SALON: A Universal Stay Point-Based Location Analysis Platform

Yue Hu[1,4], Sijie Ruan[1,4*], Yuting Ni[4], Huajun He[2,4], Jie Bao[4*], Ruiyuan Li[3,4], Yu Zheng[1,2,4]
[1]Xidian University, Shaanxi, China [2]Southwest Jiaotong University, Chengdu, China
[3]Chongqing University, Chongqing, China [4]JD Intelligent Cities Research, Beijing, China
hhuyuee@gmail.com;sjruan@stu.xidian.edu.cn;Niyuting_612@163.com;hehuajun@my.swjtu.edu.cn
baojie3@jd.com;liruiyuan@whu.edu.cn;msyuzheng@outlook.com

ABSTRACT

The prevalence of positioning technologies has fostered massive trajectory data. Stay points from trajectories indicate the visiting of moving objects to locations, which provide an opportunity to understand the locations comprehensively. Many existing works rely on stay points to analyze locations. However, they are ad-hoc solutions to tackle specific problems, and it is time-consuming and tedious to develop each application. In this paper, we propose a universal StAy point-based LOcation aNalysis platform, i.e., SALON, with the characteristics of universality, efficiency and flexibility. It can retrieve stay points using flexible conditions, associate stay points with locations, extract comprehensive location profiles and visualize the analysis results to users. Based on the combination of these functions, we demonstrate three different location analysis scenarios, i.e., illegal location discovery, popular location ranking, location temporal analysis to show its characteristics.

CCS CONCEPTS

• **Information systems → Spatial-temporal systems**.

KEYWORDS

Spatio-temporal Data Mining; Location Analysis; Interactive Exploration

ACM Reference Format:
Yu Hu, Sijie Ruan, Yuting Ni, Huajun He, Jie Bao, Ruiyuan Li, and Yu Zheng. 2021. SALON: A Universal Stay Point-Based Location Analysis Platform. In *29th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '21), November 2–5, 2021, Beijing, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3474717.3483991

1 INTRODUCTION

With the wide adoption of GPS-enabled devices, massive trajectory data has been generated every day, which represents different mobility behaviors of moving objects. Among them, an individual's staying behavior provides us an opportunity to understand the interaction between moving objects and locations. The staying behavior is usually reflected by the stay point in his/her trajectories, which represents that a moving object stays in a geographical region for a while. With those stay points, many location analysis-based applications can be facilitated, e.g., risky region discovery [1], POI recommendation [12], and urban planning [2].

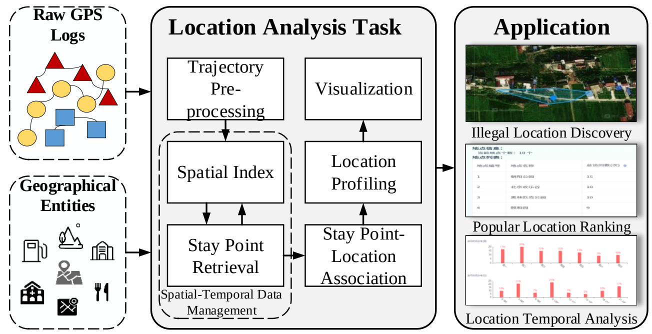


Figure 1: System overview

However, to support different applications, one has to design problem-specific solutions. Apart from that, even for similar tasks, processing steps and parameters should be tweaked, and visualization techniques should be employed to produce good results due to different data distributions. Therefore, developing those applications is extremely time-consuming and tedious. In fact, many processing steps of those works are shared in common.

In this paper, we propose a universal StAy point-based LOcation aNalysis platform (SALON) [1], which tends to provide a convenient and efficient approach to obtain insights into location analysis results through interactive exploration for product managers, government regulators, researchers, etc. We take full consideration of the properties of stay points and locations, and integrate the common processing steps for location analysis tasks. As shown in Figure 1, the platform takes raw GPS logs and geographical entities such as POI, road network as input, summarises six processes covering data pre-processing, management, interaction and visualization.

The advantages of our platform are summarized as follows: 1) Universality. The platform supports multiple analysis methods to tackle various applications. 2) Efficiency. The platform uses JUST [5] as database and builds spatial indexes for spatial data. Users can obtain analysis results in a short time. 3) Flexibility. The platform supports interactive retrieval and analysis. Users can customize parameters in these functions and explore the analysis results by different parameters.

