Smart Innovation, Systems and Technologies 356

Yiming Bie
Kun Gao
Robert J. Howlett
Lakhmi C. Jain *Editors*

Smart Transportation Systems 2023

Proceedings of 6th KES-STS International Symposium

KES International

Springer

Smart Innovation, Systems and Technologies

Volume 356

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago, DBLP.

All books published in the series are submitted for consideration in Web of Science.

Yiming Bie · Kun Gao · Robert J. Howlett ·
Lakhmi C. Jain
Editors

Smart Transportation Systems 2023

Proceedings of 6th KES-STS International Symposium

Springer

Editors
Yiming Bie
Jilin University
Jilin, China

Kun Gao
Chalmers University of Technology
Göteborg, Sweden

Robert J. Howlett
KES International Research
Shoreham-by-sea, UK

Lakhmi C. Jain
KES International
Selby, UK

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-981-99-3283-2 ISBN 978-981-99-3284-9 (eBook)
https://doi.org/10.1007/978-981-99-3284-9

Preface

Smart and sustainable transportation systems are becoming increasingly important in the face of growing concerns over climate change and air pollution. Advanced technologies such as electric vehicles, connected and automated vehicles, and big data are at the forefront of the evolutions of improving efficiency, safety, and accessibility of transport systems. Electric vehicles, in particular, have gained widespread attention as a sustainable alternative to traditional fossil fuel-powered vehicles. Connected and automated vehicles offer the potential to improve safety, efficiency, and sustainability, while big transport data provides a wealth of information that can be used to optimize transportation systems in a data-driven way. However, there are still challenges that need to be addressed, such as breakthroughs in methodologies and applications to fully leverage the potentials of abovementioned techniques and ensuring equitable access to transportation services. Innovative solutions and collaboration among various stakeholders, including industry, academia, and government, are essential to fully facilitate the prosperous development of smart and sustainable transportation systems. In this regard, the 6th International Symposium on Smart Transport Systems is going to be organized in June 2023 to provide a favorable communication environment, exchanging knowledge and collaborative platforms among researchers and practitioners in the fields of smart transport systems. The vision is to facilitate complementary collaborations among academic and industrial communities in terms of research, implementation, and applications.

In 2023, the international symposium finally accepted 21 papers from research institutions and countries worldwide. Each paper underwent a comprehensive peer-review process consisting of two rounds and at least two external reviewers and one editorial member. Reviewers and authors engaged in joint dialogues to improve papers jointly from different perspectives. The accepted papers cover a wide range of topics on smart and sustainable transportation systems. Three papers focus on optimizing and managing electric vehicle operation and speed controls to reduce operational costs and improve energy efficiency. Another four papers address traffic safety analysis and behavior modeling, including human drivers' speeding behavior, user interface, and traffic accident analysis. Four papers investigate the advancement

in utilizing autonomous vehicles, consisting of research about environmental perceptions, coordination strategy, and speed controls. Meanwhile, another four papers investigate the data requisition using emerging data collection tools and utilizing trajectories data to predict vehicles' trajectories, which are indispensable inputs for the control of autonomous vehicles. Two more papers examine traffic congestion prediction using big data and machine learning. The remaining papers cover different topics in the field of smart transport systems such as on-demand public transit, activity-based modeling for transport planning, efficient operations of trucks, etc. The authors, mainly scholars and practitioners from Sweden, Germany, Mexico, China, and Tunis, had the opportunity to present their research outcomes, exchange knowledge, and network in a relaxed and pleasant atmosphere during the symposium.

Jilin, China — Yiming Bie
Göteborg, Sweden — Kun Gao
Shoreham-by-sea, UK — Robert J. Howlett
Selby, UK — Lakhmi C. Jain

Contents

About the Editors

Yiming Bie is Professor of Transportation Engineering with the School of Transportation, Jilin University, China. He obtained his Ph.D. degree at Jilin University in June 2012. His research interests include public transportation operations, traffic control, and intersection design. He has authored or co-authored over 70 journal articles at top-tier journals such as Transportation Research Part C, Computer-Aided Civil and Infrastructure Engineering, Journal of Transportation Engineering-ASCE. The algorithms developed by him were adopted by the most popular adaptive traffic control system in China and has been implemented to more than 20 cities. Recently, his research is focused on electric bus operations in cold regions. He is Principal Investigator for two projects funded by National Natural Science Foundation of China and a few projects from other funding agencies. In 2019, he was conferred the outstanding reviewer recognition by Journal of Transportation Engineering-ASCE.

Kun Gao is an Assistant Professor in the Urban Mobility Systems research group, Department of Architecture and Civil Engineering. His research focuses on promoting sustainable mobility with focuses on electrification, shared mobility, and connected automation. Special interests are attached to establishing new approaches and tools for system planning, optimization and evaluation of emerging transport systems leveraging big data and machine learning. The overall goal is to facilitate the development of a safer, more sustainable, and equitable transportation system.

Dr. Robert J. Howlett is Executive Chair of KES International, a non-profit organization that facilitates knowledge transfer and the dissemination of research results in areas including Intelligent Systems, Sustainability, and Knowledge Transfer. He is Visiting Professor at Bournemouth University in the UK. His technical expertise is in the use of intelligent systems to solve industrial problems. He has been successful in applying artificial intelligence, machine learning, and related technologies to sustainability and renewable energy systems; condition monitoring, diagnostic tools and systems; and automotive electronics and engine management systems. His current research work is focused on the use of smart microgrids to achieve reduced energy costs and lower carbon emissions in areas such as housing and protected horticulture.

Dr. Lakhmi C. Jain Ph.D., ME, BE(Hons), Fellow (Engineers Australia), is with the University of Technology Sydney, Australia, and Liverpool Hope University, UK. She serves the KES International for providing a professional community the opportunities for publications, knowledge exchange, cooperation, and teaming. Involving around 5,000 researchers drawn from universities and companies worldwide, KES facilitates international cooperation and generates synergy in teaching and research. KES regularly provides networking opportunities for professional community through one of the largest conferences of its kind in the area of KES.

