Exploring OD Patterns of Interested Region based on Taxi Trajectories

Min Lu · Jie Liang · Zuchao Wang · Xiaoru Yuan

Abstract Traffics of different regions in a city have different Origin-Destination (OD) patterns, which potentially reveal the surrounding traffic context and social functions. In this work, we present a visual analysis system to explore OD patterns of interested regions based on taxi trajectories. The system integrates interactive trajectory filtering with visual OD patterns exploration. Trajectories related to interested region are selected by a suite of graphical filtering tools, from which OD clusters are detected automatically. OD traffic patterns can be explored at two levels: overview of OD and detailed exploration on dynamic OD patterns, including information of dynamic traffic volume and travel time. By testing on real taxi trajectory data sets, we demonstrate the effectiveness of our system by case studies.

Keywords OD Pattern · Trajectory · Filter

1 Introduction

Understanding the movement of human beings and vehicles in large cities is crucial in transportation field. Such movement, known as the Origin-Destination (OD) pattern, reveals the traffic context of a city, including the hot OD regions and regular commuting patterns among regions. Many research efforts have been put in studying urban OD patterns. For example, regular commuting patterns are extracted by studying traffic flows between OD regions (Peng et al. 2012). Regions’ land-use patterns are inferred by analysing the dynamics of traffic flow related to the regions (Pan et al. 2013). OD patterns in urban can be indicated from different types of movement data. Travelling around the city in a considerable spatial and temporal scale, taxis are viewed as the representative of urban traffic. Many traffic related researches are conducted based on taxi trajectory data, such as inferring traffic jams by studying taxis’ traffic speed (Liu et al. 2011b; Wang et al. 2013), studying traffic condition of single route (Lu et al. 2015c) or multiple routes (Liu et al. 2011a) (Lu et al. 2015a) from passing taxis. As a typical urban movement data, we use taxi trajectories to study urban OD patterns in this work.

Different from work to study the global traffic flows around the city (Andrienko and Andrienko 2011), we focus on local OD pattern analysis which may be dimmed in global analysis. Given the regions of interest, our goal is to extract related OD regions and explore the OD patterns. Specifically, we target at exploring OD patterns from two levels:

– **Overview of OD regions**: to explore where the traffic flows mainly come/go and how traffic related statistics of OD regions distribute.

– **Detail exploration on OD patterns**: to explore how traffic related measurements of traffic flows, i.e. the travel flow volume and travel time cost, change along time.

According to the above two tasks, we extend the OD-Wheel (Lu et al. 2015b) which only explores the OD patterns related to a central region, to a visual analytics system which supports exploration on OD patterns detected from general filtered trajectories. Meanwhile, the designs of the region glyph and OD-Wheel are refined. Specifically, the system integrates: (1) an intuitive visual query interface to define the region of interest and filter taxi trajectories; (2) an automatic clustering algorithm to extract OD, O/D regions from trajectories; (3) a spatial and statistical overview of regions; (4) an adapted OD-Wheel to explore the dynamic OD patterns in detail.
2 Related Work

Our work is related to trajectory filtering, OD visualization and temporal data visualization. We discuss existing works in those areas below.

2.1 Trajectory Filtering

Trajectory filtering techniques aim at extracting a subset of trajectories satisfying specific conditions. By driving a spatial geometric query object (such as a circle), filtering result is dynamic updated. Many interactive query methods have been designed. MagicLens (Fishkin and Stone 1995) is generic lens used in 2D spatial space. Lenses with different functions embed different visualizations of selected area. Krüger et al. propose TrajectoryLenses (Krüger et al. 2013) to filter trajectories. Ferreira et al. (Ferreira et al. 2013) also develop a system, but for taxi OD data, which executes OD queries and visualizes the results in multiple views. In addition to fully visual queries, filtering can also be performed via spatial visual query language. The query command can be interactively composed by predefined icons which represent either the spatial objects or operators (Aufaure-Portier and Bonhomme 1999).

