T-PickSeer: Visual Analysis of
Taxi Pick-up Point Selection Behavior

Shuxian Gu, Yemo Dai, Zezheng Feng, Yong Wang, and Haipeng Zeng
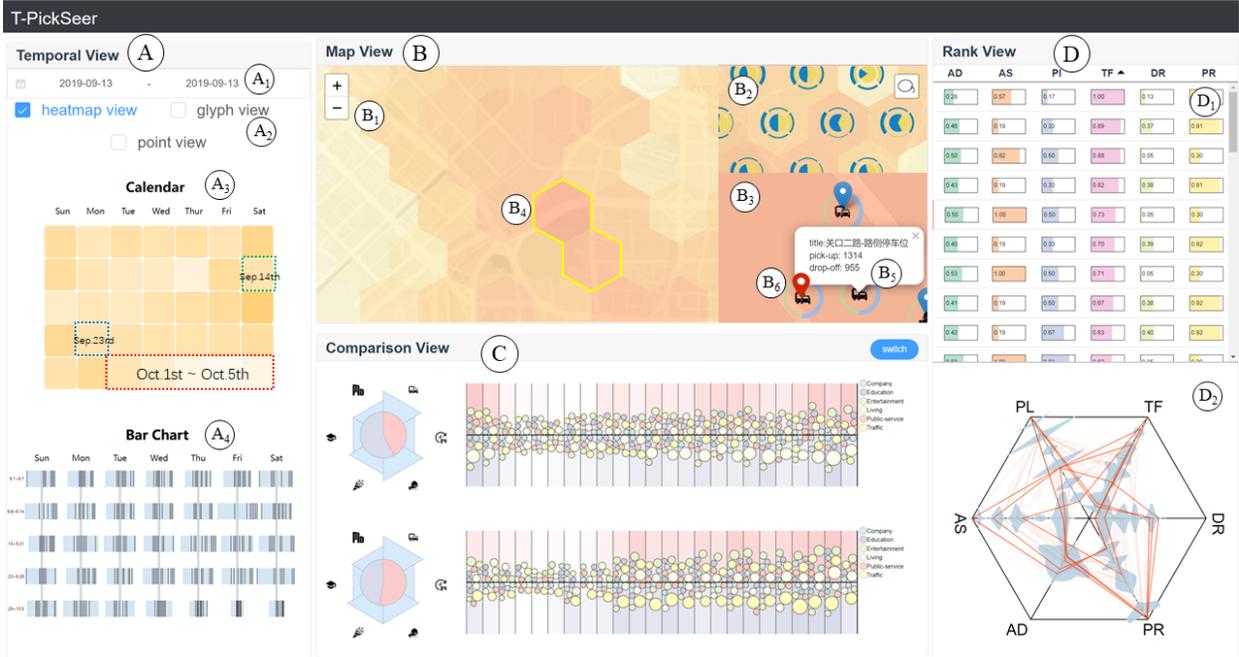
Fig. 1. T-PickSeer supports interactive exploration of pick-up point selection. (A) The temporal view comprises a time and view selection widget (A_1, A_2), a calendar heat map (A_3), and a multiscale temporal chart (A_4), enabling to configure the system and providing a temporal distribution of pick-up points. (B) The map view comprises a heatmap layer (B_1), a glyph layer (B_2), a POI layer (B_3) and provides operations for selecting regions and points (B_4, B_6), facilitating multi-scale exploration. (C) The comparison view reveals patterns between different selected regions. (D) The rank view displays the performance of different POIs. The POIs are ranked based on a series of defined criteria on D_1. D_2 further summarizes the performance of selected POIs.

Abstract— Taxi drivers often take much time to navigate the streets to look for passengers, which leads to high vacancy rates and wasted resources. Empty taxi cruising remains a big concern for taxi companies. Analyzing the pick-up point selection behavior can solve this problem effectively, providing suggestions for taxi management and dispatch. Many studies have been devoted to analyzing and recommending hot-spot regions of pick-up points, which can make it easier for drivers to pick up passengers. However, the selection of pick-up points is complex and affected by multiple factors, such as convenience and traffic management. Most existing approaches cannot produce satisfactory results in real-world applications because of the changing travel demands and the lack of interpretability. In this paper, we introduce a visual analytics system, T-PickSeer, for taxi company analysts to better explore and understand the pick-up point selection behavior of passengers. We explore massive taxi GPS data and employ an overview-to-detail approach to enable effective analysis of pick-up point selection. Our system provides coordinated views to compare different regularities and characteristics in different regions. Also, our system assists in identifying potential pick-up points and checking the performance of each pick-up point. Three case studies based on a real-world dataset and interviews with experts have demonstrated the effectiveness of our system.

Index Terms—Taxi travel behavior, pick-up point selection, visual analysis

◆

1 INTRODUCTION

• *Shuxian Gu and Haipeng Zeng are with School of Intelligent Systems Engineering, Sun Yat-sen University. E-mail: gushx3@mail2.sysu.edu.cn; zenghp5@mail.sysu.edu.cn.*
• *Yemo Dai is with Huawei Technologies Co., Ltd. E-mail: daiyemo@huawei.com.*
• *Zezheng Feng is with Department of Computer Science and Engineering, The Hong Kong University of Science and Technology. E-mail: zfengak@connect.ust.hk.*

With the rapid development of online taxi platforms (such as Uber and Didi), taxi has become a popular travel choice [37] and greatly

• *Yong Wang is with School of Computing and Information Systems, Singapore Management University. E-mail: yongwang@smu.edu.sg.*

improves the convenience of people's travel [33]. However, empty taxi cruising around the city causes a lot of energy waste every day, and taxi vacancy rate is high [23]. For example, taxi drivers often rely on personal experience to navigate the streets to look for passengers, which often takes much time though there are a large number of unfulfilled passenger orders in other areas [29]. To address this issue, some taxi companies offer online platforms for drivers to give advice on cruising. However, the information drivers can access is limited and has a lag, making the recommendation mechanism vague to them [32]. Taxi companies are in urgent need of more intelligent strategies to understand the needs of passengers, provide suggestions for drivers to cruise and improve the taxi service quality.

Analyzing the log data of pick-up point selection behavior of passengers can solve this problem effectively. First, understanding the spatio-temporal distribution of pick-up points helps taxi companies complete taxi scheduling. Through the discovery and comparative analysis of hot spots, taxi companies can issue dispatch instructions as soon as possible to effectively avoid drivers being trapped in congested traffic. Second, by analyzing the passenger travel characteristics at different pick-up points, taxi companies can provide suggestions for drivers to better plan cruise routes. Third, understanding the factors influencing passengers' pick-up point selection can help taxi companies adjust the settings of pick-up points. It can provide passengers with a comfortable ride experience. Therefore, we focus on identifying potential patterns of pick-up point selection behavior of passengers hidden in a large number of GPS records.

Various research has been done to analyze passengers' taxi travel patterns. Finding high-traffic areas in cities is a straightforward approach. Bi et al. [5] dug out the travel rules in hotspot regions. But the suggestions provided for cruising and dispatch are regional. More detailed information is lacking. With the availability of large-scale GPS data, mining the GPS data helps taxi companies gain more insights into the travel behaviors of crowds within cities. Ferreira et al. [13] developed a visual analytics system to help explore large-scale GPS data collected from taxis. But it focused more on the mobility characteristics and big data query, rather than the pick-up selection. In addition, some studies have directly explored the cruising characteristics of high-income drivers. Gao et al. [14] analyzed the cruising routes of high-income drivers. Yuan et al. [40] tried to find areas where high-revenue orders could be received. But they cannot be applied to solve our problem. Different from routes or areas analysis, our analysis is multi-scale and focuses on pick-up point selection behavior. Little research has been done on analyzing the pick-up point selection by passengers.

However, it is a challenging task to explore the pick-up point selection behavior of passengers due to three major reasons. **(1) Large-scale and multi-attribute spatio-temporal data.** The large scale of GPS data and the spatio-temporal attributes increase the difficulty of data analysis. **(2) Dynamic urban transportation.** Urban transportation is dynamic and it is not easy to extract movement patterns from large-scale data, which makes it difficult to identify pick-up point preferences. **(3) Multiple influencing factors.** Passengers' selection of pick-up point is influenced by many factors, such as travel purpose, traffic conditions [46], convenience [38] and so on. It leads to complicated passenger preferences. Due to these challenges, a fully automated analysis of pick-up point selection is difficult. Visual analysis, combined with both advanced computational power and human cognitive abilities, can be an effective solution for analyzing pick-up point selection behavior.