Chapter 1
Research Progress on Key Technologies of Environmental Perception and Detection of Transport System in Goaf of Coal Mine

Bo Lu, TaiXue Bei, Zhiwei Meng, Ying Luo, Lei Zhao, Nan Xu, Jianhua Liu, and Fayi Yan

Abstract A large number of mined out areas have been formed after the continuous mining of minerals, resulting in many potential safety hazards. With the continuous development of artificial intelligence and machine learning in recent years, many researchers have put forward many creative ideas on the intellectualization of goaf transportation system. Lidar is widely used in three-dimensional imaging technology in recent years because of its advantages of small size, high accuracy, and fast imaging. This paper systematically summarizes the contributions and problems of researchers in recent years from three aspects: autonomous navigation of UAVs, point cloud processing, and radar accuracy improvement. The development direction of goaf detection technology in the future is summarized to provide a reference for the research in related fields.

1.1 Domestic and International Status Quo

The goaf formed after the mining of mineral resources is one of the main sources of mine disasters. The goaf is a "cavity" caused by artificial mining or crustal movement. It is difficult to predict the situation below the surface. During the mining period, people and equipment may fall into the goaf at any time. Therefore, it is imperative to accelerate the research on the detection technology of the goaf and quickly master the internal structure of the goaf with modern technology [1]. At present, the traditional drilling method is mainly used for the detection of mined-out areas, high-density

B. Lu · T. Bei (✉) · Y. Luo · L. Zhao · N. Xu · J. Liu · F. Yan
School of Mechanical and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China
e-mail: 13709@sdjzu.edu.cn

Z. Meng
State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130022, China

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_1

resistivity surveying [2], Engineering geophysical prospecting [3]. The drilling method uses drilling equipment and drilling tools to detect the preset exploratory hole location, which is labor-intensive and time-consuming. The high-density resistivity method is affected by such complex factors as the goaf environment and electromagnetism, which leads to the problems of limited signal propagation distance, interference, and refraction of signals. Geophysical prospecting refers to the application of geophysical exploration technology to the exploration of goaf. Li Yongming, Pan Dongming, et al., proposed that ground penetrating radar with different frequencies is used to detect the goaf, and the general shape of the goaf is obtained from the reflected image of electromagnetic wave. Li Guangxu, Gu Hanming, et al., described that the ground penetrating radar method and seismic imaging method are comprehensively used to detect the karst cavity, and the occurrence state of the cavity is obtained.

In Europe, America, and other countries, there are much researches on the detection technology of underground goaf, mainly focusing on physical exploration. The main technical means used include the high-density resistivity method, three-dimensional seismic method, transient electromagnetic method. In Britain, France, Italy, and other countries, there are many researches on ground penetrating radar and laser detection technology, and the microgravity method and shallow seismic method are also widely used. In Russia, the seismic reflection wave method, a transient electromagnetic method, gas emission measurement method, and cross-well electromagnetic wave transmission are commonly used in geophysical exploration technology. The above methods are conducive to rapid and automatic data acquisition, but they are easily affected by the location of the mining area, the distribution distance of the detection equipment, and the depth of the goaf, resulting in poor detection results.

Compared with traditional detection methods, 3D laser scanning technology can effectively improve the work efficiency and detection accuracy of goaf safety detection. The 3D laser scanning technology adopts the laser ranging method. By measuring the 3D coordinate value, reflectivity, and texture information of each point in the goaf, it can quickly scan the surface of the goaf to obtain the 3D point cloud data. It has the characteristics of fast detection speed, high accuracy, and a strong sense of reality of the goaf model. It has great development prospects in many fields. Miller et al. [4] in 1992 proposed that an automatic laser scanning system was installed on the probe rod for the first time to detect the excavated area; Li Jielin et al. [5] in 2020 presented that the lidar will be carried on the UAV (Unmanned Aerial Vehicle) for the first time to detect the goaf with complex shape.

The way of using UAVs to carry detection equipment is gradually applied to the field of coal mine detection. Different from the indoor environment mapping, the coal mine environment is bad, the signal is poor, and there is much dust. These adverse factors put forward higher requirements for UAV autonomous navigation, point cloud quality, and radar accuracy. This paper analyzes the tasks and challenges faced by the navigation and positioning technology, point cloud processing technology, and radar accuracy improvement technology at this stage, and looks forward to its development trend in the goaf environment.

1.2 UAV Autonomous Navigation

1.2.1 Navigation and Positioning Technology

Autonomous navigation technology refers to the real-time positioning and environment perception of UAVs based on the goaf environment. In recent years, with the development of synchronous positioning and map building (SLAM) technology, multi-sensor fusion SLAM, semantics-based and depth-based learning SLAM technology have been derived from traditional SLAM technology, which greatly improves the positioning accuracy and robustness [6]. The technical framework for autonomous navigation of UAV in goaf is established through the study of SLAM technology, as shown in Fig. 1.1.

At present, the positioning technology applicable to the goaf environment is mainly divided into the following categories: positioning technology based on wireless communication, positioning technology based on machine vision, and positioning technology based on multi-sensor fusion [7].

