origins and destinations respectively but achieve 86.43% OD correlation, which reduces overlap by up to ten times comparing to other approaches. Compared with other methods, including traditional methods [1, 4], ACE algorithm with discretization [16], CCA[8], KCCA[7] and state-of-the-art multi-view clustering method [19], our method achieves minimum overlaps, concentrated regions, highly correlated patterns and full spatial coverage.

Case studies of both New York City and Beijing provide insightful findings. Specifically, our analysis discovered an opposite urban dynamics between morning and evening travel patterns in New York City. In a three-year analysis of Beijing's mobilities, we also identified mobility patterns and regional variations during these years, like the development of suburban areas and newly surfaced regions.

We summarize the contributions of our paper below:

1. A novel region-aware mobility pattern mining framework. It considers both mobility pattern extraction and OD region partition without overlap.
2. A kernel-based extension to the ACE algorithm for extracting maximal correlated features. KACE is the first algorithm to solve the HGR maximal correlation problem with continuous input and feature constraints.
3. A thorough evaluation of KACE towards both feature and cluster results with real world data. Feature evaluation shows KACE has tradeoff between correlation and distribution kurtosis thus features are easier to be clustered. Cluster results show the versatility of our approach in disentangling complicated mobility patterns comparing with both traditional methods and state-of-the-art clustering algorithms.
4. Comprehensive case studies of both New York City and Beijing with real taxi data. Urban dynamics analysis of NYC reveals people's travel patterns between different functionalities in the city. A three-year analysis of Beijing's mobility patterns reveals its city development and urban sprawl through years.

This work is an extension to [14] with more detailed literature review, thorough feature and cluster results evaluation, new case studies using POI information of NYC and three-year taxi data of Beijing. The remaining parts of the paper are organized as follows. Section 2 introduces related work, Section 3 formulates the problem mathematically, Section 4 illustrates the proposed algorithm KACE, Section 5 evaluates the algorithm on both feature and cluster results with real data, Section 6 points out the findings of case study in NYC, Section 7 compares Beijing's mobility patterns in three years and the paper concludes in Section 8.

2 Literature review

Several areas of studies are related to our work, including urban dynamic analytics, region partition using GPS data, mobility pattern mining and co-clustering problem.

2.1 Urban dynamic analytics

Traditional urban dynamic research adopt standard GIS methods to study urban development. Inostroza et al. used satellite imagery for 10 Latin American cities over a period of 20 years to characterize urban development and sprawling features measured with GIS tools [11]. However, these research do not take urban mobility into consideration. Some other studies considered both spatial dynamics and human mobility but lacked real data support. According to Priemus, the relationship between mobility and spatial dynamics is complicated and is an important research area [21]. García-Palomares adopted statistic methods on census data to study relationships between urban sprawl and commuters mobility [6]. Gao used Call Detail Records studying spatial-temporal urban dynamics to identify individual travel patterns or hot cells but lacked focus on mobility patterns [5]. To the best of our knowledge, our work is the first to compare city-wide mobility patterns across the span of several years using large-scale real world data with a focus on urban analytics.

2.2 City region partition

Studies of partitioning city into non-overlapping regions have similarities to our work but none of them partitions regions from a mobility perspective. Huang *et al.* partitioned city into regions using taxi trajectory data to reflect street connections and traffic information [9]. These regions capture urban traffic information on roads but cannot help with mobility pattern mining. Yuan et al. discovered regions of different functions using point-of-interest and mobility information [30]. Their work partitioned the urban area boundaries based on existing road network, while our method establishes regions based on mobility patterns learned from data. Qi et al. divided city into grids and used time-domain features of taxi pick-ups and drop-offs for social function identification [22]. Instead of dividing the city into mobility regions, their paper focused on accurately identifying each area's functionality. Liu et al. revealed sub-regions of Shanghai by identifying sub-networks of taxi trip network [15]. These regions are densely intra-connected but with less inter-connections, thus the identified regions cannot fully reveal the spatial dynamics of a city.

2.3 Mobility pattern mining

One important type of mobility pattern mining is to extract common patterns of OD data. Zhu et al. modified density based clustering algorithm DBSCAN in mobility mining setup to extract popular OD pairs [32]. However, it can only find salient and precise patterns and leads to overlap between O/D clusters. Another type of work adopted a two-stage process by discretizing a city into non-overlapping sets first and extracting OD patterns on that set. Toole et al. extracted meaningful stay points from call detail records, and estimated OD flow between those points [27]. Tang et al. adopted DBSCAN to cluster taxi pick-ups and drop-offs first and then used discretized OD clusters to calibrate a statistic mobility model [25]. Zhang et al. divided the urban area into small OD grids before estimating OD flows to extract semantic meaning of people's mobilities [31]. All of these work extracted mobility findings on discrete set of regions, which are POI regions [27], O/D clusters [25] or uniform grids [31] rather than original continuous data, which limit the size, shape of the clustering results and have no guarantee of the reasonable discretization and findings' granularity.

2.4 Co-clustering methods

Co-clustering is the problem of simultaneously clustering two types of correlated data, while is not widely applied to mobility pattern mining problems. Kuo et al. used non-negative tensor factorization (NTF) based co-clustering algorithm and established an OD flow [12]. However, it still has the discretization issue that binning GPS samples into grids before applying NTF. We adopted feature-based co-clustering algorithm without discretization to overcome information loss of discretization. Nie et al. proposed a state-of-the-art multi-view clustering algorithm MLAN [19], a generalization of co-clustering, which extracted an optimal distance metric from data itself and co-clustered data under the optimal metric. However, it only focused on achieving maximum correlation without trade-offs on cluster size.

3 Problem formulation

3.1 Region-aware mobility pattern mining problem

Given N taxi OD trips $T = \{(x_i, y_i), x_i = (xlat_i, xlon_i), y_i = (ylat_i, yon_i), i = 1, \cdots, N\}$[1] as input, we would like to partition N origins X into N_x origin regions and N destinations Y into N_y destination regions, i.e. each origin point x_i has an origin region label $OR(x_i)$ and destination point y_i has destination region label $DR(y_i)$ as output, such that

1. regional dependency between OR and DR is maximized;
2. points inside each cluster are close to each other.

3.1.1 Dependence measure

Rényi et al. pointed that the Hirschfeld-Gebelein-Rényi maximal correlation is the only dependence measure that satisfies all Rényi's seven postulates [23]. While the common used correlation coefficient only satisfies two of them. Maximal correlation quantizes regional dependencies to a measure between 0 and 1, and it is comparable regardless of different data size. Therefore, we adopt maximal correlation to quantify dependency mathematically.

Definition 1 *Hirschfeld-Gebelein-Rényi maximal correlation*. Maximal correlation between jointly distributed random variables X and Y is defined as:

$$\rho(X;Y) \triangleq \sup_{\substack{f:\mathcal{X}\to\mathbb{R},\ g:\mathcal{Y}\to\mathbb{R} \\ \mathbb{E}[f(X)]=\mathbb{E}[g(Y)]=0 \\ \mathbb{E}[f^2(X)]=\mathbb{E}[g^2(Y)]=1}} \mathbb{E}[f(X)g(Y)]$$

where the supremum is taken over all Borel measurable functions. Furthermore, $0 \le \rho(X;Y) \le 1$, $\rho(X;Y) = 0$ if and only if X is independent of Y, $\rho(X;Y) = 1$ if there is a strict dependence between X ad Y, i.e. $Y = f(X)$ or $X = g(Y)$.

In the region-aware mobility pattern mining problem, we would like to find feature functions $f(X)$ of origins X and $g(Y)$ of destinations Y, s.t. the correlation between $f(X)$ and $g(Y)$ is maximized for X and Y. The maximal correlation of 0 happens only when origin and destination regions are completely independent, which means origins contain no information of destinations and vice versa. While the maximal correlation of 1 indicates that the destinations can be predicted with 100% certainty given origins or vice versa.

