Study on the Method and Application of Big Data Mining of Mobile Trajectory Based on MapReduce

Jiatong Han

Hohhot Municipal Engineering Technology Service Center, Inner Mongolia, 010023, China

Abstract: In the era of mapreduce when "Internet +" is developed to "Big data x", big data has gradually become a research focus closely followed by the scientific and technological circle, industry circle, and government departments. Big data analysis for moving taxi trajectory has gradually become a research hotspot in the fields of smart city information computing and smart city construction. At present, social problems such as traffic congestion, environmental degradation, and energy shortages are seriously affecting the safe and livable development of smart cities and their sustainable development. Through deep mining, analysis and comprehensive utilization of taxi trajectory data based on geographic location in the mobile social taxi network, it provides a new idea for the analysis of complex urban public transportation problems. This paper will focus on the new data analysis method and its practical application of deep analysis and mining of mobile taxi trajectory big data based on mapreduce. It's dedicated to effectively solve the three major problems of data, including the real-time, robustness and accuracy, and provides theoretical basis and relevant practical technology for the application of urban dynamic monitoring and early warning control of complex urban public transportation network.

1. Introduction

With the progress and development of modern informatics and mathematics technology, people have gradually entered the era with the development direction of "human-machine-thing" big data fusion. In general, big data fusion refers to data system and set that human cannot accurately perceive, acquire, manage, process and provide service for other data with a traditional vocit technology and combination tools of software and hardware within a tolerable time. It has 5 features, including volume (large data volume), variety (various modalities), velocity (fast data generation), value (large data value but low data density), and veracity (real but difficult to accurately identify). In recent years, the fusion of big data has been highly concerned by international academic circles (e.g., *Nature*, the academic journals of European Society for Electronic Informatics and Computer Mathematics and others vercimnews, *Computer Journal*, *Software Journal*), industries (e.g., Google), government departments (such as the United States and China) and other information technology organizations (such as McKinsey and Gardner). [1]

2. Urban computing and big data

Big Data Research Technology is a cross-science technology research method or a tool for the accumulation and discovery of cross-science knowledge. It is closely related to the application of other statistical and artificial intelligence research methods such as data mining, statistical analysis, and search. However, it is also different from the essence and connotation of statistics and other artificial intelligence research technologies. At present, the data framework of adhadoop distributed computing is simple, fast and easy to use, and has developed into a mainstream research technology for big data analysis.

It is an unprecedented challenge and opportunity for the development of smart cities to effectively analyze and use traffic big data to meet the application needs of highly time-sensitive traffic dynamic supervision, traffic warning control and traffic convenience service.

2.1 Big data analysis and mining of mobile terminal development trajectory

In recent years, "City Computing", co-led by founders such as Xie Xing and Zheng Yu of Microsoft Asia Research Institute, has set off a technical research boom based on taxi GPS trajectory data analysis and research and mining. At present, the analysis and research of taxi-based trajectory processing data are mainly focused on "trajectory processing technology and trajectory feature analysis, planning route navigation and improving transportation service, passenger searching strategy recommendation and prediction of the probability for taking taxi, urban dynamic monitoring and traffic flow prediction, urban planning assessment and functional characteristic classification, traffic anomaly detection and accident emergency response, road condition monitoring and urban environment monitoring" and a series of scientific and technical issues.

2.2 Trajectory processing technology and trajectory feature analysis

In the trajectory data collection and management analysis technology of Swedish urban automatic calculation, trajectory motion feature management is usually a widely used trajectory data preprocessing analysis technology, mainly to study how to collect and preprocess the accuracy of rough original moving object trajectory data, etc. Where, for example, map matching, compressible feature of trajectory data, searchability of trajectory data, and mining potential of frequent movement pattern of trajectories. The focus of trajectory motion feature analysis is to study how to quickly extract reliable original mobile vehicles and objects and their trajectories from large-scale noisy original trajectory data, analyze the trajectories and characteristics of mobile vehicles (or crowds), and then perform statistics on and investigate the relationship between the trajectories and characteristics and the road topology, driving habits or other personal behaviors, and lifestyle habits. The literature is based on gpsgps trajectory movement data collected and generated by 50 taxis in 4 cities in Sweden in 6 months. Statistics and analysis of taxi passengers' travel habits and behaviors are performed. The results show that the travel distance of passengers in the riding time approximately complies with the two-segment power law distribution. Literature research results show that the distance and mileage of private vehicles on highways in several major cities are approximately comply with exponential power law distribution, which is not based on the two-segment exponential power law distribution described in the results of this literature. Based on the data of GPS traveling trajectory of urban floating vehicles, this paper establishes a model of the correlation and interaction between different spatial dynamic characteristics of cities, and performs statistics on analyzes the characteristics of urban spatial dynamics. This paper further performs statistics and analyzes the underlying interaction cause and connection between the spatial dynamic characteristics and the temporal dynamic characteristics of the city. It also uses visual techniques to

further analyze the causes and associated time relationships of urban traffic network congestion and emergencies in Beijing. [2]

