International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 17
July 2018

A Visual Analytic System to handle Large Scale Trajectory Data with Different Route Choices

G Asha Rani & S.Giri Babu

[1] *PG Scholar, Department of CSE, PACE Institute of Technology and Sciences,Vallur, Prakasam,, Andhrapradesh, India*

[2] *Associate Professor, Department of CSE, PACE Institute of Technology and Sciences,Vallur, Prakasam, Andhrapradesh, India*

.

Abstract—*There are often multiple routes between regions. Drivers choose different routes with different considerations. Such considerations, have always been a point of interest in the transportation area. Studies of route choice behaviour are usually based on small range experiments with a group of volunteers. However, the experiment data is quite limited in its spatial and temporal scale as well as the practical reliability. In this work, we explore the possibility of studying route choice behaviour based on general trajectory data-set, which is more realistic in a wider scale. We develop a visual analytic system to help users handle the large-scale trajectory data, compare different route choices, and explore the underlying reasons. Specifically, the system consists of:*

1. the interactive trajectory filtering which supports graphical trajectory query;

2. the spatial view which gives an overview of all feasible routes extracted from filtered trajectories;

3. the factor visualizations which provide the exploration and hypothesis construction of different factors' impact on route choice behavior, and the verification with an integrated route choice model. Applying to real taxi GPS dataset, we report the system's performance and demonstrate its effectiveness with three cases.

Index Terms—*Route Choice Behavior, Visual Analysis, Interaction, Route Choice Model*

1 Introduction

With the development of sensing technologies, a variety *big data* has been produced in urban space. Urban computing combines urban sensing, data management, analytics and services as an integral process, which throws light on the rich knowledge of city and improves people lives [1]. Transportation is one of the most essential urban computing applications. Many transportation systems analyse the city-wide human mobility data and other urban data (e.g. weather data etc) to understand the travel behavior [2], [3] and improve the travel experience [4]. In modern traffic networks, there are often multiple routes when travelling from one place to another. Understanding how drivers make route choices, i.e., the route choice behavior, is an interesting topic in transportation area. It not only assists the city planners in the improvement of route usage, but also helps drivers make wise travelling decisions. However, route choice behavior is not an easy problem. Drivers choose different routes considering different factors. The expected time cost is one example. Choosing the route with minimum time cost is what widely experienced in daily life. Some other factors may also influence route decision making, like the number of traffic lights, travelling comfortableness, etc. Meanwhile, the impact of factors may change over time. Drivers who care about the travelling comfortableness at weekends, might trade it off with travel efficiency on workdays. Moreover, the problem is even more complex when various factors interact with each other.

Classically, research efforts have been made to study the influence of different factors on route choices based on Stated Preference (SP) survey data [5]. SP survey collects the route preferences in hypothetical situations from respondents. Different choice considerations, such as travel safety, can be directly captured by the information in questionnaires. With SP data, various route choice models [6], [7] are developed, trying to estimate the impact of different factors on the route choice behaviour. However, such investigations are limited in scale and the surveys need to be carefully designed. Also, information obtained from investigation is quite subjective and not practically reliable enough. In more recent years, some researchers perform the analysis with the help of Global Positioning System (GPS) where GPS receivers are used to collect trajectories from volunteers. Compared to traditional investigations, it takes less effort and is more realistic. But such pilot studies are often conducted among a limited number of users in a restrained spatio-temporal scale, like only collecting morning commute trips [8], [9]. In this work, we explore the possibility of studying route choice behaviour based on more general GPS trajectory data, i.e., taxi GPS trajectories. Compared to well-designed experiments, it takes less effort to collect general taxi GPS data. Taxi trajectories are sampled in real situation and cover a wider spatial and temporal range. However, new challenges arise when it comes to route choice behaviour analysis:

• ***Extract relevant trajectories in the context of multiple routes:*** Unlike the experimental GPS trajectories which are constrained in a relatively limited spatial and temporal range, the general trajectory data could be very complex to handle. Extracting trajectories related to the routes is a big challenge to tackle.

• ***Raise hypotheses on factors that significantly influence the route choice behaviour:*** Different from verification of predefined factors in hypothesis-oriented experiments, it is a crucial challenge to decide what factors to detect from general GPS trajectories and how to indicate their impact on route choices. Visual analytics is proposed as the science of analytical reasoning facilitated by interactive visual interfaces [10]. By integrating computational and theory-based tools with innovative interactive techniques and visual representations, visual analytics enables human to participate in problem solving.

