2019 IEEE 35th International Conference on Data Engineering (ICDE)

Vaite: a Visualization-Assisted Interactive Big Urban Trajectory Data Exploration System

Chuang Yang*§‡, Yilan Zhang†‡, Bo Tang*§, Min Zhu†
*Department of Computer Science and Engineering, Southern University of Science and Technology
§PCL Research Center of Networks and Communications, Peng Cheng Laboratory
*{11510080@mail., tangb3@}sustc.edu.cn
†College of Computer Science, Sichuan University
†{zhangyilan@stu., zhumin@}scu.edu.cn

***Abstract*—Big urban trajectory exploration extracts insights from trajectories. It has many smart-city applications, e.g., traffic jam detection, taxi movement pattern analysis. The challenges of big urban trajectory data exploration are: (i) the data analysts probably do not have experience or knowledge on issuing their analysis tasks by SQL-like queries or analysis operations accurately; and (ii) big urban trajectory data is naturally complex, e.g., unpredictability, interrelation, etc. In this work, we architect and implement a visualization-assisted big urban trajectory data exploration system (Vaite) to address these chanllenges. Vaite includes three layers, from data collection to results visualization. We devise novel visualization views in Vaite to support interactive big urban trajectory exploratory analysis. We demonstrate the effectiveness of Vaite by the real world applications.**

***Index Terms*—trajectory data exploration; data visualization;**

I. Introduction

With the development of tracking techniques (i.e., GPS), the movement trajectories of urban entities (e.g., taxi, bus, passenger) are collected automatically and become available recently. Big urban trajectory data analysis plays an important role in many smart city applications. For example, traffic congestion prediction [1], finding significant places [2], [3], movement pattern analysis [4], [5]. As the larger volume and higher velocity of generated urban trajectories [6], data analysts are building exploration-driven applications [7], [8]. For interactive data exploration, data analysts will pose ad-hoc visualization queries for analyzing data or gleaning insights [9]. The ad-hoc queries do not always have the same analysis procedure. For example, profiling the situation of the central road in a city is dramatically different than clustering the traffic jam events in the city. In addition, the analysts probably do not have experience or knowledge to issue proper specific analytical queries.

Consider a data analyst looking for potential issues (e.g., traffic congestion, abnormal driving trajectories) in the streaming urban trajectories of a city. In this situation, the data analyst does not know how to issue her analytical queries to find interesting results exactly, and there are not clear analysis procedures in her mind for an analysis task. The complex nature of urban trajectories makes exploratory urban trajectories analysis even more intractable. In this work, we proposed Vaite, a visualization-assisted interactive big urban trajectory data exploration system, for the urban trajectory data exploratory analysis applications.

Vaite is built upon Spark, which is designed for big data analytical applications inherently. It consists of three layers: i) trajectories preprocessing layer, which is the data preprocessing layer, collecting urban trajectories, cleaning the data errors (e.g., value missing, wrong data records) and integrating with the underlying road map, ii) exploratory analytical layer, including a set of atomic analysis operators and several exploratory analysis models (e.g., correlation analysis, community detection), and iii) interactive visualization layer, visualizing the exploratory analysis result by Vaite visualization views (e.g., spatial bubble view, trajectory-map projection view).

Our demonstration will let ICDE participants explore the insights hidden in big urban trajectories, issue ad-hoc trajectory analysis queries, interactively perform aggregation analysis operations on the dataset (e.g., find the most popular taxi stop places, identify the traffic congestions in the city), and adaptively employ visualization views for different analysis results. Once an ad-hoc exploratory analysis query is issued, its visualization results could navigate users to discover other relevant insights.

In Section II, we present the architecture of Vaite and discuss the core layers in detail. In Section III, we discuss the novel visualization views in our system. We conclude with two case studies that illustrate the effectiveness of our system in Section IV, and a discussion of the demonstration proposal in Section V.

II. Vaite System

The Vaite system framework is illustrated in Figure 1. It includes (i) The trajectories preprocessing layer (at the bottom); (ii) The exploratory analytical layer (at the middle) is the core layer. It builds upon Spark[1]; (iii) The interactive visualization layer (at the top) is the front-end of Vaite system. It visualizes the analytical results to the users. A user inputs his/her analytical actions in this layer and Vaite processes the input analytical task interactively.