3.1.2 Constraint satisfaction

In order to satisfy the spatial constraint that *points within each cluster are close to each other*, we need to make sure the features are close when the corresponding points are close. Mathematically speaking, when origins x_i and x_j are close to each other geographically, their corresponding features $f(x_i)$ and $f(x_j)$ should be similar in value for numbers or similar in norm for vectors. We use euclidean distance between two vectors x_i and x_j to quantize physical distance of two points, and base-2 norm of $f(x_i) - f(x_j)$ to quantize feature difference. The same principle applies to the destination points, i.e.

$\|f(x_i) - f(x_j)\|_2$ increases as $\|x_i - x_j\|_2$ increases

$\|g(y_i) - g(y_j)\|_2$ increases as $\|y_i - y_j\|_2$ increases

Here, $f(x_i)$ is the feature vector of the i-th origin point x_i, likewise $g(y_i)$ is the feature vector of i-th destination y_i.

3.2 Mathematic formulation

Combining Sections 3.1.1 and 3.1.2, we formulate the **region-aware mobility pattern mining** problem as following.

Given N taxi OD trips T=$\{(x_i, y_i), i = 1, \cdots, N\}$ as input, we would like to partition origins X into N_x origin regions and destinations Y into N_y destination regions, such that

1. The HGR maximal correlation between OR and DR is maximized, i.e. the correlation between D-dimensional features $f(X)$ and $g(Y)$ is maximized for X and Y

$$\max_{\substack{\mathbb{E}[f(X)]=\mathbb{E}[g(Y)]=0 \\ \mathrm{Cov}[f(X)]=\mathrm{Cov}[g(Y)]=\mathrm{I}}} \mathbb{E}[f(X)^T g(Y)]$$

2. Points inside a region are close to each other, i.e.

$\|f(x_i) - f(x_j)\|_2$ increases as $\|x_i - x_j\|_2$ increases

$\|g(y_i) - g(y_j)\|_2$ increases as $\|y_i - y_j\|_2$ increases

Based on proximity of features $f(X)$ and $g(Y)$, we use linkage cluster [24], which is commonly applied in geographic areas, to find N_x ORs, N_y DRs and mobility patterns between origin and destination regions.

[1] Here $xlat_i$ and $xlon_i$ represent trip i's latitude and longitude of origin respectively; $ylat_i$ and $ylon_i$ represent trip i's latitude and longitude of destination respectively.

4 Algorithm

4.1 Overview

Figure 1 shows the flow chart of our algorithm. The input of our algorithm is OD pairs of taxi trips. Preprocessing raw trajectory data to extract OD pairs is necessary. Our algorithm has two steps, feature extraction with kernelized ACE and feature clustering with linkage clustering [24]. First, we extract features of origins and destinations considering both OD correlation and spatial constraints formulated in Section 3. Then we cluster O/D features separately. As features already contain information of OD correlation, clustering them separately still captures mobility patterns while obtaining non-overlapping O/D regions. The outputs of our algorithm are origin regions (OR), destination regions (DR), i.e. OR and DR label of each O and D point, and OD patterns from OR to DR with a probability indicating the percentage of trips originating from OR that end at DR.

4.2 Kernelized ACE

4.2.1 Limitations of ACE algorithm

Makur et al. solved the problem of obtaining D-dimensional feature functions $f(\cdot)$ and $g(\cdot)$ to achieve maximal correlation by Alternating Conditional Expectation (ACE) algorithm efficiently [16].

However, this ACE algorithm cannot be applied to our problem directly due to two issues. First, ACE is designed for discrete-valued data while our co-clustering problem involves continuous-valued GPS data. Second, ACE cannot place additional constraints on features. The connection between X and Y is hidden in data pairs. For example, in the Netflix [18] Prize problem of predicting user ratings for films, the similarities between movies are reflected by the ratings of different users but there is no obvious connections between movies or users themselves. On the contrary, from the mobility perspective, not only the dependencies between origins (X) and destinations (Y) should be considered, the distances between origin points or destination points should also contribute to finding geographically meaningful regions.

A simple but suboptimal solution is to discretize continuous data into discrete categories. However, the drawbacks are obvious. First, the best discretization approach is hard to find. It is unrealistic and time-consuming to adopt the best combination after traversing every possibilities of methods and parameters. Second, the step of discretization loses information. Instead of discretization, we propose to use the original data for feature extraction.

4.2.2 Kernel method

We address the limitations of ACE algorithm by applying appropriate kernels on the original continuous data. The main reason that ACE cannot deal with continuous data lies in the step of computing conditional expectations.

$$\hat{\mathbb{E}}_n[g(Y)|X=x] = \frac{\sum_{i=1}^{N} g(y_i)\mathbb{1}(x_i = x)}{\sum_{i=1}^{N} \mathbb{1}(x_i = x)} \tag{1}$$

where $\mathbb{1}(\cdot)$ is an indicator function.

Equation 1 shows how to compute empirical expectations in ACE using discrete data samples. In continuous-valued data scenario, $\mathbb{1}(x_i = x)$ can be small or only 1 thus loses information contained in the neighbourhood of x. Therefore, we apply kernel smoothing [2] technique which uses values of points nearby for expectation estimation. Intuitively, we replace the indicator function in Eq. 1 in numerator to a kernel $K(x_i, x)$ shown as Eq. 2. It is used to replace Eq. 1 in ACE as step 2a in Algorithm 1.

$$f(x) = \hat{\mathbb{E}}_n[g(Y)|X=x] \leftarrow \frac{\sum_{i=1}^{N} g(y_i)K(x_i, x)}{N} \tag{2}$$

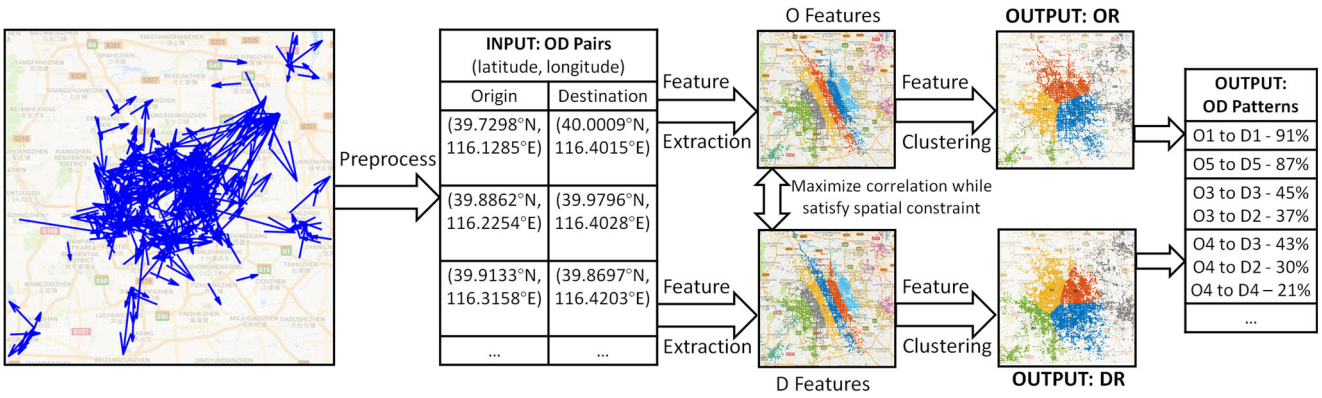


Fig. 1 Flow chart of our algorithm

Equation 2 considers $g(\cdot)$ of points around x with weights determined by kernel $K(x_i, x)$. When $K(x_i, x) = 1$, $g(y_i)$ feature of pair (x_i, y_i) is fully considered to compute $f(x)$. When $K(x_i, x) = 0$, $g(y_i)$ has no influence on $f(x)$ at all.