In order to show the advantages of our platform, we demonstrate three scenarios: 1) Illegal location discovery, 2) Popular location

*Jie Bao and Sijie Ruan are the corresponding authors.

[1]http://just.urban-computing.com/salon.html

SIGSPATIAL '21, November 2–5, 2021, Beijing, China

ACM ISBN 978-1-4503-8664-7/21/11.
https://doi.org/10.1145/3474717.3483991

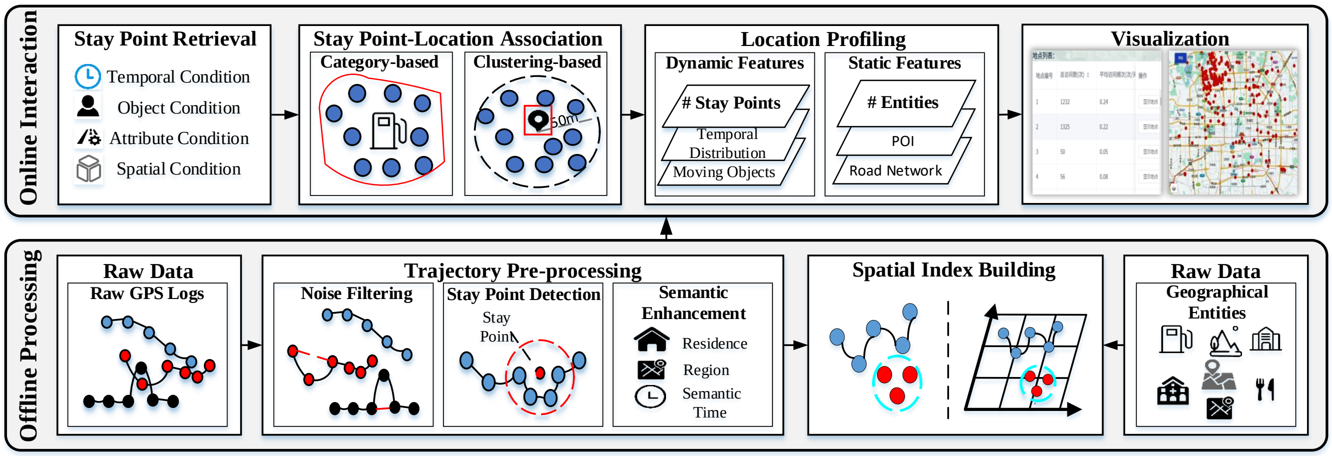


Figure 2: System framework.

ranking, 3) Location temporal analysis through the combination of different functions in our platform.

2 SYSTEM OVERVIEW

To understand the locations comprehensively, the platform designs a two-stage approach that follows the characteristics of universality, efficiency and flexibility. Figure 2 shows the framework of SALON, which contains two stages: (1) **Offline Processing** (Section 3), SALON first prepares raw GPS points and geographic entities, e.g., points of interest, and road network. For raw GPS points, they will be transformed to stay points with semantic information through noise filtering, stay points detection and semantic enhancement. Considering the efficiency of the interactive exploration of the online interaction stage, spatial indexes will be built for stay points and geographical entities. The goal of this phase is to obtain and store related data. (2) **Online Interaction** (Section 4), SALON first conducts stay point retrieval with a combination of four conditions to get candidate stay points. After retrieval, it offers two ways, i.e., category-based method and clustering-based method, to associate stay points with locations. During association, it takes consideration of multiple characteristics describing the location and provides the profiling of the location. Based on the profiles of locations, one can perform multiple location analysis. The goal of this phase is to achieve interactive exploration of location analysis by retrieval, association, profiling and visualization.

In summary, the platform is universal, efficient, and flexible to fulfill the complete process for location analysis tasks based on the stay points.

3 OFFLINE PROCESSING

In this section, we present data pre-processing and management, which builds the foundations of data interaction. This section is divided into two steps. (1) trajectory pre-processing, which transforms trajectory data to stay points with semantic information; (2) spatial index building, which manages stay points and geographical entities for efficient join and query.

3.1 Trajectory Pre-processing

Trajectory pre-processing is important, as the data noise and outliers in raw data affect the accuracy and performance of later applications. Moreover, we can also enhance the stay points with semantic meanings extracted from time, location, and the spatio-temporal patterns of the moving objects, which makes the stay points retrieval more flexible. The platform covers the following sub-tasks in the trajectory pre-processing [6–8]:

Noise Filtering. Due to sensor noise and other factors, such as receiving poor positioning signals in urban canyons, trajectories are never perfectly accurate. We employ a heuristic-based outlier detection algorithm [10] to remove noise GPS points.

