Review and Analysis on Trajectory Big Data

Xueying Wu [1, 2, 3, 4, 5, *], Liheng Xia [1, 2, 3, 4, 5]

[1] Shaanxi Provincial Land Engineering Construction Group Co., Ltd. China
[2] Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd. China
[3] Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Natural Resources China
[4] Shaanxi Provincial Land Consolidation Engineering Technology Research Center China
[5] Land Engineering Technology Innovation Center, Ministry of Natural Resources China

Abstract: A trajectory can be seen as the imprint left in space by a moving object over time. In recent years, with the widespread use of civil GPS and other positioning devices on mobile terminals and the development and popularity of location-based services and mobile social networks, a large amount of trajectory data is increasingly accumulated in daily life and served by different types of applications. A large amount of trajectory data is generated in the fields of traffic, weather, ecology and mobile services. The effective understanding and utilization of these data requires not only automatic and efficient analysis methods, but also intuitive and vivid visualization, and this paper summarizes the application of trajectory data in cities.

Key words: Trajectory big data, visualization, urban planning.

1. The overview of the trajectory data

Nowadays, with the rapid development of satellite positioning technology, LBS (Location Based Service) technology and mobile Internet, trajectory data is generated anytime and anywhere in our daily life. Briefly, trajectory data is the data information formed by sampling the position, time, direction and speed of one or more moving objects by means of GPS locator, mobile phone service, radio frequency identification, etc. [1]. It is a kind of spatial big data, which conforms to the characteristics of volume, velocity and variety.

1.1 The features of trajectory data

Trajectory data contains rich spatiotemporal and behavioral information of moving objects. In addition to the typical "3V" (volume, velocity, variety) feature of big data, trajectory data has the following characteristics due to the disuse of equipment, frequency and storage mode.

Space-time sequence. As mentioned before, the trajectory data is to collect the information of the object in the spatiotemporal environment, the trajectory point contains the space-time dynamics of the object, the sequence is also the unit of data operation.

Heterofrequency sampling. Because of the randomness of object activity trace and the difference of sampling equipment, the interval difference of trajectory data is obvious. The trajectories are not only measured in seconds and minutes, such as vehicle GPS, but also collected in hours and days on social media. This discrepancy also increases the difficulty of data processing and analysis.

Road network correlation and Poor data quality. In the traffic data, the running state of the trajectory is limited by the actual traffic network, the data is closely related to the road network, so the spatio-temporal topology information of the road network is often used to optimize the data processing. When the sampling interval reaches more than minutes, the track data of continuously moving objects present a discrete state, if the position accuracy is affected by the equipment at the same time, the data quality may not be guaranteed.

1.2 The types of trajectory data

Generally, the trajectory data include human activity trajectory, traffic activity trajectory, animal activity trajectory and natural activity trajectory. Human activities trajectory data includes both active and passive. Active recording refers to the trajectory data obtained by people in the daily social process through the exchange of photos, emails or outdoor sports enthusiasts using GPS devices to record their itinerary, while passive means that users inadvertently turn on the GPS location (e.g. bus card, credit card consumption) to converge into trajectory data. Traffic activity trajectory is mainly the movement data recorded by car GPS in the city, which can be used to analyze the demands of cars, Shared car parking, charging and so on. The activity track data of animals are mainly acquired through sensors, while the activity track of natural phenomena is studied by collecting the activity track of typhoon and ocean events.

* Corresponding author: 243681563@qq.com

According to the sampling method and driving factors, all the above trajectory data are generally divided into the following three categories, Trajectory data based on time sampling, based on position sampling and based on event-triggered sampling. The two types of trajectory data we talk about currently are GPS and cell phone signaling.
GPS is sampled according to time, and this time-based data is the information recorded at the same time interval, such as vehicle GPS, animal migration, or the data obtained by global reverse deduction, such as hurricane data [2]. vehicle GPS, for instance, these devices at a fixed frequency send their geographic coordinates to the center. Although it can cover all activities, the representativeness of the data is not strong, there may be many useless data or data omissions, for example, when the vehicle is not moving, the onboard GPS will still collect data and the data loss caused by the sensor signal loss is inevitable in the transmission process.
As for cell phone signaling data, cell is the smallest unit in the GSM system that is an area where a user can communicate with the antenna. A location region consisting of a sequence of cells can be used to locate the user when communication occurs. Mobile service providers can track down the base stations that are communicating with the phones. Therefore, when the mobile phone is used such as turned on, turned off, phone calls and connected to the Internet, the GSM system would record the user's information.

