An Open Source TrajAnalytics Software for Modeling, Transformation and Visualization of Urban Trajectory Data

Shamal AL-Dohuk1, Farah Kamw1, Ye Zhao2, Xinyue Ye3, Jing Yang4, Suphanut Jamonnak2

Abstract—We present an open source software, named TrajAnalytics, which models, transforms, and then visualizes urban trajectory data for transportation and urban study. It allows researchers, administrations, and practitioners to understand the population mobility data and to discover knowledge intuitively. A conceptual data model is presented which integrates trajectory data with geo-structures through a variety of data access queries. The model guides us to develop data processing and management capability and support various visual queries through a Web-based visualization interface. A set of visualization widgets and interaction functions promote easy user engagement. A primary user study shows the usability of the software.

I. INTRODUCTION

Thanks to advanced technologies in sensing and computing, the mobility patterns and dynamics of urban cities and their citizen are recorded and manifested in a variety of urban trajectory datasets, which include the moving paths of human, taxis, buses, fleets, cars, and so on. Understanding and analyzing such large-scale, complex data is of great importance for transportation study, urban design and planning.

The collected trajectory datasets are mostly in the form of raw list of data points and their corresponding attributes, which are not ready to use for visual analytics that requires integrated data of trajectories and geo-context. Domain researchers and practitioners need to perform trajectory data transformation and management from the raw data, before they can even utilize a visualization toolkit. Unfortunately, despite myriad visualization techniques and systems for ad-hoc urban trajectory data analytics tasks [2], there remains a gap between the demand of transportation and urban researchers and a free-access, open-source software integrating raw trajectory data processing, storage, and interactive visualization. In this paper, we introduce TrajAnalytics system which has been developed to fill the gap with the participation many domain experts.

TrajAnalytics is an open source visual analytics software, which integrates scalable modeling, transformation, management, and interactive visualization within a powerful Web-based computing platform. It provides exploratory data visualization tools for researchers, administrations, practitioners and general public to understand the data and to reveal knowledge intuitively. TrajAnalytics includes three major modules: (1) Data loading and transformation; (2) TrajBase database; (3) TrajVis visual analytics interface. The system is developed in two forms: a cloud based software and a local version, to fulfill the requirements of many real world users. It has been released for open access [1] with tutorials and user support.

TrajAnalytics is designed to seamlessly integrate data modeling and visual analytics functions. The major contributions can be summarized as:

- A Geo-Trajectory model is developed to incorporate different types of raw data and match them to geographic structures. The modules of data loading and map matching enable users to upload and transform their own data easily.
- A TrajBase database is developed to store both the trajectory data and the geographic data. We design effective spatio-temporal indexing algorithms of these data to support interactive visual queries.
- A TrajVis visual interface is designed which allows users to perform interactive queries and visualizations of multiple trajectory attributes in geo-contexts. It supports common tasks for urban trajectory study and is extensible for adding more visualization functions.

The software is tested with domain researchers and users to evaluate its usability and effectiveness.

II. RELATED WORK

Data Driven Transportation Studies: Transportation studies develop and evaluate strategies that improve safety, mobility, and sustainability in transportation systems, while data driven study is a fundamental method [23]. The floating car technique has been used by intelligent transportation systems to obtain its positional information [7, 22]. Taxis often serve as floating cars to obtain human mobility data and examine real-time traffic status and individual behaviors [11, 15, 13, 20].

Urban Trajectories Data Mining: Trajectories of human and vehicle motions are used to discover knowledge from large-scale datasets [31]. Utilizing the trajectory data has been divided into three main categories: the study of the collective behavior of a city's population, the traffic flow, and the operators (e.g. drivers) [5]. In particular, vehicle trajectory data has been used in traffic monitoring and prediction [17], urban planning [32], driving routing [28, 27], extracting geographical borders [21], service improvement...
[29], energy consumption analysis [30], and dynamic travel time estimation [18]. Researchers have successfully used the emerging trajectory data from different aspects. Data visualization is a necessary component in their research, which usually costs them much effort.

