

Application and research on trajectory big data

Siyong Chen[†], Jingjing Liao^{*, †}

Global Education Guangzhou, Guangzhou, China

* Corresponding author: 161847051@masu.edu.cn

[†]These authors contributed equally.

Abstract. In the era of big data, the formation of massive trajectory data of moving objects with rich Spatio-temporal feature information can mine data such as human activity patterns and behavioral characteristics through trajectory data processing technology. This paper analyzes the concept of trajectory data, the application process and the research results. Firstly, this paper briefly describes the trajectory data. Secondly, it outlines the supporting techniques of trajectory data, including data cleaning, integration, and transformation. Next, it introduces the classification steps of trajectory data mining, including using the segmentation method to segment the data, extract features from the segments; besides, it models and studies the predicted questions in this topic. Then it briefly introduces the application method of trajectory data mining in data privacy and the application of trajectory data in the transportation field. Finally, the paper summarizes the challenges and future research directions of trajectory data processing.

Keywords: trajectory; big data; data privacy; data processing.

1. Introduction

With the advancement of technology and society, people's lives are becoming increasingly rich. The browsing records of social networks, the frequency and route of taxis, the check-in situation of passengers in subway stations, and the migration of animals will all generate trajectory data. These data are collected and stored on the Internet in large quantities and have great research value. And how do these massive trajectory data exert practical value? This requires data mining and analysis, for example, to analyze the user's travel preferences and plan the best route for them, to determine the urban development center according to the density of urban residents, and to provide a reasonable urban planning scheme [1-3].

To improve the effect of trajectory data mining, it is usually necessary to preprocess the trajectory data. Some erroneous samples due to equipment anomalies require data cleaning, followed by a certain degree of trajectory compression. Preprocessed trajectory data can reduce the pressure of data mining. Due to the wide application of data mining and analysis, the technology of data mining has emerged as the times require. For instance, in the face of a large number of different types of users' browsing frequency data for various styles of small videos, advanced technology should be used to filter and store these data, and use the most suitable modelling method to make a concrete representation of these effective data. At the same time, it uses visualization technology (direct visualization) to visually display the characteristics of the original trajectory data [4].

There are many different paths between the same starting point and the same destination. Different paths have different road conditions, and different road conditions will lead to different driving experiences and effects, which will generate trajectory data. Developers can use this data to plan the best route for users to travel.

A large amount of trajectory data contains a lot of personal information, which may endanger human privacy to a certain extent. For example, through the correlation of data, an attacker may master the user's home address, travel habits, and physical health, and thereby cause physical or mental harm to the user. Therefore, the privacy protection of trajectory big data is one of the focuses of research [5-7].

This paper focuses on the process of how trajectory big data realizes the final value, and focuses on three aspects: the preprocessing method of trajectory big data, the mining of trajectory big data

and the application of trajectory data. This paper introduces the process of trajectory big data preprocessing by sorting out and summarizing the current preprocessing methods of trajectory data, mining and application applications, and provides a summary of the research in three aspects: data mining methods and trajectory data applications, and discuss the possible future research directions of these three aspects.

2. Overview Of Trajectory Data

In the first part of this section, the basic concepts and functions of tracking big data are introduced, and the applications of tracking big data in different fields are listed.

Trajectory big data is based on the extensive application of mobile intelligence such as various positioning devices and electronic devices and has the characteristics of time and space. Trajectory big data, as the processing object of trajectory big data, can mine the effective information in big data from the source, features, structure and other aspects of processing information technology. Therefore, it can quickly help locate the objects to obtain accurate geographical location, speed, direction and time, and can quickly update the latest data.

The effective trajectory big data includes natural activity trajectories, vehicle activity trajectories, human activity trajectories, and animal activity trajectories [8-10]. These four aspects are described in detail below:

The big data of the natural activity trajectory has always been the focus of attention. Researchers need to analyze the data information obtained from the trajectory big data and use the analyzed information to infer the activity trajectory, time and speed of natural disasters such as typhoons and sandstorms.