2. Visual analysis of trajectory data

Visual analysis combines visualization, human-computer interaction, and automated analysis. Although the technology of machine automated analysis has been relatively mature, it is necessary for human beings to define analysis tasks and patterns in data analysis. In a typical visual analysis, users can evaluate the visual results of automatic analysis of the system, modify the analysis model, and obtain new automatic analysis results.

2.1 The Methods of visual analysis

2.1.1 Direct visualization

Direct visualization is the most basic visual analysis method to depict and display the trajectory data one by one. There is no further modeling in this method, so that it accurately has kept the information and display the dynamic changes more clearly. Furthermore, direct visualization can be divided into position animation(e.g. the location of the personnel to participate in the meeting), path visualization(e.g. the path of the ship), space-time cube visualization(e.g. the spatiotemporal trajectory of personal movement), time axis visualization(e.g. changes in vehicle speed) and parallel coordinates(e.g. Parallel visualization of traffic trajectories at intersections). However, when there are too many trajectories, the manual task is heavy, visual result trajectories will block each other.

2.1.2 Cluster visualization

If the data is large, cluster visualization could be considered. This method firstly preserves the important track data, removes the duplicate data, and then makes statistics on each dimension according to the requirements. According to the different dimensions, it can be divided into space-time and attribute clustering, points-destination clustering and path clustering.
The first one can not only act on a single dimension (space, time, attributes), but also combine them, such as ship track density, the number of times a taxi takes a passenger, the activity intensity distribution of different regions and time, etc. As for points-destination clustering, in the data pre-processing stage, the trajectory data will be converted into OD data (origin - destination data). It does not record the specific movement data, and generally USES the flow chart to visually display the result, such as the flow chart of the population migration in the United States. These “flows” in chart have different properties and can change over time. Path clustering is to specify the appropriate similarity function in advance, obtain the trajectory of different paths according to the density, segmentation and other algorithms, and then display the path of each class [3].
In practice, due to the complexity of problems, the researchers also developed interactive clustering methods. The progressive system allows users to manually select trajectories and specify them as one or more classes. Semi-supervised clustering system users can also specify parameters before clustering starts, monitor in real time and manually modify the results after completion. Summary, cluster visualization can analyze a large amount of trajectory data with the help of computer, but it is difficult to study the interaction and relative motion between trajectories, and extra programming is also needed.

2.1.3 Feature visualization

If the characteristics of the data can be determined and calculated directly, in visualization, the characteristics (e.g. events, behaviors, locations) could be extracted firstly, and then plotted. It can directly reflect the characteristics which the user's most concerned about, making the automatic search more system more efficient. At the same time, it may also lose a lot of information that seems irrelevant to features. This information, on the one hand, can explain the features more clearly, on the other hand, this data processing method in feature visualization is not applicable to exploratory tasks.

2.2 Data processing techniques in visualization

Regardless of the method of visual analysis chosen, it is usually necessary to preprocess the trajectory data before visual analysis, including cleaning, segmentation, compression, and storage.

2.2.1 Hadoop distribution processing based on MapReduce model

MapReduce is a programming model that can process data sets, consisting of job tracker for assignment and task tracker module for executing user-defined tasks, and it performs calculations on Hadoop composed of thousands of servers. MapReduce divides the data into each node, the results of each node are returned to the control center regularly to update the global state, and in the continuous development, the quadtree search method is applied to the model to speed up the calculation. Hadoop platform optimized the k-means algorithm to process lots of trajectory data, in Hadoop, the square error is used to evaluate the quality, improving the efficiency and accuracy of the algorithm [4].

