

# The Application Research of Operating Vehicle GPS Big Data Mining

Yu Yao<sup>1, a \*</sup>, Chen Ding<sup>2, b</sup>

<sup>1</sup>Key Laboratory of Operation Safety Technology on Transport Vehicles Ministry of Transport, PRC,  
Research Institute of Highway Ministry of Transport,  
Beijing, China

<sup>2</sup>Department of Electronic and Information Engineering, Key Laboratory of Communication and  
Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University  
Beijing, China

<sup>a</sup>yaoyuqyzx@163.com, <sup>b</sup>dingchenqyzx@163.com

**Keywords:** big data; GPS; data mining; operating vehicle

**Abstract.** To support short-term traffic prediction, service people traveling conveniently and effectively, or achieve effective control for traffic state, mining effective information from commercial vehicle GPS data will be a very meaningful research area. The paper review data mining of commercial vehicle GPS information from different industries such as taxi, bus, and long distance bus. Then the paper makes recommendations for improvement, and provides a meaningful reference for improving traffic management level.

## Introduction

In recent years, with speeding up of the urbanization and increasing of the vehicle ownership, different type of the transportation bring people convenient, and also bring many negative effects such as traffic accident, traffic jam, and traffic pollution. In this situation, intelligent transportation system gets people recognition and attention as an advantage technology to solve the traffic problem. As one of the subsystem of the intelligent transportation system, GPS and its application attract more and more attention gradually. Each commercial vehicle (include taxi, bus and long distance bus) which has GPS can collect and upload many types of information such as equipment status, operational status, geographic information, the instantaneous speed and run azimuth by setting time intervals. The attendant problem is to store, manage, analyze, and share the massive data.

## The GPS Data Mining Application of Taxi

Taxi has become an important and common urban transportation. The taxi system based on GPS not only can provide network data of road status, but also can help to relief the traffic jam reasonably. It can provide the commercial data timely and accurately, so that people can improve the level of taxi service. Through the data mining of taxi GPS information, we can get some parameters, such as time, speed, acceleration, and location. It can help to assess the traffic state. Tracking for the running track, we can analysis the best route, and avoid the traffic jam. After all the information have been processed, people can forecast the travel preferences, and provide a better technic support to the city planning. GPS data can also help manager to understand the operational status of the taxi market and the implement effect of the market regulation.

**Traffic Status Estimation Research.** The road condition estimation is an extensive application in taxi GPS data application. Dong et al. research the estimate method of traffic status under long-term GPS data cycle and large number of vehicle. They research the method of the travel time using simulation taxi GPS data under GIS [1]. Zhang et al. propose the method to exclude the singular data by taximeter [2]. Zhang et al. use the city public information platform of Guangzhou, and set the target of storage methods queue mode of road topology. Besides, they also research the information process

method of GPS map matching by region search and border rectangle. They also research the real-time traffic information collection [3]. Jiang et al. design two-advance estimation model of average section speed by GPS data from different type vehicle running characteristics. Moreover, they design a fusion estimation of average speed [4].

**Traffic Behavior Research.** The traffic behavior research based on taxi GPS data include residents travel behavior research and driving behavior research. Li et al. according to the research of taxi GPS track and driver experience, propose an experience method of calculating the optimal route [5]. Tong et al. analysis residents travel behavior through the taxi GPS data. They research the method from working day and rest day, based on the data from the shortest distance data-matching algorithm and pretreatment method to process data. They deep analysis the total daily trip volume, travel time, and taxi unload ratio etc. [6].

**Travel OD Prediction Research.** Travel OD prediction will process GPS data firstly. It will match the GPS anchor point and the location in the map. Then it will extract the GPS position coordinate of passenger getting on or off the taxi. Using different algorithm to divide traffic area, and get the prediction result of OD matrix. Lv et al. optimize the travel data, and extract the position coordinate of passengers getting on or off the taxi. They divide the traffic area by k-means [7]. Zhang et al. collect resident travel information, such as position and time, through researching the GPS equipment data to post processing GPS data [8]. Alvarez proposes a Hidden markov model system based on GPS log and the current position to predict user's destination [9].