Stay Point Detection. This task extracts a subset of consecutive GPS points as a stay point which represents that the moving object has stayed for a while, such as waiting for passengers, loading goods. We adopt a clustering-based algorithm [4] to extract stay points. Respectively, we obtain the centroid (lon, lat), the mid-time (t_{mid}) and the difference between the end time and the start time ($duration$) as a stay point' attributes from the above subset.

Semantic Enhancement: This task enhances stay points with semantic information which can be indicators to retrieval and analysis. For stay point itself, we consider the district where it is located ($region$), the day of the week (e.g., Sunday) ($semantic\_t$) when it is generated. For each moving object (oid), we extract its frequent stay locations in different periods, e.g., night or daytime by clustering its stay points. Then for each frequent stay location, we exact its province, city and location type as the moving object's labels (oid_label).

Overall, a stay point is characterized as follows:

$SP = \langle oid, \{oid\_label\}, lon, lat, t_{mid}, duration, region, \{semantic\_t\}\rangle$

3.2 Spatial Index Building

SALON contains two types of data, respectively, the processed stay points and geographic entities. A geographic entity is a geographic feature, like a restaurant or a road segment. It occupies a position in space and its attributes and geographic location are recorded in the related database. In SALON, we use POIs and road networks as geographic entities to retrieve stay points and analyze location auxiliarily. In order to achieve interactive analysis and make the platform efficient, these massive data will be indexed and stored in a spatial database [5–7] for retrieval and data summaries.

4 ONLINE INTERACTION

In this section, we present the online interaction of the location analysis task based on stay points. This section is divided into three steps. 1) Stay Point retrieval, which is used to obtain candidate stay points; 2) Stay Point-Location Association, which focuses on the way to associate stay points with locations. 3) Location Profiling, which summarises the properties of each location and generates statistics and analysis results.

4.1 Stay Point Retrieval

In order to obtain desired stay points and apply them to different applications, we take full account of the properties of stay points and design four main selection dimensions as follows:
Spatial condition, which pays attention to each stay point' spatial features, such as the region, POIs and road network, e.g., finding stay points that are close (e.g., within 50m) to any restaurant;
Temporal condition, which focuses on a temporal feature of the stay points, e.g., finding stay points on Sundays;
Object condition, which is used to obtain stay points of moving objects with the same label, e.g., finding stay points generated by moving objects whose resident place is a university;
Attribute condition, which focuses on the duration of each stay point, e.g., finding stay points whose duration is more than 30min.

We support a free combination of these types of conditions to flexibly obtain desired stay points.

4.2 Stay Point-Location Association

In some applications, the underlying location is known, such as all gas stations in a city, while in others, the location is unknown, such as an illegal warehouse. Therefore, we devise the following two association methods.
Clustering-based Association. We adopt density-based spatial clustering of applications with noise (DBSCAN) [3] to generate frequently visited locations. The platform supports users to customly input this algorithm's parameters (*minpts* and *eps*) according to different application scenarios.
Catagory-based Association. The platform contains location data with categories from OpenStreetMap[2] and allows users to select one particular category to analyze. We focus on stay points within these locations regardless of quantity.

4.3 Location Profiling

Location profiling is an essential process of our platform, which helps us understand location better through concrete value and distributions. To meet the universality, we extract static and dynamic features for each location, details as follows:
Static features, which indicate where the location is and what is around the location. There are many kinds of static features. For example, calculating the distribution of road network around a location infers its traffic status in one city. In our platform, we identify the center point of location (longitude and latitude), distribution of POIs within 100m and distribution of road network within 200m as static features, which can well infer the location's significance and functions.
Dynamic features, which indicate access of moving objects to the location. As we focus on stay points within the location, information about moving objects they visit are natural dynamic features. For example, if a location contains hundreds of taxi stay points, the location can be regarded as a candidate drop-off and pick-up location with high popularity. In our platform, we propose six dynamic features (the number of stay points, the number of moving objects, the frequency of visits, distribution of visit periods, distribution of visit weeks/months/years, distribution of visit duration) to describe the moving objects' visiting pattern of the location.