2.3 Planning route navigation and improving transportation service

All kinds of public service subjects (including government and citizens <including passenger and taxi drivers>) that are closely related to traffic information can fully benefit from the traffic information and services brought to users by the Microsoft Asian Vehicle Traffic Track Data Mining Institute. At present, based on the institute, Yuan Jing and other relevant persons in charge of Microsoft Asia Research Institute has developed a set of t-drive sensor system, which can be used to perceive the traffic flow in real time with the help of taxi drivers equipped with GPS sensors, and design and develop the fastest driving route for ordinary users who need it. The improved version of t-drive of taxi network sensor combines taxi drivers' rich knowledge on urban roads and skilled driving experience, and further takes various influencing factors, including weather, driving habit and driving skills, and familiarity with roads in consideration through comprehensive analysis. The design of personalized fastest driving route for ordinary users is proposed. At the same time, in order to solve the problem of difficulty for hailing taxis, through comprehensive analysis for the time records of getting on and off of passengers, t-finderive system provides users with personalized two-way recommended service of lines for taxi drivers seeking passengers and services for passengers hailing taxi. It can effectively alleviate the difficulty for hailing taxis during off-peak hours, but cannot really solve the difficulties for hailing taxis during peak hours. T-share green bus system comprehensively considers many factors such as the location of taxi passengers, passenger destinations and routes that affect the normal driving of taxis, rationally planning the commute carpooling time and route of taxis. It can effectively solve the difficulty for hailing taxis in peak hours and greatly improve the safety of taxi passengers and the efficiency of commuting and carpooling by designing the management system of real-time dynamic carpooling solution. The persons in charge from the University of North Texas gpsbastani and other universities use the algorithm that analyzing big data with GPS trajectory to develop a small green bus system between taxi and green bus with high flexibility. And it provides recommendations for the reliability of opening green bus lines based on the collection and analysis of big data of the taxi's trajectory analysis.

3. Urban planning assessment and functional characteristics classification

Urban planning is usually formulated by local government departments in accordance with the goal of mid-term and long-term construction and development of a city. Thus, it is difficult to conduct quantitative inspection at fine-grained level and evaluate its importance and effectiveness. Based on the collection and analysis of data on points of interest of vehicle traffic trajectories, it is possible to more accurately learn about and effectively evaluate the overall planning and development of a city, and to find out the possible problems of taxis in urban planning in a timely manner. For example, Zheng Yu, et al. of Microsoft Asia Research Institute have collected and analyzed the connectivity data of taxis in a certain area to more accurately determine the advantages and disadvantages of the area planning, and analyzed the detours and low speed possibly caused by traffic trajectory planning in the data of points of interest of taxi traffic trajectory, as well as detected the possible shortcomings of taxi traffic trajectory planning;

Yuan Jing et al. have combined the data of traffic trajectory points of interest and analyzed people's movement behavior pattern. They identify and classify the regional traffic function and people flow in different urban areas of Beijing by referring to the topics and classification methods of point of interest in the database of in the field of natural language processing technology. For

example, Pan Gang, et al. from Zhejiang University analyze and extract the time series and connectivity change of the number of people's mobile point of interest from the collected data of taxi traffic trajectory point of interest, and find that these time series change data can accurately reveal the relationship between the function and people flow the corresponding area of the city. The data of people flow can accurately reflect the social function of the corresponding area of a city, and divide the functions of different urban areas of Hangzhou in a fine-grained manner. [3]

4. Traffic anomaly detection and emergency response

It's beneficial to strengthen and improve the ability of the local governments to respond to and prevent and control incidents in a timely manner, improve city management and effectively guarantee safety of the city by timely and accurate perception and even early warning of various emergencies (such as natural disasters, mass emergencies, traffic accidents and temporary traffic control, etc.) in the process of China's urban construction. By analyzing the abnormal mobile traffic trajectory of more than 30,000 urban taxis in Beijing, and discovering abnormal traffic events in these cities, the main idea is that when an abnormal event occurs, nearby road traffic and airflow may be disordered quickly to some extent. The causes of abnormalities shall be further predicted and analyzed based on specific traffic routes.

5. Conclusion

Looking at the latest research progress at home and abroad, as the development hotspot of an emerging scientific research and an important scientific research field, taxi movement trajectory data analysis and mining has achieved substantial and meaningful research results, which has effectively promoted the construction of intelligent transportation and the development of smart cities, and provides an effective way to solve urban problems.

References

[1] Xia Dawen. Research on Method and Application of Big Data Mining of Mobile Track Based on MapReduce [D]. 2016.
[2] Yang Jie, Li Xiaoping, Chen Tian. Group mining method based on incremental spatiotemporal trajectory big data [J]. Computer Research and Development, 2014 (S2): 76-85.
[3] Li Xin. Mining trajectory adjoint patterns based on time-space segmentation and word vector similarity [J]. Journal of Sun Yat-Sen University, 2019, 58: 5, 2019, 58 (5): 17-25