Algorithm 1 D-dimensional kernelized ACE for clustering.

Require: training samples $\{(x_i, y_i) : i = 1, ..., N\}$

1. Initialize: randomly generate and regularize $f_d(x_i)$, $g_d(y_i)$, $i = 1, \cdots, N, d = 1, \cdots, D$

repeat

2a. **Feature iteration:**

$f_d(x_i) \leftarrow \frac{\sum_{j=1}^{N} g_d(y_j)K(x_j,x_i)}{N}$, $K(x_j, x_i) = 1 - \|x_j - x_i\|_2$

$g_d(y_i) \leftarrow \frac{\sum_{j=1}^{N} f_d(x_j)K(y_j,y_i)}{N}$, $K(y_j, y_i) = 1 - \|y_j - y_i\|_2$

2b. **Regularize:** $f_d(\cdot), g_d(\cdot), d = 1, ..., D$.

$f_d(x_i) \leftarrow f_d(x_i) - \frac{\sum_i f_d(x_i)}{N}$, $f_d(x_i) \leftarrow \frac{f_d(x_i)}{\sqrt{\frac{\sum_i f_d(x_i)^2}{N}}}$

$g_d(y_i) \leftarrow g_d(y_i) - \frac{\sum_i g_d(y_i)}{N}$, $g_d(y_i) \leftarrow \frac{g_d(y_i)}{\sqrt{\frac{\sum_i g_d(y_i)^2}{N}}}$

2c. **Gram-Schmidt:**

for $d = 1$ to D **do**

for $k = 1$ to $d - 1$ **do**

$f_d(x) \leftarrow f_d(x) - \frac{\langle f_k(x), f_d(x)\rangle}{\langle f_k(x), f_k(x)\rangle} f_k(x)$

$g_d(x) \leftarrow g_d(x) - \frac{\langle g_k(x), g_d(x)\rangle}{\langle g_k(x), g_k(x)\rangle} g_k(x)$

end for

end for

until $\mathbb{E}[f_{1,2\cdots D}(x)^T g_{1,2\cdots D}(y)]$ stops to increase

3. **Output:** Region label of each point x_i and y_i

3a. $OR(x_i) \leftarrow$ Linkage cluster of $f_{1,\cdots,D}(x_i), i = 1, \cdots, N$ with maximum cluster number N_x,

3b. $DR(y_i) \leftarrow$ Linkage cluster of $g_{1,\cdots,D}(y_i), i = 1, \cdots, N$ with maximum cluster number N_y

4.2.3 Extract D-dimensional deatures

Equation 2 tells us how to extract 1D feature $f(x)$ and $g(y)$. However, mobility patterns are complicated and extracting only one dimension of feature is not adequate. Therefore, we further extract D-dimensional features in Algorithm 1. The key idea is to obtain features that are orthogonal to the previously obtained features to avoid redundant information [16]. Therefore, we apply Gram-Schmidt orthogonalization to the feature functions in step 2c.

4.3 Kernel choice

One challenge of Kernelized ACE is how to choose the most appropriate kernel function.

Recall that the spatial constraint states that features should be close to each other when data points are close. Therefore, features of points close to the specific point of interest are more important than features of points far away. In Eq. 2, kernel $K(x_i, x)$ serves as feature weights, which should satisfy the following two properties.

1. $K(x_i, x)$ is a non-increasing function of $\|x_i - x\|_2$
2. $K(x_i, x) = 1$ when $\|x_i - x\|_2 = 0$

Intuitively, data that are close to x should have higher weights for computing $f(x)$, which can be expressed mathematically in Property 1. It also satisfies $\|f(x_i) - f(x_j)\|_2$ increases as $\|x_i - x_j\|_2$ increases. Property 2 comes from the meaning of weights, where $K(x_i, x) = 1$ means feature $g(y_i)$ of trip (x_i, y_i) is fully considered while $K(x_i, x) = 0$ means $g(y_i)$ is not considered completely.

The following three kernels satisfy the aforementioned two properties and are commonly used in kernel smoothing techniques. They are

1. window kernel, $K_1(x_i, x) = \mathbb{1}(\|x_i - x\|_2 <= \gamma)$,, parameterized by $\gamma(\gamma > 0)$;
2. Gaussian kernel, $K_2(x_i, x) = e^{-\frac{\|x_i - x\|_2^2}{2\sigma^2}}$, parameterized by $\sigma(\sigma > 0)$;
3. negative linear kernel, $K_3(x_i, x) = 1 - a\|x_i - x\|_2$, parameterized by $a(a > 0)$.

The features f and g are not only influenced by different kernels, but also parameter choices. However, when we don't have prior knowledge of what we are looking for, negative linear kernel (K_3) is a good choice of identifying significant data correlation patterns for continuous data. We prove that parameters of negative linear kernel don't affect correlation's results in Theorem 4.1[2]. Therefore, negative linear kernel, $K_3(x_j, x) = 1 - \|x_j - x\|_2$, is chosen to be used in Algorithm 1.

Theorem 1 *Different intercepts or slopes for negative linear kernel do not change the results of features, i.e.* $K_1(x_i, x) = 1 - \|x_i - x\|_2$, $K_2(x_i, x) = b - \|x_i - x\|_2$ *and* $K_3(x_i, x) = 1 - a\|x_i - x\|_2 (a > 0)$ *produce the same features* f *and* g.

Based on proximity of extracted features, we cluster high-dimensional features into O/D region labels using linkage clustering [24] methods. Finally, the whole kernelized ACE algorithm is shown in Algorithm 1.

[2]Proof of Theorem 4.1 can be found in the initial version [14].

Table 1 Summary of Beijing and NYC Taxi Dataset

		Beijing 2012	Beijing 2014	Beijing 2015	NYC 2015
17:00-17:59	Total Trip Number	86311	155370	118433	213175
	Average OD Distance	4.42km	4.64km	3.63km	2.95km
	Filtered Trip Number	48900	88918	54199	127648
7:00-7:59	Total Trip Number	70755	129867	116817	208336
	Average OD Distance	5.66km	5.78km	4.71km	3.38km
	Filtered Trip Number	47250	86069	65330	137140

5 Evaluation

5.1 Data description

We evaluated Algorithm 1 and conducted case studies on two real world datasets of Beijing and New York City. Since mobility patterns depend on time-of-day, day-of-week and other temporal factors, one month of weekdays' data in November during the morning and evening rush hour 7:00-7:59/17:00-17:59 were considered.

For New York City, we used open source dataset [26] of the year 2015 published by the NYC Taxi and Limousine Commission. The data only contains OD pair information with time stamps without intermediate points. We combined trips of both yellow taxis, serving Manhattan exclusionary zone, and green taxis, serving borough areas, to evaluate taxi trips covering the whole city. Besides, only trips longer than 2km were considered.

For Beijing taxi data, we extracted occupied trip information from the raw GPS trajectories with sample rate at around 1 minute in year 2012, 2014 and 2015. Each sample contains a unique taxi number, time stamp, latitude, longitude, and occupancy indicator (occupied 1, vacant 0 or stopped −1). We determined pick-up happens when the occupancy indicator turns from 0 to 1 and drop-off happens when occupancy turns from 1 to 0. In addition, we omitted short trips to eliminate incorrectly recorded data. Only trips longer than 3km and lasting more than 1min were considered. Beijing's filtered OD distance is longer than New York's because Beijing has longer average OD distance. More details of the taxi datasets are shown in Table 1.

5.2 Feature evaluation

We compared extracted feature of KACE with two typical feature extraction methods (CCA [8] and KCCA [7]) using Beijing's evening data. CCA and KCCA correlate linear and high-dimensional relationships between two variables respectively. The features are evaluated by three metrics:

Correlation: measures correlation of f and g's different dimension of features.

