TrajRank: Exploring Travel Behaviour on a Route by Trajectory Ranking

Min Lu Zuchao Wang * Xiaoru Yuan †

Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
Center for Computational Science and Engineering, Peking University

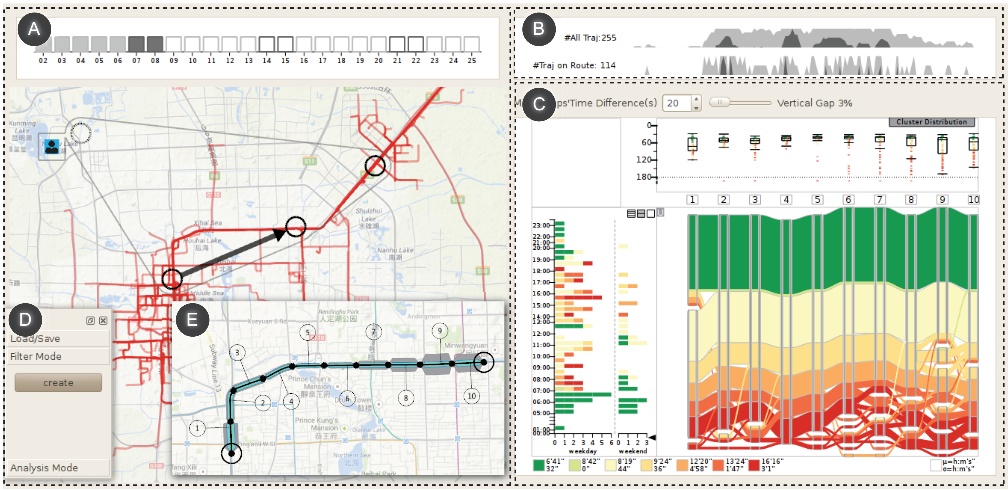


Figure 1: Interface of *TrajRank*: A. spatial-temporal view: supports trajectory filtering and route segmentation; B. horizon graph view: shows the distribution of trajectories over a day; C. ranking view: provides visualization and exploration of trajectories by ranking; D. menu panel; E. part of spatial temporal view, drawing bands over road segments encoding the average (the blue ones) and variance (the gray ones) of travel time.

Abstract

In this paper, we propose a novel visual analysis method *TrajRank* to study the travel behaviour of vehicles along one route. We focus on the spatial-temporal distribution of travel time, i.e., the time spent on each road segment and the travel time variation in rush/non-rush hours. *TrajRank* first allows users to interactively select a route, and segment it into several road segments. Then trajectories passing this route are automatically extracted. These trajectories are ranked on each road segment according to travel time and further clustered according to the rankings on all road segments. Based on the above ranking analysis, we provide a temporal distribution view showing the temporal distribution of travel time and a ranking diagram view showing the spatial variation of travel time. With real taxi GPS data, we present three use cases and an informal user study to show the effectiveness and usability of our method.

Keywords: travel behaviour, ranking visualization, filtering

Index Terms: H.5 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Graphical user interfaces;

1 Introduction

Understanding how vehicles move along a route is a basic task in traffic analysis. It can help road users better schedule their trips and avoid congestions. It can also help traffic administrators assess the traffic condition and identify traffic bottlenecks. Although real time traffic monitoring and prediction are already mature, and available in most online maps, e.g. Google Map, most of them only provide aggregated traffic condition of a road but do not reveal the micro-behaviours of vehicles. Vehicles' travel behaviour analysis is also important, which is complementary to the traffic monitoring and prediction.

In this work, we focus on exploring vehicles' travel behaviour on a route. We specifically study how the travel time change on different road segment and at different occurrence time. Our study of travel behaviour consists of the following aspects:

- **Overview of travel behaviour**, e.g. what is the mainstream and outlier behaviour?
- **Relationship between travel behaviour and road segments**, e.g. which road segment affects travel behaviour most?
- **Relationship between travel behaviour and trip's occurrence time**, e.g. how does the vehicle behave when vehicle starts at different time?

Taxis are viewed as sensors of the city traffic situation. Therefore we treat them as reasonable sampling of real traffic flow and use them to study the vehicles' travel behaviour.

Several critical issues should be considered in this study. The first issue is to extract trajectories moving on a certain route from massive taxi GPS dataset. A user interface has to be defined to support easy interactive filtering. The second issue is to summarise the travel behaviour of these trajectories. Since analysing trajectory individually is not scalable, summary is important. Still the summary should be intuitive and easy to understand. The third issue is to correlate travel behaviour with occurrence time and road segment. We especially need to consider the temporal granularity and spatial granularity (road segmentation). They are not fixed and are subject to change according to different analysis demands.

*e-mail: {lumin.vis, zuchao.wang}@gmail.com
†e-mail:xiaoru.yuan@pku.edu.cn

Considering those issues, we develop *TrajRank*, a visual analysis method based on trajectory ranking. We first provide a suite of interactions to support trajectory selecting and road segmentation. Then we summarise the travel behaviour of these trajectories based on their travel time rankings along the route. This is achieved with a ranking diagram design, which merges trajectories with similar rankings on all road segments as bands. Each band corresponds to one kind of travel behaviour. To correlate travel behaviour with occurrence time and road segment, we further provide a temporal view, a spatial view and a box-plot for multi-perspective exploration.

The major contributions of this work are as follows:

- We propose a ranking based visual analysis method to study taxi travel behaviour on a route.

- We develop an interactive system to support ranking based explorations of taxi travel behaviour, and we evaluate it with use cases and a user study.

2 Related Work

In this section, we review related work in four categories: taxi data analysis, travel time visualization, interactive trajectory filtering and ranking data visualization.

2.1 Taxi Data Analysis

Many kinds of problems can be studied with taxi data, including land-use classification [21], route recommendation [31], outlier detection [33] and traffic prediction [5]. In this work, we use taxi data to study travel time pattern and its influence factors. It is mainly related to traffic outlier analysis and traffic prediction.