In this work, from the perspective of visual analytics, we propose a visual analytics system which leverages human interaction and judgement in the trajectory data mining process [11] to tackle the above challenges: with a suite of graphical filters, trajectories between regions of interest are queried interactively; based on filtered trajectories, feasible routes are constructed automatically; with a list of factors derived from general GPS trajectory data, route choice distributions over those factors are visualized, which supports to explore and raise hypotheses on potential influence; then the hypotheses are further verified by the statistical model to draw reliable conclusions. The contributions of this work are:

• We explore the possibility of analyzing multiple route choice behaviour based on general GPS data.
• We develop a visual analytic system to explore the route choice behaviour with real GPS data.

For the remaining part, we first report related work in Section 2. Then in Section 3, we give an overview of the data background, analytic tasks and overall system pipeline. Details of route generation and visual design are explained in Section 4 and Section 5. We report the system's performance in Section 6 and validate its effectiveness in Section 7. In Section 8, we have a discussion on the system. Finally comes the conclusion.

2 Related Work

In this section, we have a discussion on the related work: route choice behaviour analysis in transportation field, research progress in visual analytics of trajectories, rank-based visualization and route visualization.

2.1 Route Choice Behaviour Analysis

Route choice behaviour has been widely studied in the transportation area. In early years, most researches are based on statistical investigations or experiments. By analysing a total of 2182 home-to-work records in Seattle, Mannering et al. [12] find that 26% people do not always use the same route. To find the reasons, Khattak et al. [13] study 700 commute trips collected via questionnaires, and find that both congestion and the perception of alternative routes increase the probability of route changes. With respect to personality, males, young people and experienced drivers are more likely to change routes, as concluded by Xu et al. [14] in a study of 247 morning home-to-work trips. In these works, statistical inquiries play an important role, where questionnaires are carefully designed to obtain problem-related information involving personal details. However, investigations are limited in both the sample range and its validity. Realism is also a problem given the divergence between recalled and observed circumstances. To obtain more authentic information, some researchers base their studies on GPS data in recent years. Li et al. [8] study morning route choice patterns based on a GPS dataset collected from 182 vehicles in 10 days. Factors like age, departure time and income level are found convincingly influential. More recently, Vacca et al. [9] study route switch behaviour between the same OD (i.e. Origin-Destination) pair by tracking the participants with portable GPS devices. Some dominant factors are revealed, such as traffic light number (per km), highway percentage, perception of time, etc. Compared with investigations, GPS records provide more truthful measurement of route choice behaviour, with lower costs and higher precision. However, subject to the analytical requirement of individual characteristics, the data is still problem related and range-limited. Instead, our system is designed for general GPS data covering a much larger range (tens of thousands of taxies). One similar work is proposed by Pan et al. [15]. They extract regular routing patterns from the historic taxi trajectories and detect the anomaly routing behaviour that significant differ from the original patterns. Based on social media data, they focus on exploring semantic meaning of the travel anomalies. Different from their semantic exploration, our work focus on comparing the properties of

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 17
July 2018

multiple routes and exploring the regular factors that impact route choice behaviour, such as the departure time. Meanwhile, what's provided in our system can support interactive data customization and real-time processing according to different analytical demands.

2.2 Trajectory Visual Analysis

In trajectory mining field, Zheng [11] survey various mining techinques, including outlier detection, pattern mining etc. From the apsect of visual analysis, Andrienko et al. [16] present a taxonomy of generic analytic techniques based on possible types of movement data. For trajectories, there are three kinds of visual explorations [17]: direct depiction, pattern extraction and visual aggregation. Direct plotting could simply fail because of visual cluttering. Pattern extraction methods employ automatic analysis to extract underlying data pattens [18], e.g. the traffic jam propagation graph extraction [19]. Aggregation methods visualize movement groups to reveal the high-level movement graph. Guo [20] and Andrienko [21] et al. construct geographical regions and visually aggregate the in-between movements as flows. Besides aggregation between regions, travel behaviour within interchange region can also be visualized. Guo et al. [22] provide a circular design to explore movement at a road intersection. Zeng et al. [23] derive a visualization from Circos [24] to display interchange traffic flow at subway transition stations. Lu et al. [25] aggregate trajectories along a single route and rank them by the time cost along the road segments, to reveal mainstream and outliers. Liu et al. [26] study the route diversity between locations and provide a clock like radial layout to display temporal statistic distribution. Different from analysing individual trajectories in Liu et al.'s work, our method provides analysis based on the extracted topology structure. Zeng et al. [27] visualize the mobility of routes starting from a single source in public transportation system and provide the comparison among different routes. Similar to their routes' comparison, our work provides comparison among multiple routes. Alternative to analyse trajectories as a whole, some works perform local analysis of the filtered trajectories of interest. Andrienko et al.'s book [28, Chapter 4.2] summarizes the different kinds of filtering, including the spatial, temporal filtering etc. Marios et al. [29] introduce spatial query which specifies a spatiotemporal pattern as a sequence of distinct spatial predicates. Vieira et al. [30] design the trajectory query using regular expression over a spatial alphabet of regions. Different from those textural query languages, visual query languages explore the spatial data graphically. Ferreira et al. [31] propose a visual query model to filter trajectories by their origins and destinations. Trajectory Lenses [32] supports users to interactively filter the trajectories by manipulating the lenses on the map. Similar to Trajectory Lenses, we design a suite of circular filters in this work. Compared with Trajectory Lenses, our design not only supports more spatial constraints but also allows the direction assignment.