Validity: is the correlation coefficient between feature distance $\|f_i - f_j\|_2$ and coordinates distance $\|x_i - x_j\|_2$. A bigger value means that the features satisfy the spatial constraint better.

Distribution Kurtosis: measures the "tailedness" of a distribution. A larger value means the distribution has heavier tail, more outliers and higher peak. High kurtosis leads to difficulties in clustering features.

The results are shown in Table 2. For CCA, it found linear projections of latitudes and longitudes but only extracted two dimensions of features. For KCCA, Gaussian kernel was applied here with best kernel parameter chosen and regularization for comparison. Both CCA and KCCA result to high feature correlation even better than KACE. However, they have much bigger kurtosis and lead to larger clusters as shown in Table 3 of cluster results. Besides, we also found that the first two dimensions of features extracted by KACE are similar to CCA's two dimensional features with 0.9556 and 0.9522 similarity respectively. The similarity is measured by the cosine of angle between two feature vectors. Since ACE is a non-linear generalization

Table 2 Feature comparisons of KACE, CCA and KCCA

	Correlation			Validity		Kurtosis (f)		
	D1	D2	D3	f	g	f_1	f_2	f_3
KACE	0.85	0.82	0.76	0.95	0.95	2.39	1.99	7.79
CCA	0.87	0.82	/	0.98	0.98	5.03	3.54	/
KCCA	0.88	0.84	0.84	0.89	0.89	5.64	3.97	14.59

Table 3 Comparisons with other methods based on spatial coverage, in-cluster distance and regional dependency

Methods	Spatial Coverage	Average Origin in-cluster Distance	Average Destination in-cluster Distance	Regional Correlation	Origin Overlap	Destination Overlap
KACE	100%	2.98km	3.21km	0.8643	0.33%	0.22%
ACE-5 × 5	100%	5.06km	5.21km	0.8068	0.78%	0.64%
ACE-20 × 20	100%	8.48km	8.78km	0.8646	0.75%	0.61%
MLAN	100%	11.82km	12.58km	1	4.43%	4.19%
CCA	100%	4.38km	4.78km	0.8480	0.34%	0.22%
KCCA	100%	4.99km	6.12km	0.8576	0.32%	0.35%
K-Means++	100%	4.26km	4.42km	1	54.26%	50.75%
DBSCAN	25.75%	0.60km	0.63km	1	39.21%	35.85%

of CCA [16], our method KACE captures both linear and non-linear correlated features. Only in this case, the top two most correlated features happens to be linear ones as CCA's features and a non-linear feature in the third feature dimension which CCA cannot extract. The results show that KACE captures major correlations of data while has tradeoff with less outliers.

5.3 Model analysis

We analyzed variations of regional correlation from two aspects: feature dimension and number of clusters.

Feature Dimension. For discrete variables, an upper bound of feature dimension is the variable's cardinality. While for continuous variables, we can extract as many features as we want. However, there is a trade-off between feature dimensions and computational power.

Cluster Number. Number of clusters also influences the OD patterns and the understanding of the city. Though many regions provide detailed analysis of mobility patterns, they lack holistic understanding and guiding information for city management authorities.

Regional Correlation. We obtain discrete O/D regions and compute regional correlations as correlation metric of our model. We define a matrix as $B = \frac{P_{O,D}}{\sqrt{P_O}\sqrt{P_D}}$ [16], where O/D represents origin/destination regions' label respectively. Correlations are singular values of B in a descending order, i.e. $corr_i = \sigma_i$, where $\sigma_0 = 1 \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{K-1}$ are singular values of B matrix , in which $K = min\{|\mathcal{O}|, |\mathcal{D}|\}$ [16]. Instead of using just the maximal regional correlation ($corr_1$) as the correlation metric, we used an average of top-5 correlations, called *average regional correlation* defined in Eq. 3. Because maximal correlation applied in real-world data may encounter the problem that it weighs too much on the most salient dependent pattern and fails to identify other informative ones. Therefore, we took next four correlations into consideration for identifying more patterns and avoiding some insignificant ones.

$$corr = \frac{\sum_{i=1}^{i=5} corr_i}{5} = \frac{\sum_{i=1}^{i=5} \sigma_i}{5} \tag{3}$$

We studied how regional correlations changed with the increase of feature dimensions for different cluster numbers and showed results in Fig. 2. We extracted origin and destination regions using the first $D(1 \leq D \leq 5)$ dimensions of KACE feature and set region number from 20 to 60. We first showed Beijing's results in Fig. 2a. The correlation under different number of clusters generally increases with more features added in, but has a turning point at minimum optimal feature dimension. Specifically, the correlations raise significantly for the first three dimension of features and reach a stable stage (when

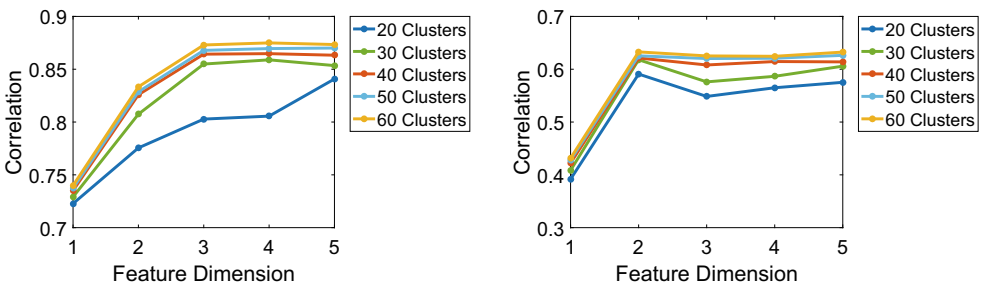


Fig. 2 Relationship between regional correlation to number of clusters and feature dimensions in Beijing and NYC

(a) Beijing, 17:00-17:59

(b) New York City, 17:00-17:59

cluster number is more than 30). Therefore, the minimum optimal feature dimension for obtaining mobility patterns in Beijing should be three. The results of New York's data show similar findings in Fig. 2b. The difference is that the minimum optimal feature dimension is two. Therefore, in our case studies of Beijing and NYC, we used three and two dimension of features for each city respectively. With number of clusters increasing, the regional correlation increases gradually but the increasing step decreases for both cities. It tells us that separating for more clusters benefits only a little for regional correlations. Therefore, we chose a cluster number of 40 for Beijing and 30 for NYC to achieve a comparatively higher correlation with fewer clusters in our case studies.

5.4 Cluster performance analysis

We compared KACE with both traditional methods and state-of-the-art co-clustering algorithms on the Beijing dataset using four metrics:

Spatial coverage, measures the percentage of points identified and clustered by the given method.

Average in-cluster distance, measures average pair-wise distance within the same region. A smaller value is better.

Regional correlation, average of the top-five correlations among all pairs of OD regions defined in Section 5.3. We use this metric to measure how well the clustering methods retain trip information.

Overlap, estimates the spatial ambiguity among different origin or destination regions. This metric is computed using the KNN classification error of cluster labels based on origin or destination coordinates, which should be very small for non-overlapping data [3]. In particular, the KNN model (K=5) was trained and tested on a 9-1 split of the origin or destination points, i.e. 10-fold cross-validation classification error was used as overlapping indicator.