For traffic outlier studies, Zheng et al. [33] detect flawed traffic connections between different urban regions. Then they summarize the flawed connections based on mining frequent patterns and association rules. Wang et al. [28] detect traffic congestions on each road, and structure them with congestion propagation graph. In this paper, we not only study the travel time outliers along a route, but also the mainstream patterns.

For traffic prediction studies, Castro et al. [6] have built a city scale traffic flow model for traffic prediction, where the capacity of each road segment is estimated from historical taxi data. Huang et al. [14] focus on predicting travel destination. To cope with ”data sparse problem” in prediction, they synthesize new trajectories from decomposed sub-trajectories. In comparison, our focus is to summarise historical traffic pattern, which can potentially help to make more accurate traffic predictions.

2.2 Travel Time Visualization

Travel time is a crucial index to measure urban transportation quality. It has been widely studied in visualization community. An overview of the travel time within a whole city can be provided by the time-distance transformation [8] technique. Such a technique chooses a central location and distorts the map so that the distance of any location to the central location is proportional to the travel time. However, this technique is limited to one snapshot or the average condition of the urban traffic. It is not able to show its dynamics. Wu et al. [30] instead propose *BoundarySeer* to visualize the dynamics of the reachable boundary, e.g. the regions accessible from a central region in 30 minutes. With multiple linked views, they are able to analyse the spatial temporal patterns of boundary change.

Our work does not try to analyse the travel time within a whole city. Instead, we focus on a specific route. Route level travel time analysis have been addressed by a few existing works. For example, Zeng et al. [32] have built a travel time model on public transportation system. They are able to estimate the travel time on each route, and show its temporal variation and spatial decomposition of travel time. Liu et al. [19] have studied the route diversity problem. Given routes, they show the temporal variation of trajectory number and travel speed. Tominski et al.'s trajectory wall [25] is based on stacking different trajectories in 3D space, which is able to show the spatial variation of movement attributes, including travel time. The fundamental difference between our work and these existing works is that we use a ranking based methodology. We study the travel time ranking change on each road segment.

2.3 Interactive Trajectory Filtering

As our taxi dataset is in city scale, and each time we only focus on a specific route, we need to filter the trajectories on this route. Trajectory filtering has been widely studied, and different kinds of filters has been summarized in Andrienko et al.'s book [2, Chapter 4.2]. The filtering tools can be implemented in different ways. One way is based on visual query languages [7], where users specify query composition by direct manipulation on icons and menus [1]. In this work we implement fully interactive filters similar to TrajectoryLenses [18]. It allows users to extract trajectories from a common origin and a common destination. The origin and destination are defined by interactive lenses. Ferreira et al.'s system [20] for New York taxi exploration also has a similar design to filter trajectories by their origins and destinations.

2.4 Ranking Data Visualization

Ranking is a common analysis method. It is frequently used in visualization [23, 22]. In many cases, analysts can get multiple rankings with different criteria. Kidwell et al. [16] propose to generate an overview of these rankings with MDS projection [29] and heatmap techniques. In order to compare these rankings, Behrisch et al. [4] present a small multiple view of circular glyphs. Each glyph supports the comparison of a pair of rankings. Gratzl et al. [10] consider to rank items with multiple attributes, assuming that the ranking criteria is a linear function of those attributes. Their LineUp design not only shows the difference between different rankings, but also indicates the cause of such difference.

Ranking can change with time, and analysis of such dynamic change is crucial. Batty [3] has designed Rank Clocks to show the change of city population rankings across several centuries. Their design is similar to a parallel coordinates [15]. When large number of items are ranked, the above visualization can be very cluttered. Therefore, Shi et al. propose RankExplorer [24], in which they segment the rankings into several groups and use a ThemeRiver [13] to show their temporal changes. Specially designed glyphs are embedded in the ThemeRiver to show the number of items that jump between different ranking groups. In order to avoid cluttering, we also use a grouping strategy like RankExplorer [24]. However, we not only group travel times on each road segment, but also group trajectories based on the travel times on all road segments.

3 System and Data

In this section, we provide an overview of our system and discuss the data used in the experiment.

3.1 System Overview

TrajRank provides an interactive visual analytic method for taxi travel behaviour exploration on a route. The major idea is ranking. Figure 2 shows its workflow. *TrajRank* takes taxi GPS data and road network data as input. In the offline pre-processing stage, GPS trajectories are cleaned and matched to road network. A quad-tree trajectory index is built on map to facilitate filtering. In the runtime visual analysis stage, trajectories can be filtered by a suite of spatial-temporal filters and *TrajRank* detects the most passed route. Then this route is divided into road segments. For each trajectory, travel time on each road segment is computed. On each road segment, the travel time of different trajectories are clustered into groups,

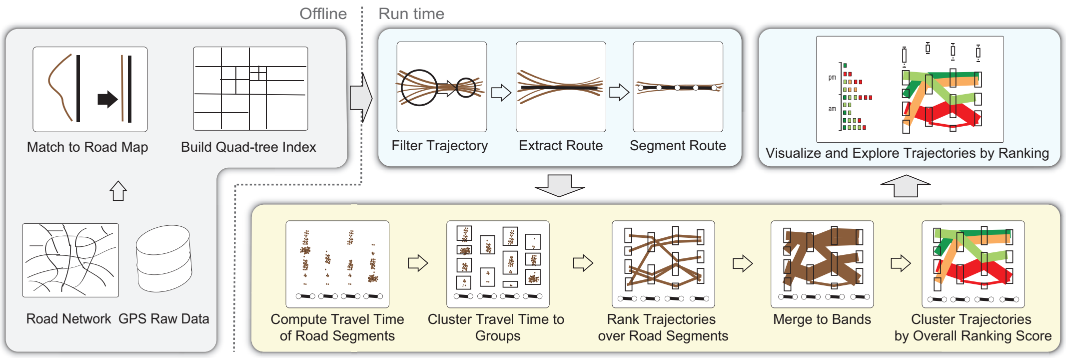


Figure 2: The Workflow of *TrajRank* consists of an offline preprocessing stage and a runtime visual analysis stage.

and these groups are ranked by average travel time in ascending order. A ranking score will be calculated for each trajectory, based on which the trajectories are clustered into groups. Each trajectory group represents a kind of travel behaviour. The ranking is visualized and supports interactions for further exploration.