Nikhil Deshpandea et al. [8] proposed that in the environment a wireless sensor node and a wireless sensor network are built. The directional antenna on the robot receives the signal strength. Finally, the orientation estimation based on the particle filter is used for positioning. The robot successfully realizes autonomous navigation and obstacle avoidance, which provides a reference for UAV flight. Lv Hualong of Tianjin University designed a UAV autonomous perception and positioning system based on binocular vision and improved the algorithm, which can improve the speed

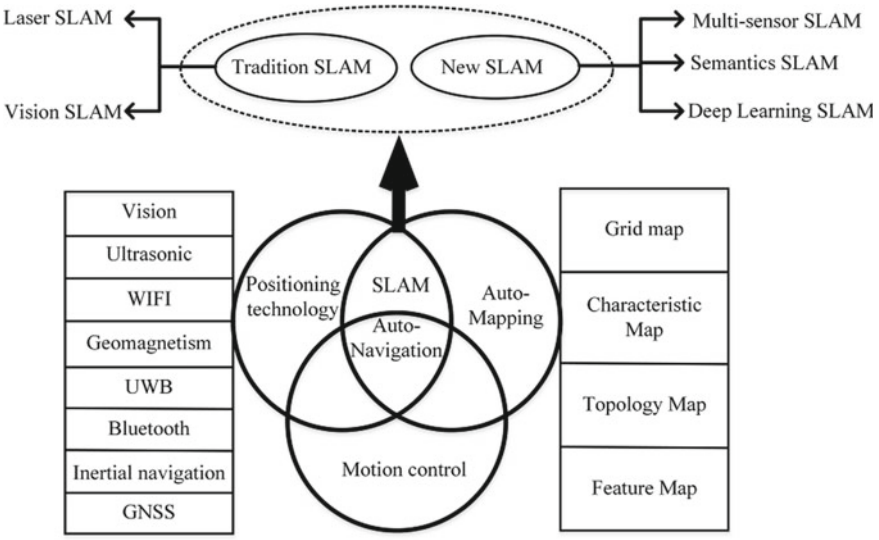


Fig. 1.1 UAV autonomous navigation technology framework

Table 1.1 Comparison of positioning methods

Signal carrier	Positioning method	Accuracy	Advantage	Shortcoming
Visual signal	Visual positioning [10]	Submicron scale	Wide detection range	High processor requirements
Inertial navigation	Dead reckoning	Meter	Data stability	With accumulated error
Bluetooth	Ranging intersection method	Meter	Simple principle and low power	Limited positioning range
WIFI	Fingerprint matching method	Meter	Low cost and easy to build	Limited positioning range
Radar	Pole method [11]	Submicron	High precision	Low application and high cost

and accuracy of UAV trajectory planning [9]. At present, indoor positioning can be roughly divided into two categories, positioning based on the external signal source and positioning based on the natural signal source. External signal source positioning includes WiFi, Bluetooth, RF with low positioning cost and convenient and fast deployment. Natural signal source positioning includes inertial navigation positioning and geomagnetic field positioning. Table 1.1 shows the comparison of positioning methods.

When using a single positioning method, there are some shortcomings, such as low efficiency, poor accuracy, being greatly affected by the environment, and poor robustness. In recent years, many researchers began to use multi-sensor fusion positioning technology to ensure the reliability of UAV operation and maximum obstacle avoidance [12].

1.3 Laser 3D Imaging Radar Technology

1.3.1 Principle and Composition

The laser 3D imaging system is mainly composed of a 3D laser scanner, a computer, power supply, and bracket. Its main component is a three-dimensional laser scanner, which is composed of a laser transmitter, a receiver, a timer, a filter, a circuit board, and a CCD. It was used initially in the field of surveying and mapping. Because it can quickly acquire the three-dimensional coordinate data of the target surface, many experts and scholars have gradually transferred the three-dimensional laser scanning technology from single-point measurement to surface measurement in recent years, in the coal mining, construction, and military fields. Accident handling and other fields have been widely used. Its principle is laser ranging. The principle of three-dimensional laser scanner based on the pulse method is shown in Fig. 1.2, with a

measuring range of up to 5 km. UAVs are often used to carry out goaf detection. The detection platform is shown in Fig. 1.3. Through a large number of point cloud data, a three-dimensional coordinate reflectivity and other information on the surface of the object to be measured, the three-dimensional model of the object to be measured can be quickly presented.

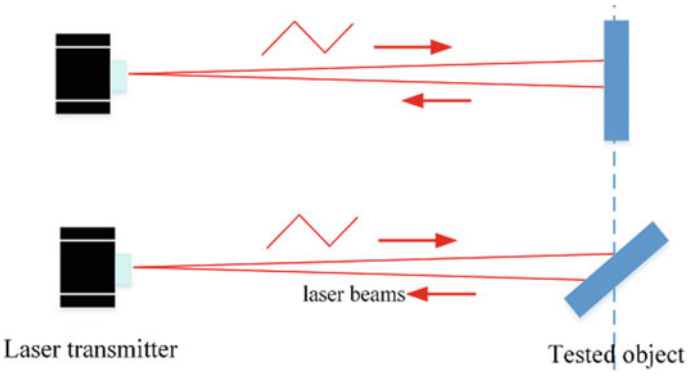


Fig. 1.2 Schematic diagram of pulse laser ranging

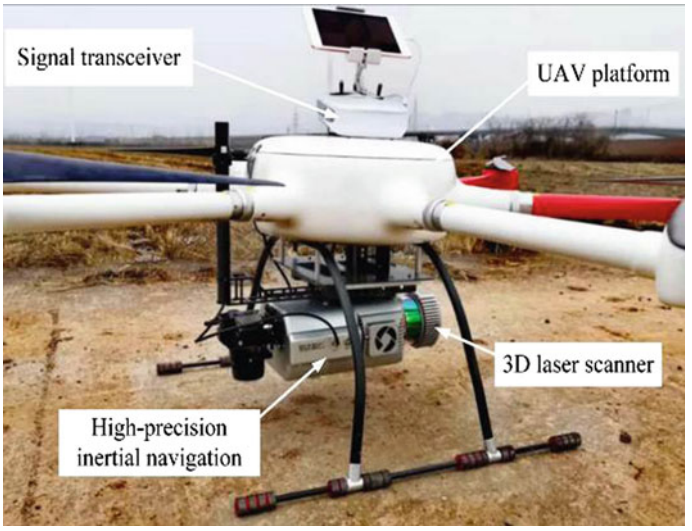


Fig. 1.3 UAV detection platform

1.3.2 Point Cloud Processing

A large number of data points on the surface of the empty area obtained in the UAV detection and mapping are called point clouds. With the continuous improvement of the performance of the lidar, the efficiency of obtaining point clouds is also increasing. The complex environment of the mined-out area and the accuracy of the equipment will make the obtained point cloud data noisy, so it is necessary to denoise and register the point cloud.