A vehicle track is a vehicle equipped with a GPS positioning device, which can be used for vehicles, ships, etc. The track big data will obtain the latest positioning information to help the vehicle drive and record the accurate information of the vehicle driving within the moving range. In addition, the trajectory data can also be used for traffic analysis, such as detecting road accidents, destroying road sections or excavating congested road sections.

The trajectory of human activities is mainly obtained through the positioning devices of mobile phones and electronic equipment, as well as through the network. The acquisition of human activity trajectory data is generally divided into two situations, one is active acquisition and the other is passive acquisition. The active acquisition is when people actively share their location information. The other is that people unintentionally turn on location services that automatically read location information.

Animal activity trajectories can help researchers obtain the migration locations, migration times, and daily habits of animals.

3. Trajectory Data Preprocessing

It is necessary to review, screen and sort out the preliminary collected data before classification or grouping. The reasons for the need for data preprocessing are as follows: the first is because most of the attribute values are missing or only contain aggregated data; the second is because the wrong or biased cluster values are included; The third one is because of the normative and logical nature of the data records is not compliant or inconsistent with other data sets.

After data preprocessing, the collected data contains high-quality data such as integrity, logical accuracy, credibility, and accuracy. It can make those messy data into complete and orderly data, thus making it easier for researchers to Process and analyze these data. The processing methods of preprocessing mainly include data cleaning, data integration, data transformation, and data reduction.

3.1. Data cleaning

Data cleansing is to eliminate redundant data. Error data will be deleted or corrected, or missing data will be filled in. The steps of data cleaning include the following points: missing value processing, abnormal value processing, deduplication processing, and noisy data processing.

The noisy data processing indicates that there are errors in the data, and there is a huge gap between the final data values and the expected value. Noise data processing includes median or means filtering, cluster analysis methods, human-machine combined detection methods, regression methods, etc.

Median or mean filtering is a kind of nonlinear signal processing technology based on ranking statistics theory, which can effectively suppress noise. The principle of median or mean filtering is to replace the value of a point in a digital image or digital sequence with the median value of each point value in a field of the point and then make the pixel value close to the true value to eliminate noise. The median filtering process is similar to the mean filtering, and the median filtering is more robust than the mean filtering when dealing with extreme errors. Both of them can handle a single noise point but cannot handle multiple consecutive noise points very well.

Cluster analysis method: Similar data will be grouped for data cleaning according to a certain standard. Fig. 1 is an example of cluster analysis, the plus sign in the case represents the cluster center, and the circle represents the same data. Each identical data is clustered together according to the corresponding center.

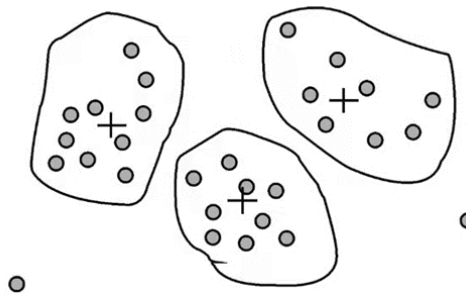


Figure 1. A example of Cluster Analysis.

Regression method: There are two methods of regression, one is unitary linear regression and the other is multiple linear regression. Univariate linear regression is the best straight line to find two attributes. It can find another attribute from one attribute. Multilinear regression is the ability to find multiple attributes and fit the data into a multidimensional surface, which can effectively remove noise. The characteristic of this method is that it cannot only show the significant relationship between the independent variable and the dependent variable but also show the influence intensity of multiple independent variables on a dependent variable. Fig. 2 is an example of the regression method. Different points find common features, and the straight line derived from the features is $y = x + 1$.