The above workflow is supported by a carefully designed user interface. As shown in Figure 1, the interface consists of four views: a spatial-temporal view, a horizon graph view, a ranking view and a menu panel. In the spatial-temporal view, users interactively define spatial-temporal filters and configure of route segmentation. The horizon graph view displays temporal distribution of selected trajectories over a day. The ranking view supports trajectory ranking analysis. It consists of three components: a ranking diagram, an occurrence temporal distribution view (temporal distribution view for short) and a modified box-plot. The ranking diagram visualizes trajectories ranking over road segments. The temporal distribution view displays the distribution of trajectory groups with respect to occurrence time. The modified box-plot gives a statistical description of travel time on each road segment. Such statistics is also shown in the spatial view and encoded by the width of each road segment band.

3.2 Data

Our GPS dataset is a real taxi dataset recorded in the city of Beijing. In 24 days, from March 2nd to 25th, 2009, the GPS trajectories of 28,519 taxis are collected, consisting of 379,107,927 sampling points. The data size is 34.5 GB. Each sampling point contains record of $time$, $latitude$, $longitude$, $speedmagnitude$, $direction$, plus a boolean $CarryPassengerState$, indicating whether there are passengers in the taxi. The sampling rate is one point per 30 seconds. To perform our study, we also use Beijing's road network dataset, available from OpenStreetMap's jXAPI [9].

Following the preprocessing steps in an existing paper [28], we clean both GPS dataset and road network dataset, and perform map matching to map the trajectories to the road network. We only use trajectories with passengers inside. This gives us 1.8 GB GPS data. According to the map matching result, we relocate all matched sampling points on roads. As the sampling rate is low, we also insert many sampling points in each trajectory to make it follow the roads. The final data size is 12.1 GB.

4 Visual Design

In this section, we first explain the idea of ranking. Then we present design rationales and visual encodings in the ranking view.

4.1 Ranking

Given a set of taxi trajectories $\mathbf{X} = \{x_1, x_2, ..., x_m\}$ passing a selected route, we model each trajectory as a vector $A = \{a_1, a_2, ..., a_n\}$, where $\mathbf{a_i}$ is the travel time on the ith road segment. Although abstraction methods such as MDS [29] can potentially summarize such high dimensional data as a few clusters, it is not intuitive to interpret the meaning of these clusters. We consider to use ranking, which structures collections of items based on the value of their attributes.

In TrajRank, we first rank trajectories on each road segment. As can be seen later, directly visualizing the rankings of all individual trajectories would cause heavy visual clutter. So we do some clustering. For each road segment, we cluster travel time into groups $\mathbf{G}$ and rank these groups. We use a hierarchical clustering algorithm [12, Chapter 14.3.12]. Such an algorithm requires a distance threshold D_{min}, as the minimum average travel time gap between groups. It can be understood as *"significant travel time difference"*. The selection of D_{min} value will be discussed later.

Travel time on different road segments may be in different scales, thus not comparable. Therefore within each road segment, we calculate an outlier index c_i for each group $g_i \in \mathbf{G}$. The outlier index is essentially a normalization of the travel time, defined as follows:

$$C(g_i) = \begin{cases} \left\lceil \frac{avg(G_{mainstream}) - avg(g_i)}{avg(G_{mainstream}) - avg(G_{min})} * N_{level} \right\rceil, avg(G_{mainstream}) > avg(g_i) \\ 0, avg(G_{mainstream}) = avg(g_i) \\ \left\lfloor \frac{avg(G_{mainstream}) - avg(g_i)}{avg(G_{max}) - avg(G_{mainstream})} * N_{level} \right\rfloor, avg(G_{mainstream}) < avg(g_i) \end{cases} \quad (1)$$

where avg is a function to calculate average travel time, $G_{mainstream}$ is the group with the most trajectories, G_{max} and G_{min} are groups with the maximum and minimum average travel time. N_{level} determines the number of outlier levels. In our system, we choose $N_{level} = 5$.

Based on this definition, the outlier index for a group g_i can be explained as follows:

$$C(g_i) = \begin{cases} (0, N_{level}], g_i \text{ is a positive outlier} \\ 0, \ g_i \text{ is the mainstream} \\ [-N_{level}, 0), g_i \text{ is a negative outlier} \end{cases} \quad (2)$$

where positive outlier indicates less travel time than mainstream, and negative outlier indicates more time.

We further propose a score S to evaluate the overall ranking of trajectory x_i. It is defined as following:

$$S(x_i) = \sum_r w_r * c_r \quad (3)$$

where r is the index of road segment, c_r is the outlier index of the travel time group of trajectory x_i on road segment r. w_r serves as a weighting factor, equal to interquartile range(IQR) [26]. IQR is a common measure of statistical dispersion. It is defined as $IQR = Q3 - Q1$, where $Q1$ is the 1st Quartile, and $Q3$ is the 3rd Quartile. If the dataset has a low IQR, its distribution is concentrative, otherwise dispersed. We use IQR to measure the abnormality of outliers happening on a road segment. Outliers that happen on road segments with concentrative travel time distribution seem more accidental than those happening on road segment with dispersed distribution. By using *IQR* as the weighting factor, we emphasize on outliers caused by regular factors such as morning peak instead of irregular factors such as a sudden traffic accident, so as to bring out the ordinary travelling behaviour on a route.

With the score S, trajectories are clustered into trajectory groups. Each trajectory group corresponds to one kind of travel behaviour.

4.2 Ranking Diagram

Ranking diagram visualizes trajectories' rankings over road segments. For an easy comparison over different segments, we adopt a line-based visual, inspired by parallel coordinates [15], parallel sets [17] and LineUp [10]. Specifically, road segments are represented as axes, trajectories by polylines with vertices on axes. In our design, we have the following considerations:

- **C I: Keep continuity of trajectories**: essential for correctly tracing it over road segments.
- **C II: Reduce visual clutter**: should be scalable to hundreds of trajectories.
- **C III: Display ranking change over road segments**: reveal the relation between travel behaviour and road segments.
- **C IV: Overview of travel behaviour**: show mainstream and outlier behaviour.
- **C V: Support multilevel analysis**: handle different analysis granularities in spatial aspect.
- **C VI: Keep small features visible**: so as not to miss outliers.