**Trajectory Visualization:** A large number of approaches have been proposed to visually explore movement data [2]. Many of them are focused on the origins and destinations of the trajectories, such as flow maps [19], Flowstrates [5], OD maps [25], and visual queries [8]. Other works visualize trajectories using various visual metaphors and interactions, such as GeoTime [12], TripVista [9], FromDaDay [10], and more [14], [9], [8], [24], [26], [4]. However, few existing works have created available software for public users. TrajAnalytics is aimed to provide researchers and practitioners with a convenient platform for trajectory visualization.

**III. GEO+TRAJECTORY DATA MODEL**

The data model is proposed to guide TrajAnalytics design and operations. It defines trajectory data, geo-structures and their relationship, and data access queries over them.

**Point and Trajectory:** A continuous trajectory may have a sequence of consecutive points which are sampled within a small time interval. Each point includes spatial location, time, speed, and other attributes (Fig. 2). Here the spatial location usually is a pair of longitude and latitude recorded by GPS devices. Sometimes, a trajectory may not include consecutive sample points but instead consists of unevenly distributed points. This often happens when the points are samples by WiFi or check in/out devices. Moreover, some trajectories only have the start/end points, such as the O/D (origin/destination) data of taxi trips. Fig. 2 shows a trajectory $c$, consisting of linked points $p_1, ..., p_n$.

**Geographical Structure:** Urban space consists of multiple types of geographical structures. In analytical tasks, most commonly used structures include streets, regions, and POIs (Points of Interest). A region may be a grid cell, a zip coded region, or a user-defined area. It has the name and bounding polygon. A street, usually represented as a geometric polyline, is part of urban street network. A POI is the common expression for a place which may have accurate geometric borders. But most cases, it is considered as a centroid with a center point and small radius. Demographics, business, social and other information can be linked to these geographic structures.

**Data Registration and Matching:** The trajectory data should be registered to the geo-structures. This is a fundamental task in GIS systems with map-matching algorithms. As a result, each GPS point in Fig. 2 is assigned to one street segment. Moreover, they can also be matched to regions and POIs. These important operations are basic steps in TrajAnalytics to support domain users in performing data pre-processing.

**Data Query and Joint Query** In visual analytics tasks,
trajectory data need to be accessed by given geospatial and temporal constraints, which can be described as

$$E(G, T, \pi) \rightarrow \Phi.$$  

As depicted in Fig. 3, the query conditions involve the geographic structure $G$ and the time constraint $T$ applied with the query mode $\pi$ to the trajectory data $E$. The result subset $\Phi$ are then sent to visual analytics module. Here, $\pi$ usually can be of multiple modes such as:

- **Pass mode**: Find trajectories traversing $G$ in $T$;
- **Start mode**: Find trajectories starting from $G$ in $T$;
- **End mode**: Find trajectories ending at $G$ in $T$;
- **Contain mode**: Find trajectories whose path are contained inside $G$ in $T$;

In most real-world tasks, data query tasks often have multiple tasks, where the individual queries $\pi_i$ are combined with joint conditions. As illustrated in Fig. 4, Union, Difference and Intersection can be used to generate final query results.

IV. VISUAL ANALYTICS TASKS

The query results are utilized in a variety of data analytics tasks with visual tool. We summarize the tasks in several categories:

- **Trajectory Study**: Users query trajectories with given spatial-temporal conditions, and then visualize them (in different ways) on map, which is integrated with different types of charts, glyphs and diagrams for their temporal or spatial distributions. One special case of the trajectory study is the study of O/D data, which refers to origin and destination of the trajectories. Similarly, O/D points are visualized on the map and combined with visual charts.

- **Street Study**: After querying trajectories, the results are used to acquire a variety of attributes on street segments. These attributes mostly are about traffic information such as traffic speed and traffic flow (amount of passing, starting, ending trajectories). Aggregation values such as average, maximum, minimum of the speed may be computed. The computed attributes may also be related to other types of information incorporated in the raw trajectory data.