To address the above problems, we propose a visual analytics system to handle large volumes of time-varying traffic data, aiming at visual exploration of passengers' behavior for pick-up point selection in different regions. Analysts at taxi companies can use our system to obtain a series of useful discoveries, which can assist in the taxi operation management and dispatch guidance. Specifically, we adopt a hierarchical exploration approach comprising city, region, and point scales. We combine GPS data with POI data to provide insight into pick-up point analysis. A novel comparison view is designed to facilitate the comparative analysis of different regions. In summary, the primary contributions of this paper are as follows:

- We propose an interactive visual analytics system for analysts

in taxi companies to analyze passengers' pick-up point selection behavior at multiple scales (city, region, and point), featuring pattern exploration and potential pick-up points discovery.

- A novel design is developed for easily comparing different pick-up point selection patterns. An augmented beeswarm graph is adopted to show numerous trips and corresponding POI information. Further, a stacked bar chart option is provided for better comparison of large-scale pick-up point selection data.

- Three case studies using a real-world dataset, together with expert interviews, are conducted to evaluate the effectiveness of T-PickSeer in empowering interactive exploration of passengers' pick-up point selection behavior.

2 RELATED WORK

This section reviews related research on the analysis of pick-up selection (Section 2.1) and visual analytics for traffic data (Section 2.2).

2.1 Analysis of Pick-up Selection

The study of pick-up selection of passengers is beneficial to the management of taxi operation and future urban planning, which is of great significance to the urban structure, policy making, resource allocation, and so on. These studies can be mainly divided into two parts: pattern analysis and potential pick-up exploration.

The studies of pattern analysis are mainly committed to finding behavior patterns for pick-up selection and analyzing influencing factors. Traditional methods (e.g., statistics and regression) have been proved effective and many useful conclusions have been drawn [15]. But these suggestions are general and have trouble dealing with dynamic spatio-temporal data. More detailed information is desired. Then some research turns to the discovery and analysis of hotspots. Bi et al. [5] dug out the travel rules of different hotspots. However, further analysis of the differences between the hotspots is lacking.

Apart from these pattern analyses, some algorithms for finding and recommending potential pick-up locations are proposed. Various improved clustering algorithms are another commonly used technology. Berdeddouch et al. [4] utilized technologies of K-Means and regression to discover potential pick-up locations that are easier to find passengers. Zhang et al. [43] improved spatio-temporal clustering based on K-Means and tried to find a set of personalized pick-up locations taking drivers' preferences into consideration by combining pick-up data and POI's attributions. Xu et al. [38] considered the information on taxi trajectory and improved the algorithm based on HotSpotScan and Preference Trajectory Scan algorithms. These methods take into account a variety of factors to recommend pick-up points. However, a deeper analysis of the importance of different factors is necessary. Machine learning methods, such as deep learning [18], reinforcement learning [39], DeepFM [34] also play an important role in potential location discovery. However, the model's output and access to information are limited, which confuses the user about the results. Although the above research has achieved good performance, these studies lack transparency.

In summary, existing studies are difficult for analysts who lack mathematical domain expertise to analyze. In addition, the current methods usually generate pick-up hotspots by various algorithms while further analysis of the differences between the hotspots is lacking, which also needs interpretation. It is difficult for them to intuitively display the temporal and spatial variation of the pick-up point. In this paper, we focus on the visual analytics of the taxi pick-up data and try to find different patterns of pick-up selection, which help with the setting and recommendation of the pick-up point. We designed multi-scale analysis flow and interactive operations to analyze pick-up point selection, which can help users to conduct multidimensional analysis and judgment themselves.

2.2 Visual Analytics of Traffic Data

Thanks to the development of fruitful advanced location-sensing technologies for collocating a vast amount of traffic data, the status of the

moving objects, is recorded from both spatial and temporal dimensions [12]. The analyses of these data have made many contributions to solving urban problems, such as bus route planning [35], and taxi operation [20, 47]. However, due to the complexity of urban problems and the multidimensionality of traffic data [11], some of these methods may not perform well without the involvement of domain experts. The combination of data visualization and urban computing methods enables experts to explore traffic data interactively [9].

Data query focuses on developing new visual query models to quickly query traffic data and explore traffic information. Representative works include [2, 13, 19, 31]. Filtering, sampling, aggregation, etc. combined with interaction [6] can quickly present the results the user wants. What's more, the main task that most studies focus on is pattern mining, committed to enabling analysts to obtain patterns and insights from big data [22]. Zeng et al. [41] explored the relationship between human mobility and POIs. Deng et al. [8] studied the cascades of spatial contagions. At last, different from pattern mining, some studies have tried to take advantage of the information extracted from pattern exploration for decision making. Weng et al. [35] proposed a visual analytics system to generate optimal transit routes interactively. Liu et al. [21] solved the problem of comparing solutions rapidly for billboard placements. Visualization helps to obtain useful information and interaction is allowed to participate in the process of generating decisions, making the system perform better than the algorithms.

The success of visual analytics is inseparable from the help of three factors: the proper visual representation of data, good comparison design, and human interaction. Firstly, various proper and novel visualization designs of spatio-temporal data have been proposed to facilitate the efficient completion of data analysis tasks. Visualization of spatial properties is often map-based [13, 22], such as heat maps. Rendering and aggregation are effective solutions to visual clutter [11, 44]. For temporal properties, axis-based design is a common visual form [13]. ThemeRiver [17] and horizon map [28] can compare multiple properties over time. What's more, calendar [27] and radial layout [2, 45] can be a good option for periodicity, which can provide a distinct comparison of different periods. Additionally, a good comparison design is essential to grasp the difference. Juxtaposition is the most commonly used method for comparison [21, 26, 35]. Superposition overlaps multiple objects to show the difference [13]. Explicit encoding presents differences visually. For instance, the grid heat map encoded by size and color in [3] showed the pattern similarity between different time periods. At last, interaction operation helps with combining human knowledge. Basic interactions like clicking and boxing can promote exploration from overview to detail, such as the flow matrix view in [35], which can be clicked for more information. What's more, interactions such as sketch [3] can make exploration more interesting, enabling people to interact with systems.

Although these studies have been proven effective, they cannot directly used for our visual analytics tasks. For the taxi data used in this paper, relatively little research has been done on pattern analysis of passengers' behavior for pick-up selection. It is still a challenge to explore enormous pick-up points. Problems such as visual redundancy and poor scalability are easy to appear. Therefore, we combined the multiple visualization techniques and designed a novel visual analytics system, which enables users to interactively explore pick-up selection behavior at multiple scales. In particular, a juxtaposition comparison view was designed to interact with plenty of pick-up points and gathering locations, depicting different patterns in different regions.

3 DATA AND ANALYTICAL TASKS

This section first describes the data processing procedures and the derived output. Next, we derive a list of analytical tasks by working with domain experts.

3.1 Data Description

In this research, T-PickSeer is constructed from three types of data. i) We utilize the road network of Shenzhen and geographical data from

Table 1. The Metadata of Each OD Trip Record.

Field	Description
ID	The id of the taxi
stime	Trip start time
slocation	Location of pick-up point
etime	Trip end time
elocation	Location of drop-off point
cO	Category of the nearest POI of the O point
cD	Category of the nearest POI of the D point

the open street map[1]. ii) We retrieve Points of Interest (POIs) data from the Baidu Map Service[2]. The dataset includes 190362 POI locations where each record contains the longitude, latitude, name, address, and functionality of a structure in the urban environment. iii) We use the taxi GPS data from September 1, 2019 to October 5, 2019 in Shenzhen city. The raw data records the location of each taxi with a total size of over 200 GB. Each record consists of a timestamp, taxi id, longitude, latitude, speed, travel direction, and occupancy status.

3.2 Data Processing

- **POI classification:** The raw POI data is categorized into 20 industry categories and 158 detail categories. To better focus on the POI information related to our scenario, we reclassified POI data into six categories: company, education (i.e., schools and universities), entertainment (i.e., restaurants, shopping malls, and tourism), living, public service (i.e., government agency) and traffic(i.e., station and parking lot).