Similar to TrajectoryLenses, we develop graphical filtering tool to help user select taxi trajectories. However, our system supports more complex settings on geometric constraints.

2.2 OD Visualization

OD data can either be collected from discrete movement data (Wang et al. 2014) or aggregated from continuous trajectory data (Andrienko and Andrienko 2011). Such OD data can be visualized with a set of techniques, including flow map, OD matrix and OD map. Flow map (Thompson and Lavin 1996) shows the origins and destinations as nodes on a geographic map. The directed links connecting the origin and destination nodes represent the traffic flows between the pair of OD. Flow map suffers from serious visual clutter. Possible solutions to alleviate such visual clutter include edge-filtering (Rae 2009), edge-bundling (Holten and van Wijk 2009), node hierarchy (Guo 2009) or allowing only one origin or one destination (Phan et al. 2005). OD matrix (Andrienko and Andrienko 2008) visualizes the OD data in an abstract space. It represents origins and destinations as rows and columns of a matrix. Each cell encodes an OD, whose flow magnitude is represented as color of the cell. Due to lack of spatial information, OD matrix are usually used in conjunction with a geographic map. OD map (Wood et al. 2010) makes a balance between clutter free and spatial information preservation. It partitions a geographic region to subregions with a 2D grid and defines ODs between the subregions.

In our work, we visualize OD data from two views respectively. In spatial view, we focus on the spatial overview of ODs. In abstract view, the dynamics of traffic flow volume and travel time of ODs are explored within a hybrid linear-circular design.

2.3 Temporal Data Visualization

Visualizing temporal data is very common in visualization field. Animation is one of the most familiar visualization techniques. However, it is widely believed that animation does not work well for analysis tasks (Robertson et al. 2008). Therefore, many static methods are developed.

If the data have some known periodicity, then calendar view (van Wijk and van Selow 1999) and spiral view (Weber et al. 2001) can be good options. However, for general temporal data, timeline is the most popular method. To compare different temporal data, timelines can be juxtaposed (McLachlan et al. 2008) or superimposed (Hochheiser and Shneiderman 2004) for visual comparison. For example, Guo et al. (Guo et al. 2011) embeds traffic directions into superimposed ThemeRiver (Havre et al. 2000) to compare different type vehicles over time. Wang et al. (Wang and Yuan 2014) align the trajectories to time and stack them in juxtaposition space for comparison. Temporal data visualization techniques have been used to analyze OD data. In Ferreira et al.’s taxi trip exploration work (Ferreira et al. 2013), once users select multiple ODs and execute the query, the dynamics of trip number on each OD will be shown in a timeline. Boyandin et al.’s Flowstrates (Boyandin et al. 2011) analyze the refugee flow with a specially designed three part interfaces. Two maps are used for origin and destination selection, while a heatmap is embedded between them to compare the flow volume between different ODs. Here the heatmap is in fact a juxtaposition of timelines.

In our work, we use a circular design for flow volume comparison, as that in KronoMiner (Zhao et al. 2011). In addition, we also embed a linear timeline component for travel time comparison.
3 Overview

In this section, we first introduce the related terminologies and define the problem to be solved in the work. Then the tasks are clarified further and system’s overview is given.

3.1 Terminologies and Problem

To facilitate the discussion, we discuss some common terminologies in this work. Trajectory is a list of positions recording the movement in temporal order. Origin/Destination (O/D) refers to the beginning/ending position of the movement. Origin/Destination Region is the region where lots of trajectories start or end, i.e. the cluster of Origins/Destinations. OD Region is the region where it is both an Origin Region and also a Destination Region. Region of Interest is generalized as interest region where trajectories travel through in this work. It can either be a specific region, or a compound one of several regions combined in certain logic operations.

In this work, we aim to explore the ODs of trajectories related to region of interest. Comparing to previous highly relevant work (Lu et al. 2015b)(Figure 1(b)), to-be studied trajectories are filtered by a suite of spatial filter (Section 4), which include but not be limited to those starting from/ending in a central region, e.g. the orange trajectories in Figure 1(a)). For instance, the OD patterns of trajectories passing a certain crossing is in the scope of this work but not in the previous work.