2.3 Rank-based Visualization

Ranking as an operation to organize data in order is widely used in visualization, especially when comparing data items over multiple attributes. Because of the linear property of ordering, ranking technique is usually integrated into line-based visualizations .Parallel Coordinates visualizes multivariate data by connecting items' actual value over multiple attributes, which embeds the ranking implicitly. Instead of actual value, Bump Charts explicitly visualizes data by order and connect order change with slopes. One more recent ranking design is LineUp , which not only visualizes the ranking changes, but also encodes the cause of the rank. Similar to those ranking techniques, we rank routes over attributes for comparison. However, in our case, we need to deal with dynamic route attributes, e.g. the travel time cost attribute of a route which ensembles the time costs from all trajectories. Some ranking visualizations deal with dynamic changes by expanding the time dimension. Batty designs Rank Clocks to show the change of city population rankings across several centuries, which is similar to Parallel Coordinates but represents different time as axes. On the other hand, keeping time continuous, Shi et al. propose RankExplorer, in which they segment the rankings into several groups and use a ThemeRiver [39] to show their temporal changes. Instead of expanding time, we aggregate the dynamic route attribute samplings by trajectories and propose a ranking visualization for attributes with single value and multiple values.

2.4 Route Visualization To visualize a path, a well-known technique in geographical application is the space-time cube , which visualizes the dynamic changes of geographical and temporal changes of a path in 3D. Tominski et al. propose stacking bands in hybrid 2D/3D view to visualize the trajectory attributes. These direct drawing method would bother with visual clutter when the number of routes increases. With the metaphor of lenses, Karnick et al. [32] place magnifying lenses on the significant points along a route, to encode more details. The other way is to do distortion. Agrawala [38] creates route maps that are similar to humandrawing maps. The route is distorted and simplified to highlight important features, which makes the route map more readable. Alternatively, Sun et al. [37] distort the map to broaden the roadsof interest so that temporal information can be embedded. For easy perception of the geospatial information of routes, we keepthe map view undistorted. With careful design, the topologicalinformation of multiple routes is encoded. For sufficient analysis among routes, Zeng et al. [27] presentan is time flow map view in a parallel is otime fashion. There are similar flow diagrams [36], [37] when broadening the horizontal representation to temporal dimension. We derive the abstract routeview from flow diagram to show the topology structure.

3.OVERVIEW

In this section, we first introduce the background of data and tasks of this work. Then we present pipeline of the visual analytics system.

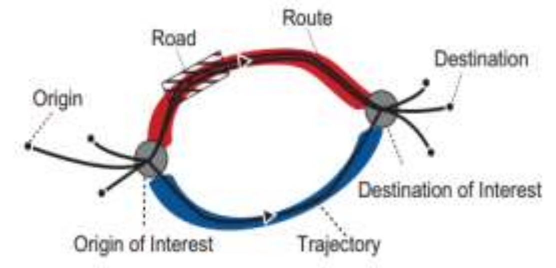

Fig. 1. Illustration of Related Traffic Concepts

3.1 Data Background

In the following, we list down common terminologies used in this work to facilitate our discussion. They are illustrated in Figure 1:

• *Trajectory* records a list of positions that an object travels in temporal order.

• *Origin/Destination (O/D)* refers to the beginning/ending position of the movement.

• *Road* is the physical connection between one location to another, where vehicles can travel on.

• *Route* is a sequence of roads that vehicles travel through.

• *Origin/Destination of Interest (OoI/DoI)* refers to the beginning/ending position of movement that analyst is interest in.