- **Trip extraction:** Taxi GPS data records a sequence of key points that a taxi passed by, which is unclear for the pick-up point (Origin) and drop-off point (Destination) of a single trip. So we need to extract each OD trip from raw GPS data. It consists of three steps: i) Data cleaning. Some records that are out of the study area or occupancy status changes abnormally, are eliminated at first. ii) Pick-up and drop-off points identification. It can be extracted according to the change in occupancy status. It means that someone gets off when the taxi's status changes from 1 to 0, the opposite means someone getting on. iii) Attribute addition. To facilitate the analysis of pick-up point selection with surrounding POIs, we attach the category attribute of the nearest POI to each OD point. On average, about two million trips are extracted each day. The metadata of each trip record is shown in Table 1.

- **Geographical partition:** In order to facilitate pattern exploration, we divide the study area into equal-sized grids. Firstly we consider the shape of the grid. Hexagon is chosen because it is recommended as a better alternative to be used as a statistical unit [25]. The main advantages are i) that hexagons are the only geometric shape for regular tessellations that shares a real border with every neighbor and not only a single point with some neighbors [1] and ii) hexagons have minimal visual ambiguity, and have a positive effect on the memory performance of object location [10]. Then we tested the grid at different resolutions (200 m, 400 m and 1 km), finding that 400-meter hexagons provide representative samples which can guarantee the details of the pick-up point distribution and is in line with the passenger's willingness to walk the distance. After partition, we assign taxi movements to each hexagon based on their original location.

3.3 Task Analysis

To develop a feasible and practical approach for analyzing and improving pick-up points with visual analytics, we work closely with our collaborative experts (E1, E2, E3) to derive the analytical tasks. E1 is a Ph.D. candidate specializing in urban visualization and is also

---

[1]https://www.openstreetmap.org/

[2]https://api.map.baidu.com/lbsapi/getpoint/index.html
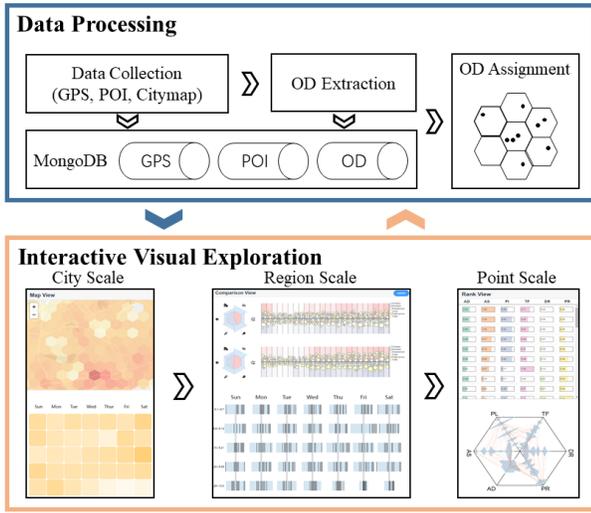
Fig. 2. A visualization system pipeline for multi-scale analysis of behavior for pick-up point selection. Our system consists of two phases: data processing and interactive visual exploration. In the data processing phase, we perform well-established methods to process data and index them spatially in the database. In the interactive visual exploration phase, four coordinated views are provided to support three-scale exploration.

one of the co-authors. E2 is an analyst in a taxi company with over 10 years of working experience and has been involved in several taxi management projects. E3 is a researcher who has long been engaged in urban computing and visual analytics. From the feedback of these experts, we summarize a set of system requirements as shown below:

T1 Obtain an overview of global traffic distribution in a city. The users need to grasp the spatio-temporal distribution of pick-up points over the city. A summary of daily traffic volumes provides information for time checking (T.1.1). Then the distribution of pick-ups in different regions helps to understand busy regions for further exploration (T.1.2).

T2 Provide the spatio-temporal patterns for regional traffic. After understanding the global distribution of pick-up points, users need regional analysis. Multiscale temporal pattern at the hourly, daily, and weekly periods is noteworthy (T.2.1). Besides, it is necessary to explore the traffic within a region (T.2.2), such as the traffic volumes of pick-ups and drop-offs, and the direction of traffic. E2 points out that drop-offs are also important information, meaning that there are potential passengers subsequently or more drivers are here. These help the driver make decisions about cruising.

T3 Compare the behavior of pick-up point selection with different regions. Visual comparison of situations across different regions should be supported. It is necessary to compare different locations diversely to find distribution similarities and differences, facilitating the discovery of preference patterns for pick-up point selection.

T4 Show detailed information for point analysis. In a focused region, users expect to explore the preference of passengers in choosing different pick-up points. A set of criteria is necessary to explore and compare the performance of different pick-up points.

4 SYSTEM OVERVIEW

T-PickSeer is a web-based visual analytics application constituted of two phases, namely, data processing and interactive visual exploration, as illustrated in Fig. 2. In the data processing phase, T-PickSeer processes datasets offline and stores them in the MongoDB database. We extract OD trips from raw GPS data and index them spatially in the
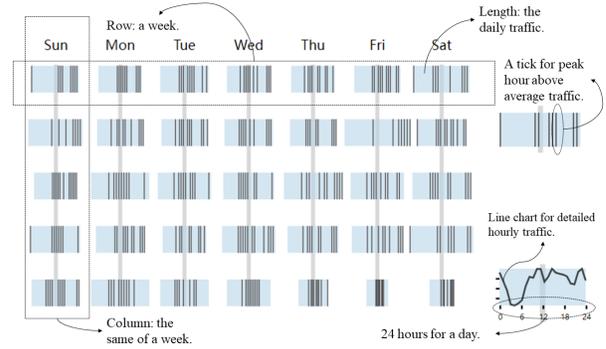
Fig. 3. The design of the multiscale temporal chart, shows temporal patterns at different granularities. Each row represents a week, and each column represents the same day of the week. It supports traffic volume comparison weekly, daily, and hourly.

database. Then we divide the study area into grids and assign OD trips to each grid based on location.

The interactive visual exploration phase consists of three stages of multi-scale exploration. We organize the interface by the city-, region-, and point-scale analyses. Starting from the city scale, a heatmap in map view (Fig. 1B$_1$) and a calendar chart (Fig. 1A$_3$) are designed to provide a spatio-temporal overview of city traffic distribution (**T1**). Users can have a basic understanding about data. Then narrowing down the exploration to region scale, we design a multiscale temporal chart for regional exploration (**T2**) and a comparison view (Fig. 1C) to compare region patterns in different situations (**T3**). At last, users can analyze the preference for pick-up point selection at point scale. A rank view (Fig. 1D) is provided to compare different points and summarize their performance (**T4**). The visualization modules are implemented in D3.js[3] and Leaflet.js[4] for different rendering requirements, and they are integrated using Vue.js[5] framework.

5 VISUALIZATION

In this section, we provide a detailed description of the visualization designs in our system.

5.1 Temporal View

The temporal view (Fig. 1A) is designed to configure the system and provide multi-scale temporal information (**T1, T2**). At the top (Fig. 1A$_1$), a date selector is provided for users to select traffic on a particular day they are interested in. Then a view selector (Fig. 1A$_2$) is provided to configure the visibility of different views flexibly through a set of checkboxes. To obtain an overview of global traffic distribution over time for date selection (**T.1.1**), a calendar heatmap (Fig. 1A$_3$) is designed to summarize the daily traffic volume of the whole city. Each rectangle represents a day and the color is encoded as the volume of the daily traffic. The darker rectangle indicates the larger traffic volume in a day.