1.3.2.1 Point Cloud Denoising Technology

The noise of the point cloud will cause the loss of image edge details. The noise reduction of the point cloud is an effective method to improve detection accuracy. Three-dimensional point cloud denoising algorithms mainly include methods based on partial differential equations, methods based on signal processing, methods based on neighborhood filtering, and methods based on projection. Table 1.2 shows the common denoising methods and their differences. Abbreviations: MLS (Moving Least Squares), LOP (Local Optimal Projection), WLOP (Weighted Locally Optimal Projection), FLOP (Feature-preserving Locally Optimal Projection operator), NPD (Neural Projection Denoising).

Noise reduction of point clouds has been studied by most scholars, but most of them are grid models. The grid model itself will also generate some noise, and the algorithm itself will lead to distortion and contraction of the results, so it is not ideal. Establishing a 3D point cloud model for direct filtering has certain advantages over the Grid model, and how to improve the robustness of its filtering algorithm still has great challenges. With the rapid development of in-depth learning in recent years, since in-depth learning is data-driven and not controlled by a single model, it has significant advantages in processing point cloud data. Some scholars convolute the volume grid of the point cloud transformation process, while others propose to convolute point cloud data directly, such as PointNet、PointNet++ [13], NPD is precisely trained in the early stage by means of deep learning, which greatly improves the noise reduction accuracy and is the future development trend.

1.3.2.2 Point Cloud Registration Technology.

Point cloud registration is also called point cloud splicing, which is the primary work in the data processing. In order to get a complete 3D point cloud model, point cloud registration technology is needed because part of the real image information is also removed when point cloud denoising is performed to remove interference points. At present, there are two popular algorithms: the iterative closest point (ICP) algorithm and the normal distributions transform (NDT) algorithm. Tables 1.3 and 1.4 respectively list their research progress and make a summary and comparison.

Table 1.2 Comparison of common noise removal methods

Classification	Literature	Usage method	Performance
Methods based on partial differential equations	Clareng	Local finite element	Ideal for surface light adaptation
	Lange et al	mcvf	Better enhancement of set features
	Xiao	Dynamic mean curvature flow	Superior to bilateral filtering method
Method based on signal processing	Taubin	Laplace filtering	Ensure amplitude and phase
	Pauly and Gros	Spectrum analysis	Ensuring boundary curvature and stability
A method based on neighborhood filtering	Fleishman	Improved bilateral filtering	Suppressing shrinkage
	Choudhury	Trilateral filtering	Better than non-local noise reduction
	Miu Y	Non-local denoising methods	Feature preservation
Projection-based approach	Lipman	LOP	Better than MLS
	Huang	WLOP	No normal estimation
	Liao	FLOP	Accelerate LOP
	Duan C	NPD	Low mean square error

A single algorithm has more or fewer shortcomings. For example, the ICP algorithm takes too much time and has too many iterations, which leads to the algorithm falling into the local optimal problem. When the initial position and attitude are not ideal, only the local optimal configuration can be achieved, and it is also vulnerable to noise. Compared with the ICP algorithm, the NDT algorithm has faster speed, a larger convergence domain, and better robustness, but it has the problem of unreliable attitude solutions due to sparse point clouds. A single radar also has some disadvantages, such as fewer data and sparse point clouds. Therefore, algorithm fusion and sensor fusion are hot research directions in the future.

Table 1.3 ICP and its improved algorithm

Literature	Algorithm	Year	Improvement points
Besl [14]	ICP	1992	Point to point
Masuda	ICP	1996	LMedS replace LS
Lu F [15]	IDC	1997	User polar coordinate distance
Trucco [16]	RICP	1999	Adopt random sampling method
Godin [17]	ICPIF	2001	Take color and curvature as European invariant features
Luo xianbo	ICP	2004	Incoming label point
Minguez J [18]	MbICP	2006	New distance metric
Yoshitaka H [19]	Intensity-ICP	2007	Introducing point cloud strength information
Segal	G-ICP	2009	Face to face, use the covariance matrix
Hong S	VICP	2012	Data distortion correction
Yang J L	GO-TCP	2013	Branch and bound method
Alismail H	CICP	2014	A continuous time estimation method
Serafin J	NICP	2015	Highlight local characteristics
Li Renzhong	ISS-ICP	2017	Feature extraction using ISS algorithm
Pavlov	AA-ICP	2018	Introducing Anderson acceleration
Liang Siyao	SICP	2021	Adding dynamic iteration factor

Table 1.4 NDT and its improved algorithm

Literature	Algorithm	Year	Improvement points
Biber et al. [20]	NDT	2003	2D planar point cloud data matching
Magnusson M	3D-NDT	2007	Apply to 3D space
Magnusson M	P2D-NDT	2009	Gaussian distribution
Stoyanovy D	D3D-NDT	2012	Normal distribution
JariSaarinen et al	NDT-OM	2013	Recursive update strategy
Zhang xiao et al	NDT	2014	SURF algorithm
Chen Qingyan	3D-NDT	2015	Introducing curvature features
Fu Yiran	NDT	2018	Improved ICP fusion NDT
Mu Lili	NDT	2021	Genetic algorithm with NDT fusion