5 DEMONSTRATION

As shown in Figure 4, the main page contains four sections. The *Data Interaction* panel supports four types of stay point retrieval and two types of location analysis for users to obtain desired stay points and apply them to different scenarios. The *Data visualization* is used to display stay points and locations on the map. The button in the lower-left corner is designed for map switching. The *Data statistics* panel displays the overall dataset summaries. The *Data Distributions* panel is used to visualize the spatial-temporal distributions of stay points, including temporal distributions, POI distributions, road network distributions, etc.

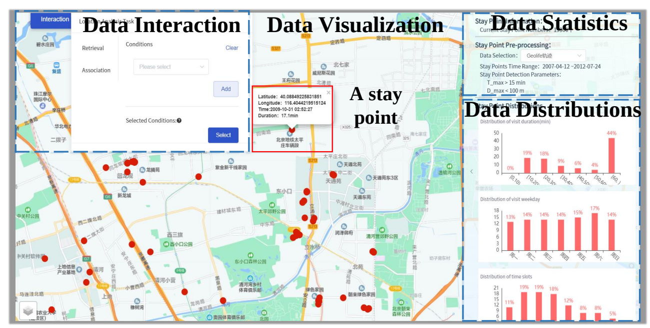


Figure 3: System Viewing Interface.

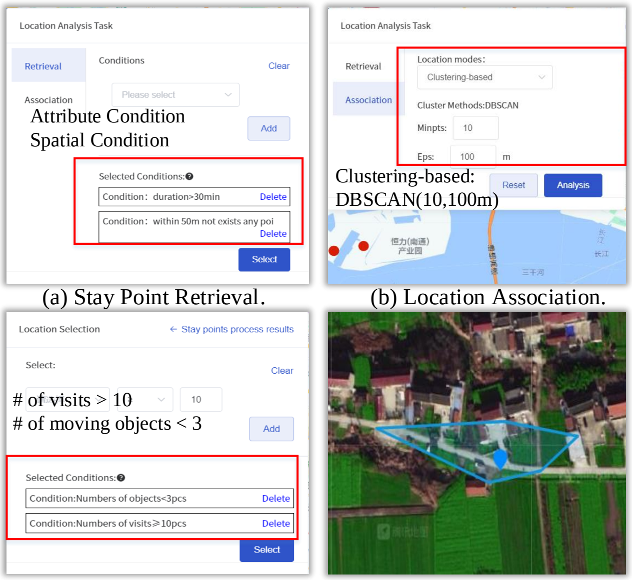
Our platform SALON will be demonstrated using three types of trajectory data (pedestrians, taxis and hazardous material vehicles) and displaying three application scenarios as following:

5.1 Illegal Location Discovery

We can select the HCT dataset, which contains almost 1 month of GPS records for 2,854 hazardous chemical material trucks and some of them may load and unload hazardous chemical materials at unregistered locations illegally. Obviously, trucks may stay in a certain location for a while when they load and unload materials, thus, we can discover those illegal locations based on stay points of HCT trucks. We first extract stay points with 15min, 100m. Then, as shown in Figure 4(a), we set two conditions to seek suspicious stay points: 1) the duration of the stay point is longer than 30min (an attribute condition) and 2) there does not exist POIs within 50m of the stay point (a spatial condition). In the association step, we adopt the clustering method to cluster these suspicious stay points

[2]http://download.geofabrik.de/asia.html

for obtaining candidate illegal locations, as shown in Figure 4(b). Next, we filter these candidate locations to find locations whose number of visits is more than 10 and the number of objects is less than 3, as shown in Figure 4(c). Finally, as shown in Figure 4(d), we show a suspicious illegal location with the highest number of visits. Moreover, this location is indeed an illegal warehouse for chemicals based on the on-field investigation.



(a) Stay Point Retrieval. (b) Location Association.

(c) Location Selection. (d) An Illegal Location.

Figure 4: Processes of obtaining an illegal location.

5.2 Popular Location Ranking

We switch the dataset to T-drive[9], which contains 7 days of GPS records for 10,357 taxis. In this example, we explore where taxis often stay in the evening nearby restaurants. Like the first scenario, We first obtain stay points described above, which satisfy the following criteria: 1) the duration of the stay point is less than 30 minutes, 2) the time of the stay point is within 17:00-21:00 and 3) there exist POIs with the type of restaurant within 50m. Then we use clustering analysis mode. As shown in Figure 5, we display the ranking of popular locations, according to the number of visits and then combine the map to show the top two locations specifically.