Following these considerations, as shown in Figure 3, we propose ranking diagram. In this design, road segments are represented as vertical axes. From left to right, it follows the spatial order along the route, i.e., the leftmost axis is the first road segment. On each axis, travel time groups are represented by rectangles with gray frame. The groups above have shorter average travel time. Trajectories are drawn as continuous curves across corresponding rectangles on axes(**C I**).

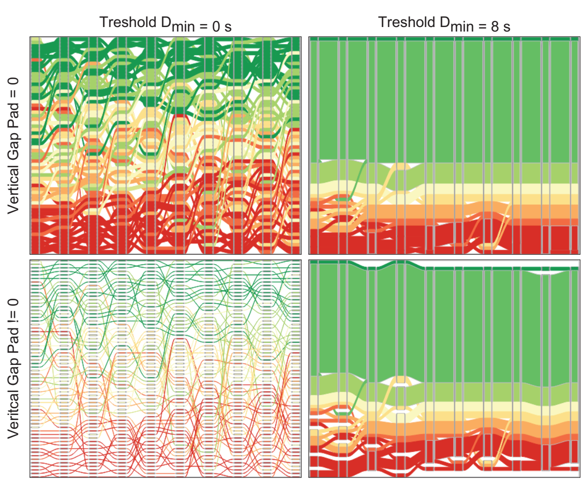


Figure 3: Ranking Diagrams: different settings on threshold D_{min} and vertical gap.

As trajectories are clustered into groups by their overall ranking score S, we merge trajectories in the same group into one band. This reduces unnecessary jumping within a travel time group, thus reducing visual clutter (**C II**). To emphasize ranking change (**C III**), bands are rendered in a braided manner. That is, bands dropping in ranking are drawn behind the rising ones.

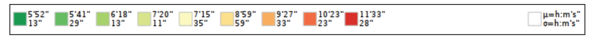


Figure 4: Color Legend: the colors for trajectory groups are from green to red, corresponding to highest score to lowest score.

Different trajectory groups imply different kinds of travel behaviour. To give an overview of travel behaviour (**C IV**), trajectory groups are sorted in descending average ranking score order and a diverging color scheme (RdYlGn) from ColorBrewer [11] is assigned. Figure 4 shows a color legend, where green color encodes high average ranking score and red encodes low score. Additionally, for each trajectory group, its average travel time (μ) is labeled on the upper right, and standard deviation (σ) on the lower right, both in the format of $h : m's''$. For ease of reading, an instruction glyph is drawn on the rightmost position.

As discussed in Section 4.1, the threshold D_{min} in travel time clustering implies different levels of clustering within a road segment. Thus by choosing different D_{min}, multilevel analysis can be supported(**C V**). Figure 3 shows visualizations with different choice of D_{min}. Diagrams on left show individual trajectories and diagrams on right show trajectories merged into bands.

To keep the visibility of small travel time groups (**C VI**), we allow users to adjust the vertical gaps between rectangles. Broadening the gap makes the groups more visually distinguishable. In the upper row of Figure 3, there's no vertical gap. The groups are compactly placed, so the change of vertical position of band indicates the ranking change. However, it is hard to differentiate different groups. In the lower row, there's a vertical gap. Although bands in this form become wiggly, groups are more distinguishable. It is easier to notice small groups.

4.3 Occurrence Temporal Distribution View

Complementary to the ranking diagram, the occurrence temporal distribution view visualizes the travel behaviour with respect to trips' occurrence time. We adopt a histogram based design similar to Liu et al.'s work [19]. In this way, we can show the temporal distribution of different ranking groups. Although other methods such as pixel table [28] and calendar view [27] can show more details, they are restricted to show the trajectory number in only one group or the total. In our design, we have the following considerations:

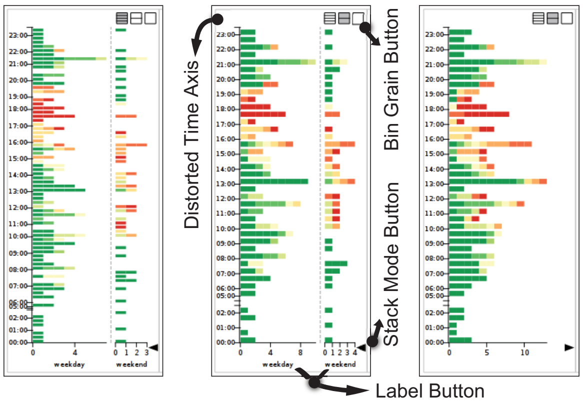
- **C I: Deal with uneven temporal distribution**: as few taxis travel at night, a linear timeline would waste display space at nighttime.
- **C II: Discretize the time understandably**: discretization is necessary for statistics, and it should follow certain convention.
- **C III: Support exploration of periodicity**: to reveal of periodic pattern of human activity.

Figure 5 shows temporal distribution views with different settings. The vertical axis is occurrence time and horizontal axis is the number of trajectories. Trajectories are drawn as rectangles. The color coding is consistent with the ranking diagram.

To deal with the uneven distribution of trips over a day (**C I**), time axis is distorted to reserve vertical space.

To support multi-level observation, statistics can be performed per 15 minutes, per half an hour and per hour (**C II**). Figure 5(a) is at a finer temporal granularity than Figure 5(b). Users can switch the temporal granularity by the three buttons on the right-top.

To explore the periodicity of human activity, we distinguish trajectories at weekdays and weekends (**C III**) by default (Figure 5(a)(b)). Users can also merge them (Figure 5(c)). Additionally, the weekday and weekend labels also act as temporal filters to show weekdays' or weekends' trajectories only.



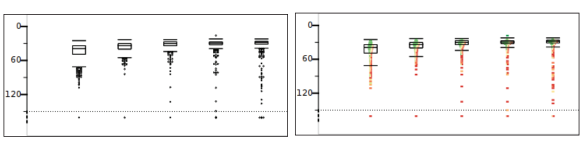
(a) (b) (c)

Figure 5: Occurrence Temporal Distribution View: (a) fine granularity, separating weekday/weekend; (b) coarse granularity, separating weekday/weekend; (c) coarse granularity, weekday and weekend data merged.