1.3.3 Radar Accuracy Improvement

The high concentration of dust particles in the goaf environment refracts the laser energy to a certain extent, which is the main reason for the decline of the lidar accuracy. In addition to improving the accuracy of point cloud processing, some scholars have also done a lot of research on the lidar itself. Li Xianglian effectively combined the liquid crystal phased array and the single slit streak tube for the stability of the mechanical beam, which improved the scanning accuracy and distance. Chen Zhen

put forward the concept of polarization modulation, using this method to transmit a pulse can obtain a complete three-dimensional map. Wu Yuhao used UAV to carry lidar, IMU, and GPS to detect the shoal environment and used OpenCV to extract stripes from the collected images, which improved the average resolution to the centimeter level, providing a new idea for goaf detection. Cui Zihao considered the variability of the marine environment and studied the triangulation method and streak tube method of airborne lidar. Among them, the triangulation method is applicable to short distances, and the streak method is applicable to long distances. Westfeld et al. [21] proposed that underwater 3D laser detection has been studied, and the results show that triangulation distance measurement has higher accuracy for short-range imaging, which can reach above a millimeter level. Yao Jinliang proposed a method to improve ranging accuracy by registering and superimposing range image sequences for airborne radar.

1.4 Development Trend

For the unmanned working face in the goaf, it is necessary to ensure the real-time transmission of information and the accuracy of positioning. At present, positioning technology at home and abroad has great limitations. The main development trends can be summarized as follows:

(1) UAV dynamics model. The goaf environment is complex and changeable and faces the risk of machines falling into the goaf at any time. Therefore, when building the dynamic model, it is necessary to reasonably determine the instantaneous speed, acceleration, and other motion parameters of the detector, and improve the endurance time. Before conducting field detection, it is necessary to establish a dynamic model that can position itself, and then apply it to practice after a successful simulation.
(2) Algorithm Fusion and Multi-sensor Fusion. There are more or less shortcomings in a single algorithm or sensor. Developing a UAV detection system with multi-sensor algorithm fusion technology and combining machine learning can improve the reliability and stability of the system.
(3) Study on the navigation system of goaf. At present, the positioning of the detection system is mainly realized through global path planning. When there are unpredictable obstacles in the path, it is impossible to avoid them in time. Therefore, the research on the local path planning of the underground UAV in the future is very important. Multi-sensor fusion and point cloud fusion are the trends of future research.
(4) Multi-machine collaboration. Although a single UAV has been widely used in various fields, the underground environment of the mining area is complex, and a single UAV cannot meet its detection needs. Multi-machine systems can

significantly improve detection accuracy and efficiency. In the future, multi-machine cooperative detection and its fusion algorithm will be a hot research direction.

1.5 Conclusions

The goaf structure is complex and the environment is uncertain, so it must be surveyed and the internal structure must be mastered before the formal development. The key technologies and basic principles of goaf environment perception and detection are summarized.

UAV plays an important role in the field of goaf detection, pushing the mining area exploration technology to "unmanned," avoiding the problem of personal safety faced by traditional detection. This paper systematically summarizes the early mine exploration technology and key technologies of UAV detection, clarifies the development of navigation positioning and point cloud processing technology in recent years, and gives the research trend of UAV detection technology, which provides a reference for related fields.

It is the basis of goaf detection to use multi-sensors and lidar to sense the environment. It is a more efficient and faster way to use multi-computer cooperation to make visual modeling. Single detection method cannot meet the requirements of the new situation, and it is often difficult to determine the nature of geological anomalies. Therefore, a combination of multiple detection methods should be carried out according to different engineering characteristics.

Acknowledgements This research is supported by the Doctoral Research Fund of Shandong Jianzhu University (X18067Z), National Scholarship Fund (202008370131), and the Graduate Innovation Fund of Jilin University (2022048).

References

1. Han, C., Du, C., Zu, F., et al.: Factors influencing the stability of a slope containing a coal seam in a goaf. Appl. Sci. (Basel) **12**(22), 11699 (2022)
2. Cao, B., Wang, J., Du, H., et al.: Research on comprehensive detection and visualize of hidden cavity goaf. Sci. Rep. **12**(1), 22309 (2022)
3. Liu, C., Liu, J., Yue, J.: Development status and key problems of Chinese mining geophysical technology. J. China Coal Soc. **39**(1), 19–25 (2014)
4. Miller, F., Potvin, Y., Jacob, D.: Laser measurement of open stope dilution. CIM Bull. **85**, 96–102 (1992)
5. Li, J., Yang, C., Hu, Y.: Application research of UAV-lidar in detection of underground goaf. Met Mine **12**, 168–172 (2020)
6. Xu, Q., Li, K., Wang, J., Yuan, Q., Yang, Y., Chu, W.: The status, challenges, and trends: an interpretation of technology roadmap of intelligent and connected vehicles in China (2020). J. Intell. Connect. Veh. **5**(1), 1–7 (2022). https://doi.org/10.1108/jicv-07-2021-0010