**Taxi Operation Management Research.** The taxi operation management is a very important part of the taxi GPS data research. Using the data mining technology can provide the necessary technologic supporting for the enterprise operation. Weng et al. use the GPS data to collect taxi operation incident and status information. They improve the system of wide-area floating car data processing method, and propose the taxi operation index calculation model based on floating car data [10]. Yang et al. propose the taxi capacity planning model based on the average operating income, unloaded ratio, average trip distance, average running velocity, mean work time, and other city environment estimated parameters. Besides, they also use the taxi operation income to restrain the model [11].

## The GPS Data Mining Application for Bus

In recent years, bus as an important part of the urban traffic, have already become a necessary transportation way for people traveling. For data mining using GPS data, analyzing and researching bus operation situation, analyzing bus condition, and researching the bus OD, they are all important to plan the bus, establish the timetable, and improve the public transport service quality.

**The analysis of Service Time.** The research of operating time division began in the 1980s. Salicru et al. conclude the research [12]. Carey researches randomness of one-stage operation time, and merger it into total operating costs. The researches of period division are less in China [13]. Xu et al. use the greedy strategy to divide operation time based on the vehicle positioning data. Moreover, in their recent paper, they use the K-means clustering algorithm to analyze the bus operation time. They combine with the characteristic of the bus sample data, and improve the K-means clustering algorithm. They improve the traditional cluster center selection method, and design a cluster center updating algorithm that using triangle inequality to decrease unnecessary distance calculation and the cluster updating algorithm based on fuzzy clustering [14].

**The Analysis of Bus Condition.** Now a day, for the vehicle ownership increases year by year, the environment pollution become more and more seriously. Therefore, researches of the bus performance can give an important technology support to energy conservation and emission reduction or other aspects of the research. Hu et al. collect the Harbin typical bus lines operating conditions data, use PCA and cluster technology to process condition data, build the bus condition of Harbin. It can provide some certain basis for matching public bus power system and optimizing the control strategy [15]. Wang et al. collect the bus system data of Beijing. Moreover, they use cluster to preliminary analyze the sample data. Then, they confirm the necessary data sample, and build the bus condition. After that, they use the condition to contrastive analyze the dependency of this condition and Europe traveling condition [16].

**Bus OD Research.** The amount of bus OD is very important basic data for urban bus network planning and management. The traditional method needs a wide range of personal travel situation investigation. Because of the big difficulty and large cycle length, acquiring the investigation data cannot reflect the passenger flow distribution correctly. Moreover, the update cycle of data is too long to optimize and adjust the bus network. However, through extract bus GPS and IC card information, using the technology of data mining and data fusion, we can build bus OD computational formula model much more effectively. Xu et al. propose a data fusion theory based on bus GPS and bus IC card. They use the total quantity control preferably solve the total quantity error problem when calculating the people number of getting of the bus. Moreover, they build single line bus OD generation computational model [17]. Wu build the data warehouse structural system based on bus IC card data, and build the data warehouse structural system based on bus IC card data and bus GPS data. Besides, the data warehouse structural system he build include data warehouse module, data mining tools module, data logical analysis module, and user control module, these four constructions [18]. Zhang introduce the dynamic bus passenger flow data collection method, and on this foundation, she builds the data warehouse based on bus IC card data and bus GPS data. After that, she proposes a data mining system application platform with four constructions. Based on the SPSS Clementine foundation platform, she builds the platform, which applies to bus IC card [19].

### **The GPS Data Mining Application of Coach**

With the development of the transportation, whether passenger or freight transportation largen rapidly. The safety problem of the operating vehicle more and more become a focus problem which noticed by people. For the discordant problem of each factor for traffic route, the vehicle safety problem become more and more seriously. With the rapid development of the GPS technology, an increasing number of the operating vehicle install the GPS security monitoring and control system. The GPS system returns a large number of data continuously. It provides a foundation for management function department to acquire decision support information. Through extracting the GPS data and using data mining technology, it can provide a useful technology support for dynamic monitoring.