At the bottom of the temporal view (Fig. 1A$_4$), we design a multi-scale temporal pattern chart to better reveal the temporal distribution of pick-up points in the focused region (**T.2.1**). It shows changes of traffic at different time granularities. Fig. 3 shows the design details. The layout of blue rectangles is similar to the calendar heatmap, with each row representing a week, and each column representing the same day of the week. For each rectangle, we use length to show the daily traffic volume. Furthermore, the hourly pattern is also worthy of attention. Users are concerned about changes in hourly traffic. First, we only highlight the peak hours with higher traffic than the average per day. We split each rectangle's length into 24 equal pieces to represent the 24 hours in a day. We encode these hours with ticks and placed them in each rectangle from left (00:00) to right (24:00). For better comparing

---

[3]https://d3js.org/
[4]https://leafletjs.com/
[5]https://vuejs.org/
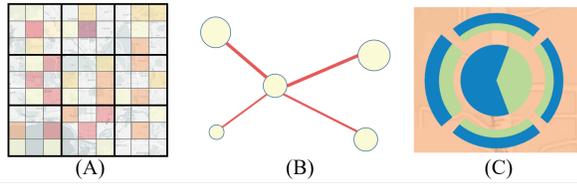
Fig. 4. Designs of the glyph view. (A) and (B) are the design alternatives of the glyph view, which may introduce visual clutter problems. (C) is the final adopted design.

and analyzing, the rectangles in each column are centered, with dashed gray lines indicating 12:00. Then E2 suggested that he would also like to know the details of how traffic flows change within one day, which ticks cannot show. So we provide a choice to check traffic changes hourly in a day with a line chart, which can be viewed by clicking on the rectangle. The graph is initialized to the traffic volume for the entire city and then changes with the selected region.

5.2 Map View

The map view (Fig. 1B) provides a spatial overview of pick-ups (**T.1.2**) and supporting regional exploration (**T.2.2**).

Heatmep view. Traffic in a large-scale city typically comprises massive trips. The heatmap (Fig. 1B$_1$) is developed to assist with mastering the global distribution of OD trips quickly and identifying regions of interest (**T.1.2**). It encodes the density of pick-ups at each grid. The darker, the higher number of pick-up points. From the heatmap, users can easily find crowds from a coarse-grained perspective.

Glyph view. To reveal traffic pattern in a focused region (**T.2.2**), we attach a glyph view (Fig. 1B$_2$) to the heatmap. As shown in Fig. 4 C, the pie chart inside represents the comparison between the pick-ups (blue) and drop-offs (green) of each grid. Two outer rings visualize the drop-offs (green) and pick-ups (blue) in the grid by the geographical directions, with thickness presenting the flow size. Users can select regions with the lasso tool (Fig. 1B$_4$).

Design alternatives. We considered several alternative solutions during our glyph design process. We first tried the OD matrix [36] to illustrate the movement of OD trips. As shown in Fig. 4A, we tried to mesh the city map and map the original grids of the entire city into every single grid. The color was used to encode the volume of each grid area to other grids. But it has several drawbacks. First, it encodes the relative position of the destination, so it is difficult to fix the position of the destination intuitively. Second, the dense grid and colors make it hard to spot patterns. Moreover, we tried the flow map [30] as shown in Fig. 4B. However, it introduced a visual clutter problem when there were too many grids. So in the end we chose the glyph design (Fig. 4C), which was also recognized by the experts. It can effectively display the pick-up and drop-off information and flow direction in different grids.

Point view. To assist visual linkage and smoother operations, we present a point view (Fig. 1B$_3$) to provide a spatial context for point analysis. Different types of POIs are marked on the map with various icons and a donut graph is placed around the POI to summarize the nearby pick-ups (blue) and drop-offs (green) for taxis. If users click the icon, the detailed information for the POI will be displayed (Fig. 1B$_5$).

5.3 Comparison View

To compare the behavior of pick-up point selection in different regions (**T3**), the comparison view (Fig. 1C) coordinates two graphs for two regions at the same time.

Fig. 5 shows the design details. For each region, a glyph on the left (Fig. 5A) summarizes the information for the region. The middle pie chart displays the proportion of all-day pick-ups (pink) and drop-offs (purple) in the region. The outer arc bars represent different categories of POI with height encoding the number. Afterward, users expect to explore the preference of pick-up selection, such as the preferred period and locations. Inspired by the 1-D beeswarm graph[6], we design a novel 2-D visual metaphor to represent the pattern in a region (Fig. 5B). As

[6]https://observablehq.com/@fil/experimental-plot-beeswarm
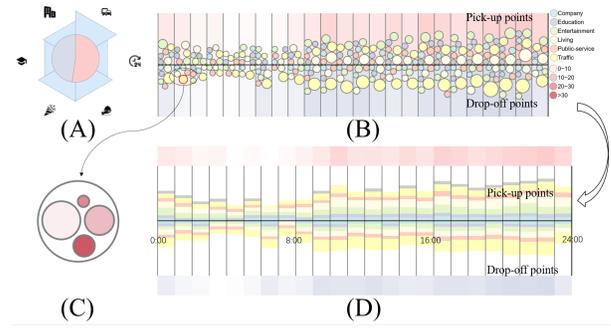
Fig. 5. Design details for the comparison view. A glyph shows the comparison of pick-ups and drop-offs and the overall distribution of POIs (A). Two kinds of options are provided to switch for details (B, D). A circle pack displays the information of trip duration (C).

shown in Fig. 5B, the x-axis represents time, and the y-axis represents the pick-up or drop-off points. The background represents the pick-up (top) and drop-off (bottom) hourly. Then we count the number of pick-ups and drop-offs by an hour and aggregate them according to their POI attributes, trying to present the time-varying flow and spatial properties of the pick-up point. These points for each hour are encoded into circles with different colors indicating different POI attributes and placed in the corresponding grids. The area of the circle indicates the quantity. Furthermore, more detailed information about the corresponding trips is needed. Experts point out that drivers are concerned about the travel duration of an order. Therefore, a circle pack (Fig. 5C) is provided to summarize the travel duration by clicking on each circle. We divide trips into four categories according to the time spent on the trip: 0-10 minutes, 10-20 minutes, 20-30 minutes, and more than 30 minutes. Each circle represents a category, with the area of the circle indicating the number of trips. The darker, the longer time spent on the trip. Then force simulation is used to keep all the circles tightly connected without overlapping for readability. In this way, the beeswarm graph can provide an overview of the temporal distribution and also enable detailed local inspection. In addition, the number of circles is sometimes large, affecting the efficiency of the comparison analysis. With this in mind, we provide checkboxes on the right to support point filtering by POI attribute, facilitating focus on points of interest.

The new beeswarm graph we design can effectively display the spatio-temporal information of the region and help to interact with a large number of trips, facilitating pattern discovery. However, according to the feedback from E1, when the selected region is too large, the circles overflow occurs. So we provide the option of a stacked bar chart (Fig. 5D) for users to switch. Its layout is similar to the beeswarm graph, which can more clearly show the changes over time. The original background is pushed to the top and bottom.

Design alternatives. For the beeswarm graph, we have considered aggregating data by the taxi ID to present hourly information, hoping to alleviate large-scale OD data. As shown in Fig. 6, we use each circle to represent a taxi, with the area of the circle representing the number of orders and the color showing the average duration. However, the visual presentation is confusing due to the still large number of taxis. It is difficult for users to effectively interact with circles. So we assign the POI attribute to each pick-up point according to the nearest POI. Then we aggregate the pick-up points according to the POI attributes, and distribute the pick-up points and the drop-off points separately on the positive and negative axes of y, which effectively solves this problem. At the same time, it provides an opportunity to explore the spatial characteristics of the pick-up points.

5.4 Rank View

After obtaining the patterns of the region, point-scale exploration (**T4**) is necessary to compare the performance of each pick-up point and explore the preference for passengers in different pick-up points. We design the rank view (Fig. 1D) to compare and analyze pick-up points,
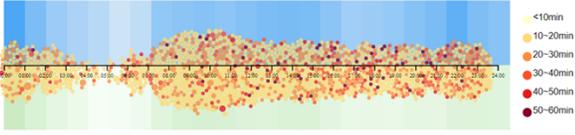
Fig. 6. Alternative design for the comparison view, which is difficult for users to interact with circles.

(A)   (B)

(C)   (D)

Fig. 7. (A) and (B) show the selected regions for pattern comparison. (C) and (D) are the corresponding temporal patterns over a month in region A and B. The traffic in region A is busy during the day while the traffic in region B is busy at night.

which can help with driver cruise guidance. In reality, POIs are usually recommended as pick-up points because they are easy to find. Considering the large number of historical pick-ups, as well as the habit of passengers taking a taxi, we use POI to replace historical pick-ups for analysis. Users can also add points on the map as a supplement. They are collectively called candidate points.