Multiple Routes are all the travelling routes between thesame pair of *OoI* and *DoI*. Note that*OoI*/*DoI* is not necessarily the O/D. *OoI*/*DoI* is set atthe region of interest where multiple routes begin/end.Different from the predefined factors in the controlled experiments, factors in our case are directly derived from general GPS dataset. In Table 1, factors are categorised into two groups:*route-related factors* and *trajectory-related factors*. For each route,some inherent attributes probably act as factors in route decision making, e.g. the length of route, the number of traffic lights along the route and route importance. Specifically, the route importancerefers to the average road level of the route. The road level is calculated from the *highway* tag in Open StreetMap [38] which indicates what type the road is, such as a trunk or residential road .Besides those static attributes, there are important dynamic factors which change over time. such as the time cost. So the time cost distribution of passing vehicles is viewed as a route-related factor and its average and variance are studied.

On the other hand, each trajectory has individual differences that potentially affect the route choice, e.g. departure time in a day, departure day and trajectory's length. For example, drivers travelling in peak time and off-peak time may make different route decisions. It is also possible for drivers to select different routes when travelling in different distances from O to D.

3.2 Analytic Tasks

In this part, we clarify analytic tasks to explore route choice behaviour with general GPS trajectories. According to the typology of visualization tasks [49], our analytic tasks are designed from high-level to low-level. Given a pair of *OoI* and *DoI*, firstly, anoverview is given to present multiple feasible routes. Then the route-related factors and trajectory-related factors are calculated and visualized. It helps to build hypotheses about how factors impact the route choices. With the hypothesis concertain routes and factors, the system should be capable to examine the proposed hypotheses to tell if the impact is significant. Based on these requirements, the design tasks are summarized as following:

• *Overview of multiple route choices (T1)*: give an overview of all feasible route choices between *OoI* and *DoI*.

• *Exploration on the route-related factors' impact on route choice (T2)*: describe each route by route-related factors and compare routes in terms of route-related factors.

• *Building hypotheses on the impact of trajectory-related factors on route choice (T3)*: explore the route choice distribution over trajectory-related factors and propose hypotheses on the potential impact.

• *Evaluating the impact of trajectory-related factors (T4)*: build a statistic model to examine whether the impact is significant or not.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 17
July 2018

3.3 System Overview
To support the above tasks, we propose a visual analytic system integrating automatic processing, visualization and interaction. Figure 2 shows the system's pipeline. In the preprocessing stage, trajectories are cleaned. A quadtree spatial index is built to facilitate filtering in massive trajectories. In run-time stage, trajectories between a pair of *OoI* and *DoI* are filtered using a suite of graphical filters. With those filtered trajectories, all feasible routes are extracted. A topology graph of the routes is constructed using a grid-based algorithm. For each route or trajectory, related factors (discussed in Section 3.1) are derived. Then those routes and factors are fed as input to the visualization module. The visualization module consists of three parts. The spatial view gives a geographical overview of the multiple routes (T1). The route-related factor view displays the route-related factors in a ranking diagram. Users can compare them across different routes (T2). The trajectory-related factor view visualizes different route choices over trajectory-related factors. This view supports the proposal of hypothesis (T3). Then users can input their hypotheses. A choice analysis model, i.e. Multi nominal Logit model (MNL) [20], is used for the verification (T4). After modelling, the results are displayed back in the trajectory-related factor view, to tell whether the impact is significant or not. To combine all aspects, the three views cooperate in a brushing and linking manner, i.e. entities selected in one view are updated in other views. At last, users can launch a new loop of analysis by resetting the filtering.

4 Multiple Routes Generation

For the massive taxi GPS trajectories, the system provides a suiteof graphical fitlers. They support to query trajectories intuitively with spatial and temporal constraints. With the filtered trajectories,a grid-based algorithm is proposed to extract all feasible routes automatically.

4.1 Trajectory Filtering
From the temporal aspect, a two-level temporal filter is provided:date and time. Date range is set in the date filter. Time range in a day is set in the time filter, whose granularity is 10 minutes. With these two different temporal granularities, the temporal filter allows users to query trajectories in a periodic pattern, such as the commute trips in the morning. From the spatial aspect, we design the filter similar to TrajectoryLenses [32]. The filter covers a circular area and filters trajectories with 6 spatial constraints. The 6 constraints are defined according to the spatial relationship between trajectory and the underlying circular area: *origin*, *destination*, *origin/destination*, *passing*, *inclusive*, and *exclusive*. The concepts are shown in Figure 3(b). For example, a filter with the *origin* constraint filters trajectories starting from the circular area. Besides the spatial constraints, there are some other geometric constraints, e.g. the center position and radius of the circular area. For usage simplicity, constraint configuration is embedded into the circular filter. As Figure 3(a) shows, when hovering on a certain region, certain function is waked and the corresponding handle is shown. For example, hovering in the center of the circle invokes the moving function and a + handle is visible. Clicking and dragging the + changes the center of the filter. Complex queries can be built which combines different filters in an intersection manner. Moreover, for two or more filters, directions can be assigned between filters to select trajectories following certain flow directions. For the ease of constraint perception, constraints are explicitly encoded in the circular filter. Figure 3(b) shows the circular filters with 6 spatial constraints respectively. In this work, the first two filters are detected as the *OoI* and *DoI* by default.