For single-view methods, the classic K-Means++ [1] and DBSCAN [4] were adopted as baselines. We directly clustered OD trips using these two methods by treating trips as points in four dimensional space, i.e. latitudes and longitudes of origins and destinations respectively. Then evaluated the results in O/D view. For co-clustering methods, we compared with state-of-the-art multi-view clustering algorithm MLAN [19]. We also compared with the performance of the original ACE algorithm, by discretizing GPS data using 5×5 and 20×20 grids. Results of using features extracted by both CCA [8] and KCCA [7] are also listed. For fair comparison, we set the cluster number to 40 for all methods as shown in Table 3. The single-view methods, K-means++ and DBSCAN, regional correlations are at the maximum value 1. However, the resulting clusters have a significant amount of spatial overlap. All the co-clustering methods have less overlap. MLAN has great regional correlation, tolerable overlap but fails to identify detailed regions as the average in-cluster distance is greater than 10km. Because MLAN only focuses on achieving best correlation without trade-offs on cluster size, which lead to large regions and fails to provide detailed guidance for city management. The original ACE algorithm improves upon the results of MLAN by introducing cluster size trade-offs manually with discretization, though it is not as good as KACE. Other feature extraction methods, CCA and KCCA, result to bigger regions as their features are denser and harder to differentiate for the clustering method. Our algorithm, KACE, outperforms all the other methods for overall performance, which has minimum overlap, best spatial coverage, small region size and good correlation.

6 Case study: New York City urban dynamics analysis

6.1 OD transition probability

As pointed out in Section 5.3, we used first two dimension of features with 30 regions for the NYC case study. After getting O/D regions, we obtained OD mobility patterns with transition probability $P(O, D)$.

$$P_{O,D}(o, d) = \frac{\text{\# of trips started from region } o \text{ and ended at } d}{\text{Total \# of trips}}$$

For easier visualization and interpretation, we permutated the region labels O and D such that the OD patterns with highest probability lie on the diagonal of the transition matrix in orders. This problem is known as the **Munkres Assignment Problem** which can be solved in polynomial time by the Munkres Algorithm [17]. In our setup, we want to find an assignment of O and D labels, such that the sum of probabilities from region O_i to D_i is maximized. Therefore, if the OD pattern is one-to-one, the OD transition probability matrix should be diagonal. Stronger diagonality represents better OD correlations. Figure 3 shows OD mobility patterns with transition probability in NYC.

6.2 Urban dynamics and region functionalities

Figure 4a shows the NYC region partition results. For visualization, we colored adjacent regions differently to distinguish from surrounding regions; O/D regions with the same ID number shared the same color. We can see that **NYC's mobility regions align with its city topology.** Both the origin and the destination partition identify existing districts in NYC such as Manhattan Borough, Bronx and

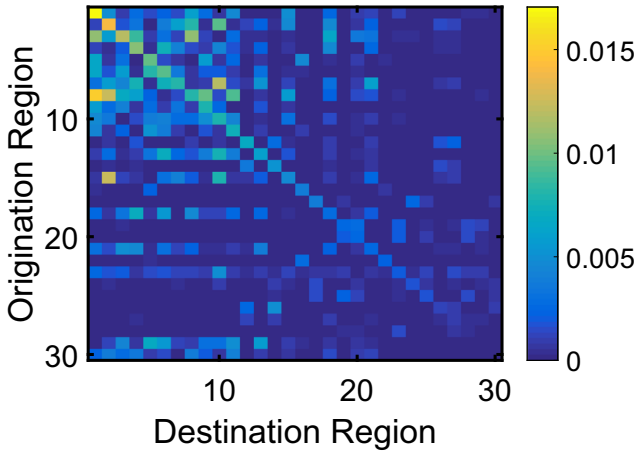


Fig. 3 NYC mobility patterns with OD transition probability of 30 regions, workdays of November 2015, 17:00-17:59

Brooklyn. For example, the outlines between O5 and O16, O2 and O12, separate Manhattan Borough from Bronx and Brooklyn respectively. On the Manhattan island, the regions reveal the block city topology.

Each region has different functionalities in terms of Point-Of-Interest (POI). We obtained NYC's POIs from NYC Open Data [20] and only adopted POI labels established before 2016 to be consistent with the taxi trip data. Seven functionalities are considered in our analysis as shown in Fig. 4b. For example, JFK international airport is the most popular POI in region O24, thus more than 80% of O24's POI distribution is transportation; while region O26 is a major residential area.

Combining with mobility patterns, we analyzed how people move between areas of different functionalities. We computed the POI transition probability $P(F_1, F_2)$ as following

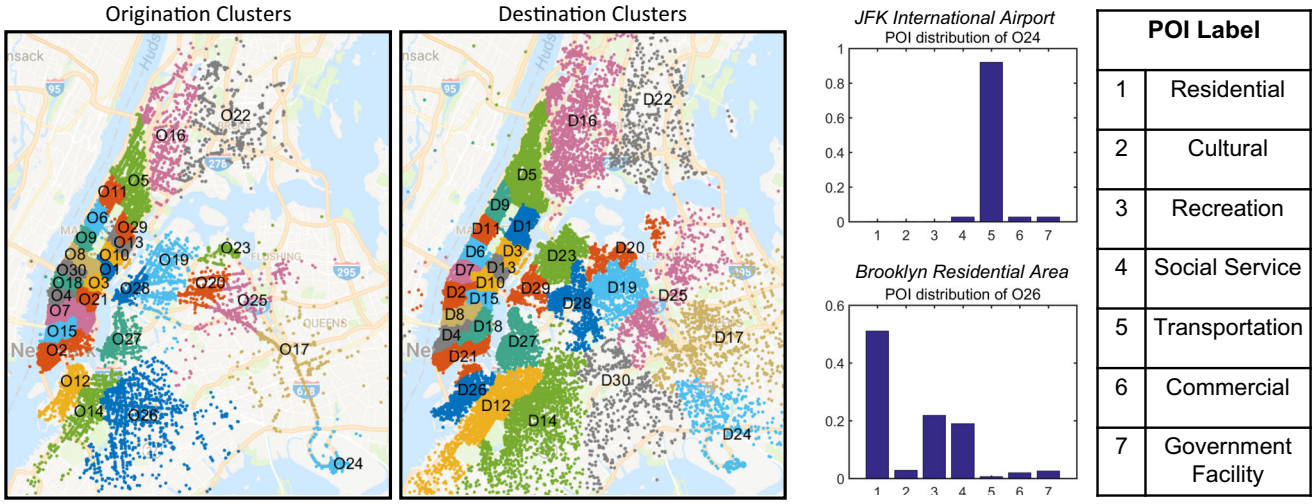
$$P(F_1, F_2) = \sum_{O,D} P(F_1|O)P(F_2|D)P(O, D)$$

in which $P(F_i|O)$ is the percentage of POI label F_i out of all the POIs in region O, $P(O, D)$ is the OD transition probability defined in Section 6.1. Figure 5 plotted all the major urban dynamics. We found that **people move in opposite directions in the morning and evening.** In the evening, people move from Government Facility and Commercial areas to Residential area; while in the morning, trips are made in the opposite direction originated from Residential areas to Government Facility and Commercial areas. Moreover, **popular mobility movements are the same regardless of morning or evening.** Major dynamic patterns like movements between Recreation and Residential are popular in both morning and evening. The only difference is that in the morning both directions have similar probability of moving between Recreation and Residential while more people go from Recreation to Residential (0.027) than the reverse direction (0.024) in the evening.