‡ indicates equal contributions

[1] https://spark.apache.org

2375-026X/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDE.2019.00233

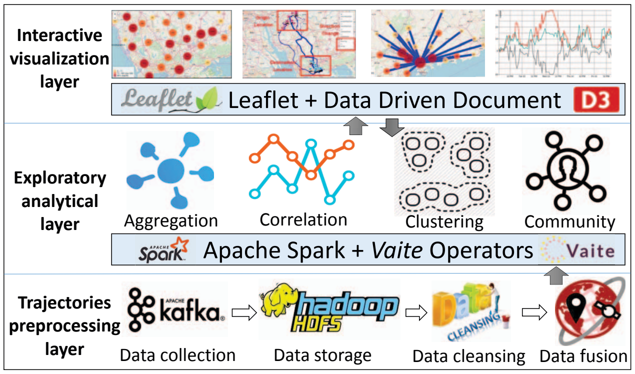

Fig. 1. Vaite system framework

A. Exploratory Analytical Layer

Exploratory analytical layer is the core layer of Vaite. It consists of two phases: (I) Computation engine. It includes a set of atomic urban trajectory analytical operators, e.g., trajectory event extractor, sub-trajectory segmentation operator. (II) Exploration engine. It is a set of exploration analytical models, e.g., aggregation analysis, entities clustering, which takes inputs from computation engine and returns the analytical results to the interactive visualization layer. We have suggested several atomic operators and analytical models in Vaite. Nevertheless, it is extensible as follows.

Atomic Operator: Atomic operators are the basic units in computation engine. The extensibility of atomic operators is customization. For instance, some users want to analyze the low-speed car driving, which leads to road congestion. The Vaite provided stop event extractor can be customized to low-speed event extractor by setting the speed of moving object less than 5km/h and the movement distance larger than 100 meters.

Analytical Model: Analytical model takes inputs from the computation engine and returns analytical outputs to the visualization layer. Consider the above example, after extracting all low-speed events, Vaite incurs clustering model to analyze the low-speed events by spatial and temporal clusters. The clustering result will be visualized in the visualization layer. Hence users gain insights from the visualized low-speed event result, i.e., the government could identify the busiest road through the result. Data analysts may propose their "analytical model" with their domain knowledge, e.g., movement pattern analysis model.

B. Interactive Visualization Layer

It is the user interface of Vaite system. It visualizes the analytical results from the middle layer to the end user and provides interactive analysis interfaces. It takes users' interactive analytical actions, e.g., mouse movement, keyboard input as inputs and parses these analytical actions to the exploration analytical layer. Users can also specify the analytical target in visualization figures, e.g., selecting the largest community in stop events communities to conduct further analysis. It includes a suite of customized visualization views for trajectory exploratory analysis tasks, which will be elaborated in Section III.

III. Visualization Design

In this section, we discuss the novel visualization views in Vaite for interactive visualization-assisted urban trajectory exploratory analysis. For each view, we first describe its visualization requirements, then elaborate the visualized result, and conclude with its application scenarios.

A. Spatial Bubble View

Many trajectory analysis applications consider a set of movements or events in the constraint places, for example, traffic jam detection [1], and significant place identification [2]. We devise a spatial-aware event aggregation view (Spatial Bubble View) for that kind of analysis tasks, as shown in Figure 2(a). Comparing with bubble view in Leaflet [2], our Spatial Bubble View have several improvements (e.g., generalized aggregation operator) to support complex trajectory exploratory analysis applications.

We demonstrate the usage of bubble views in Figure 2(a). The stop events of one-week taxi trajectories in Shenzhen, one of the largest cities in China, are visualized in Figure 2(a-1). The color and size of each bubble show the number of stop events in it. For instance, there are 2454 stop events in Guangzhou this week, and only 602 stop events in Huizhou. In order to gain fast insights about the stop events overview, the zoom-out result shows in Figure 2(a-2). Consider a specific bubble, i.e., Huizhou in Figure 2(a-3). It can be further zoom-in and reveal the details of stop-event clusters in it (cf. Figure 2(a-4)). Spatial bubble view can be widely used in many advanced and complex exploratory analysis applications, e.g., identifying significant places, finding the taxi communities.