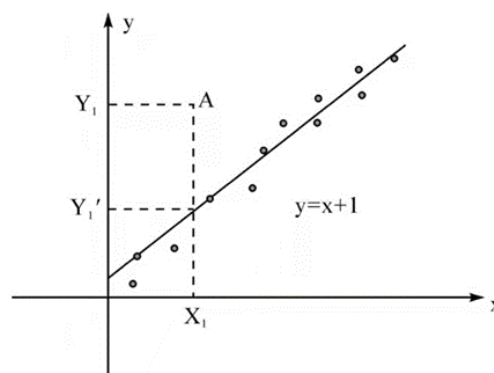


Figure 2. A Example of regression method

3.2.Data integration

Data integration refers to the fusion of data sources to discover related information and the same meaning from different data from data sources [9-12]. Among them, data integration work plays an important role in big data management and query [2].

Data management is the process of effectively collecting, storing, processing and applying data using computer hardware and software technology. A good data management strategy can improve data quality and access efficiency.

Data integration technology can help users provide high-quality data management strategies, and can obtain clear data structure and correlation information between data from massive data.

With the advent of the era of big data, the data to be processed is rapidly expanding. In traditional index data, only one data can be queried, which is inefficient. Data integration technology helps to build high-quality indexes, which can query multiple data sources, thus improving efficiency [2,13].

Data integration is to organize and store all data of different characteristics and source code in the same database. Taking For example, assume that there are only Table 1 and Table 2 in the initial dataset. The information in Table 1 is the title of the song, the name of the artist and the song language. In Table 2 are the singer's name and the singer's nationality. Combining the data from these two tables, we can learn more. For example, we can know that the nationality of the singer Justin Bieber of "STAY" is Canada. Get richer information through data integration.

Table 1. Information About Songs

Song name	Singer name	Language
STAY	Justin Bieber	English
Runaway	AURORA	English
พุดจิ้ง	MEYOU	Thai language

Table 2. Singer Information

Singer's name	Nationality
Justin Bieber	Canada
AURORA	Norway
MEYOU	Thailand

3.3.Data conversion

Data conversion is the process of merging, cleaning, integrating, and then converting data from one representation form to another.

The steps of data transformation include the following points: Smoothing processing, aggregation processing, data generalization processing, normalization processing, attribute construction processing. Commonly included contents of data transformation include: Data range conversion, data granularity conversion, data type conversion, data semantic conversion, row-column conversion, data discretization, data standardization, extraction of new fields, etc.

The steps of data conversion can be summarized as follows:

Smoothing processing helps to reduce noise in data. Common data smoothing methods include Gaussian filtering, median filtering, locally weighted regression of loess and Lowe's and Savitzky-Golay filtering, etc. The average value of noise is 0, and smoothing can only be used in scenes where the true value has little or no linear change. The window of the smoothing method needs to be selected according to the specific data, which is very important, otherwise, the effect is not good.

Aggregation processing is the operation of summarizing or summarizing data. This method is often used to construct data cubes or perform multi-granularity analysis on data.

Data generalization replaces lower-level data objects with higher-level concepts.

Normalization projects data about attributes on a small scale. The extracted data is not only correct but also helpful for people to improve the efficiency of extracting the data.

Attribute construction processing is to construct new attributes from existing attributes to help data processing.

3.4.Data reduction

In the face of large-scale and high-latitude data sets for data mining, it usually takes a lot of time to find models and patterns [3,14-15]. The answer to high-latitude problems is data reduction. The main purpose of the data reduction technique is to obtain a reduced data set from huge data sets, which will not affect the final mining analysis results. Reducing the data set can keep the original data set and also improve the efficiency of data mining. The steps of data reduction mainly include the following points: Data aggregation, dimension reduction, etc.

The data cube is a kind of technical organization for data analysis and indexing. It is a powerful tool for handling big data. The processing of metadata by data cube can greatly improve the efficiency of data queries.

Dimensionality reduction is mainly used to detect and eliminate dull, weakly correlated, or redundant attributes or dimensions. The characteristic of data reduction is to reduce the size of the data to be mined without affecting the final result of mining analysis. Data mining generally consumes a lot of time and energy, but the method of data reduction can improve the efficiency of data mining. For example, a company now has a huge product datasheet, and due to time constraints, it needs to quickly analyze the accurate results of this product. At this time, using data reduction is one of the best methods.