7. Zhu, J., Easa, S., Gao, K.: Merging control strategies of connected and autonomous vehicles at freeway on-ramps: a comprehensive review. J. Intell. Connect. Veh. **5**(2), 99–111 (2022). https://doi.org/10.1108/jicv-02-2022-0005
8. Deshpande, N., Grant, E., Henderson, T.C., Draelos, M.T.: Autonomous navigation using received signal strength and bearing-only pseudo gradient interpolation. Robot. Auton. Syst. **75**, 129–144 (2016)
9. Wu, J., Qu, X.: Intersection control with connected and automated vehicles: a review. J. Intell. Connect. Veh. **5**(3), 260–269 (2022)
10. Wen, Y., Wang, H., Jin, Y., Cao, J.: Integrated computer vision algorithms and drone scheduling. Commun. Transp. Res. **1**, 100002 (2021). https://doi.org/10.1016/j.commtr.2021.100002
11. Li, J., Zhong, R., Hu, Q., Ai, M.: Feature-based laser scan mathcing and its application for indoor mapping. Sensors **16**(8), 1265 (2016)
12. Olovsson, T., Svensson, T., Wu, J.: Future connected vehicles: communications demands, privacy and cyber-security. Commun. Transp. Res. **2**, 100056 (2022). https://doi.org/10.1016/j.commtr.2022.100056
13. Nong, X., Bai, W., Liu, G.: Airborne LiDAR point cloud classification using PointNet++ network with full neighborhood features. PLoS ONE **18**(2), e0280346 (2017)
14. Besl, P., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) **14**(3), 239–256 (1992)
15. Lu, F., Milios, E.: Robot Pose Estimation in unknown environments by matching 2D range scans. J. Intell. Rob. Syst. **18**(3), 249–275 (1997)
16. Trucco, E., Fusiello, A., Roberto, V.: Robust motion and correspondences of noisy 3-D point sets with missing data. Pattern Recogn. Lett. **20**(9), 889–898 (1999)
17. Godin, G., Laurendeau, D., Bergevin, R.: A method for the registration of attributed range images. In: The third International Conference on 3D Digital Imaging and Modeling (2001 3DMI), pp. 179–186. IEEE (2001)
18. Minguez, J., Montesano, L., Lamiraux, F.: Metric-based iterative closest point scan matching for sensor displacement estimation. IEEE Trans. Robot. **22**(5), 1047–1054 (2006)
19. Yoshitaka, H., Hirohiko, K., Akihisa, O., et al.: Mobile Robot localization and mapping by scan matching using laser reflection intensity of the SOKUIKI sensor. In: The 32th Industrial Electronics Society Conference (2006 IECON), pp. 3018–3023. IEEE (2006)
20. Biber, P., Straber, W.: The normal distributions transform: a new approach to laser scan matching. In: The 16th Intelligent Robots and Systems Conference (2003 IROS), pp. 2743–2748. IEEE (2003)
21. Westfeld, P., Maas, H., Richter, K.: Analysis and correction of ocean wave pattern induced systematic coordinate errors in airborne LiDAR bathymetry. ISPRS J. Photogramm. Remote. Sens. **128**, 314–325 (2017)

Chapter 2
LSTM-Based Vehicle Trajectory Prediction Using UAV Aerial Data

Baozhen Yao, Qian Zhong, Heqi Cui, Sixuan Chen, Chuanyun Fu, Kun Gao, and Shaohua Cui

Abstract Accurately predicting the trajectory of a vehicle is a critical capability for autonomous vehicles (AVs). While human drivers can infer the future trajectory of other vehicles in the next few seconds based on information such as experience and traffic rules, most of the widely used Advance Driving Assistance Systems (ADAS) need to provide better trajectory prediction. They are usually only of limited use in emergencies such as sudden braking. In this paper, we propose a trajectory prediction network structure based on LSTM neural networks, which can accurately predict the future trajectory of a vehicle based on its historical trajectory. Unlike previous studies focusing only on trajectory prediction for highways without intersections, our network uses vehicle trajectory data from aerial photographs of intersections taken by Unmanned Aerial Vehicle (UAV). The speed of vehicles at this location fluctuates more frequently, so predicting the trajectory of vehicles at intersections is of great importance for autonomous driving.

B. Yao · Q. Zhong · H. Cui · S. Chen
State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, School of Automotive Engineering, Dalian University of Technology, Dalian 116024, P. R. China

C. Fu
School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China

K. Gao · S. Cui (✉)
Department of Architecture and Civil Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden
e-mail: shaohuac@Chalmers.se

S. Cui
School of Transportation Science and Engineering, Beihang University, Beijing 100191, P. R. China

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_2

2.1 Introduction

In recent years, research on autonomous driving technology has become very popular. For autonomous vehicles, reasonable behavioral control is significant. Cui et al. [1] proposed a cooperative constrained control algorithm based on the backstepping technique to improve AVs' stability. Of course, reasonable control depends not only on the current state of surrounding vehicles but also their future trajectories. It is essential to develop trajectory prediction techniques. Although many results have been achieved in current trajectory prediction-related research, there are still many challenges, such as complex traffic environments and the uncertainty of human drivers. Human-driven vehicles and autonomous vehicles will coexist on the road for a long time, and the control of autonomous vehicles in mixed traffic environments is crucial for traffic safety and traffic flow stability, which has been thoroughly studied by many scholars [2–5].

Vehicle motion prediction models have been extensively studied by many. Lefevre et al. [6] provided a comprehensive overview of trajectory prediction methods. Vehicle trajectory prediction methods can be divided into three types: physical model-based, maneuver-based, and interaction-based.

The Gaussian noise method reduces the vehicle to a simpler physical model. Polychronopoulos et al. [7] proposed a novel approach based on Switching Kalman Filters (SKF), which represents the uncertainty of the corresponding vehicle state under different driving behaviors using a Gaussian distribution with multiple peaks. The Monte Carlo method randomly samples the inputs to the physical model of the vehicle, then inputs the physical model to generate possible future trajectories of the vehicle, and finally filters the generated trajectories according to vehicle dynamics, road structure, and other conditions [8]. The direct method takes the current state parameters of the vehicle into the kinematic and dynamic models of the vehicle and calculates the vehicle state at the next moment [9].

Maneuver-based vehicle trajectory prediction methods generate matching future trajectories of vehicles based on the driver's behavioral intention estimated by the prediction model. Dizan Vasquez et al. [10] used the trajectory graph method to represent all prototype trajectories as multiple graphs. They learned the prototype trajectories corresponding to different driving behaviors using a Topology Learning Network (TGN). Hu et al. [11] argued that drivers' operational intentions would differ in traffic scenarios, resulting in different driving strategies. Therefore, vehicle trajectory prediction research was conducted based on the combination of motion information and classification intentions.