3.2 Task

Starting from the two tasks introduced in Section 1, we specify the tasks supported by our system further, as following:

- Extraction of O/D clusters related to region of interest (T1): to customize the region of interest and then extract the O/D, OD regions based on the trajectories travelling through.
- Spatial overview of O/D, OD regions (T2): to give an overview of how the regions distribute in spatial space.
- Statistical overview of O/D, OD regions (T3): to show the statistical overview of regions.
- Traffic volume comparison among regions (T4): to compare traffic volume among OD clusters.
- Detail exploration of OD pattern dynamic (T5): to explore traffic flow’s dynamic measurements in detail, including travel time cost etc.

3.3 System Overview

To support exploring OD patterns of interested region, the system composes of four major parts: data preprocessing, trajectory filtering, OD clustering and visual analysis on OD patterns. System’s pipeline is shown in Figure 2.

Following the preprocessing steps in an existing paper (Wang et al. 2013) (Wang et al. 2015), we cleaned both GPS dataset and road network dataset, and performed map matching to map the trajectories to the road network. To facilitate trajectory filtering, a spatial quad-tree index is built on trajectory data set.

In the beginning of run-time stage, trajectories can be filtered interactively from spatial and temporal aspects. With a suite of circular graphical filters, user is able to configure the region of interest by multiple filters. With region of interest settled done, trajectories satisfying certain spatial constraints are filtered. Additionally, a two-layer time filter supports to set temporal constraints in date and time scales. After filtering, the selected trajectories are fed as input to an adaptive
4 Filter

In this section, we first introduce filter model and then present visual design of spatial and temporal filters.

4.1 Filter Model

Our filter model selects trajectories from spatial and temporal aspects. The filter model is built on a series of atomic queries, which are formalized in Table 1. With these atomic queries, complex filtering can be composed.

From spatial aspect, considering the spatial relationships between trajectory and region, region and region, the atomic spatial queries can be further divided into two classes: location and direction. Six location options are provided according to six possible relationships between trajectory and region: origin, destination, origin/destination, passing, inclusive, and exclusive. Particularly, exclusive option provides the function of filtering out trajectories passing a certain area. On the other hand, with two or more filters, direction can be assigned among them to filter trajectories following certain flow direction.

From temporal aspect, a date-time temporal concept model is used. Not only regular continue time setting, it also allows users to set time constraints in daily periodicity, for example, setting a date range from March 2 to March 5 and time range from 8:00 to 12:00.

The filter model allows user to set constraints iteratively. Complex query can be made based on querying result from previous queries. For example, user can set several location queries to filter trajectories travelling through all these regions. As mentioned in Section 3.1, the region of interest can be defined by either a spatial filter or multiple filters.

4.2 Filter Design

Similar to TrajectoryLenses (Krüger et al. 2013), we choose circle as the basic filtering shape, whose points on boundary have equal distance to center point. The circular filter is embedded in spatial view so that it is intuitive to directly manipulate the filter in spatial context. There are two design considerations:

**Semantic Visibility** For the ease of parameter perception, parameters are explicitly encoded in circular filter. As shown in Figure 3(a), text below informs the radius of underlying region. The top part shows spatial filter with current location constraint (e.g. inclusion constraint here). Figure 3(b) shows filters with different location options. For the assigned direction, an arrow is linked between filters as Figure 3(c) shows.

**Usage Simplicity** Parameter tuning is integrated into the circular filter, which is made free from keyboard and menu. As Figure 3(d) shows, different functions are waked when hovering in certain regions and corresponding manipulating...
Table 1 Filter Model: spatial and temporal filtering are supported.