Faced with large-scale data sets for data mining, the most important thing is often not the selection of specific mining techniques, algorithms or software, but the selection and preparation of data [3]. That is, data preprocessing. One of the most important aspects of data preprocessing is data reduction. Reducing the scale of the data set based on the mining results can improve the efficiency of data mining without affecting the final mining analysis results. Data cube summation and dimension reduction can be used in combination. First, use the data cube summation to select a suitable subset of variables, and then use the combined transformation obtained by dimension reduction to further reduce the dimension of the data [3,16].

In engineering practice, data that has not been preprocessed by trajectory data is unclean data containing missing values, duplicate values, etc. Therefore, the data needs to be preprocessed for data analysis. In this section, the trajectory preprocessing method is described. The trajectory preprocessing mainly includes data cleaning, data integration, data conversion, and data reduction.

4. Processing And Analysis Of Trajectory Data

Trajectory data mining is a process of processing a large amount of messy data and mining valuable knowledge and information hidden in the trajectory data. It is an effective technique to analyze and extract important information in the data [4]. This section will summarize the classification of trajectory data mining and the research problems of data mining.

4.1.The classification of trajectory data

To distinguish different states of a trajectory or segment, the trajectory data needs to be classified. In addition, there is a lot of trajectory data, and it is necessary to mine more trends and values in the trajectory data through classification. Trajectory classification is generally divided into three steps.

Use the segmentation method to split the trajectory into segments

Extract features from segments of sample points

By establishing model segments or sampling points, Dynamic Bayesian Network (DBN), Hidden Markov Model (HMM), Conditional Random Field (CRF), etc. The methods listed above are commonly used.

At first, researchers used the Gaussian Mixture Model (GMM) to process the trajectory data [5], but the GMM was unable to obtain the temporal relationship and order between entities. To solve this

defect, the researchers used the HMM algorithm to process the trajectory data. Later, researchers put forward the SD-HMM model, which is an improved model of HMM [6]. It has better adaptability in dealing with the cognition of human activity and reduces the overall error rate.

4.2.The data prediction

The prediction is a major research part in trajectory data mining. The main research direction is traffic flow prediction of behavior prediction. Traffic flow prediction is applied to traffic dredging to avoid stampede accidents due to the concentration of people, and behavior prediction is applied to provide corresponding recommendation services according to user interests and preferences. Barabasi and Brockmann et al. studied the activity patterns of MIT students and found that people's travel behaviors, like animals, have certain regularity in time and space, which makes it possible to predict the user's location.

This section focuses on the content of trajectory data location prediction.

Position prediction based on contextual content of trajectory data: If there is a correlation between the user's past visited location and the current visited location, the location can be predicted by analyzing the associated probability of the context. To this end, Song et al. proposed a special tree structure to store context information, and nodes in the tree record the probability of reaching the corresponding node from the root node. When the current location of the user is known, the child node with the highest probability of visiting the tree is obtained by searching from the corresponding root node, and the location corresponding to the child node is the prediction result.

Most of the location prediction methods based on the context of trajectory data are suitable for GPS data and are mostly used in individual prediction modeling: Through the study of human urban travel data, Cho et al. found that users' travel activities are correlated with home distance and social relations. Therefore, Cho et al. used Gaussian distribution to model user travel behavior and estimated the user's travel behavior at work and home. Location probability distribution, taking social and temporal factors into account and incorporating these factors into the distribution function. It is an important method to predict the user's position by using the patterns in the trajectory data. Pattern mining of trajectory data includes sequential patterns, frequent patterns and periodic patterns.

Location prediction model based on social connections: If the data is sparse, the research of position prediction will encounter many obstacles. When there is little or no user data in the past, it is difficult for the forecasting model to obtain effective results. Therefore, a new research idea was developed: By using the social relationship of users, the influence of sparse trajectory data and cold start on the forecasting models is reduced.

The General Correlation Model (GSC) is a model proposed by Gao et al. to analyze the distance between social connections and the geographical location of friends. Making use of social relations is an important method to reduce the influence of data sparsity on prediction models.