**Over Speed Point Analysis.** In the numerous factors of the road transportation safety accidents, operating vehicles over speed driving is one of the main reasons. It is very meaningful for strengthening the road transport safety management that deep analyzing and checking operating vehicle over speed multipoint section, reveals over the spatial distribution pattern of speed incident comprehensively. Liu et al. aim at the problem for the process of operating vehicle over speed point data mining, propose a new cluster method. According to the GPS real-time monitoring data, the method analyzes over speed multiple points. Through search all over speed incident in neighborhood of the over speed point, the method looks for the over speed density sub-threshold point or location, and build density reachable maximum value over speed point cluster, the method looks for the over speed density sub-threshold point or location, and build density reachable maximum value over speed point cluster. In the same time the method use simple adjacency list to replace R\*-tree, and simplify the process of building data structure, reduce memory usage [20].

**Driver Behavior Research.** GPS data include speed, location, running time and so on. We can use this information to analyze driver behavior through data mining, thus detect the abnormal driving behavior or traffic offence. In this way, can we guarantee the operating vehicle safety driving, and be convenient for enterprise or industry to supervise. Niu et al. research the dangerous driving behavior of operating driver. They use the fuzzy clustering and hierarchical clustering to analyze the level dividing of dangerous driving behavior. Besides, they discuss the conclusion of two method through the example of operating driver dangerous behavior [21]. Mao researches the identification of fatigue driving. Through collecting the information of speed, accelerated speed, and other sensor data, using PCA and cluster analyze the fatigue driving [22]. Siyuan Liu et al. through GPS data on taxi, use clustering algorithm to find taxi fraudulent behavior [23].

## Proposals for the Existing Research

Big data mining research of the operating vehicle GPS is a multidisciplinary application, including traffic engineering science, geographic information science, data mining science, database theory and so on. According to the summary and analysis of existing references, this paper proposes two recommendations for improvement as following.

For taxi, the existing references limitedly focus on the research for extracting the operating information instead of improving the intelligent management of taxi by data mining. At present, taxi management systems, built in Beijing, Shanghai and other big cities, are only in the primary display and simple analysis phase; therefore fully mining the wealth of GPS information of taxi is the only way for intelligent management. This paper suggests analyzing the change of taxi traffic demand, according to the taxi distributed condition on the road network; thus, reasonably layout the taxi stops to add a taxi service mode and passenger waiting form. In addition, with the rapid development of intelligent transportation system and web of internet technology, the taxi intelligent scheduling and reservation system has gradually gone deep into the city life. Factors such as service level of taxi by phone, passenger-waiting time and the reliability of reservation information will determine the life or death of the system application.

For coach, safety is the lifeline of development of road passenger transport industry, and energy saving is the lifeblood of efficiency of road passenger transport industry. Existing references are mostly from a security point of view and rarely attempt to improve operational economy by digging information deeply. This paper suggests adding a CAN data acquisition module to the GPS device on the coach and collecting the real-time fuel consumption and dynamic vehicle conditions in the whole process. It is achievable to find out the reasons from both coach and the driver by mining the CAN and GPS data, which cause high fuel consumption, and then develop effective measures to reduce fuel consumption.

## Acknowledgment

This work is supported by research projects of Soft Science "Comparative Research on Security Management Policies of Road Transportation in China and Western Countries", Ministry of Transport(2014-312-223-410), Apply and Fundamental Research Project "application and research of road transport dynamic monitoring data analysis method", Ministry of Transport(2014-319-223-240).

## References

- [1] Tong Xiaohua, Chen Jianyang, Travel Time Estimation and Simulation Model Based on Taxi-Based GPS and GIS, *J. Journal of Tongji University(Natural Science)*. 12 (2005) 1604-1607.
- [2] Zhang Hesheng, Zhang Yi, Wen Huimin, Hu Dongcheng, Estimation approaches of average link travel time using GPS data, *J. Journal of Tongji UniversityJournal of Jilin University (Engineering and Technology Edition)*. 3 (2007) 533-537.
- [3] Zhang Wei, Xu Jianmin, Zhang Zi, The Dynamic Traffic Information Collection and Processing Methods of ITS Common Information Platform Based on GPS Equipped Taxi Data, *J. Journal of Highway and Transportation Research and Development*. 5 (2007) 125-129.
- [4] Jiang Guiyan, Chang Ande, Li Qi, Yi Feng, Estimation Models for Average Speed of Traffic Flow Based on GPS Data of Taxi, *J. Journal of Southwest Jiaotong University*. 46 (2011) 638-644.
- [5] Li Qingquan, Zeng Zhe, Path-finding through flexible hierarchical road networks: An experiential approach using taxi trajectory data, *J. International Journal of Applied Earth Observation and Geoinformation*. 1 (2011) 110-119.
- [6] Tong Xiaojun, Xiang Nanping, Zhu Dingju, The Analysis of City Dweller's Travel Choice Behavior Based on Taxi GPS Trajectory Data, *J. Computer & Telecommunication*. 1 (2012) 56-59.