Firstly, a set of criteria is required to evaluate the performance of these points. According to literature research [7, 38, 46] and expert opinions, we propose a set of criteria. Our criteria are as follows. For each candidate point i: i) AD: Accessibility is assessed by the average distance from nearby historical points within coverage to that point. D_{ij} in Formula (1) means the distance from the historical point j to i. D is the radius of the coverage area. n is the total number of historical points. Finally, candidate points with better accessibility have higher scores. ii) AS: Average traffic speed in the vicinity represents the traffic smoothness near the candidate point. A high score indicates a safe and comfortable ride experience. iii) PL: POI level is the proportion of nearby POI categories in all, indicating the convenience of taking a taxi. iv) TF: Transfer convenience, impacting passenger source, means the proportion of transportation facilities in all POIs. v) PR and vi) DR evaluate the probability of passenger arrival and driver discovery respectively. In Formula (2), NP_i and ND_i mean the number of historical pick-up points and empty taxis. Then we calculate the quantity per unit length and unit time. Finally, we normalize all scores to [0, 1].

$$AD_i = \sum_{0<j<n} (1 - D_{ij}/D)/n \qquad (1)$$

$$PR_i = NP_i/L/T, DR_i = ND_i/L/T \qquad (2)$$

Afterward, a rank list (Fig. 1D$_1$) with six evaluation criteria is designed for comparing candidate points. The list presents each candidate point in a row and arranges its criteria in six columns. We encode the scores by the length of bars and provide a click action on the header of each column to rank the points by different criteria. Finally a radar graph (Fig. 1D$_2$) is placed under the list to provide a summary depiction of all candidate points. Each axis from the center to the outer edge represents a score from 0 to 1. The scores for each candidate point are connected by light orange lines. And the violin plots colored in blue are attached to each axis, showing the score distribution of all candidate points for that criterion, which emphasizes the pick-up preference. Users can click points of interest on the rank list. Then the scores of the corresponding points will be highlighted with thicker orange lines in the radar graph (Fig. 1D$_2$).

5.5 User Interactions

Rich interactions are provided for users to explore pick-up point data, which are summarized as follows.

- **Multi-scale navigation** helps users navigate effectively across different scales. In the temporal view, users can set date (Fig. 1A$_1$) and configure the visibility of views flexibly through a set of checkboxes (Fig. 1A$_2$). In the map view, users can select regions with the lasso tool (Fig. 1B$_4$) or set a point by clicking on the map (Fig. 1B$_6$).

- **Highlighting** enables users to focus on the information of interest, which is supported in the map view and rank view. For example, the selected region and point will be highlighted in the map view (Fig. 1B$_4$, B$_6$) and the rank view (Fig. 1D$_2$).
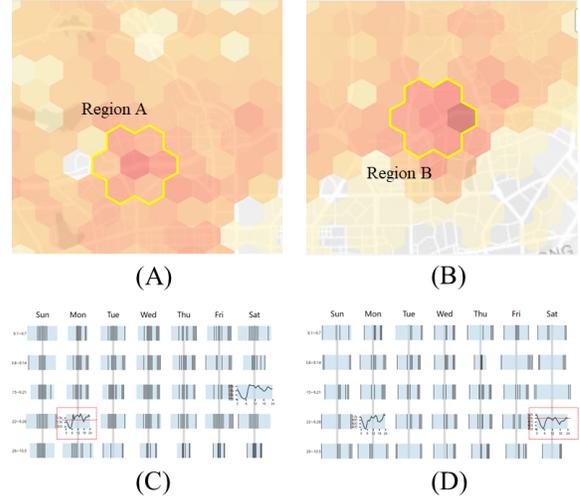
- **Linking** connects four views in the system. For example, after setting points on the map, new points will be added at the end of the rank list. When users click on a row in the rank list, it can be visually linked to the location on the map and the score on the radar chart.

6 EVALUATION

In this section, we demonstrate the effectiveness and usability of T-PickSeer to accomplish the visualization tasks in Section 3.3 and discover insights through three case studies and expert interviews with the aforementioned collaborating experts (E1, E2 and E3, who have been introduced in Section 3.3). The dataset used in these three case studies is a 5-week taxi GPS dataset in Shenzhen, from September 1st to October 5th, 2019.

6.1 Case study

6.1.1 Taxi pick-up behavior comparison across different regions

As an expert in urban visualization, E1 was interested in using our system to explore the pick-up point selection patterns in different regions of the city. After loading data to T-PickSeer, the temporal overview of pick-ups across the city was displayed in the calendar heatmap (Fig. 1A$_3$) first. E1 observed darker rectangles on weekends than weekdays, indicating higher traffic volume on weekends.

He chose a Monday (September 23, highlighted with a blue rectangle in Fig. 1A$_3$) and immediately found two hotspot regions (Fig. 7A, B) on the map view. After checking, E1 found that they were both in the downtown area, with massive traffic. After selecting these regions with the lasso tool, he observed the changes in the multiscale temporal chart (Fig. 7C, D). For region A, Fig. 7C showed that the ticks were concentrated in the middle, indicating more peak hours in the daytime. While the ticks in Fig. 7D were distributed on both sides, meaning more traffic in the evening in region B. Afterwards he examined the comparison view (Fig. 8) to compare two regions. The left glyphs for two regions (Fig. 8A$_1$, B$_1$) were similar, showing that they had similar distributions of POIs with roughly equal pick-ups and drop-offs. In the right beeswarm graphs for two regions, he found that the various circles above the x-axis (Fig. 8A$_3$, B$_3$) were similar in size. It was hard to for E1 to find passengers' preference in specific POI categories when they choose the pick-up point. However, for the circles below the x-axis (Fig. 8A$_4$, B$_4$), the white and green circles were significantly larger than others, indicating the number of drop-offs near "Living" and "Entertainment" was increased. "Living" and "Entertainment" POIs were the preferred drop-off points. E1 thought that the drop-off points could strongly reflect the purpose of the trip, which may be the pick-up
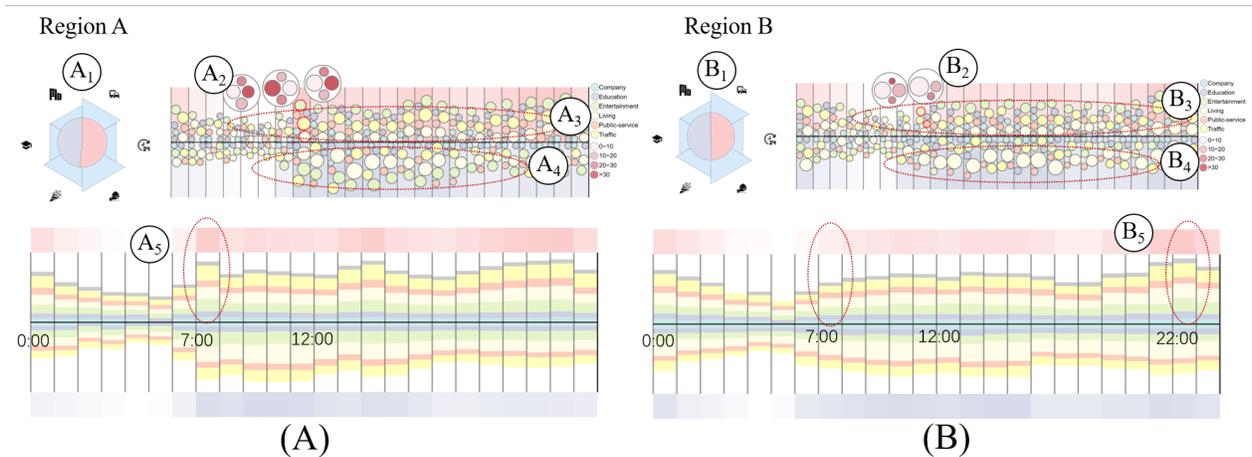
Fig. 8. Pattern comparison between two selected regions. (A) shows two presentations of region A on a weekday. Region A has more long-duration trips in the morning and more trips for entertainment. (B) provides two graphs of region B on a weekday. Region B has obvious characteristics of the evening peak.

point of the next trip. "It can provide useful suggestions for driver cruise," said E1.