4.4 Modified Box-plot View

We use box-plots to give a statistics summary of travel time on each road segment. They are added on top of the ranking diagram, as shown in Figure 6(a). The height of box is by convention *IQR*, which is also the weighting factor *w*. To utilize space better, we make some modifications on the classical box-plot. First, we jitter the outliers to alleviate point overlap. Besides, we put all extreme outliers below the dash line to save vertical space. To correlate it with the ranking diagram, a distribution histogram can be covered over the box-plot, whose color is consistent with the ranking diagram, as Figure 6(b) shows.

Additionally, as Figure 1E shows, we also encode the same information on the map. The median travel time corresponds to the width of inner blue band, and *IQR* corresponds to the width of outer gray band. We think median and IQR are incomparable, so we encode them independently.



(a) (b)

Figure 6: Modified Box-plot View: (a) box-plot only; (b) distribution histogram covered.

5 Interaction

In this section, we present the interactions in our system, including trajectory filtering, route segmentation, brushing and linking.

5.1 Trajectory Filtering

We support interactive trajectory filtering in the spatial view. Similar to TrajectoryLenses [18], users can dynamically extract trajectories by dragging circular shaped lenses. As Figure 7 shows, mouse hovering on different regions of the lens evoke different functions. For example, traffic flow direction can be defined by dragging a line between two lenses.

Temporally, trajectories can be filtered by days, as shown on the top of Figure 1A. Each square represents a day. The ones with black frames are weekends.

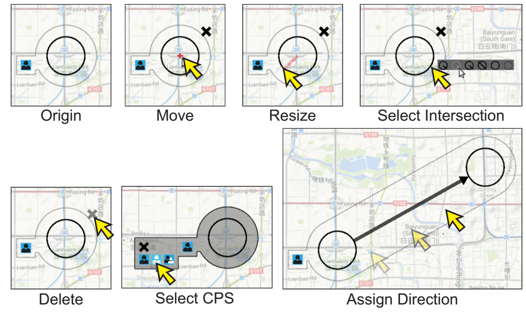


Figure 7: Trajectory Filtering Interactions. CPS is short for Carry Passenger Status.

5.2 Route Segmentation

When trajectories are filtered, the most frequently passed route are selected by default, and split into ten road segments. Route segmentation can be fine tuned by users, as Figure 8 shows. For example, dragging the cursor upwards decreases segmentation granularity and downwards increases it. Two road segments can be merged by deleting the node between them. On the contrary, one road segment can be split by adding a new node. Moreover, the route's endpoints and the nodes can also be adjusted by dragging.

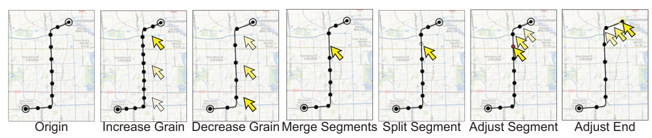


Figure 8: Route Segmentation Interactions.

5.3 Brushing and Linking

TrajRank composed of multiple linked views. Users can make selections in different views. For example, in the ranking view, users can select trajectories by clicking on a trajectory band or a cell in color legend. In occurrence temporal distribution view, users can set a time range by dragging on the rectangles, or pressing the weekends and weekday buttons. These selections will be updated in other views. Besides, in the spatial view, users can play an animation to recover the movement of selected taxi trajectories, with start time aligned together. This helps them verify their findings.

6 Experiment Results

In order to test the effectiveness of our method, we have implemented a prototype system. The system is written in C++, with Qt framework. The rendering is performed with both OpenGL and Qt Graphics View. We run the system on an Intel(R) Core(TM)2 2.66 GHz Laptop with 4 GB RAM and a NVIDIA Geforce GTX 470 GPU. Based on this implementation, we have performed quite a lot of exploratory studies on the Beijing taxi dataset. In the following part of this section, we report three use cases.

6.1 Case 1: Overview of Travel Behaviour over Space and Time

Our system provides an overview of taxi driving behaviour on each route. As shown in Figure 9(a), we have selected 6 typical routes in Beijing and visualized the spatial variation of their travel time with bands in map view. Route *A* starts from north 4th ring and ends at the west 2nd ring. The travel time variance is significantly larger on the west 2nd ring. Route *B* and *E* start from highways and end at the 3rd and 4th ring respectively. Similar to route *A*, travel time variance increases once taxis get on the main ring road, meaning that there is less change of the travel time when traveling on highway than traveling on urban road. Besides, it usually takes longer at the turning corners than other road segments. Route *C* and *F* are in downtown. Their travel time are quite even distributed along the routes. Some small variance of travel time may be shaped by the neighbouring commercial regions. Starting from Beijing International Airport, route *D* follows the Airport highway. The 5th segment stands out with much larger variance in travel time. Perhaps this is due to a toll station there.

The temporal distribution of travel time is shown in Figure 9(b). For route *B* and *E*, trajectories with low scores (red color) appear mainly in the morning. In contrast, for route *A*, *C* and *F*, trajectories with low scores appear mainly in the afternoon. Especially for route *F*, on weekdays taxi travels with low ranking behaviour around 18:00. Meanwhile, the taxi flow volume decreases on route *F*. Route *D* has very few taxis with low ranking scores, and these taxis appear all over a day. Perhaps this can be explained by their regular stoppage at the toll station.