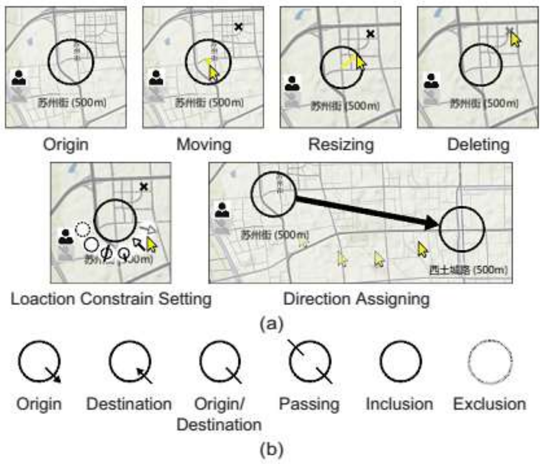

Fig. 3. Circular Filter: (a) different functions are invoked by hovering on corresponding regions. Direction between filters is assigned by dragging from one to another. (b) the circular filters with 6 spatial constraints.

4.2 Multiple Routes Extraction
With the filtered trajectories from *OoI* to *DoI*, we employ a general grid-based algorithm to extract multiple routes automatically. The basic idea is to cover the trajectories by grid and then build up the multiple route graph among cells of the grid. Figure 4 illustrates the process of route extraction. Figure 4(a) shows the filtered trajectories between *OoI* and *DoI*. At the beginning, a uniform grid is

covered over the boundary box of filtered trajectories, which divides the space into cells (Figure 4(b)). Trajectories are segmented by the cells and each of them can be denoted by the sequence of passing cells (Figure 4(c)). Each cell collects the segments from trajectories which intersect with it. Then for each cell that contains segments, we derive the average direction from trajectory segments inside it. The directions are further approximated as horizontal or vertical ones (Figure 4(d)). The horizontal direction is more likely the left-right going than the up-down going and the vertical one is more likely the up-down going. To remove the zigzag between two cells, two types of ambiguous cells are detected: the neighbour cells with horizontal direction which are side-by-side horizontally; the neighbour cells with vertical direction which are side-by-side vertically. The detected cells are merged (Figure 4(e)). After that, routes are formed by linking the centroids of cells (Figure 4(f)). Cells with more than one in/out degree are detected as the splitting/merging nodes (Figure 4(g)). The multiple route graph is constructed with these nodes and the routes connecting them. Finally, multiple routes are encoded visually (Figure 4(h)), which will be introduced in Section 5.1.

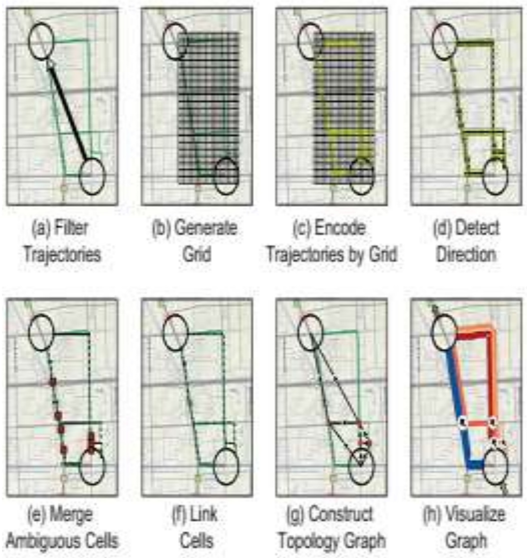

Fig. 4. Multiple Routes Construction: by covering a grid over trajectories, multiple route graph is built upon travelled cells.

5 Visual Design

In this section, we present design of visualizations in our system. Corresponding to tasks introduced in Section 3.2, the interface mainly consists of three parts: the route spatial view, the routerelated factor view and the trajectory-related factor view.

5.1 Route Spatial View

With the specified *OoI* and *DoI*, multiple routes are obtained by the algorithm introduced before. To provide an overview of all the feasible routes (T1), the spatial view is designed with following considerations:

• Represent *OoI* and *DoI* (CI): to locate the areas of *OoI* and *DoI*.

• Display multiple routes (CII): to visualize the feasible routes between *OoI* and *DoI*, including both the popular ones and the seldom travelled ones.