Next, he switched to the stacked bar chart (Fig. 8A₅, B₅) to explore time variation of traffic. This time he focused on pick-ups above the x-axis. He observed the higher bar and darker rectangle at 7:00 AM in Fig. 8A₅, showing a more obvious peak hour at 7:00 AM than region B. He felt interested in the rush hour at 7:00 AM. To further explore trips at 7:00 AM, he switched back to the beeswarm graph and checked interested circles. Comparing Fig. 8A₂ and Fig. 8B₂, the dark red circles of circle packs in Fig. 8A₂ were larger than others, meaning more long-duration trips (>30min) at 7:00 AM in region A than B. In contrast, short-duration travel accounted for the main part in region B (Fig. 8B₂). Learned from E1, there were many people need to travel a long distance to go to work in region A. Although there were plenty of pick-up points in both areas, "region A has a huge demand for picking up in the morning, and longer trips bring higher incomes for drivers," said E1.

The above research confirmed that T-PickSeer can effectively help to explore pick-up selection patterns in different regions. E1 said, "It is effective to support exploration at multiple scales, which helps a lot for taxi dispatch."

6.1.2 Taxi pick-up point selection in the living region

As an analyst in a taxi company, E2 works on taxi dispatch. He was interested in the characteristics of passengers in selecting pick-up points, wondering if there are any regularities and preferences.

E2 checked the calendar heatmap and selected a normal weekday (September 23) for exploration. He quickly discovered the dark red regions with high traffic volume in Shenzhen. For the large amount of taxi demand, E2 was eager to understand the preference of pick-up point selection to provide detailed suggestions for taxi dispatch. He selected the darkest grid (Fig. 9A), which contained a large living quarter. After clicking the "point view" in Fig. 1A₁, the map view immediately zoomed to the point scale and the rank view ranked all POIs automatically. At E2's suggestion, we set D in Formula (1) to be 500 meters to measure the pick-up situation near POI. He checked the surrounding conditions of POI in the map view firstly. As shown in Fig. 9B, the blue rings at the intersection were obviously thicker than those around the living quarter. It indicated that the pick-ups were concentrated at the further intersection, where there were abundant POIs. As he had thought there would be more pick-ups near the living quarter, he was surprised by the results. With doubt, he continued to inspect the rank view, hoping for an explanation.

As shown in Fig. 9C, the violin plots (blue) of PL, TF, and DR were distributed on the outer edges of the radar chart. In summary, most of the pick-ups in the region had high scores and good performance.
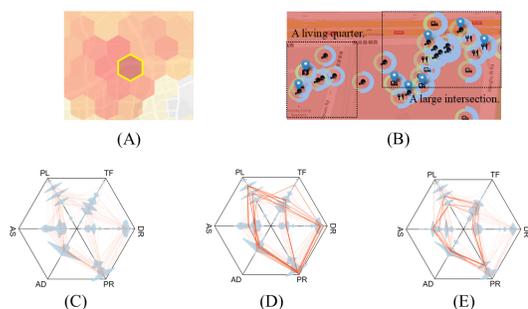
Fig. 9. Exploration for preference of pick-up point selection in a living quarter. (A) shows the selected region in Luohu District. (B) shows that there are more pick-ups at the larger intersection. (C) displays the overall performance of all candidate points. (D) and (E) display the performance of selected points. Passengers prefer to pick up at POI-rich places.

Afterward, he ranked all points by PR score (passenger arrival rate). He clicked the rank list and checked points with high PR scores, with corresponding scores highlighted in the radar chart with thicker orange lines (Fig. 9D). These points had high PL, TF and DR scores, showing that most passengers preferred to wait for taxis in areas with rich POIs and convenient transportation. E2 said, "Passengers think that the probability of taking a taxi is higher here. In fact, there are indeed more taxis passing by here." However, the lower AS (average speed) and AD (average distance) scores meant that it was easy to get congested and passengers had to walk for a long distance. Therefore, E2 ranked the points according to the AD scores from high to low and clicked on points with high AD scores for further exploration. The scores of these points were highlighted in Fig. 9E. He found that some of these points had lower DR and PL scores, indicating fewer empty taxis and POIs nearby. The probability of taking a taxi at the nearby points was lower. Therefore, "passengers have to walk a certain distance to the further intersection to take a taxi, which brings a bad ride experience," said E2. It resulted that the pick-up points were clustered at the further intersection rather than the living quarter. He suggested, "to improve passengers' riding experience, it is better to recommend drivers to cruise around the living quarter, which also benefits traffic management."

E2 praised, "T-PickSeer is helpful for my work. It helps to identify abnormal areas quickly and provide information to make decisions."

6.1.3 Pick-up point selection around tourist spots

E3 was interested in the problem of taking a taxi in scenic spots. Therefore, he selected a holiday (September 13, highlighted with a green rectangle in Fig. 1A₃) and found a tourist region in the west of the
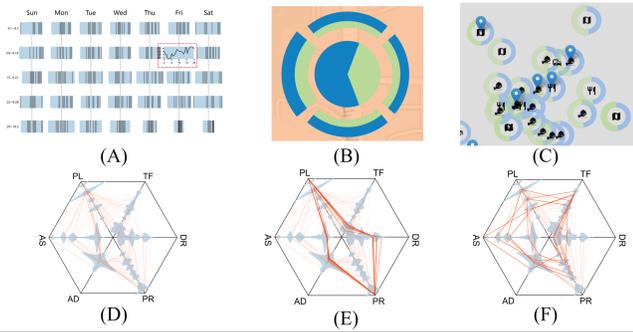
Fig. 10. Exploration for preference of pick-up point selection in a tourist region. (A) shows the change of traffic volume in the region over a month. Rush hours are concentrated in the afternoon and evening. (B) indicates that more traffic comes from and goes to the western regions. (C) shows that there are more drop-off points than pick-ups. (D) displays the score distribution of all points. (E) and (F) shows scores of selected points. There are fewer pick-ups in places with high TF score because there are lots of parking lots and passengers don't like to take taxis here.

city (Fig. 1B₄), with many parks, gymnasiums, and amusement parks. He first inspected the regional pattern. As shown in Fig. 10A, pick-up points started to gradually increase in the afternoon. In Fig. 10B, the rings on the left are thicker than the others, indicating that there were more trips from and to the western regions, where there were many apartments and universities. "There may be potential passengers there." he said.

Then he switched to the point view for further exploration. He found that the green rings are thicker than blue ones (Fig. 10C), indicating more drop-offs than pick-ups around POIs in this region. Then he examined the radar chart (Fig. 10D) and found the violin plot of DR (driver discovery rate) was lower than PR (passenger arrival rate), meaning the number of taxis cannot meet the demand of passengers. To further explore the demand of passengers, he checked points with high PR scores, hoping to find passengers' preferences for pick-up selection. As highlighted in thick orange lines (Fig. 10E), he observed that the PL (POI level) scores of these points were high while TF (transportation facility level) scores were low. It indicated that passengers preferred pick-up points where there were many POIs but few transportation facilities. To find out the reason, he then ranked all points according to the TF score from high to low and clicked on them one by one. As shown in Fig. 10F, some points with high TF had low DR scores, meaning there were few empty taxis near these points. Afterward, he examined the corresponding points on the map view and found that many of them were parking spots (Fig. 1B₅). E3 guessed that self-driving tours may account for a certain proportion. This is the reason why drivers are unwilling to cruise here. But in fact, there is still unmet demand for taxis. For unmet passengers' needs, it was inappropriate to recommend remote locations to passengers. "It is better to create a good cruising environment for taxi drivers. The nearest POIs in the tourist spot are easier to find and have the shortest walking distance for passengers." E3 suggested.

In summary, E3 found it interesting to explore preferences of pick-up point selection with our system, T-PickSeer. He could easily identify factors that passengers are concerned about and find current deficiencies. He said, "T-PickSeer can be of great help in providing suggestions for better taxi service."

6.2 Expert Interview

We interviewed three aforementioned experts (E1, E2, and E3) individually and collected their feedback. Each interview lasted about an hour. First, a fifteen-minute tutorial was provided to participants, which outlined the functions, visual designs, and interactions of our system. Then, participants were allowed to freely explore provided data with our system for about forty minutes. After that, we collected participants' feedback and suggestions. The feedback from the experts was valuable based on their expertise, which is summarized as follows.

Visual designs. The experts confirmed that our system is well-designed and could be easily understood by users with different backgrounds. In particular, E1 praised the comparison view. " It is of great help to discover the spatial property characteristics of pick-up points," said E1. E3 favors the multiscale temporal chart. He said, " It provides abundant temporal patterns at different granularity, which is helpful."