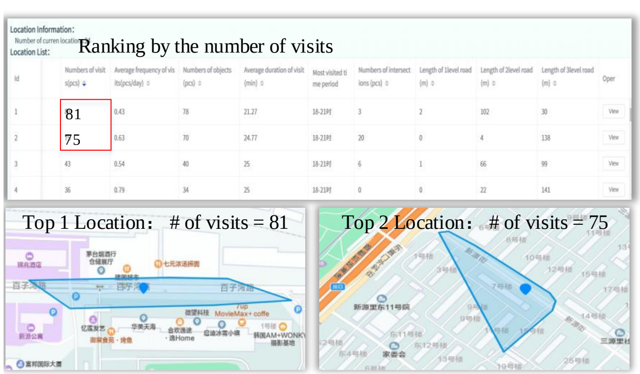


Figure 5: Ranking of popular locations for taxis.

Popular Location Ranking can be widely used in taxi management, in which the driver can have full knowledge of popular locations for receiving passengers to avoid wasting time and resources.

5.3 Location Temporal Analysis

We use Geolife [11] data, which contains 5 years of GPS records for 183 moving objects. Due to the long period of the data, we can observe location temporal variation, such as year, month, day. As shown in Figure 6, we give the example of Beijing Happy Valley. This location is from the category-based association and the parameter is *theme_park*. We find June is the peak month for the location and more visits on Fridays and Saturdays by tracking the moving objects' visiting time.

Location Temporal Analysis helps us understand a location's cyclical characteristics and development. It is widely adopted in urban planning.

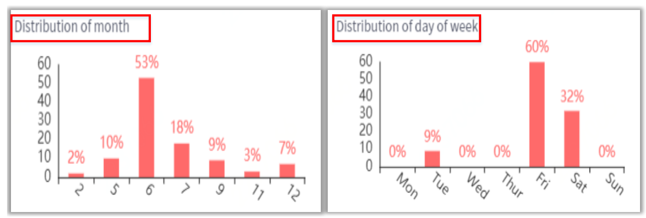


Figure 6: Example of location temporal distribution.

ACKNOWLEDGMENTS

The work was supported by the National Key R&D Program of China (2019YFB2101801), NSFC (61976168, 62076191).

REFERENCES

[1] Su Bae, Aravind Ravi, Kajanan Sangaralingam, Nisha Verma, Anindya Datta, and Varun Chugh. 2019. Suspicious Location Detection Using Trajectory Analysis & Location Backfilling–A Scalable Approach. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2007–2010.

[2] Edward Barbour, Carlos Cerezo Davila, Siddharth Gupta, Christoph Reinhart, Jasleen Kaur, and Marta C González. 2019. Planning for sustainable cities by estimating building occupancy with mobile phones. *Nature communications* 10, 1 (2019), 1–10.

[3] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *kdd*, Vol. 96. 226–231.

[4] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. 2008. Mining user similarity based on location history. In *ACM SIGSPATIAL*. 1–10.

[5] Ruiyuan Li, Huajun He, Rubin Wang, Yuchuan Huang, Junwen Liu, Sijie Ruan, Tianfu He, Jie Bao, and Yu Zheng. 2020. Just: Jd urban spatio-temporal data engine. In *ICDE*. IEEE, 1558–1569.

[6] Ruiyuan Li, Huajun He, Rubin Wang, Sijie Ruan, Tianfu He, Jie Bao, Junbo Zhang, Liang Hong, and Yu Zheng. 2021. TrajMesa: A Distributed NoSQL-Based Trajectory Data Management System. *TKDE* (2021).

[7] Ruiyuan Li, Huajun He, Rubin Wang, Sijie Ruan, Yuan Sui, Jie Bao, and Yu Zheng. 2020. Trajmesa: A distributed nosql storage engine for big trajectory data. In *ICDE*. IEEE, 2002–2005.

[8] Sijie Ruan, Ruiyuan Li, Jie Bao, Tianfu He, and Yu Zheng. 2018. Cloudtp: A cloud-based flexible trajectory preprocessing framework. In *ICDE*. IEEE, 1601–1604.

[9] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. 2010. T-drive: driving directions based on taxi trajectories. In *ACM SIGSPATIAL*. 99–108.

[10] Yu Zheng. 2015. Trajectory Data Mining: An Overview. *ACM Trans. Intell. Syst. Technol.* 6, 3 (2015), 29:1–29:41.

[11] Yu Zheng, Xing Xie, Wei-Ying Ma, et al. 2010. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.* 33, 2 (2010), 32–39.

[12] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining interesting locations and travel sequences from GPS trajectories. In *WWW*. 791–800.