- **Region Study**: Urban regions, such as zip code areas, are also the target for urban study from trajectories. The query results thus can be mapped to spatial regions to compute multiple attributes, such as the amount of trips starting/ending in a region, the average speed in given times, etc. They can be visualized by region view on the map and by different visual metaphors.

- **POI Study**: POIs are important so that the query results' trajectories can be projected to POIs and visualized to show different attributes.

- **Flow Study**: In addition to region/street/POI attributes, the query results are often used to compute values related to pairwise relationships between them, such as the population flow from one region to another region, or from one type of POIs to another type of POIs. Time dimension is always an important factor in these studies. In visualization tools, interactive brushing and filtering over times is indispensable. TrajAnalytics is developed to provide an open source software tool for practitioners to conduct these tasks.

V. TRAJECTORY DATA MANAGEMENT

A. Data Loading and Map Matching

TrajAnalytics supports users to upload their raw trajectory datasets in the comma-separated values (CSV) format. The software converts the input data to a Trajectory Data (TD) table in TrajBase. Sampling points are processed to form trajectories in TD table, instead of the raw sampling points. Then, TrajAnalytics provides two different ways to match the trajectory data with geographical units, streets and regions, respectively.

1) **Street Map Matching**: TrajAnalytics automatically finds a bounding region of the trajectories and allows users to adjust the region. Then, street geometry data inside the region is downloaded from OpenStreetMap. A street network table is created in TrajBase which has a list of street segments including a unique ID and the geometry (polyline) of the segment. After mapping the trajectories in TD table to the streets, A TDS (Trajectory Data on Streets) table stores trajectories linked to street IDs. Figure 5 (a) shows the downloaded street network in green and the sampling points in red with a New York taxi dataset.

2) **Region Map Matching**: When a street network is not available or not to be used, TrajAnalytics supports region map-matching. Currently, two types of regions are supported: zip code regions and rectangular/hexagon grid regions. Trajectory sampling points are aggregated on these regions. A TDR table (Trajectory Data on Regions) is created to store the trajectory data with region IDs. Figure 5 (b) and (c) show the zip code regions and rectangular regions (in green) with the New York taxi dataset.

B. TrajBase Functions

TrajBase is a scalable spatial database that is developed to store and manage trajectory data with TD, TDR, and TDS tables. It is implemented on the PostgreSQL platform with spatial extension PostGIS. TrajBase supports fast computation over various data queries in a remote and Web-based computing environment. It is optimized to answer spatial and temporal queries. Three type of efficient indices (B-Tree, GIST and Gin) are added over the data to enhance the queries' speed over big trajectory data.

When users interactively conduct queries over map, TrajBase extracts trajectory data from tables based on the given spatial and temporal constraints. Then the results are aggregated over streets or regions. Users can choose to
show points, trajectories or aggregated information, such as average (min, max) speed, flow on roads or in regions.

VI. TRAJVIS VISUALIZATION

TraJVis is designed to support online visual analysis with fast speed and easy user interaction. A set of coordinated views form the interface as illustrated in Figure 1. Users can conduct exploratory visual analysis through the informative and intuitive interactions. On the interactive map (Figure 1 (b)), users can easily select regions with different shapes to form queries. The results are shown on the map in points, heat maps, or trajectories, while users can alter the display methods. The visual report view (Figure 1 (c)) shows a set of charts and diagrams for quantitative analysis of the query results. On the query control panel (Figure 1 (a)), users can manage multiple queries and export result reports. Next we show how the visualization tools are used. A dataset of taxi trajectories in Porto, Portugal [16] is used to show examples.

1) Trajectory Visualization: Trajectories are directly shown as connected polylines on the map (Figure 1 (b)). Users can select to show the speed or other attributes on each trajectory by line width or color. Meanwhile, start (pickup) and end (drop-off) locations are visualized as green and red points, respectively. Figure 1 (b) shows all taxi trips that passed a specified region in Porto between June 16, 2014 and June 22, 2014.

2) Heatmap Visualization: Start (Pickup) and end (Drop-Off) points can be visualized as heatmap. Their densities are shown by the colors. Figure 6 shows the heatmap of start and end points of Figure 1 (b).