• Indicate traffic flow directions (CIII): to show the travelling directions along routes, especially at the intersections.

• Summarize the routing (CIV): to summarize the major route choices by merging similar routes

After defining the filtering conditions (see Section 4.1), the *OoI* and *DoI* circular filters are settled on the map. To indicate *OoI* and *DoI* filters, inward and outward arrows are attached to the circular filters respectively (CI) (Figure 5(a)). Each extracted route is visualized as a band, whose width encodes the number of passing trajectories. A logarithmic mapping is used to maintain the visibility of seldom travelled routes (CII). Routes are stacked together when sharing the same roads. When hovering, a tooltip is shown to facilitate selection of the bands (Figure 5(b)). The number of travelled trajectories is also displayed in the tooltip. The current hovered route is highlighted both in the spatial view and the tooltip. Users can easily switch their focuses in the tooltip, in case some route is too small to choose on the map. Considering that directions in straight roads are self-evident, we only indicate the traffic directions at the crossings of roads (CIII) using glyphs. The size of glyph encodes the volume of passing traffic flows. The arrow inside the glyph implies the average traffic flow direction at the crossing. In order to summarize the complex routing, we divide the routes into a few groups (CIV). Each group contains a mainstream route and some alternative routes. We first choose some most popular routes as the mainstreams. Specifically, the route whose traffic volume is larger than third quartile $Q3$ of the whole traffic volume distribution are regarded as the mainstreams. The maximum number of mainstream routes is limited to 5 in order to avoid excessive dividing. With mainstreams determined, the remaining are assigned to the mainstream routes according to the topology similarity. In our case, we denote a route as a sequence of its road crossings, and use the edit distance [21] to measure the similarity between routes, which counts the minimum amount of switches required to transform from one sequence to the other. We show the grouping results in a

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 17
July 2018

topology graph to help understand the routing (Figure 5(a)). Each mainstream with its similar alternative routes are considered as a group.

Qualitative colors [22] are used to differentiate different groups. Within each group, all routes are colored similarly, with the lightness inversely proportional to the route popularity. Figure 5(b) shows the color legend of the two groups in Figure 5(a). The color legend is consistent over all views.

5.2 Route-related Factor View

Inspired by ranking visualizations (e.g., LineUp [36]), we design a ranking-based visualization to support exploration on route-related factors' impact on route choice behaviour (T2). The ranking-based visualization helps users interpret how the factors affect route choices. There are several considerations we have taken in the design:

• Accommodate different factor types (CI): to visualize both static and dynamic factors

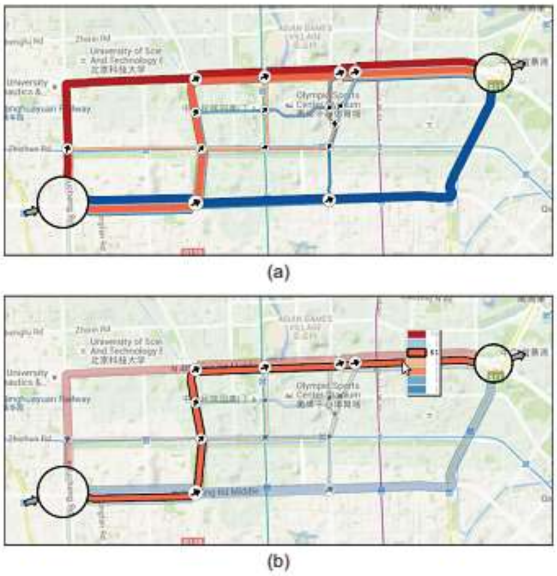

Fig. 5. Route Spatial View: (a) geographical overview of multiple routes: the route width encodes the amount of traffic flow. The arrow glyph at each intersection indicates the average flow direction. (b) one highlighted route in the road segment tooltip: all stacked routes are displayed in the tooltip to facilitate selection.

6 IMPLEMENTATION AND PERFORMANCE

A prototype system is developed to verify the effectiveness of our method. In this section, we first introduce the experiment dataset and implementation detail. Then we report the system's performance.

6.1 Input Data

We take the GPS dataset recorded in Beijing as the experiment data. The data is collected from 28,519 taxis in 24 days, from March 2*nd* to 25*th*, 2009. The data size is 34.5 GB in total and consists of 379,107,927 sampling points. The sampling rate is every 30 seconds. Each sampling point contains the following attributes: *time*, *latitude*, *longitude*, *speedmagnitude*, *direction* as well as a boolean *CarryPassengerState*. *CarryPassengerState*. is a tag indicating whether the taxi carries passengers or not. In this work, we only use the trajectories with passengers, each of which can be identified by ID. Besides the taxi GPS dataset, the road network data is collected from OpenStreetMap's jXAPI. Following an existing paper [19], trajectories are cleaned and matched to the road network in the data preprocessing step. The final data size is 12.1 GB.