<table>
<thead>
<tr>
<th>Category</th>
<th>Attribute</th>
<th>Type</th>
<th>Formalization</th>
<th>Description (filtered out)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td>Location</td>
<td>Origin</td>
<td>traj.origin ∈ region_defined</td>
<td>trajectories originating from defined region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Destination</td>
<td>traj.destination ∈ region_defined</td>
<td>trajectories destination for defined region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Origin/Destination</td>
<td>traj.origin ∈ region_defined ∪ traj.destination ∈ region_defined</td>
<td>trajectories originating from or destination for defined region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Passing</td>
<td>traj.location ∈ region_defined ∩ traj.origin ∩ traj.destination ∈ region_defined</td>
<td>trajectories passing through defined region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inclusion</td>
<td>traj.location ⊈ region_defined</td>
<td>trajectories intersecting with defined region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exclusion</td>
<td>traj.location ⊈ region_defined</td>
<td>trajectories not intersecting with defined region</td>
</tr>
<tr>
<td>Temporal</td>
<td>Date</td>
<td>Date Range</td>
<td>traj.date ≥ begin_date ∩ traj.date ≤ end_date</td>
<td>trajectories travelling in defined date range</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>Time Range</td>
<td>traj.time ≥ begin_time ∩ traj.time ≤ end_time</td>
<td>trajectories travelling in defined time range</td>
</tr>
</tbody>
</table>

5 Overview of OD Clusters

In this section, we first introduce the extraction of regions from filtered trajectories and then present the visual design of OD overview and detail pattern exploration.

5.1 O/D, OD Region Extraction

After filtering trajectories of interested region, we adopt the iterative DBSCAN algorithm Pan et al. (2013) on origins and destinations respectively to obtain O clusters and D clusters (T1). The convex hull underlying each cluster is detected as O or D region by Graham’s scan (Graham 1972). Regions within a certain distance (Threshold D) are merged into a larger region. Specifically, if the merged region contains both O and D regions, it is an OD region.
5.2 Overview

Overview of regions consists of spatial overview (T2) and statistical overview (T3). As Figure 4 shows, two types of spatial overview are provided. Figure 4(a) shows the original clusters’ convex hulls, which gives precise spatial description of regions. However, constrained by absolute spatial shape, it is not convenient to encode more information upon it. Hence, as Figure 4(b) shows, spatial view with glyphs is designed as a trade-off between abstract and spatial space.

![Fig. 4 Spatial Overview: (a) regions are visualized in convex hull. (b) glyphs are used to represent the detected regions. (c) glyph design for O/D, OD Regions](image)

As Figure 4(c) illustrates, each region is represented by a square glyph. The region’s traffic volume is mapped to the size of square. The square’s frame color is used to distinguish whether the region is O or D or OD region. Taking O regions as examples, a sequential blue list is assigned to them from dark to light in descending traffic volume order. So is the red list for D regions and green list for OD regions. To make it more intuitive, an arrow in the right-bottom explicitly encodes the direction of traffic flow, such as an going-out direction in O region glyph. For each region, to emphasize on the traffic flow related to other detected regions, the volumes of related traffic flows are visualized in color dithering manner inside the glyph. That is, the ratio of certain color inside one glyph is proportional to the traffic volume related to corresponding region. Additionally, the average traffic volume distribution in a day can be unfolded by clicking the dot handle in the right-top. From left to right, the average traffic volume distribution is displayed in hour granularity from 00:00 to 24:00, smaller value with larger lightness. A gray vertical line is drawn every four hours to help with hour reading. When folded, the number of trajectories is labelled aside to the dot handle. Particularly for OD glyph, there are two dots, one for the coming-in and other for the going-out flow.

Additionally, the statistical overviews of the traffic volume distribution over time and the distributions of average travel distance and travel time are provided. Aware of extreme values, As Figure 5 shows, we wrap the extreme values and count them as a whole in the two ends of bar plot.