Usability & Effectiveness. All three experts agreed that our proposed method is helpful and effective to analyze the pick-up points selection of passengers. E1 mentioned, "This approach supports multi-scale exploration including the city, region, and point scale. The hierarchical analysis inspires my research, and the comparison view does really help in comparing patterns of selected regions. E2 praised that T-PickSeer is helpful to make use of large amounts of GPS data collected from taxis. "T-PickSeer is an effective tool that can provide an intuitive visualization, this improves the decision interpretation. Compared to the current tools, it allows me to easily discover more detailed information to guide drivers on the cruise," he commented. E3 said that "T-PickSeer provides visualization designs which are also easy to follow. It is friendly to the analysts lacking professional skills in statistics and machine learning methods."

Suggestions. During the interview, three experts also provided fruitful suggestions on improving T-PickSeer. Both E1 and E2 mentioned that the region segmentation function in the map view still needs more improvement. In the current version, T-PickSeer uses a hexagonal grid to segment the urban area based on the latitude and longitude, E1 and E2 suggested that it would be better to consider the spatial contextual information during the region segmentation. For example, they hope that a building or a unit (park, campus) can be assigned to a grid, which makes sense. Instead of a unit being divided into several parts, each part is in a separate grid. E3 said that it would be more effective if T-PickSeer integrated more recommendation approaches like mainstream deep learning models. In addition, he suggested that T-PickSeer could compare these different recommendation methods and give the corresponding optimal method under different scenarios.

7 Discussion

In this section, we discuss lessons learned, the implications, and identify the limitations of our system and propose directions for future studies.

Lesson learned. We have learned three lessons. First, visual analytics is helpful to analyze pick-up point selection behavior in such big taxi data. Generally, research mainly focuses on regional analysis or pick-up point recommendation. In this paper, we apply visual analytics to analyze passengers' behavior for pick-up point selection at multiple scales. By working with domain experts, we find visual analytics is especially helpful for comparing and reasoning different behavior of pick-up point selection. We can easily find some parts interesting or abnormal, and conduct further exploration. Second, to analyze such big spatial-temporal taxi data, a three-scale exploration strategy is effective. We start with the city scale, and then compare different regions, and further explore different pick-up points, which is smooth and helpful for exploration. Third, extensive communication with end users is important in determining the system's requirements and implementing the visualization forms. They provide useful insights that make our systems more responsive to the needs of real-world applications. Experts subscribe to the hierarchical analysis approach and help us refine our tasks. For instance, E2 states how the analysis of drop-off point is essential for assisting the driver cruise when exploring the pick-up point selection behavior. In addition, users are unfamiliar with visual analytics. We can only implement concise visualization designs and effective tools by communicating frequently.

Implications. This study presents a visual analytics system, assisting analysts in taxi companies to explore passengers' pick-up point selection behavior. In terms of taxi companies, our system helps taxi companies regularly analyze data for better decision-making. The usage scenarios show several abnormal areas with high taxi demand or poor travel experience, which can provide insights to analysts and guide the scheduling of taxis. For example, it can coordinate the relationship between drivers and passengers and improve the income of drivers by reasonable incentives to guide drivers' cruises. The company needs to

improve the environment of taxi riding, i.e., setting up riding areas or pick-up points at appropriate locations, especially in places with dense traffic such as scenic spots and stations. In terms of drivers, there are several suggestions for them to cruise. First, plan cruise routes according to peak times in dense traffic areas. It helps prevent missing crowds and avoid congestion on the way. Second, focus on the surrounding POI types. For example, passengers in some areas prefer "Living" and "Entertainment" POI.

Partition of the urban area. As mentioned from E1 and E2, in T-PickSeer we partition the urban area equally into hexagons regardless of the contextual information. This is enough for the current method to some extent. However, the current segmentation methods may not be effective, if we want to integrate more mainstream deep learning methods in the future. Furthermore, the previous literature [42] mentioned that traffic aggregations might depend on the shapes and scales of the spatial partition units, for example, the MAUP [24] [16]. Therefore, further exploration between the partition of the area and the result for recommendation is still our future task.

Analysis of temporal pattern. E3 pointed out that in the multiscale temporal chart, the ticks at the same time cannot be aligned as the rectangle length represents both flow and time. This puts a burden on the analysis of temporal patterns. Figuring out how to align the ticks will be our future work.

Scalability. T-PickSeer is designed for analyzing the pick-up point selection behavior of passengers. With a large number of taxi GPS data, it is easy to cause scalability issues. For better in-depth exploration, we have adopted a three-scale exploration strategy, namely city, region, and point scales. While the map design and the beeswarm graph will show visual clutters when we explore the information of a large area. We have provided another option, the stacked bar chart, to ease visual clutters. To better handle the scalability issues, we plan to explore more designs and adopt some automatic methods to filter unnecessary information.

Generality. Although we focus on analyzing taxi data, our analytical pipeline and strategy can be easily adopted to other similar spatial-temporal data, such as bus data, and telco data. The design for comparison can also be applied to other applications for comparing temporal data.

Evaluation. Our system, T-PickSeer, is currently evaluated with only three expert users. To better evaluate the usability and effectiveness of our system, a long-term study with more domain experts are needed, which is left for future work.

8 CONCLUSION

In this paper, we propose T-PickSeer, an interactive visual analytics system that helps users visually explore passengers' behavior of pick-up point selection. Several well-designed visualizations and interaction techniques are combined to facilitate multi-scale exploration at city, region and point scales. Coordinated contrast views are provided to compare different patterns in different regions. We propose a set of criteria for examining the performance of each pick-up point. In the end, we demonstrate the effectiveness of our system with three case studies on a real-world taxi dataset in Shenzhen and interviews with three domain experts. The results show that our system is useful to explore the pick-up points selection behavior of passengers and provide guidance for empty taxi cruising. In the future, we will integrate recommendation algorithms for better suggestions for users.

REFERENCES

[1] J. Adamczyk and D. Tiede. Zonalmetrics-a python toolbox for zonal landscape structure analysis. *Computers & Geosciences*, 99:91–99, 2017.

[2] S. Al-Dohuki, Y. Wu, F. Kamw, J. Yang, X. Li, Y. Zhao, X. Ye, W. Chen, C. Ma, and F. Wang. Semantictraj: A new approach to interacting with massive taxi trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):11–20, 2016. doi: 10.1109/TVCG.2016.2598416

[3] S. AL-Dohuki, Y. Zhao, F. Kamw, J. Yang, X. Ye, and W. Chen. Qutevis: Visually studying transportation patterns using multisketch query of joint traffic situations. *IEEE Computer Graphics and Applications*, 41(2):35–48, 2021. doi: 10.1109/MCG.2019.2911230

[4] A. Berdeddouch, A. Yahyaouy, Y. Bennani, and R. Verde. Recommender system for most relevant k pick-up points. In *Proceedings of the International Conference on Artificial Intelligence & Industrial Applications*, pp. 277–289. Springer, 2020. doi: 10.1007/978-3-030-51186-9_20

[5] S. Bi, Y. Sheng, W. He, J. Fan, and R. Xu. Analysis of travel hot spots of taxi passengers based on community detection. *Journal of Advanced Transportation*, 2021, 2021. doi: 10.1155/2021/6646768

[6] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan, and R. Maciejewski. Vaud: A visual analysis approach for exploring spatio-temporal urban data. *IEEE Transactions on Visualization and Computer Graphics*, 24(9):2636–2648, 2018. doi: 10.1109/TVCG.2017.2758362

[7] Y. Chen, Q. Fu, and J. Zhu. Finding next high-quality passenger based on spatio-temporal big data. In *Proceedings of the International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pp. 447–452. IEEE, 2020.

[8] Z. Deng, D. Weng, Y. Liang, J. Bao, Y. Zheng, T. Schreck, M. Xu, and Y. Wu. Visual cascade analytics of large-scale spatiotemporal data. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2486–2499, 2021.

[9] Z. Deng, D. Weng, S. Liu, Y. Tian, M. Xu, and Y. Wu. A survey of urban visual analytics: Advances and future directions. *Computational Visual Media*, 9(1):3–39, 2023.

[10] D. Edler, J. Keil, A.-K. Bestgen, L. Kuchinke, and F. Dickmann. Hexagonal map grids–an experimental study on the performance in memory of object locations. *Cartography and Geographic Information Science*, 46(5):401–411, 2019.