- [7] Lv Yuqiang, Qin Yong, Jia Limin, Dong Honghui, Jia Xianbo, Sun Zhiyuan, Path-finding through flexible hierarchical road networks: An experiential approach using taxi trajectory data, *J. Logistics Technology*. 9 (2010) 86-88.
- [8] Zhang Junfeng, *Research on OD Survey Based on the Global Positioning System*, Beijing Jiaotong University, Beijing, 2011.
- [9] ALVAREZ GARICIA. Trip destination prediction based on past GPS log using a Hidden Markov Model, *J. Expert Systems with Applications*, 37 (2012) 166—8171.
- [10] WENG Jian-cheng, LIU Wen-tao, CHEN Zhi-hong, RONG Jian, Research on Floating Car Data Based Taxi Operation and Management, *J. Journal of Beijing University of Technology*. 36 (2010) 779-784.
- [11] YANG Yingjun, Zhao Xiangmo, Schedule Model of Urban Taxi Quantity Based on Taxi Running Information, *J. China Journal of Highway and Transport, China Journal of Highway and Transport*, 25 (2012) 120-125.
- [12] Salicru M, Fleurent C, Armengol J M, Timetable-based operation in urban transport: Runtime optimisation and improvements in the operating process, *J. Transportation Research Part A*, 45 (2011), 721-740.
- [13] Carey M, Reliability of interconnected scheduled services, *J. European Journal of Operational Research*, 79 (1994), 51-72.
- [14] Shen Yindong, Zhang Tonghui, Xu Jia, Homogeneous Bus Running Time Bands Analysis Based on K-means Algorithms, *J. Journal of Transportation Systems Engineering and Information Technology*. 2 (2014) 87-93.
- [15] Hu Chen, Wu Xiaogang, Li Xiaojun, Wang Xiaoguang, Construction of Harbin City Driving Cycle, *J. Journal of Harbin University of Science and Technology*. 1 (2014) 85-89.
- [16] Wang Chu, Han Xiukun, Ge Yushan, Wang Meng, Tan Jianwei, Liu Zhihua, Construction of Typical Driving Cycle for Public Bus in Beijing City, *J. Automotive Engineering*. 8 (2010) 704-706.
- [17] Xu Jianmin, Xiong Wenhuan, YouFeng, The Method of single-line Public transportation OD Base on GPS and Intelligent Card, *J. Microcomputer Information*. 24 (2008) 221-223.
- [18] Wu Xiangguo, *Urban Public Transportation Trip OD Matrix Inference and Application Based on Bus IC Card Data and GPS Data*, Shandong University, Shandong, 2011.
- [19] Zhang Yu, *Dynamic Public Transit Origin-Destination Estimation Based on Data Mining*, Beijing Jiaotong University, Beijing, 2010.
- [20] Niu Zengliang, Li Haibin, Wang Wenfeng, Bai Lixia, Study on Risky Behavior of Commercial Drivers Based on Clustering Analysis, *J. Journal of Shandong Jiaotong University*. 1 (2014) 19-23.
- [21] Liu Weining, Zeng Chanjuan, Sun Dihua, Clustering Analysis of Overspeed Spots for Commercial Vehicles Based on DBSCAN, *J. Computer Engineering*. 5 (2009) 268-270.
- [22] Mao Zhe, *Research on Identification of Fatigue Driving Behavior*, Wuhan University of Technology, Wuhan, 2009.
- [23] Siyuan Liu, Lionel M. Ni, Ramayya Krishnan. Fraud Detection From Taxis' Driving Behaviors, *J. IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, 63 (2014) 464-472.