7 Case study: Beijing urban analytics by mobility pattern comparisons in Year 2012, 2014 and 2015

7.1 Comparisons of regional correlation

Figure 6 illustrates that **mobility patterns become more concentrated from 2012 to 2015** as regional correlations increase gradually. For evening's OD patterns, regional



(a) Region partitions of NYC with 30 clusters, workdays of November 2015, 17:00-17:59 (b) POI labels and examples of region functionalities

Fig. 4 NYC region partitions and POI functionalities

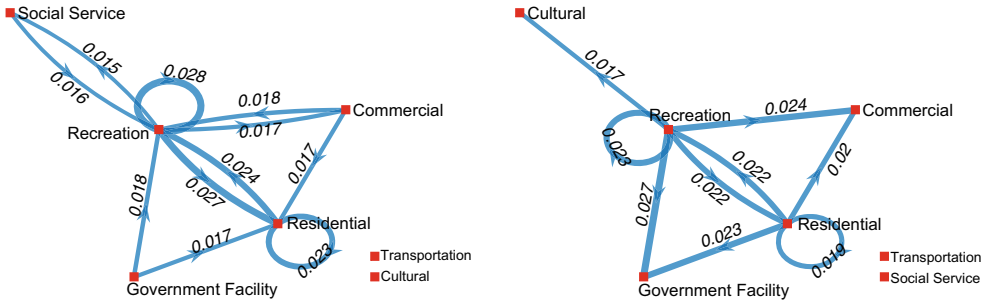


Fig. 5 Urban dynamics of NYC, workdays of Nov. 2015

(a) 17:00-17:59 (b) 7:00-7:59

correlation has a 39.51% increase from 2012 to 2014; while only 7.84% increase from 2014 to 2015. This indicates that mobility patterns change significantly from 2012 to 2014 but only a little from 2014 to 2015. Moreover, **evening mobility patterns are more regular than morning patterns in 2014 and 2015** as evening's correlations are always higher than morning's. Therefore, we made further analysis of mobility patterns over three years using data during 17:00-17:59, in order to make fair comparisons between years.

7.2 Macroscopic analysis

From a macroscopic perspective, we partitioned the city of Beijing into two regions as shown in Fig. 7, which is also the minimum number of regions, as partitioning a city into one cluster results in its own.

Top-level zoning of Beijing has been changed, from North-South division to Inner-Outer division Beijing is divided into North-Eastern and South-Western regions in 2012 (Fig. 7a). While in 2014 (Fig. 7b) and 2015 (Fig. 7c), the South-Western corner of Beijing, i.e. Fangshan District, is isolated from other areas, as it is far away from central areas and the development of that district is not as evolved as other areas. Thus, Beijing is separated into inner city and outer city, which includes regions like the Fangshan District. We also found that the interaction between inner city and Fangshan District is little, since only less than 0.1% of trips go from inner city to Fangshan District. The reverse direction, from Fangshan District to inner city, is more active but the probability of movement is also small, with about only 3% trips originated from Fangshan ended at inner city.

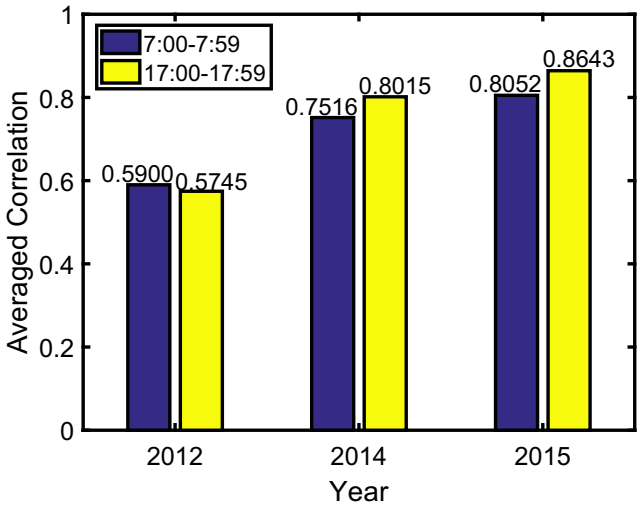


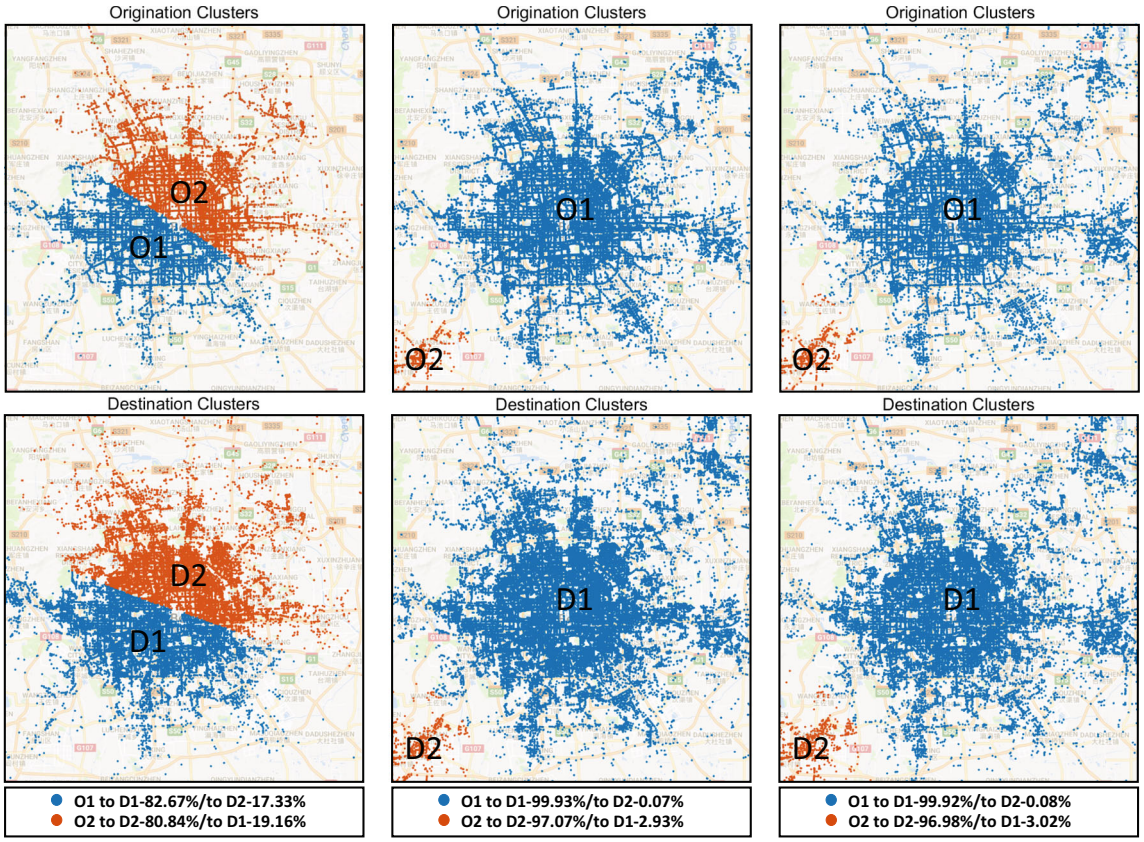
Fig. 6 Regional correlation of three year in Beijing

7.3 Microscopic analysis

In this section, we focused on analysing more detailed region divisions. Each year has 40 O/D clusters for fair comparisons as shown in Fig. 8. For clarity, we only displayed salient patterns where ORs have more than 20% trips ended at the DR, i.e. $P(D = d|O = o) > 0.2$.

New regions emerged and neighboring areas may become similar in Beijing Apart from OD patterns, the region partitions reveal urban changes among years. A new region Dongba emerged, as O32 and D32 in 2014 and O33 and D40 in 2015, which is a new residential area. Moreover, D1, D12 and D24 in 2012 merged into a large area D12 in 2014, which is labeled as D10 in 2015. This region has been developed into a mainly residential area at the west corner in Beijing. These changes show the development from 2012 to 2014 and 2015. It appears that region partition of 2015 is more similar to 2014's than 2012's. In order to verify this observation, we established a numerical method to compare between different years' region partition.