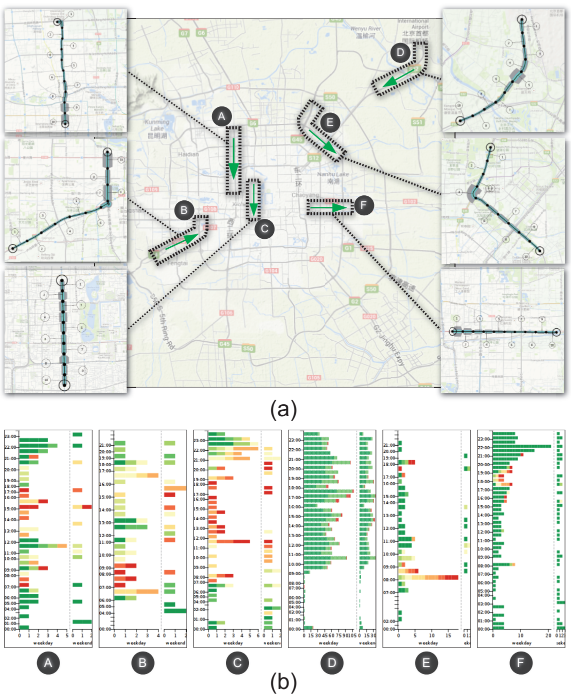


Figure 9: Travel Behaviour over Space and Time: (a) 6 typical routes in Beijing are selected and their corresponding bands are shown in the spatial view; (b) the temporal distribution views of these 6 routes.

6.2 Case 2: Understanding Travel Behaviour with Even Route Segmentation

With TrajRank, we enable users to explore the details of travel time ranking over road segments. In Figure 10(a), a westward route on the north 4th ring is selected, and 857 trajectories following the route are extracted. The horizon graph (Figure 10(b)) shows there is no obvious peak in the taxi flow volume. With $D_{min} = 30$ s, trajectories are clustered into 8 groups. As the color legend in Figure 10(c) shows, the group's ranking decreases from green to red and the average travel time increases from around 7 minutes to 18 minutes. In the temporal distribution view, clusters with low rankings are in the morning of weekdays, and in the late afternoon of both weekdays and weekends.

To study different clusters in detail, we highlight each group respectively in Figure 10(d). The 1st cluster has high ranks over all road segments. The 2nd cluster ranks low at the 4th road segment. The 3rd cluster contains few trajectories and unclear ranking

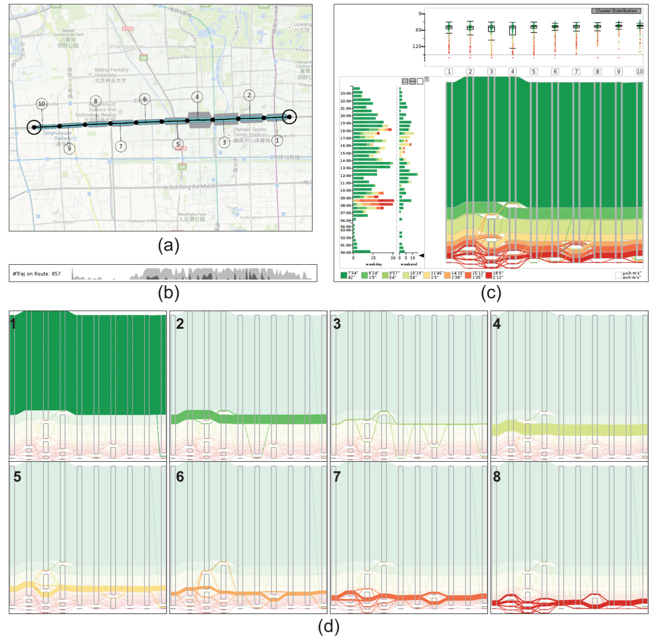
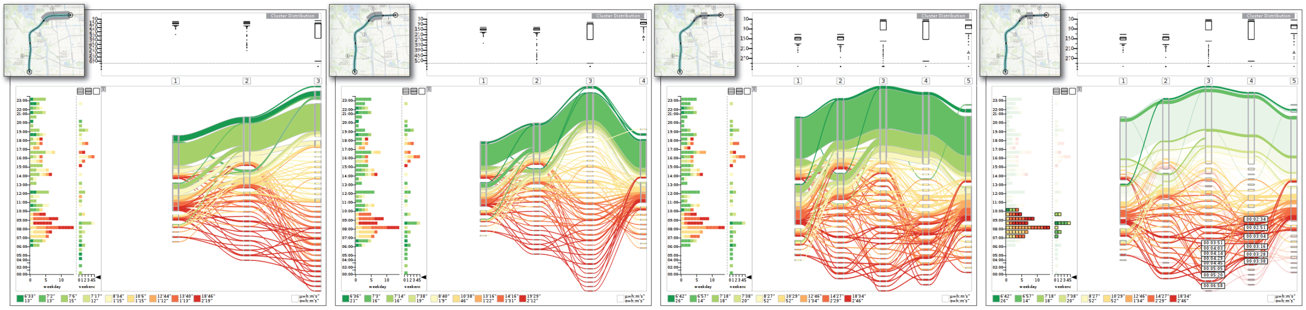


Figure 10: TrajRank with Even Segmentation: (a) a westward route is selected on the 4th ring; (b) the horizon graph shows the temporal distribution of trajectories on that route; (c) ranking diagram shows the trajectory ranking; (d) 8 clusters with different travel behaviors are highlighted respectively.

change. The 4th cluster has low ranks at the 3rd and 4th road segments. In contrast, the 5th yellow cluster has low rank at the 3th road segment. Trajectories in the 6th, 7th and 8th clusters apparently have low ranks at several road segments. For the last few groups with low rankings, we find that low rankings at different road segments appear together, which may indicate that delay on these road segments are strongly correlated. In the temporal distribution view, we also find that red and yellow trajectory groups appear together.