B. Customized Standard Views

In this section, we present a suite of customized standard views in Vaite for visualization-assisted trajectory exploratory analysis.

Star topology view: It is a traditional visualization tool to illustrate one-to-multiple objects' relationship. In Vaite, we adopt it for the analysis of the correlation among places/areas/bubbles, e.g., the relationship among different traffic jam road segments. Users can customize their correlation models and visualize them by star topology view in Vaite. For example, Figure 2(b-1) illustrates the relationship among traffic congestion bubbles.

Pixel-based view: Figure 2(b-2) is a pixel-based view of traffic jam events, i.e., the taxis had a relative slow speed (less than 5km/h) at Dongmen Middle Road. The pixel-based view includes 7×24 grids or pixels, with various degree of color to indicate the number of aggregated events in each time period. Through this view, data analyst could identify the outlier or

[2]https://leafletjs.com/

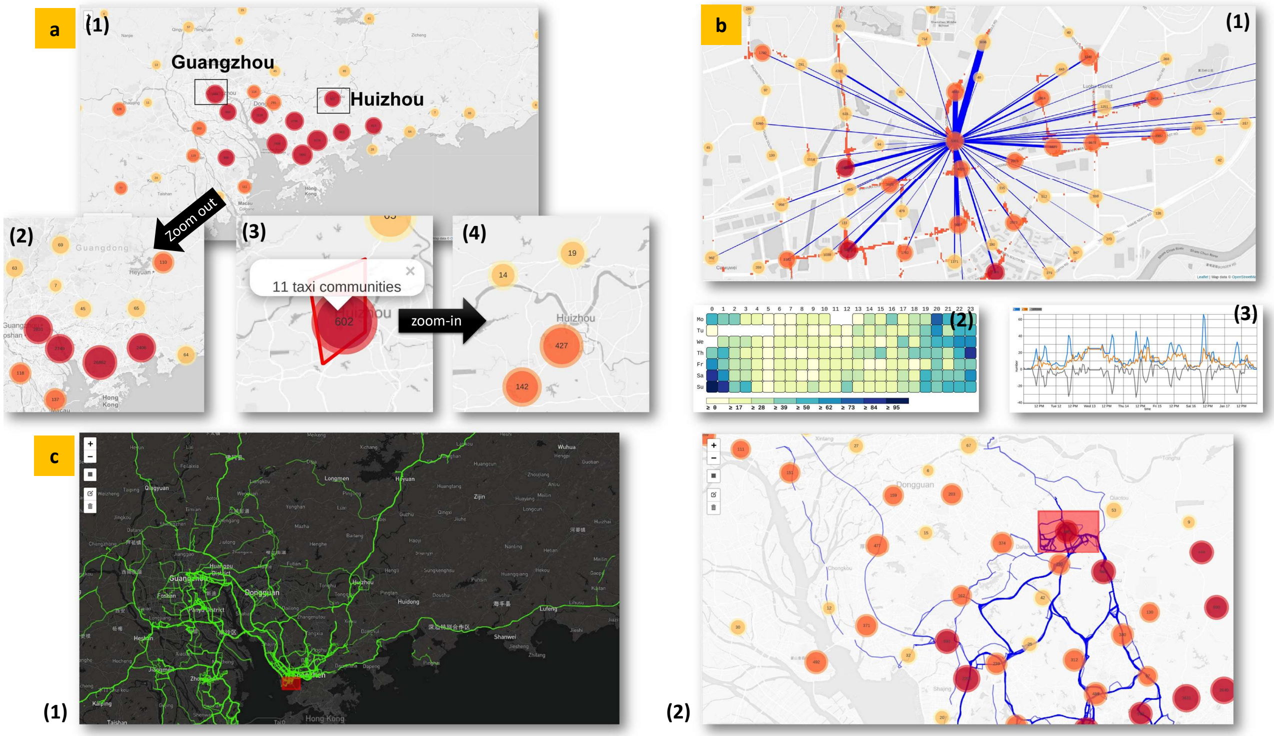

Fig. 2. Visualization views in Vaite

discover trends among these traffic jam events, e.g., it indicates that 20:00-2:00 is a peak period in Dongmen Middle Road.