3) Street Attributes Visualization: Users can select to visualize an aggregated attribute on streets such as speed or flow on the map. It can be represented by line width of color of streets. Figure 7 shows the average speed (a) and flow (b) over the road network of Porto based on the trajectories shown in Figure 1 (b).

4) Region Attributes Visualization: Similarly, a set of attributes are computed on regions by aggregating the trajectories from a query. Users can visualize an attribute by the color of these regions on the map. Figure 7 shows the average speed (c) and flow (d) over the grid cells of Porto.

VII. TRAJANALYTICS SOFTWARE IMPLEMENTATION

TraJAnalytics is developed as a cloud-based software which can be accessed through web browsers. It supports users to upload raw trajectory data, and perform road based
or region based map matching. The generated data tables are stored in TrajBase database on cloud server for remotely real-time access. TrajVis interface is implemented using multiple Javascript libraries such as D3.js for information visualization and leaflet.js for map-based visualization. TrajAnalytics is also provided as a local version for users to work with the software on their own machines after configuring their own server. The software is published online for free access at [1] with full tutorials and user guides.

![Image](image.png)

**Fig. 7.** Visualize Attributes on Streets/Regions: (a) Streets Traffic Speed (b) Streets Traffic Flow (c) Regions Traffic Speed (d) Regions Traffic Flow

![Image](image.png)

**Fig. 8.** Domain users from different fields.

**VIII. EVALUATION AND USER STUDY**

We conducted a user study of TrajAnalytics with a group of 35 domain experts and practitioners from different countries. Their ages range from 21 to 63 years old and most have PhD degrees. Their working experience ranges from 1 to 30 years in a variety of fields related to human and vehicle trajectory study, as shown in Fig. 8. They were asked to use TrajAnalytics with help of our online tutorials and videos. Then, they filled a questionnaire. The questions included: if TrajAnalytics is generally useful in your field? and how different modules are useful respectively? They gave a score (0-10) for the system and modules. Fig. 9 shows the average scores from 35 domain users. Overall, They gave a positive evaluation with all scores above 7.8 out of 10. The TrajVis interface got the highest usability score at 8.9. Please note that it is not compared with similar systems, since to the best of our knowledge, there has no other open source trajectory visualization software available. In future work, we will compare our tools with general-purpose information visualization software (e.g., Tableau) and visual analytics technologies of spatio-temporal data (e.g. V-Analytics [3]).

In addition, we asked the users to provide comments and suggestions of the system. They mostly agreed that TrajAnalytics is a good tool for domain study, such as “…is easy to learn and use and it needs less memory”; “…is useful for analyzing traffic jam, some geographic phenomena, like migration, trading, and behavior”; “…provides data support for studying the smoothness of the road and evaluating the rationality of road planning.”. Most of the researchers agreed that the data loading and processing modules including the roads and regions map matching are effective and very useful tools, since “…there is only few software providing the function of map matching”.

Moreover, both cloud version and local version are evaluated. Almost half of domain users preferred the cloud version as it does not require installation and configuration, while the others liked the local version for its stability and because sometimes the data needs to be kept confidential.

The users also pointed out some limitations and gave valuable suggestions to improve the system such as:
• Enriching the visualization and analysis functions by adding some extended features such as data analysis, spatial analysis, and machine learning methods.
• Supporting comparative study of trajectory data from different countries and regions.
• Providing a simpler way to install and configure the local version.
• Migration to a distributed platform to accommodate larger datasets.
• Providing extended programming support.
• Creating a complete guidebook.

We will enhance the software based on these suggestions.

IX. CONCLUSIONS

In this paper, we present TrajAnalytics, an open source software for visually exploring urban trajectories. It facilitates an easy access gateway so that urban trajectory data can be processed and visually analyzed by domain researchers with less effort and overhead. It could advance a broad spectrum of applications by enabling researchers, administrations, practitioners and general public to visually analyze the emerging trajectory data.

X. ACKNOWLEDGMENTS

This work was supported in part by the U.S. NSF under Grant 1555031, Grant 1555081, and Grant 1739491.

REFERENCES