The interaction-based vehicle trajectory prediction method models the interaction between vehicles and generates predicted future trajectories of the target vehicle based on the obtained multi-vehicle historical trajectories. Goli et al. [12] proposed a Gaussian Process Regression (GPR) model that learns from the historical trajectory data of the target vehicle and the surrounding environment vehicles and shares the trained data among vehicles using a vehicle cloud by collecting vehicle information

through static sensors and information communication between the vehicle and the infrastructure.

Deep learning methods based on Long- Short-Term Memory (LSTM) neural networks have been frequently used in recent years to study the trajectory prediction of autonomous vehicles. The network model could perform a long sequence of information processing and was effective for the long-term prediction of autonomous vehicle trajectories [13]. Pan et al. [14] used LSTM and attention mechanism to learn vehicle intention and position change for trajectory prediction. Hsu et al. [15] proposed a convolutional-cyclic framework (CNNLSTM) combining a convolutional LSTM and deep convolutional neural networks (CNN) to predict vehicle intentions by detecting the tail light status of surrounding vehicles.

2.2 Problem Statement

Our data is the trajectory of vehicles within an intersection. The intersection is small, so we only consider the process of a vehicle waiting for a signal at the intersection and then starting to leave. Our goal is to predict the future trajectory of the vehicles at the intersection based on the historical trajectory information detected by the sensors, which can then be used for future autonomous vehicle control. For convenience, we use a fixed coordinate system with the top left corner of the road center as the origin, as shown in Fig. 2.3.

The inputs to our model include the historical trajectory of the vehicle and several processed features.

$$X = \left[X^{(t-t_n)}, ..., X^{(t-1)}, X^{(t)}\right] \tag{2.1}$$

where

$$X^{(t)} = \left[x^{(t)}, y^{(t)}, v_x^{(t)}, v_y^{(t)}, a_x^{(t)}, a_y^{(t)}\right] \tag{2.2}$$

Tn is the number of frames before moment t, i.e., n time steps of data. x and y are the predicted vehicle's coordinates at time t. v_x, v_y, a_x, and a_y are the velocity and acceleration at the corresponding coordinates after processing at moment t, respectively.

The output of the model is the future speed of the target vehicle.

$$Y = \left[y^{(t+1)}, ..., y^{(t+t_m)}\right] \tag{2.3}$$

where

$$y^{(t)} = \left[v_x^{(t)}, v_y^{(t)}\right] \tag{2.4}$$

We use a deep learning approach to complete the mapping between inputs and outputs. We train a function $f(X)$ that reveals the intrinsic connection between the inputs and outputs so that the predicted result $\widehat{Y} = f(X)$ is as close as possible to the actual result while ensuring that it is not over-fitted.

In this paper, we do not consider the target vehicle's interaction with other vehicles as there is little lane changing, overtaking, etc., in the relatively small area of the intersection, and the goal of our approach is to train a network to predict the future trajectory of an individual vehicle. A significant difficulty in trajectory prediction is designing models that generalise historical trajectory information well [16, 18, 19]. A second difficulty is trajectory prediction for complex driving environments, such as intersection vehicle trajectory prediction, where speeds change frequently or continuously.

2.3 Data Presentation

This paper uses vehicle trajectory data at the Beijing Guanggu Avenue—Weishan Road intersection. The video was captured at the intersection utilising a UAV, and then all vehicle trajectory data was extracted from the video recordings using high-accuracy video processing aids. For the intersection vehicle trajectory point data analysis, the south inlet lane stop line of the following intersection was uniformly set as the x-axis of the coordinate system, the west inlet lane stop line was selected as the y-axis of the coordinate system, and the origin of the coordinate system was the intersection point of the two stop lines, as shown in Fig. 2.1. We measured the physical extent of the central area of the intersection (the rectangle BDHJ in Fig. 2.1), which has a length and width of 33 m and 24.7 m, respectively.

We collected five minutes of vehicle trajectory data, containing over 250 trajectories at 0.1 s intervals. The processed data is presented as (X, Y) coordinates, and based on the coordinate information and the corresponding number of frames, it is easy to identify the trajectories of vehicles in the same lane and to match the vehicles in front of and behind them.

As we are using a UAV to capture video and analyze it through software to obtain the vehicle trajectory, this trajectory contains much noise. Therefore, the vehicle velocity calculated with the noisy position information will have large fluctuations, affecting the acceleration calculation. It is, therefore, necessary to smooth the data obtained by the video processing software. A third-order Savitzky-Golay filter smooths the entire trajectory data [17]. The window length of the filter is 21, and the smoothed data is used to calculate the corresponding velocities and accelerations.

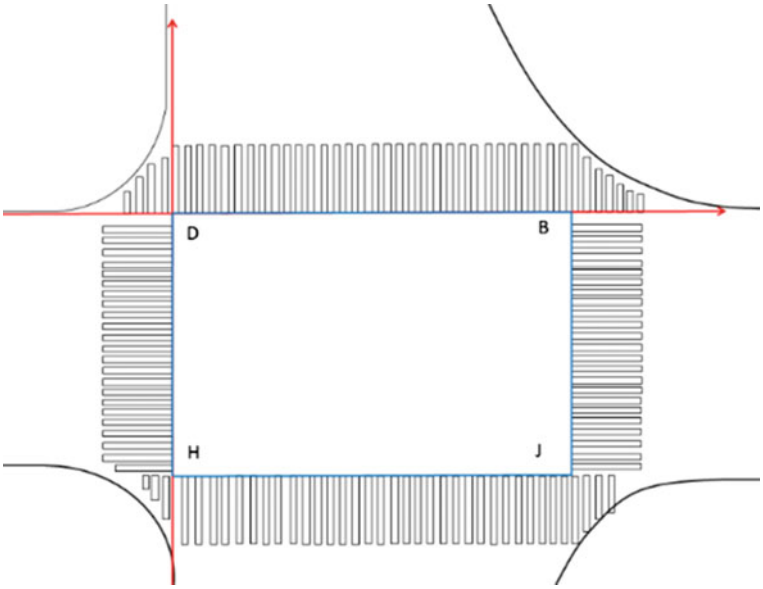


Fig. 2.1 Intersection coordinates diagram

2.4 Introduction to the Model

LSTM is a special implementation of Recurrent Neural Network (RNN), which is considered to be decisive in time series prediction. The main feature of an LSTM cell is the presence of a cell memory inside this node which makes the cell effective in slowing down the data loss rate. The LSTM cell contains internal input gates, output gates, forgetting gates, and various input nodes, etc.