Time series view: Time series view is widely used in many applications to show the trend of an object over time. we integrate this view into our system and improve it to analyze time series. For example, we visualize the vehicle flow over time on a road segment of Guangshen Yanjiang Expressway in Figure 2(b-3). Clearly, the maximum vehicle flow occurs at 9:00 every day.

C. Trajectory-Map Projection View

Last but not least, we present a novel trajectory visualization view, trajectory-map projection view, which can be used by many exploratory analysis tasks. For example, it can be employed for route recommendation by revealing the most popular road between two different places. Figure 2(c-1) visualized all the vehicle trajectories starting from Shekou (cf. red polygon in Figure 2(c-1)) in Shenzhen. The biggest challenge of trajectories visualization is how to improve the response time for visualizing massive trajectories in the road map. Visualizing massive scatter points is very time-consuming, e.g., the industry standard Tableau visualization system takes over 4 minutes on a high-end server to generate a scatter plot for a 50M-tuple dataset which is already resident in memory, as elaborated in [9]. We applied a standard uniform sampling technique to address it.

To explore more insights from the trajectories, we can apply the trajectory-map projection view over spatial bubble view. For example, we can specify one bubble (as the red polygon in Figure 2(c-2)) as the original place of trajectories. The trajectories start from it and end at other bubbles are shown in Figure 2(c-2). It shows the exact taxi movement trips among all these bubbles. Such kinds of information assist users to discover the correlations among two different places via the visualized taxi movement trips.

In summary, we provide a suite of novel visualization views for interactive trajectory analysis in top-layer of Vaite.

IV. Case Studies

In this section, we demonstrate the effectiveness of Vaite system by two real-world trajctory exploratory analysis applications. We used one week urban trajectories in Shenzhen, China. The raw trajectory data size is almost 7 GB per day, the total data size is 40.3 GB after passing the trajectory preprocessing layer in Vaite.

Identify taxi communities: In order to explore the interesting facts among taxi drivers, the data analyst in a taxi company probably wants to know where the drivers wait for passengers. The analyst extracts the stop events, i.e., the location of a given taxi does not change within 5 minutes, by incurring event extractor among all taxi trajectories. The stop events are visualized by heatmap, as shown in Figure 3(a). Obviously, the more stop events, the deeper color in the map. The analysts

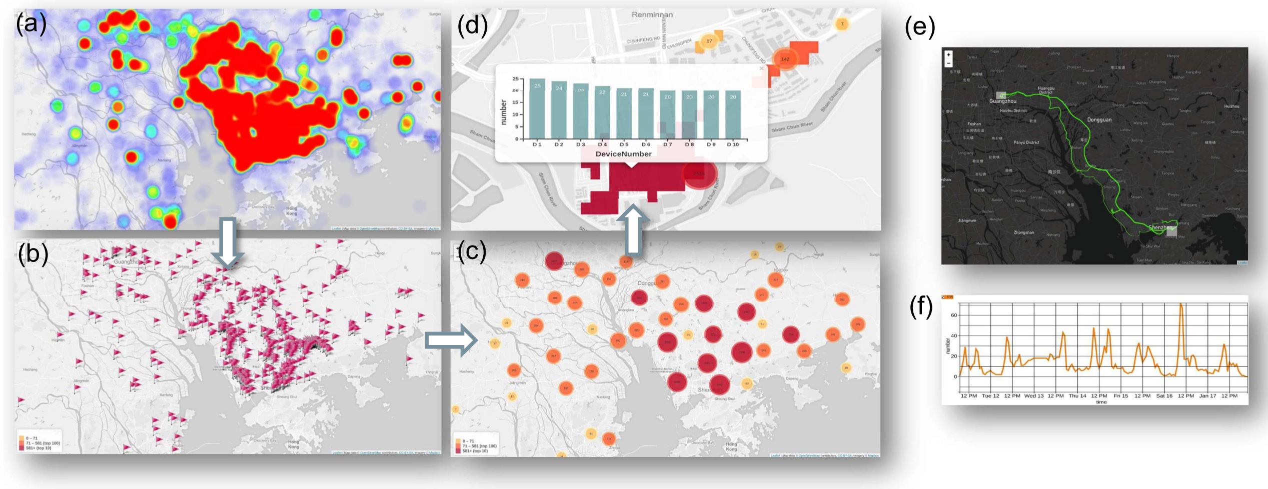

Fig. 3. Case studies in Vaite

adjust the values of parameters for stop events interactively, and use heatmap to visualize the corresponding results.