2.2.2 Spark and Storm platforms

Spark is a kind of computing framework similar to Hadoop MapReduce. Its memory distribution data set can better process trajectory data that with high requirements for iterative computing power. Storm platform has a fault-tolerant real-time system, which solves the processing needs of data that have high real-time requirements and cannot be stored in advance, such as intelligent traffic and LBSN (Location-based Social Network) location recommendation.

3. Application areas of trajectory data

Trajectory data can help researchers mine human movement patterns and activity patterns, and help cities develop into smart cities.

3.1 Urban life

The trajectory data can mine the user's interests, activity patterns, etc. For example, at a given time, the weather and the previous passenger drop-off point, predicting the location of the next passenger, or predicting the time required by the taxi at a specific time and place, etc. This not only brings convenience to the taxi industry, but also helps passengers to shorten the waiting time or arrange their travel according to the predicted waiting time. In addition, through the analysis of similar trajectory data, when residents travel, they can receive more accurate push to the location and landscape in line with their interests and hobbies, arrange their schedule more efficiently.

3.2 Urban planning

In terms of urban traffic, real-time monitoring of urban traffic can be conducted according to the trajectory data of vehicle GPS. Through the analysis of these data, we can clearly see the situation of urban traffic, such as the degree of congestion and the main time of congestion. Then researchers combine it with route navigation to intelligently plan the most suitable route according to users' needs. In addition, the trajectory data can also predict the demand for parking and riding, and planners can make targeted adjustments in urban renewal.

In urban development, using the trajectory data analysis of the urban population of travel purpose, methods and paths, the evolution regularity of the residents' activity and urban geography can be revealed. This can provide auxiliary decision-making for urban public space layout, land use, infrastructure and other planning. For example, the hot areas of residents' activities should be equipped with corresponding infrastructure to increase the comfort level, and the population density and prosperity of the area that will be analyzed will be used as a reference for advertising and commercial location selection.

3.3 Urban public health

The interaction among people and the movement of people are often one of the important factors that determine the spread of epidemic diseases. As early as 2012, Pindolia used mobile phone call data and malaria transmission map to show the impact of population movement on malaria transmission, and proposed the prevention and control mechanism [5].And just this year, when the Novel Coronavirus (COVID-19) outbreak broke out, China also used the trajectory data of mobile phone signals to monitor and find some people who passed through the epidemic area or who might be in contact with the suspected people, notify these people as soon as possible and quarantine themselves for observation at home. To a certain extent, secondary transmission is avoided. Not only that, combined with Internet technology, everyone can see the confirmed cases around on the mobile phone during the epidemic, increasing the transparency of information and raising the awareness of the people to prevent and control the epidemic. These measures have greatly reduced the impact of the new coronavirus on the national epidemic.

Although the analysis and application of trajectory data have begun to change our lives, it is still in the development stage, which requires more reasonable algorithms to improve the efficiency of data analysis, and can produce more intuitive, dynamic and real-time updated visual effects. In addition, as mobile phones and GPS have become an indispensable part of daily life, the collection and analysis of trajectory data have the problem of user privacy leakage. How to obtain more useful data while protecting privacy is also a problem worthy of attention.

References

1. Li X, Han J, Kim S, et al. (2007). “Roam: Rule-and motif-based anomaly detection in massive moving object data sets”, Proceedings of the seventh siam international conference on data mining. Philadelphia, PA: Siam: 273-284.
2. Nanni M. (2002). “Clustering methods for spatio-temporal data”. University of Pisa.
3. Pelekis N, Andrienko G, Andrienko N, et al. (2012) “Visually exploring movement data via similaritybased analysis”. Journal of In-telligent Information Systems, 38(2), 343–391.
4. Tom White. (2012) “Hadoop: The Definitive Guide, 3rd ed”. O’Reilly Media, Inc. Sebastopol, Calif, USA, 2012.
5. Pindolia D K, Garcia A J, Wesolowski A, et al. (2012) “Humanmovement data for malaria control and elimination strate-gic planning”. Malaria Journal, 11(1),205.