Adjusted Rand Index [10] and Normalized Mutual Information [29] are typical indexes to evaluate the clustering performance with the ground truth. Although we don't have a ground truth for region division, different years' partitions can be compared by these indexes to evaluate region partition's similarities. However, two issues remain to be addressed. First, spatial coverage of 2012, 2014 and 2015 are different and some suburban areas are



(a) 2012 (b) 2014 (c) 2015

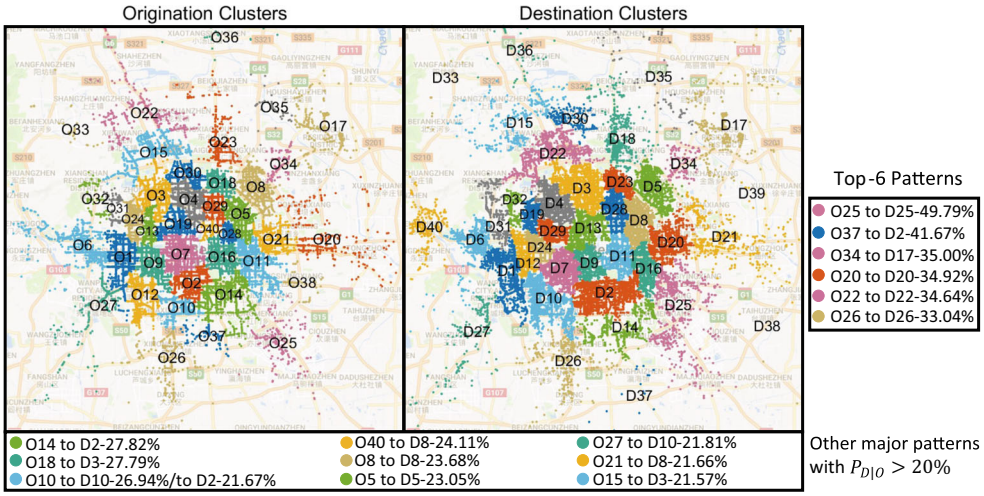
Fig. 7 Three-year OD patterns with two clusters in Beijing, workdays in November, 17:00-17:59

not sampled in 2012's data. Second, even for the same spatial coverage, different points in different years represent different locations, which cannot be compared directly using cluster labels. Therefore, we proposed to capture the partition scheme of each year by k-nearest neighbor (KNN) method, and predicted the partitions on the same dataset to be compared with. Here, we adopted k=5 and modeled each year's origin and destination partition by different time slot. Because the spatial coverage of 2012 is limited, we used KNN to model clustering results of 2014 and 2015, and made predictions on 2012's data to compare with 2012's region partition. For comparisons between 2014 and 2015, we modeled results of 2015 and made predictions on 2014's data. The result is shown in Table 4.

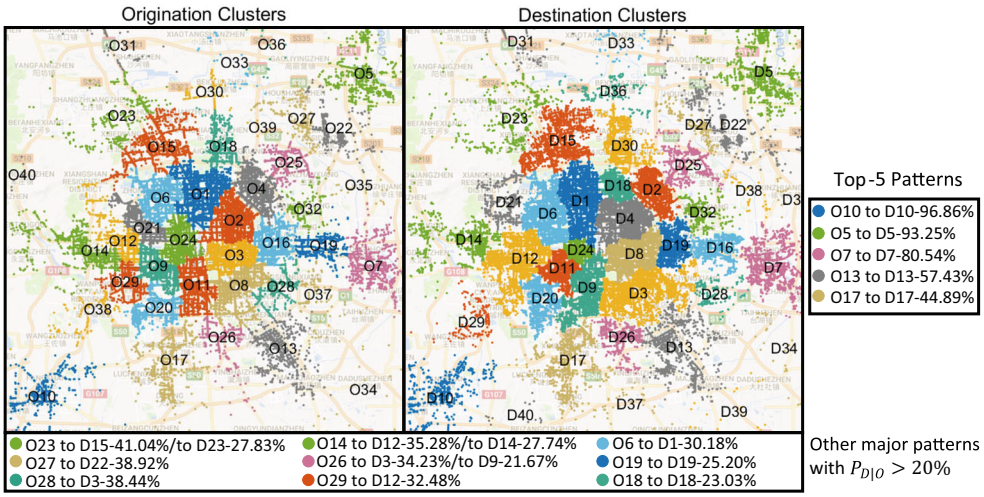
The two indexes may result in different values but the trend is the same for both of them. A bigger value means more similar partitions. For the morning peak hour, we found that close years have more similar region divisions for both origins and destinations. In particular, the indexes between region divisions in 2014 and 2015 are larger than those between 2012 and 2014, 2012 and 2015. For evening peak hour, the destination partition follows the same rule. However, a rare exception is the origin partition, i.e. 2015's partition is more similar to 2012's than 2014's in the evening.

Suburban districts in Beijing develop from 2012 to 2015 From top OD patterns of each year in Fig. 8, we found that people tend to move inside suburban districts by taxis. This trend grows stronger from 2012 to 2015 because of the development and better functional division of rural areas. From the data of 2012, we observe that the coverage of outer part (South-Western and North-Eastern corner) of Beijing is much less than 2014's and 2015's indicating an outward expansion in the year 2014 and 2015.

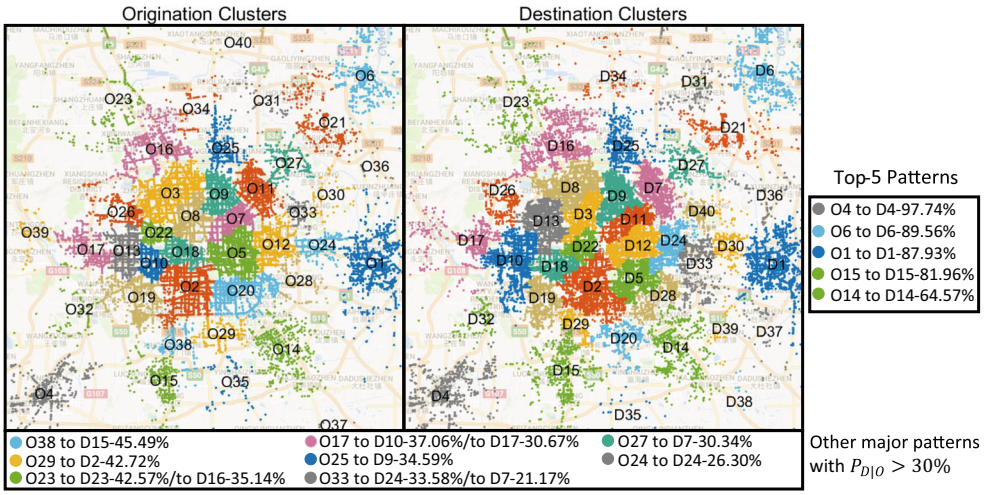
The top patterns in 2012 including O25 to D25 (Yizhuang to Yizhuang), O20 to D20 (Tongzhou to Tongzhou) and O26 to D26 (Daxing to Daxing) have more than 30% probability of internal movements. However, in 2014, the probability of OD trips originated and ended at those remote areas increases significantly, shown in Fig. 8b. Salient patterns of O10 to D10 (moving inside Fangshan District) and O5 to D5 (moving inside Shunyi District) in 2014 are not even included in 2012's patterns, indicating outward expansion



(a) 2012



(b) 2014



(c) 2015

Fig. 8 Three-year OD patterns with 40 clusters in Beijing, workdays in November, 17:00-17:59

Table 4 Comparisons of region partition between year 2012, 2014 and 2015 in Beijing, 7:00-7:59 and 17:00-17:59, workdays of November

		7:00-7:59		17:00-17:59	
	Years	ARI	NMI	ARI	NMI
OR	2012 and 2014	0.5234	0.7720	0.5414	0.7927
	2012 and 2015	0.5650	0.7961	0.5255	0.7885
	2014 and 2015	0.6041	0.8217	0.7285	0.8496
DR	2012 and 2014	0.5454	0.7938	0.6197	0.7986
	2012 and 2015	0.4950	0.7603	0.5787	0.7906
	2014 and 2015	0.6310	0.8425	0.7409	0.8293

of Beijing and newly surfaced suburban regions. Morover, Shunyi District as in O17 of Fig. 8a has evolved into three regions of O5 (downtown area of Shunyi), O22 (Beijing International Airport area) and O27 (airport surrounding area) in Fig. 8b with different region functions. Similarly, O20 (Tongzhou) area in 2012 has expanded outward and formed into O19 (Guanzhuang) and O7 (Tongzhou) areas, which is a sign of functional division. Besides, both O13 (Yizhuang) and O17 (Daxing) have matching regions O25 (Yizhuang) and O26 (Daxing) in 2012, but with around 10% probability increase of internal movements.