In order to further explore the facts among these stop events, he/she could apply a clustering operator, i.e., incurs a clustering algorithm to analyze them. These clusters could further scale up or down and the aggregated results are shown by spatial bubble view in Figure 3(c). The bubble size reflects the size of the taxi driver communities, the larger bubble size, the more stop events. For example, the bubble size of *Huanggang Port*, *Luohu Port* and *Shenzhen Bao'an International Airport* are the top-3 stop events clusters. Moreover, the analysts could identify the top-10 taxis in *Luohu Port*, as the histogram shown in Figure 3(d).

Traffic congestion exploration: A driver is planning to *Guangzhou* at the coming weekend, he would like to explore the traffic congestion condition from his home (*Luohu Port*) to *Guangzhou*. The driver uses mouse draw two polygons over *Luohu Port* and *Guangzhou*. The exploratory analytical layer incurs a sub-trajectory extraction operator and finds all sub-trajectories from *Luohu Port* to *Guangzhou*. These sub-trajectories are visualized by trajectory-map projection view in Figure 3(e). The driver chooses a convenient path according to the path color as it shows the congestion degrees. One more step further, the driver could analyze the real-time traffic flow of a chosen road, i.e., the light colored road in Figure 3(e), the visualization result is shown in Figure 3(f). It helps the driver to decide when he starts the journey will have the best traffic condition.

V. Demonstration Proposal

This demonstration would be the first public demonstration of Vaite to our community. An attendee will experience Vaite from the perspective of a data analyst (who wants to find insights from underlying urban trajectory data). In order to verify the effectiveness of Vaite, attendees are invited to issue analytical queries/tasks. We also will demonstrate a fully functional implementation of Vaite. We will show how Vaite helps data analysts to explore the big urban trajectory data effectively and efficiently. The demonstration will also highlight the importance of the interactive speed/fantastic visualization views in supporting exploration-driven trajectory analysis.

Acknowledgment

This work was supported by the National Science Foundation of China (NSFC No. 61802163), Guangdong Natural Science Foundation (Grant No. 2018A030310129), and PCL Future Regional Network Facilities for Large-scale Experiments and Applications (PCL2018KP001). We also thank Qu You for his contribution on front-end development.

References

[1] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. v. d. Wetering, “Visual traffic jam analysis based on trajectory data,” *TVCG*, vol. 19, no. 12, pp. 2159–2168, 2013.

[2] G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel, “Scalable analysis of movement data for extracting and exploring significant places,” *TVCG*, vol. 19, no. 7, pp. 1078–1094, 2013.

[3] G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Kisilevich, and S. Wrobel, “A conceptual framework and taxonomy of techniques for analyzing movement,” *J. Vis. Lang. Comput.*, vol. 22, pp. 213–232, 2011.

[4] B. Tang, M. L. Yiu, K. Mouratidis, and K. Wang, “Efficient motif discovery in spatial trajectories using discrete fréchet distance.” EDBT, 2017.

[5] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan, and R. Maciejewski, “Vaud: A visual analysis approach for exploring spatio-temporal urban data,” *TVCG*, vol. 24, no. 9, pp. 2636–2648, 2018.

[6] I. Stoica, “For big data, moore’s law means better decisions,” 2013. [Online]. Available: https://amplab.cs.berkeley.edu/for-big-data-moores-law-means-better-decisions/

[7] S. Idreos, O. Papaemmanouil, and S. Chaudhuri, “Overview of data exploration techniques,” in *SIGMOD*, 2015, pp. 277–281.

[8] B. Tang, S. Han, M. L. Yiu, R. Ding, and D. Zhang, “Extracting top-k insights from multi-dimensional data,” in *SIGMOD*, 2017, pp. 1509–1524.

[9] Y. Park, M. Cafarella, and B. Mozafari, “Visualization-aware sampling for very large databases,” in *ICDE*, 2016, pp. 755–766.