6.3 Case 3: Understanding Travel Behaviour with Uneven Route Segmentation

By adjusting the route segmentation, user can narrow down travel behaviour analysis to smaller road segments. In Figure 11, we select a route on the 4th ring. It starts at the southwest direction and ends at the northeast direction. We set the time range as one week, and extract 201 taxi trajectories. As Figure 11(a) shows, we first divide the route into 3 road segments with equal length. In the ranking diagram, the overall ranking score is correlated to the ranking on the 3rd road segment. In the box-plot view, we find it probably takes 2 minutes to travel on the 1st and 2nd road segment while the travel time on the 3rd looks significantly longer. The variance on the 3rd road is also larger. Further, we add a new node to split the 3rd segment into two halves. The left half has a large variance while the right half doesn't. The trajectory clustering is updated, which is again consistent with their rankings at the new 3rd road segment. In the next step, the 3rd road segment is split further. The result is in Figure 11(c). Rankings on the 3rd and 4th road segments are consistent to the overall ranking score. For example, the red cluster ranks lowest on the 3rd and 4th road segments. In the temporal distribution view, the red cluster is mainly in the morning on weekdays. In Figure 11(d), trajectories in the morning are highlighted to compare travel behaviour in the morning of weekdays and weekends. The green and light green trajectories at weekends rank at the top, while the red to yellow trajectories on weekdays rank at the bottom. We show the average time of the red and yellow clusters in



(a) (b) (c) (d)

Figure 11: TrajRank of Uneven Segmentation: (a) the route is segmented into 3 road segment with equal length; (b) one road segment is split, now there are 4 segments; (c) one road segment is further split, now there are 5 segments; (d) trajectories in the morning are highlighted.

the weekdays in Figure 11(d). We find although the length of the 3rd and 4th segments are nearly the same, these trajectories have longer travel time on the 3rd segment than the 4th.

7 User Study

We have performed an informal user study to evaluate the system. We recruited 7 students and all of them have some experience in visualization. We first explained to them the basic concepts of our work and demonstrated the system functionalities. Then we asked them to complete three exploration tasks with our system. Finally, we collected their feedbacks with a questionnaire and a free discussion about the advantages and disadvantages of our system.

Our tasks are designed to guide users to explore the taxi travel behaviors. In the first task, we asked users to analyze the ranking diagram with even segmentation, shown in Figure 1. This task is similar to Case 2. In the second task, we ask users to analyze the ranking diagram with uneven segmentation, shown in Figure 11. So it's basically a reproduction of Case 3. In the third task, we ask users to perform filtering and segmentation themselves, then construct their own ranking diagram and explore it. For all tasks, we ask users to focus on the three aspects of travel behavior exploration mentioned in Section 1.

Our questionnaire contains 17 questions in three categories: visual design, interaction and overall system function. The visual design questions are closely related to the design considerations mentioned in Section 4.2 and Section 4.3. The answers are based on a five-level Likert Scale, where 1 means "Strongly Disagree", and 5 means "Strongly Agree". All users finish their questionnaires. The results are summarised in Table 1, which is very encouraging. For most questions, the average rating is above 4. That means our system has fulfilled its design requirements and supports the major tasks. The only question with rating less than 4 is Q10, about comparing travel time on different road segments. We have interviewed the users rating 2 and 3 for this question. One of them was not so familiar with boxplot and has misinterpreted it. Another one says that the boxplots are so small, it's hard to measure their vertical differences, esp. when comparing non-adjacent boxplots.

During the free discussion, most users consider the functionality and interaction most satisfactory. One of them mentioned: "It's clear to get the relationship between road segments and travel behaviour." Another one said: "Color band is beautiful and provides good overview." For what they don't like, many users complained about the colors of trajectory group. When there are many groups, their colors are too similar and hard to differentiate. They also felt pity that there are no deeper analysis of the travel behavior.

8 Discussion

While the idea of ranking has been frequently used in visualization, for the first time it is applied to trajectories. This is the major difference with previous works on similar topics. Such a ranking based method allows users to compare the behaviour of taxis based on their relative position, instead of the absolute time. Besides, our visual design allows users to study the micro-behaviour of individual taxis, and compare them at a very fine spatial granularity. This is again different from existing works focusing on a whole route [19] or requiring to aggregate all trajectories [28].

Our method can be potentially extended to study other movement data, especially in the field of transportation and sport. For example, analysing the travel time ranking of buses on a specific route can help administrators study the delays, i.e. when and where buses fail to arrive on time. Analysing the ranking of race cars during a competition can be more interesting, because the goal of race car drivers is to reach the destination as soon as possible. The interactions between different race car drivers, and the tactics involved can all be studied.

In this work we are working with taxi GPS dataset. We envision our users to be drivers and transportation analysts. However, our current focus is to develop a visual analysis method rather than an application. Therefore our exploration tasks are summarized from

Table 1: Summary of the evaluation questions and results.

Category		Questions	Rating
Visual Design	Ranking Diagram	Q1: For a trajectory or travel group, easy to trace its rankings over different road segment (C I, C III)	4.3
		Q2: There's no heavy visual clutter preventing ranking change analysis (C II)	4.6
		Q3: Easy to gain an overview of different trajectory groups (C IV)	4.6
		Q4: Multi-level analysis supports to provide both high level summary and local level details (C V)	4
		Q5: Easy to notice small group (C VI)	4.1
	Occurrence Temporal Distribution View	Q6: Distortion on time axis has no effect on reading time (C I)	4.6
		Q7: Easy to understand the time discretization (C II)	4.4
		Q8: Easy to explore the difference of weekdays and weekends (C III)	4.7
	Modified Box-plot View Band Encoding on Map	Q9: Easy to compare the travel time variance of road segments	4.7
		Q10: Easy to compare the average travel time of road segments	3.3
Interactions	Filtering	Q11: Natural to evoke different functions	4.4
		Q12: Satisfied with feedback while filtering	4.4
		Q13: Easy to manipulate the filters	4.2
	Segmentation	Q14: Easy to do different segmentation	4.6
System		Q15: Intuitive to gain an overview of the travel behavior of a route	4.3
		Q16: Strong support to explore the relationship between travel behavior and road segment	4.6
		Q17: Strong support to explore the relationship between travel behavior and travel occurrence time	4.4

existing works in travel time analysis, and the participants in our user study are common students. Currently we are not working with domain experts. In the future, when we intend to address a specific domain question, it would be critical to redefine the tasks based on domain requirements. In that case, we would also need to invite domain experts for evaluation.

Our method has some limitations. According to our definition of the ranking score, we emphasize the regular outliers such as morning peak, and relatively ignore the irregular outliers such as a sudden accident. Moreover, the overall score does not have a clear semantic meaning. We find it very related to the average travel time, so we use a red to green color scale for the trajectory groups. The red color corresponds to low score, which is approximately low speed, or bad traffic condition. However, the score is not rigidly correlated to travel time, and the travel time range of different trajectory groups can overlap. Besides, as discovered in the user study, the ranking diagram can not handle too many groups. Otherwise the group colors will be hard to distinguish.