[11] Z. Feng, H. Li, W. Zeng, S.-H. Yang, and H. Qu. Topology density map for urban data visualization and analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):828–838, 2020. doi: 10.1109/TVCG.2020.3030469

[12] Z. Feng, H. Qu, S.-H. Yang, Y. Ding, and J. Song. A survey of visual analytics in urban area. *Expert Systems*, p. e13065, 2022. doi: 10.1111/exsy.13065

[13] 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 taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, 2013. doi: 10.1109/TVCG.2013.226

[14] Y. Gao, P. Xu, L. Lu, H. Liu, S. Liu, and H. Qu. Visualization of taxi drivers' income and mobility intelligence. In *Proceedings of the Advances in Visual Computing*, pp. 275–284. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-33191-6_27

[15] W. Ge, D. Shao, M. Xue, H. Zhu, and J. Cheng. Urban taxi ridership analysis in the emerging metropolis: Case study in shanghai. *Transportation Research Procedia*, 25:4916–4927, 2017. doi: 10.1016/j.trpro.2017.05.368

[16] C. E. Gehlke and K. Biehl. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, 29(185A):169–170, 1934.

[17] 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 *2011 IEEE Pacific Visualization Symposium*, pp. 163–170. IEEE, 2011.

[18] Z. Huang, G. Shan, J. Cheng, and J. Sun. Trec: an efficient recommendation system for hunting passengers with deep neural networks. *Neural Computing and Applications*, 31(1):209–222, 2019. doi: 10.1007/s00521-018-3728-2

[19] Z. Huang, Y. Zhao, W. Chen, S. Gao, K. Yu, W. Xu, M. Tang, M. Zhu, and M. Xu. A natural-language-based visual query approach of uncertain human trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1256–1266, 2020. doi: 10.1109/TVCG.2019.2934671

[20] W. Jiang and L. Zhang. The impact of the transportation network companies on the taxi industry: Evidence from beijing's gps taxi trajectory data. *IEEE Access*, 6:12438–12450, 2018. doi: 10.1109/ACCESS.2018.2810140

[21] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):1–10, 2016. doi: 10.1109/TVCG.2016.2598432

[22] M. Lu, J. Liang, Z. Wang, and X. Yuan. Exploring od patterns of interested region based on taxi trajectories. *Journal of Visualization*, 19(4):811–821, 2016. doi: 10.1007/s12650-016-0357-7

[23] B. Mu and M. Dai. Recommend taxi pick-up hotspots based on density-based clustering. In *Proceedings of the International Conference on Computer and Communication Engineering Technology (CCET)*, pp. 176–181. IEEE, 2019. doi: 10.1109/CCET48361.2019.8989132

[24] S. Openshaw. The modifiable areal unit problem, catmog 38. In *Geo Abstracts, Norwich*, 1984.

[25] R. Rempel and A. Carr. Patch analyst extension for arcview: version 3. *Available on line at: http://flash. lakeheadu. ca/~ rrempel/patch/index. html*, 2003.

[26] Q. Shen, W. Zeng, Y. Ye, S. M. Arisona, S. Schubiger, R. Burkhard, and H. Qu. Streetvizor: Visual exploration of human-scale urban forms based on street views. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):1004–1013, 2018. doi: 10.1109/TVCG.2017.2744159

[27] C. Silva and M. Saraee. Predicting road traffic accident severity using decision trees and time-series calendar heatmaps. In *Proceedings of the IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET)*, pp. 99–104. IEEE, 2019. doi: 10.1109/CSUDET47057.2019.9214709

[28] A. Suh, M. Hajij, B. Wang, C. Scheidegger, and P. Rosen. Persistent homology guided force-directed graph layouts. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):697–707, 2019. doi: 10.1109/TVCG.2019.2934802

[29] L. Tang, F. Sun, Z. Kan, C. Ren, and L. Cheng. Uncovering distribution patterns of high performance taxis from big trace data. *ISPRS International Journal of Geo-Information*, 6(5):134, 2017. doi: 10.3390/ijgi6050134

[30] T. Von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):11–20, 2015.

[31] F. Wang, W. Chen, F. Wu, Y. Zhao, H. Hong, T. Gu, L. Wang, R. Liang, and H. Bao. A visual reasoning approach for data-driven transport assessment on urban roads. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 103–112. IEEE, 2014. doi: 10.1109/VAST.2014.7042486

[32] R. Wang, C.-Y. Chow, Y. Lyu, V. C. Lee, S. Kwong, Y. Li, and J. Zeng. Taxirec: Recommending road clusters to taxi drivers using ranking-based extreme learning machines. In *Proceedings of the SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–4, 2015.

[33] T. Wang, Y. Zhang, M. Li, and L. Liu. How do passengers with different using frequencies choose between traditional taxi service and online car-hailing service? a case study of nanjing, china. *Sustainability*, 11(23):6561, 2019. doi: 10.3390/su11236561

[34] X. Wang, Y. Liu, Z. Liao, and Y. Zhao. Deepfm-based taxi pick-up area

[35] D. Weng, C. Zheng, Z. Deng, M. Ma, J. Bao, Y. Zheng, M. Xu, and Y. Wu. Towards better bus networks: a visual analytics approach. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):817–827, 2020. doi: 10.1109/TVCG.2020.3030458

[36] J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with od maps. *The Cartographic Journal*, 47(2):117–129, 2010.

[37] Z. Xiong, J. Li, and H. Wu. Understanding operation patterns of urban online ride-hailing services: A case study of xiamen. *Transport Policy*, 101:100–118, 2021.

[38] X. Xu, J. Zhou, Y. Liu, Z. Xu, and X. Zhao. Taxi-rs: Taxi-hunting recommendation system based on taxi gps data. *IEEE Transactions on Intelligent Transportation Systems*, 16(4):1716–1727, 2015. doi: 10.1109/TITS.2014.2371815

[39] Y. Yang, X. Wang, Y. Xu, and Q. Huang. Multiagent reinforcement learning-based taxi predispatching model to balance taxi supply and demand. *Journal of Advanced Transportation*, 2020, 2020. doi: 10.1155/2020/8674512

[40] C. Yuan, X. Geng, and X. Mao. Taxi high-income region recommendation and spatial correlation analysis. *IEEE Access*, 8:139529–139545, 2020. doi: 10.1109/TKDE.2017.2772907

[41] W. Zeng, C.-W. Fu, S. M. Arisona, S. Schubiger, R. Burkhard, and K.-L. Ma. Visualizing the relationship between human mobility and points of interest. *IEEE Transactions on Intelligent Transportation Systems*, 18(8):2271–2284, 2017.

[42] W. Zeng, C. Lin, J. Lin, J. Jiang, J. Xia, C. Turkay, and W. Chen. Revisiting the modifiable areal unit problem in deep traffic prediction with visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):839–848, 2020.

[43] M. Zhang, J. Liu, Y. Liu, Z. Hu, and L. Yi. Recommending pick-up points for taxi-drivers based on spatio-temporal clustering. In *Proceedings of the International Conference on Cloud and Green Computing*, pp. 67–72. IEEE, 2012. doi: 10.1109/CGC.2012.34

[44] H. Zhou, P. Xu, X. Yuan, and H. Qu. Edge bundling in information visualization. *Tsinghua Science and Technology*, 18(2):145–156, 2013. doi: 10.1109/TST.2013.6509098

[45] Z. Zhou, J. Yu, Z. Guo, and Y. Liu. Visual exploration of urban functions via spatio-temporal taxi od data. *Journal of Visual Languages & Computing*, 48:169–177, 2018. doi: 10.1109/TVCG.2013.226

[46] W. Zhu, J. Lu, Y. Li, and Y. Yang. A pick-up points recommendation system for ridesourcing service. *Sustainability*, 11(4):1097, 2019. doi: 10.3390/su11041097

[47] F. Zong, T. Wu, and H. Jia. Taxi drivers' cruising patterns—insights from taxi gps traces. *IEEE Transactions on Intelligent Transportation Systems*, 20(2):571–582, 2019. doi: 10.1109/TITS.2018.2816938

recommendation. In *Proceedings of the International Conference on Pattern Recognition*, pp. 407–421. Springer, 2021. doi: 10.1007/978-3-030-68821-9_36