5. Application of trajectory data in privacy protection

With the advent of the era of big data, mobile communication and location-aware technologies are also gradually developing. These two technologies will generate and collect a large number of moving objects' position and motion trajectory data in various fields [9,17]. Trajectory data can provide people with convenient services. For example, taxi-hailing software provides travel services for tens of millions of people every day, accumulating massive amounts of trajectory data in the process. However, trajectory data contains a large amount of private data, which can easily be leaked into the public interconnected space if not handled carefully.

For example, through a certain location frequently visited by users, malicious attackers can obtain users' private data through reasoning. For example, if a person frequents the school, a malicious attacker can infer that the person may be a student or a teacher, and can also obtain the person's work unit. Malicious attackers can obtain a large amount of private data through computational reasoning

of locations and trajectories in social networks. To ensure that the data is not leaked, the movement trajectory data is very important in terms of privacy protection.

At present, there are three methods of trajectory data privacy protection: fake trajectory data method, inhibition method and generalization method.

5.1.Fake trajectory data method

The fake trajectory data method is to add a certain amount of virtual trajectory data to the original one, which can make the original trajectory data resist attacks and also reduce the probability of leakage. This method requires a small amount of calculation and is prone to excessive data storage, which reduces the effectiveness of the data [10]. It is generally used when the amount of data is not large.

There are random generation methods and rotation simulation generation methods for adding virtual trajectory data. The random generation method refers to randomly generating a piece of virtual data similar to the original trajectory data between the starting point and the endpoint of the trajectory data. The method of rotation simulation generation is to rotate based on the original trajectory data and then to get the virtual trajectory data. Combining the random generation method with the rotation simulation generation method, by determining the k intersection of the false trajectory and the real trajectory, the trajectory between the intersection can be randomly generated, which is called the k -cross simulation method [10].

5.2.Inhibition method

The Inhibition methods can selectively publish data and choose not to publish sensitive data and frequently accessed data. The nodes with the lowest frequency of the same kind can be used instead of the nodes with the risk of leakage thus achieving the suppression of the nodes with privacy leakage risk [11]. The Inhibition method requires less computation, but it will lose information and the data utilization rate is very low. The inhibition method will remove parts of the original data during use, and it will also have some influence on data mining.

5.3.Generalization Method

The generalization method is to generalize the sampling points obtained on the trajectory into anonymous regions to protect privacy. The main generalization methods are trajectory data preprocessing and trajectory k -anonymization processing.

Trajectory preprocessing is to put all the data of the same period in the same equivalence class. In practice, it is not possible to keep the sample placement point between each trajectory at the same moment, therefore, to minimize the amount of information loss. Some tracks can be synchronized or deleted by increasing the number of tracks in the equivalence class to ensure that tracks are similar in time [4]. Or use a greedy algorithm to find another equivalence class, and after repeated synchronization processing, combine the found equivalence class into a new equivalence class [10].

Trajectory k -anonymity technology is the most popular method of trajectory privacy protection based on the generalized method. Find similar k trajectories around real users to construct an anonymous set. If there are more anonymous tracks in the same area, which are similar to but different from the original tracks, it can cause great confusion, and can also hide anonymity well to protect privacy. Clustering and sorting methods are usually used to construct the trajectory k -anonymous set [12]. In the clustering process, the distance between trajectories [13-17] is used as a measure of the similarity between trajectories, and an anonymous set of similar k trajectories is found for clustering, and then the trajectories are constructed using the position arrangement algorithm. combining these two methods can construct trajectory k -anonymity sets.

The advantage of the fake trajectory data method and the suppression method is that the amount of calculation is small, so the implementation is relatively easy. However, these two methods also have shortcomings. Among them, the false trajectory data method has the problem of data distortion, and the suppression method has the problem of loss of information and poor data utility. At present,

the privacy protection of moving trajectory data is a research hotspot in the field of space security, but the technology is not mature enough. Many researchers have done a lot of algorithms and technical research on track privacy, but there are still many key issues that need to be further studied.