9 Conclusion and Future Work

In this work, we introduce a visual analysis method *TrajRank* to explore taxi travel behaviour on a route. The central idea of our design is ranking. For a set of taxis on the same route, we compare their travel behaviour by the relative travel time ranking on each road segment. We further calculate a ranking score to group these trajectories. We have a carefully designed visual interface to reveal the spatial-temporal distribution of travel time. With rich filtering and segmentation interactions, *TrajRank* enables users to flexibly explore taxi travel behaviour on a route. Finally, we demonstrate the effectiveness and usability of our method with three use cases and a user study.

In the future, we would like to extend single route analysis to multiple routes analysis. We will consider complex route structures, e.g., parallel, merging or splitting. We especially want to compare the travel behaviour on alternative routes. Besides, we want to improve the road segmentation method. Currently the route is segmented in equal-distance manner by default. In future, we would consider segmentation according to underlying traffic context, e.g., traffic intersections. Finally, we would like to test our ranking method on other movement data, such as F1 race-car data.

Acknowledgements

The authors wish to thank Datatang and OpenStreetMap for providing the data, and the anonymous reviewers for their valuable comments. This work is supported by NSFC No. 61170204 and the National Program on Key Basic Research Project (973 Program) No. 2015CB352500.

References

[1] C. Ahlberg, C. Williamson, and B. Shneiderman. Dynamic queries for information exploration: An implementation and evaluation. In *Proc. ACM SIGCHI*, pages 619–626, 1992.

[2] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual analytics of movement*. Springer, 2013.

[3] M. Batty. Rank clocks. *Nature*, 444:592–596, 2006.

[4] M. Behrisch, J. Davey, S. Simon, T. Schreck, D. Keim, and J. Kohlhammer. Visual comparison of orderings and rankings. In *Proc. EuroVA*, 2013.

[5] M. Ben-akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani. Dynamit: a simulation-based system for traffic prediction. *DACCORD Short Term Forecasting Workshop*, 1998.

[6] P. S. Castro, D. Zhang, and S. Li. Urban traffic modelling and prediction using large scale taxi gps traces. In *Proc. ACM Pervasive*, pages 57–72, 2012.

[7] T. Catarci, M. F. Costabile, S. Levialdi, and C. Batini. Visual query systems for databases: A survey. *Journal of Visual Languages and Computing*, 8:215–260, 1997.

[8] J. W. Clark. Time-distance transformations of transportation networks. *Geographical Analysis*, 9(2):195–205, 1977.

[9] I. Dees. Openstreetmap jxapi. http://wiki.openstreetmap.org/wiki/Xapi.

[10] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2277–2286, 2013.

[11] M. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40:27–37, 2003.

[12] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning (2nd ed.)*. Springer, 2009.

[13] S. Havre, E. G. Hetzler, P. Whitney, and L. T. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Trans. Vis. Comput. Graph.*, 8(1):9–20, 2002.

[14] J. Huang, A. Y. Xue, Y. Zheng, R. Zhang, X. Xie, and Z. Xu. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In *Proc. IEEE ICDE*, pages 254–265, 2013.

[15] A. Inselberg. Multidimensional detective. In *Proc. IEEE InfoVis*, pages 100–107, 1997.

[16] P. Kidwell, G. Lebanon, and W. Cleveland. Visualizing incomplete and partially ranked data. *IEEE Trans. Vis. Comput. Graph.*, 14(6):1356–1363, 2008.

[17] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Trans. Vis. Comput. Graph.*, 12(4):558–568, 2006.

[18] R. Krüger, D. Thom, M. Wörner, H. Bosch, and T. Ertl. Trajectorylenses - a set-based filtering and exploration technique for long-term trajectory data. *Comput. Graph. Forum*, 32(3):451–460, 2013.

[19] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. Ni. Visual analysis of route diversity. In *Proc. IEEE VAST*, pages 171–180, 2011.

[20] F. Nivan, P. Jorge, T. V. Huy, F. Juliana, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city cab trips. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2149–2158, 2013.

[21] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li. Land-use classification using taxi gps traces. *IEEE Trans. Intell. Transp. Syst*.

[22] A. P. Sawant and C. G. Healey. Visualizing multidimensional query results using animation. In *Proc. IS&T/SPIE Electronic Imaging*, pages 680–904, 2008.

[23] J. Seo and B. Shneiderman. A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections. In *Proc. IEEE InfoVis*, pages 65–72, 2004.

[24] C. Shi, W. Cui, S. Liu, P. Xu, W. Chen, and H. Qu. Rankexplorer: Visualization of ranking changes in large time series data. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2669–2678, 2012.

[25] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2565–2574, 2012.

[26] G. Upton and I. Cook. *Understanding Statistics*. Oxford University Press, 1996.

[27] J. Van Wijk and E. Van Selow. Cluster and calendar based visualization of time series data. In *Proc. IEEE InfoVis*, pages 4 –9, 140, 1999.

[28] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering. Visual traffic jam analysis based on trajectory data. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2159–2168, 2013.

[29] P. C. Wong and R. D. Bergeron. Multivariate visualization using metric scaling. In *Proc. IEEE Visualization*, pages 111–118, 1997.

[30] W. Wu, Y. Zheng, H. Qu, W. Chen, E. Groller, and L. M. Ni. Boundaryseer: Visual analysis of 2d boundary changes. In *Proc. IEEE VAST*, pages 143–152, 2014.

[31] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. T-drive: driving directions based on taxi trajectories. In *Proc. ACM SIGSPATIAL GIS*, pages 99–108, 2010.

[32] W. Zeng, C.-W. Fu, S. Arisona, A. Erath, and H. Qu. Visualizing mobility of public transportation system. *IEEE Trans. Vis. Comput. Graph.*, 20(12):1833–1842, 2014.

[33] Y. Zheng, Y. Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *Proc. ACM UbiComp*, pages 89–98, 2011.