![Fig. 5 Wrapped Statistical Distributions](image)

6 Adapted OD-Wheel

OD-Wheel (Lu et al. 2015b) is adapted to explore the OD clusters detected from interested trajectories, which is not restricted to OD clusters related to a central region. Given the O/D, OD regions, the adapted OD-Wheel is designed to explore dynamic change of clusters (T4) and compare OD patterns among clusters (T5).
6.1 Visual Design

The main idea of OD-Wheel is to warp a part of linear view to circular one. O, D clusters are placed in descending traffic volume order from two ends of circle respectively. Particularly in this work, OD region is decomposed into a pair of O and D clusters, which counts the in and out traffic flow to the region respectively. A link is built between the pair, whose color is consistent with the OD glyph in spatial overview. Additionally, an arrow towards/outwards the circular center is used to distinguish O and D clusters further. And its size is proportional to the cluster’s traffic volume. In much more detail, the dynamic traffic volume of each cluster is visualized as bar plot along radial time axis. Sharing the same axis with circular view, undistorted linear layout benefits more precise dynamic analysis. For example, as Figure 6 shows, analysis on travel time cost or dynamic traffic volume can be performed in the linear view. However, it is easy to extend to other dynamic variation analysis.

![Fig. 6 Visual Design of Adapted OD-Wheel: linear and circular view displays the temporal distribution of travel flow volumes. Two possible dynamic linear visualizations are in the linear view.](image)

6.2 Interactions on OD-Wheel

Several interactions are developed for easy comparison among clusters. A cluster can be relocated by dragging-and-moving the gray annulus around the circle. Clicking black button (><) between pair of neighbouring clusters stacks the two clusters to a common middle line, as the right bottom part in Figure 6 shows. To facilitate the comparison, differences between the two clusters along time are explicitly encoded by black bars. By clicking the black button (<> outside the circle, the stacked bars can be unstacked.

When mouse hovering on bars, its time interval pops out as hint. Circular view and linear view cooperate in brush-and-link manner. Hovering on the arrow glyph in circular view highlights corresponding cluster in the linear view. Similarly, lasso selection on trajectories in linear view highlights corresponding ones in circular view and spatial view.

7 CASE STUDY

We apply the system to real taxi trajectories to show its effectiveness. Our GPS dataset is a real taxi dataset recorded in the city of Beijing in 24 days, from March 2nd to 25th, 2009. Estimated from a government report Beijing Transportation Research Center (2010), they include 43% of all licensed taxis in Beijing, and account for 7% of the traffic flow volume within Beijing’s 4th Ring. We only use trajectories carrying passengers. After preprocessing, the final dataset is 12.8 GB.
7.1 OD Spatial Overview of Interested Region

In this case, we demonstrate an OD spatial overview for an interested road. As Figure 7 shows, the interested road is defined by two filters on the 2nd ring road. The O/D, OD regions of trajectories travelling through this road are detected and visualized in Figure 7(a) and Figure 7(b). (Because regions in Figure 7(b) are far away from Figure 7(a) on the map, so we cut them out and place aside to save space.) Most of the regions are along the 2nd ring. Two OD regions (Figure 7(b)) locate at the two terminals of Beijing Airport. Unfolding the average distributions of regions marked by 1, 2 as Figure 7(c) shows, we found that there is an early morning peak (6:00-9:00) and evening peak (18:00-21:00) at 1 and 2 respectively. Unfolding coming-in and going-out traffic flows at airport, the two terminals have similar OD patterns over day. For the coming-in traffic flow, it becomes active in early morning and keeps active during the day time. On the contrary, the out-going flow becomes active from the noon until evening.

![Figure 7](image_url)

**Fig. 7 OD Spatial Distribution of a Road:** (a) the detected O/D, OD regions in down-town. (b) the two detected OD regions at Beijing Airport. (c) unfolded distribution panel of 1 and 2 regions in (a) and 3, 4 regions in (b).