6. Application Of Trajectory Data In Transportation Field

With the development and maturity of trajectory data analysis technology, this technology has been applied in the transportation field. This chapter will introduce the trajectory analysis technique in route recommendation and the application of trajectory data in the taxi industry.

6.1.Path recommendation

With the development of social platforms, many users will share their life on social platforms, such as YouTube, ins, etc., and describe when, where and how they are going to travel. At the same time, the data generated in the sharing can be sorted and analyzed, and finally, personalized travel route recommendations for target users can be formed. This section briefly summarizes how to apply the analysis of trajectory data to the path recommendation, and introduces the MSTP-LSTM algorithm required in the path recommendation.

The Introduction of Path Recommendation: People sharing their lives and describing their favourite travel methods on social platforms will form many new trajectory data. With the emergence of new trajectory data comes new challenges. For example, it is necessary to map the personalized trajectory data of the same user on different platforms to a unified space and judge the actual intention of the user through the sharing posted by users on social platforms. Researchers usually use Multi-modal [18] and Topical Package Model to address these challenges. In addition, researchers also judge users' travel preferences from three main factors: travel cost, travel time and the travel duration.

The path generation needs to be predicted based on the user's past location, so the path generation problem can be regarded as a sequence problem. LSTM plays a great role in dealing with sequence problems. Use 'LSTM' to enter the user's historical location. According to these positions, the related path will be returned as a candidate path. After sorting, a series of corresponding recommendation results will be generated.

MSTP-LSTM algorithm: MSTP-LSTM includes an online module and an offline module (Fig.3). The online module is responsible for generating location points and recommending routes. On the one hand, by mining the average cost, travel time and travel duration of travel notes published by users on social platforms [19], we construct the Topical Package Model. Then, based on the topological package model, the user's topological space and path topological space are constructed to describe the user's travel preferences and path attributes respectively. On the other hand, in the part of online planning, position enhancement is carried out on the position points input by the user to obtain the position sequence, and the position sequence is mapped to the path space. Some paths are returned as candidate paths based on the mapped location sequence. Through path recommendation, candidate paths are ranked by metrics in the Topical Package Model. Finally, a series of corresponding recommendation results are generated. The figure shows the framework of the MSTP-LSTM algorithm. The offline module is responsible for mining POIs to the Topical Package Model and training LSTM.

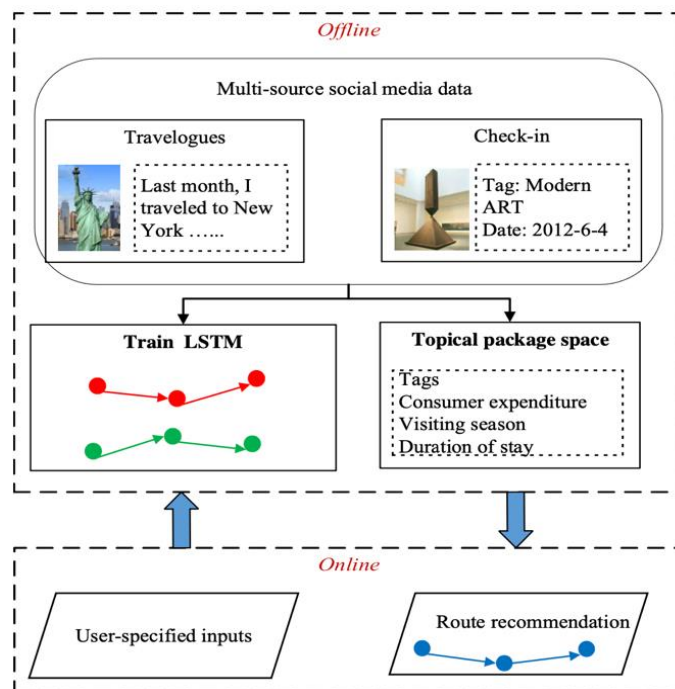


Figure 3. MSTP-LSTM algorithm framework [19]