6.2 Implementation

The system is mainly written in C++, with Qt framework. The rendering is performed with both OpenGL and Qt GraphicsView framework. A third-party library Graphviz is used to do the topological graph layout in route visualization. A MatLab extension is integrated in the system to perform the MNL analysis. As introduced in Section 4, trajectory filtering supports to narrow the scope of analysis down to trajectories related to certain OoI/DoI pair, which are fed into the extraction of multiple routes and the further visual analysis. Several strategies are adopted to facilitate the filtering.

In the preprocessing stage, trajectories are indexed by a spatial quadtree, which divides the 2D spatial region recursively and adaptively based on the distribution of trajectories' sampling points. Each quadtree node stores the IDs of trajectories that intersect it. In the run-time stage, a filtering operation is conducted in three steps: in the coarse filter step, the system fetches trajectories from the quadtree nodes where the circular filter locates; in the intermediate filter step, the top N trajectories (e.g. N = 100 in this work) which satisfy the filtering constraints are returned and rendered; in the fine filter step, trajectories satisfying the filtering constraints in the whole dataset are filtered. To ensure interactive filtering, the fine filtering step is not performed during dynamic filtering. For example, during the procedure of moving or resizing the circular filter, only the top N trajectories are returned and rendered. Once the filter is settled down, the fine filtering step is performed. Meanwhile, when multiple filters are applied, the filtering is conducted based on the previous filtered trajectories recursively, where the query space is much smaller than the whole data set.

7 Conclusion

In this paper, we explore the possibility of studying route choice behaviour based on taxi GPS trajectories. Compared to classical route choice analysis method, our general GPS based solution covers larger temporal-spatial range as well as larger number of samples. In this work, we list the factors that can be derived from trajectories, which defines the boundary of this general GPS data based solution. With this, we present a visual analytic system which supports tasks from route choice overview to verify factors' impact on route choice. The system's visualizations and interactions are designed carefully according to task-oriented considerations. The system allows interactive visual exploration in massive trajectories and factors exploration with route choice model. With Beijing taxi GPS trajectory dataset, we demonstrate three case studies to show the system's effectiveness. In the future, we would like to apply the system to more datasets. For example, applying to trajectory datasets in different areas, we probably are able to compare the route choice behavior of drivers over different regions. Meanwhile, we would like to improve and extend our system regarding the current limitations., there are two possible research directions. Considering that the input factors are fixed, we will improve the system to support the creation of factors. For example, OD distribution can be one of the possible trajectory related factors. Another interest point is to extend to the system with route advisory function. By taking the analysis of route choice, it is possible to recommend routes by taking different factors into consideration and measure the fitness of route.

REFERENCES

[1] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, p. 38, 2014.

[2] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatiotemporal causal interactions in traffic data streams," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discover and data mining*, 2011, pp. 1010–1018.

[3] J. Yuan, Y. Zheng, and X. Xie, "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*, 2012, pp. 186–194.

[4] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-drive: enhancing driving directions with taxi drivers' intelligence," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 220–232, 2013.

[5] W. Adamowicz, P. Boxall, M. Williams, and J. Louviere, "Stated preference approaches for measuring passive use values: choice experiments and contingent valuation," *American journal of agricultural economics*, vol. 80, no. 1, pp. 64–75, 1998.

[6] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta, "A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks," pp. 697–711, 1996.

[7] V. Henn, "Fuzzy route choice model for traffic assignment," *Fuzzy Sets and Systems*, vol. 116, no. 1, pp. 77–101, 2000.

[8] H. Li, R. Guensler, and J. Ogle, "Analysis of morning commute route choice patterns using global positioning system-based vehicle activity data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1926.1, pp. 162–170, 2005.

[9] A. Vacca and I. Meloni, "Understanding route switch behavior: An analysis using gps based data," *Transportation Research Procedia*, vol. 5, pp. 56–65, 2015.

[10] J. J. Thomas and K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[11] Y. Zheng, "Trajectory data mining: an overview," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 3, p. 29, 2015.

[12] F. Mannering, S.-G. Kim, W. Barfield, and L. Ng, "Statistical analysis of commuters' route, mode, and departure time flexibility," *Transportation Research Part C: Emerging Technologies*, vol. 2, no. 1, pp. 35–47, 1994.