7.2 Detail Exploration on Volume Dynamic Change

Besides the OD overview, our system supports to explore the dynamic OD patterns in detail by the adapted OD-Wheel. In this case, the interested region is set at the west-north corner on the 4th ring road. 5124 trajectories passing this region are filtered from March 9 to March 11. The spatial distribution of O/D, OD regions are as Figure 8(a) shows. The dark green OD region has the largest traffic volume. For all the other regions, there are dense dark green inside their cluster glyphs. That is, there are high ratio of traffic flows between those regions and the largest OD region. It is particularly obvious for the second largest OD region, which is east-north to the largest one in a short distance. Unfolding the in and out traffic volume distribution of the largest OD region (Figure 8(a)), the average distributions over day can be observed respectively. Examining them in OD-Wheel for detail (Figure 8(b)), the pair of in and out traffic volume has definitely different distribution shapes in the linear view. To compare them day by day clearly, the two are stacked. The black difference bars clearly show that there is larger out rather than in traffic flow at beginning of a day and then larger in traffic flow in the afternoon and evening.

7.3 OD Pattern Dynamics Exploration

Alternative to traffic volume, there can be other traffic measurement in the linear view. In this case, we take the travel time cost as an example of OD dynamic pattern, while our method can be easily extended to other measurements, such as travel distance. We set the interested region at the north-west corner on the 2nd ring road, which is one of the biggest transportation hub in Beijing. 5762 trajectories are filtered from March 9 to March 10. As Figure 9(a) shows, the largest OD region is around the transportation hub, which conveys over 1000 taxis in and out every day. Others regions are significantly smaller than the largest one. Comparing the traffic volume of the largest OD region in the OD-Wheel, we found that the traffic flow of in-direction is larger than that of the out-direction in the morning and inversely in the afternoon. The travel time cost of the two directions are visualizes as dot plots in the linear view. Averagely, most of
the travel time cost from or to this region is within 30 min. However, in longer travel time range, the in-direction spans wider than the one of out-direction. The link-and-brush function between linear view and map view supports users to check trajectories in detail. As Figure 9(b) shows, trajectories with small time cost and with large time cost are selected in the linear view and plotted on the map. Those with small time cost (within 10 min) travel in very short distance and those with large time cost (more than 40 min) reach the 5th ring of Beijing.

Comparing with the high relevant work (Lu et al. 2015b), we refine the OD region glyph (Section 5.2) after collecting feedbacks from several users in visualization field. One of the users commented on the convex hull glyph in work (Lu et al. 2015b) “the number label circle interferes the perception of detected OD region”. After we changed it into text square, the user said it is less confusing. Another is the visual encoding of direction of a region’s traffic flow. Previously in work (Lu et al. 2015b), warm and cold color series are used to encode the O and D region respectively. However, users complained that it is not intuitive. It is more intuitive using arrow whose orientation represents the direction. Additionally, besides to the spatial information, users gave positive feedbacks on the average temporal distribution, which helps them quickly target at interesting temporal patterns.
9 Conclusion

In this paper, we present a visual analytics system to explore the OD patterns of interested region. Instead of global OD pattern analysis, we tackle this problem starting from trajectory filtering. With graphical filters, the system supports to set spatial constraints and filter related trajectories. Based on these trajectories, OD clusters are extracted and corresponding regions are detected. To analyse the OD patterns, we adapt a visual design, OD-Wheel, to support the comparison among regions, which are not limited to a central region as before. With effective interactions and intuitive visual hints in comparison, the adapted OD-Wheel is capable of comparing dynamic patterns in detail. Finally, applying the method to real taxi GPS trajectories, the system’s effectiveness is demonstrated by several cases.

There are some possible research directions in future. OD-Wheel can be generalized as a visual tool to explore large scale time series data. More interactions needs to be added on OD-Wheel. For example, a magnifying function could be added for time series with small value. The other is to consider the traffic interchange flows among multiple regions. OD flow or matrix may be integrated into OD-Wheel so that it is possible to explore the traffic flows between pair of two certain regions.

10 Acknowledgement

The authors wish to thank the anonymous reviewers for their valuable comments. This work is supported by NSFC No. 61170204. This work is also partially supported by NSFC Key Project No. 61232012 and the National Program on Key Basic Research Project (973 Program) No. 2015CB352500. This work is also funded by PKU-Qihu Joint Data Visual Analytics Research Center.

References

Beijing Transportation Research Center (2010). Annual report of beijing transportation development.