In the year of 2015, the suburban regions almost remain the same while the probability of suburban population staying within their respective regions keeps increasing. The most salient increase is the internal movement in Daxing, which increases from 44.89% O17 to D17 in 2014 to 81.96% probability of O15 to D15 in 2015.

Outer city mobility patterns have overlap from 2012 to 2015 O22 to D22 (34.64%) in 2012, O23 to D15 (41.04%) in 2014 and O23 to D16 (35.14%) in 2015 are from Xibeiwang to Qinghe. O23 to D23 (22.69%) in 2012, O18 to D18 (23.03%) in 2014 and O25 to D9 (34.59%) in 2015 are from Beiyuan commercial area to Datun residential area. O20 to D20 (34.92%) in 2012, O19 to D19 (25.20%) in 2014 and O24 to D24 (26.30%) in 2015 from Guanzhuang to Sihui. Although these patterns are not in the suburban area, they still are far from central areas. On the contrary, for the inner city, both region partition and mobility patterns change constantly in year 2012, 2014 and 2015.

8 Conclusion

Mobility pattern mining and city region partition are important socioeconomic problems. However, few methods combine them together while these two problems are closely related. Therefore, we proposed a region-aware mobility pattern mining scheme by jointly partitions city regions while extracting OD patterns. Kernelized ACE, is developed to extract features that both satisfy regional spatial constraints and achieve maximal OD correlations. Experimental results on Beijing's taxi data show that our approach extracts precise features and outperforms all the other methods including both traditional and state-of-art algorithms. We also analyzed how to choose optimal feature dimension and number of clusters in our algorithm. Our case studies on New York City and Beijing provide insightful findings. Regions obtained from a mobility perspective could reveal different regions' functions and how people move between areas of different functionalities in NYC. A three-year consecutive analysis of Beijing reveals the city's development and urban sprawl. Future directions of this work involve taking other factors into consideration, like temporal index. Besides, we could further develop it for applications like mobility service demands prediction.

Acknowledgements The work is funded by Shenzhen Municipal Development and Reform Commission (SDRC[2015]1872), Natural Science Foundation of China (61807021), Shenzhen Science and Technology Research and Development Funds (JCYJ20170818094022586), Shenzhen Innovation and Entrepreneurship Project for Overseas High-Level Talents (KQJSCX20180327144037831).

References

1. Arthur D (2007) Vassilvitskii, s.: k-means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, pp 1027–1035
2. Breiman L, Friedman JH (1985) Estimating optimal transformations for multiple regression and correlation. J Am Stat Assoc 80(391):580–598
3. Ding C, He X (2004) K-nearest-neighbor consistency in data clustering: incorporating local information into global optimization. In: Proceedings of the 2004 ACM Symposium on Applied Computing, ACM, pp 584–589
4. Ester M, Kriegel HP, Sander J, Xu X et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd, vol 96, pp 226–231
5. Gao S (2015) Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. Spat Cogn Comput 15(2):86–114

6. García-Palomares JC (2010) Urban sprawl and travel to work: the case of the metropolitan area of madrid. J Transp Geogr 18(2):197–213
7. Hardoon DR, Szedmak S, Shawe-Taylor J (2004) Canonical correlation analysis: an overview with application to learning methods. Neural Comput 16(12):2639–2664
8. Hotelling H (1936) Relations between two sets of variates. Biometrika 28(3/4):321–377
9. Huang X, Zhao Y, Ma C, Yang J, Ye X, Zhang C (2016) Trajgraph: a graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. IEEE Trans Vis Comput Graph 22(1):160–169
10. Hubert L (1977) Nominal scale response agreement as a generalized correlation. Br J Math Stat Psychol 30(1):98–103
11. Inostroza L, Baur R, Csaplovics E (2013) Urban sprawl and fragmentation in latin america: a dynamic quantification and characterization of spatial patterns. J Environ Manag 115:87–97
12. Kuo CT, Bailey J, Davidson I (2015) A framework for simplifying trip data into networks via coupled matrix factorization. In: Proceedings of the 2015 SIAM International Conference on Data Mining, SIAM, pp 739–747
13. Li Z, Hong Y, Zhang Z (2016) An empirical analysis of on-demand ride sharing and traffic congestion
14. Lian J, Li Y, Gu W, Huang SL, Zhang L (2018) Joint mobility pattern mining with urban region partitions. In: Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, ACM, pp 362–371
15. Liu X, Gong L, Gong Y, Liu Y (2015) Revealing travel patterns and city structure with taxi trip data. J Transp Geogr 43:78–90
16. Makur A, Kozynski F, Huang SL, Zheng L (2015) An efficient algorithm for information decomposition and extraction. In: 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (allerton), IEEE, pp 972–979
17. Munkres J (1957) Algorithms for the assignment and transportation problems. J Soc Ind Appl Math 5(1):32–38
18. Netflix: Netflix prize data set. http://archive.ics.uci.edu/ml/datasets/Netflix+Prize
19. Nie F, Cai G, Li X (2017) Multi-view clustering and semi-supervised classification with adaptive neighbours. In: AAAI, pp 2408–2414
20. NYC IT and Telecommunications Department (2016) Poi nyc open data. https://data.cityofnewyork.us/City-Government/Points-Of-Interest/rxuy-2muj
21. Priemus H, Nijkamp P, Banister D (2001) Mobility and spatial dynamics: an uneasy relationship. J Transp Geogr 9(3):167–171
22. Qi G, Li X, Li S, Pan G, Wang Z, Zhang D (2011) Measuring social functions of city regions from large-scale taxi behaviors. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), IEEE, pp 384–388
23. Rényi A (1959) On measures of dependence. Acta Math Hungar 10(3-4):441–451
24. Sokal RR (1958) A statistical method for evaluating systematic relationship. University of Kansas Science Bulletin 28:1409–1438
25. Tang J, Liu F, Wang Y, Wang H (2015) Uncovering urban human mobility from large scale taxi gps data. Physica A:, Statistical Mechanics and its Applications 438:140–153
26. Toddwschneider (2017) Unified NYC Taxi and Uber data. https://github.com/toddwschneider/nyc-taxi-data/
27. Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, González MC (2015) The path most traveled: Travel demand estimation using big data resources. Transportation Research Part C:, Emerging Technologies 58:162–177
28. UN Department of Economics and Social Affairs PopulationDivision (2015) World urbanization prospects: The 2014 revision
29. Vinh NX, Epps J, Bailey J (2010) Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. J Mach Learn Res 11:2837–2854
30. Yuan J, Zheng Y, Xie X (2012) Discovering regions of different functions in a city using human mobility and pois. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp 186–194
31. Zhang W, Li S, Pan G (2012) Mining the semantics of origin-destination flows using taxi traces. In: Ubicomp, pp 943–949
32. Zhu B, Huang Q, Guibas L, Zhang L (2013) Urban population migration pattern mining based on taxi trajectories. In: 3Rd International Workshop on Mobile Sensing: the Future, Brought to you by big Sensor Data, Philadelphia, USA