6.2. Application of trajectory data in the taxi field

Advances in trajectory data analysis techniques are also helping taxi companies plan their employees' work hours and pickup locations. Make your work time-efficient work time instead of aimlessly waiting for guests and wasting costs. After analyzing [20], the researchers analyzed the characteristics of passenger travel time, such as the trend of passenger travel volume over time, the time distribution of travel starting points and the peak hours of travel. These data can help taxi companies to select the right time to increase the number of taxis. Allowing passengers to ride on weekends, during rush hours or after work, moreover, is also beneficial to the profitability of taxi companies.

It is also possible to analyze areas where taxi users are densely distributed through trajectory data, and more taxis can be arranged in these areas. In analyzing the spatial distribution of passenger-carrying hotspot areas, the researchers utilized a three-dimensional map with a top view (Fig. 4) as a display. The top view also uses three different colours to divide the requirements for different levels of taxis in the hotspot area [19].



Figure 4. Distribution of passenger hot spots within 14:00-16:00 in Shanghai

7. Conclusion

With the development of mobile Internet and the progress of big data technology, trajectory data processing has become one of the research focuses. This paper mainly summarizes the four technologies and methods of track data preprocessing, track data mining, track data privacy protection and track data application in the transportation field. This paper mainly introduces the current trajectory data mining scheme and privacy protection strategy, analyzes the development of trajectory data mining and trajectory data privacy protection technology, and extends it to the application of trajectory data in the transportation field. This paper provides help for researchers to quickly and comprehensively understand the key technologies of trajectory data processing and summarizes the application of trajectory data in such fields as trajectory data preprocessing trajectory data mining, trajectory data privacy protection and trajectory data induction. To provide ideas and methods for researchers in the era of big data, the processing of massive trajectory data is complicated systematic work, and the challenges faced by future trajectory data research can be summarized and prospected from the following aspects.

7.1. Trajectory data preprocessing:

The median or mean filtering technology in noise processing can deal with a single noise point, but when dealing with multiple consecutive noise points, neither mean filtering nor median filtering can achieve good results. In this paper, the preprocessing methods of trajectory data are deeply studied and systematically discussed. Combined with the research content of this paper, the author thinks that there is something worthy of further research. How to make the median or mean filtering technology in noise processing well handle single noise point and multiple continuous noise points, and achieve good results at the same time. At the same time, artificial intelligence and big data technologies are combined to obtain new technologies for trajectory data preprocessing.

7.2. Trajectory data mining:

In the direction of trajectory data mining, this paper summarizes three technical methods of GMM, HMM and DBN to process trajectory data. Methods for prediction problems in trajectory data mining are also presented. The technical method of GMM cannot obtain the temporal relationship and order between entities. To solve this defect, researchers use the HMM algorithm to process trajectory data.

7.3. Trajectory data privacy protection:

The advantage of the fake trajectory data method and the suppression method is that the amount of calculation is small, so the implementation is relatively easy. However, shortcomings also exist. Among them, the false trajectory data method has the problem of data distortion, and the suppression method has the problems of lost information and poor data utility. At present, the privacy protection of mobile trajectory data is a research hotspot in the field of space security, but the technology is not mature. Many researchers have done a lot of algorithm and technology research on trajectory privacy, but there are still many key issues that need more in-depth research. In the future, if a more comprehensive method for privacy protection of trajectory data can be developed, it can not only make people safer but also promote the development in the era of big data.

7.4. Trajectory data in transportation:

This paper introduces the application of trajectory analysis technology in route recommendation and trajectory data in the taxi industry. Through sorting and analysis in a large amount of data, route recommendations for target users will be formed. The use of trajectory data in the cab industry can bring maximum benefits to the company by analyzing the trajectory data and arranging the location and number of cab trips by analyzing passenger travel volume and travel trends. In the future, more high-dimensional and complicated trajectory data will be generated, and the difficulty of analysis will

be further increased, therefore, it is necessary to continue to explore more advanced technology methods to solve these challenges.