[13] A. J. Khattak, J. L. Schofer, and F. S. Koppelman, "Effect of traffic information on commuters' propensity to change route and departure time," *Journal of Advanced Transportation*, vol. 29, no. 2, pp. 193–212, 1995.

[14] C. Xu, W. Wang, Z. Li, and C. Yang, "Comparative study on drivers' route choice response to travel information at different departure time," *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR)*, vol. 3, pp. 97–100, 2010.

[15] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in *Proceedings of the 21st ACM SIGSPATIAL International Conference on*

Advances in Geographic Information Systems, 2013, pp. 344–353.

[16] G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Kisilevich, and S. Wrobel, "A conceptual framework and taxonomy of techniques for analyzing movement," *Journal of Visual Languages & Computing*, vol. 22, no. 3, pp. 213 – 232, 2011.

[17] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz, "Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research," *Information Visualization*, vol. 7, no. 3, pp. 173–180, 2008.

[18] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 330–339.

[19] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2159–2168,2013.

[20] D. Guo, "Flow mapping and multivariate visualization of large spatial interaction data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1041–1048, 2009.

[21] N. Adrienko and G. Adrienko, "Spatial generalization and aggregation of massive movement data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 2, pp. 205–219, 2011.

[22] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan, "Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection," in *Proceedings of IEEE Pacific Visualization Symposium (PacificVis)*, 2011, pp. 163–170.

[23] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu, "Visualizing interchange patterns in massive movement data," *Computer Graphics Forum*, vol. 32, pp. 271–280, 2013.

[24] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra, "Circos: an information aesthetic for comparative genomics," *Genome research*, vol. 19, no. 9, pp. 1639–1645,2009.

[25] M. Lu, Z. Wang, and X. Yuan, "Trajrank: Exploring travel behaviouron a route by trajectory ranking," in *2015 IEEE Pacific Visualization Symposium (PacificVis)*, 2015, pp. 311–318.

[26] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. M. Ni, "Visual analysis of route diversity," in *Proceedings of IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2011, pp. 171–180.

[27] W. Zeng, C.-W. Fu, S. Arisona, A. Erath, and H. Qu, "Visualizing mobility of public transportation system," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1833–1842, 2014.

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

[29] M. Hadjieleftheriou, G. Kollios, P. Bakalov, and V. J. Tsotras, "Complex spatio-temporal pattern queries," in *Proceedings of the 31st international conference on Very large data bases*. VLDB Endowment, 2005, pp. 877–888.

[30] M. R. Vieira, P. Bakalov, and V. J. Tsotras, "Querying trajectories using flexible patterns," in *Proceedings of the 13th International Conference on Extending Database Technology*, ser. EDBT '10, 2010, pp. 406–417.

[31] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city cab trips," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2149–2158, 2013.

[32] R. Kr¨uger, D. Thom, M.W¨orner, H. Bosch, and T. Ertl, "Trajectorylenses - a set-based filtering and exploration technique for long-term trajectory data," *Computer Graphics Forum*, vol. 32, no. 3, pp. 451–460, 2013.

[33] M. O. Ward, G. Grinstein, and D. Keim, *Interactive data visualization: foundations, techniques, and applications*. CRC Press, 2010.

[34] A. Inselberg, "Multidimensional detective," in *Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97)*, 1997, pp. 100–107.

[35] E. R. Tufte, "Envisioning information." *Optometry & Vision Science*, vol. 68, no. 4, pp. 322–324, 1991.

[36] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit, "Lineup: Visual analysis of multi-attribute rankings," *IEEE Transactions on*

Visualization and Computer Graphics, vol. 19, no. 12, pp. 2277–2286, 2013.
[37] M. Batty, "Rank clocks," *Nature*, vol. 444, pp. 592–596, 2006.
[38] C. Shi, W. Cui, S. Liu, P. Xu, W. Chen, and H. Qu, "Rankexplorer: Visualization of ranking changes in large time series data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2669–2678,2012.

Author's Profile

Ms. G Asha Rani pursuing M. Tech in Computer Science and Engineering from PACE Institute Of Technology and Sciences affiliated to the Jawaharlal Nehru Technological University ,Kakinada.

Mr.S.Giri Babu is Dedicated to Teaching Field from the last 12Years. He has Guided 20 P.G Students and 30 U.G Students. At Present He is Working as Assoc.Professor in PACE Institute Of Technology and Sciences, Vallur, Prakasam(Dt), AP, India.He is Highly Passionate and Enthusiastic about his Teaching and Believes that Inspiring Students to Give of his best in order to Discover what he Already knows is better than Simply Teaching.