References

- [1] Q. Gao, F. L. Zhang, R. J. Wang, et al. "Trajectory big data: a review of key technologies in data processing". J. Softw, 2017, 28(4), pp. 959-992.
- [2] S. Wang, Y. W. Peng, H. Lan, et al. "Survey and prospect: Data integration methodologies." J. Softw, 2020, pp. 893-908.
- [3] T. Zhang, B. Yang. "Big data dimension reduction using PCA," 2016 IEEE international conference on smart cloud (SmartCloud). IEEE, 2016, pp. 152-157.
- [4] R. Sowmya, K R.Suneetha "Data mining with big data, " 2017 11th International Conference on Intelligent Systems and Control (ISCO). 2017, pp.246-250.
- [5] F. I. Bashir, A A. Khokhar, D. Schonfeld. "Object trajectory-based activity classification and recognition using hidden Markov models." IEEE transactions on Image Processing, 2007, 16(7), pp.1912-1919.
- [6] J. C. Nascimento, M. A. Figueiredo, Marques J S. "Trajectory classification using switched dynamical hidden Markov models," IEEE Transactions on Image Processing, 2009, 19(5), pp.1338-1348.
- [7] G. Dimitrakopoulos, P. Demestichas "Intelligent transportation systems." IEEE Vehicular Technology Magazine, 2010, 5(1), pp. 77-84.
- [8] L. Santos, K. Khoshhal, J. Dias. "Trajectory-based human action segmentation." Pattern Recognition, 2015, 48(2), pp. 568-579.
- [9] W. Yin, Y. Sun, J. Zhao. "Personalized Tourism Route Recommendation System Based on Dynamic Clustering of User Groups, "2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). 2021, pp.1148-1151.
- [10] C. Chen, S. Zhang, Q. Yu, et al. "Personalized travel route recommendation algorithm based on improved genetic algorithm," Journal of Intelligent & Fuzzy Systems, 2021, 40(3), pp.4407-4423.
- [11] G. Cui, J. Luo, X. Wang "Personalized travel route recommendation using collaborative filtering based on GPS trajectories. International journal of digital earth," 2018, 11(3), pp. 284-307.
- [12] A. Z. Almusaylim, N. Z. Jhanjhi. "Comprehensive review: Privacy protection of user in location-aware services of mobile cloud computing." Wireless Personal Communications, 2020, 111(1), pp.541-564.
- [13] I. Memon, L. Chen, Q A. Arain, et al. "Pseudonym changing strategy with multiple mix zones for trajectory privacy protection in road networks." International journal of communication systems, 2018, 31(1), pp. e3437.
- [14] J. Jia, H. Huang "A trajectory (k, e)-anonymous algorithm against trajectory similarity attacks." Comput. Eng. Sci., 2019, 41(05), pp. 828-834.
- [15] Z. Huo, Y. Huang, X. Meng "History trajectory privacy-preserving through graph partition." Proceedings of the 1st international workshop on Mobile location-based service. 2011, pp. 71-78.
- [16] Q. Gao, F. Zhou, K. Zhang, et al. "Adversarial human trajectory learning for trip recommendation." IEEE Transactions on Neural Networks and Learning Systems, 2021, pp.1-13.
- [17] T. Peng, Q. Liu, D. Meng, et al. "Collaborative trajectory privacy preserving scheme in location-based services." Information Sciences, 2017, 387, pp.165-179.
- [18] S. Gao, J. Ma, C. Sun, et al. "Balancing trajectory privacy and data utility using a personalized anonymization model." Journal of Network and Computer Applications, 2014, 38, pp.125-134.
- [19] J D. Mazimpaka, S.Timpf "Trajectory data mining: A review of methods and applications. " Journal of spatial information science, 2016(13), pp.61-99.
- [20] H. Ji-hua, H. Ze, D. Jun "A hierarchical path planning method using the experience of taxi drivers." Procedia-Social and Behavioral Sciences, 2013, 96, pp.1898-1909.