Input gate: At moment t, when information is input from the input layer, it will first pass through the input gate, which will make a judgment and then decide whether to input the information into the memory cell, and the specific calculation formula is shown as follows.

$$i_t = \alpha\left(w_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{2.5}$$

ω_i is the bias term of the input gate and represents the weight matrix of the input gate.

Output gate: The role of this gate is to decide whether the information at the current moment should be extracted from the memory cell.

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{2.6}$$

Forgetting gate: All the memory cells at moment t will be decided by the forgetting gate. This decision will update the values in all memory cells. A part of the values

will be retained, while a part of the values will be forgotten.

$$f_t = \alpha\left(w_f\left[h_{t-1}, X_t\right] + b_f\right) \tag{2.7}$$

ω_f is the weight matrix of the forgetting gate, $\left[h_{t-1}, X_t\right]$ represents the joining of two vectors, which in turn form a new vector, b_f is the bias term of the forgetting gate, α represents the sigmoid function, and is the weight matrix of the forgetting gate.

After passing through the three gates inside the LSTM, the essential features will be saved, and the smaller ones will be cleared. This approach can help the system to increase the memory length and thus improve the ability to handle long-time series problems.

In this paper, we use the network structure shown in Fig. 2.2. The network structure consists of three layers of LSTM networks with 64, 32, and 32 LSTM cells, respectively, with a dropout layer in the middle of each LSTM layer. The dense layer is used at the end of the network.

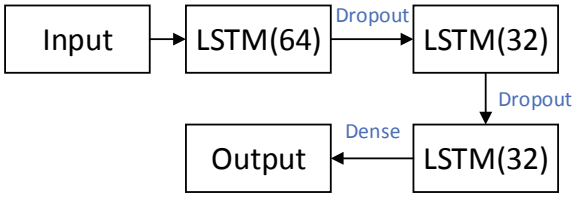


Fig. 2.2 The proposed LSTM model

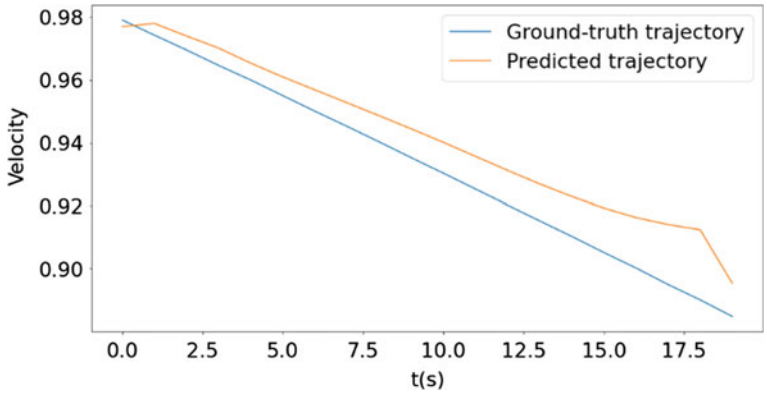


Fig. 2.3 Comparison of predicted and actual trajectories

2.5 Results

In this section, we divide the UAV aerial photography data into a training and test set, accounting for 80% and 20%, respectively. Only the training set is used in the model training process; we use the root of the mean squared error (RMSE) of the predicted trajectory relative to the actual future trajectory to evaluate the training results. A test set is used to test the training results. The ultimate goal is that the trained model can accurately predict vehicles' continuously changing trajectories at intersections.

The input to the model is the historical trajectory for the first twenty frames of the current moment, and the output is the trajectory information for the next 1 s. Given the small size of the aerial photography dataset, the data were divided into batches of 20 trajectories each, and the trajectory data within a batch was shuffled. The data from each batch was then fed into the model for training. Our proposed prediction model was implemented using TensorFlow. The RMSE converged after about 30 epochs of training. We use the r^2-score as the evaluation criterion for the test set. The r^2-score takes the value of [0,1], and the r^2-score for the test set is calculated to be 0.9, indicating that the predicted trajectory is very close to the actual trajectory and the model has good prediction results. The prediction result of one of the trajectories is shown in Fig. 2.3. It should be added that the velocities in the vertical coordinates in Fig. 2.3 are the values after normalisation, so the velocity range is limited to between 0 and 1. The speed range in m/s is 0–12.1 m/s.

2.6 Conclusions and Future Work

In this paper, we trained a deep learning network based on LSTM to predict the future trajectories of vehicles at the intersection. The model shows good prediction results. Accurate prediction of future trajectories within one second makes perfect sense for autonomous vehicles, especially at intersections where the speed of the vehicles varies widely and is susceptible to interference.

In this paper, we only considered the historical trajectory information of the target vehicle and did not consider the influence of other factors on the trajectory. Therefore, future work will incorporate more info to cope with more scenarios, such as vehicle interactions, traffic lights, and traffic signs.

Acknowledgements This research was supported in part by the Fundamental Research Funds for the Central Universities, in part by the AI Center (CHAIR) and Area of Advance Transport at Chalmers University of Technology, in part by the National Natural Science Foundation of China under Grant 52102393, in part by the Academic Excellence Foundation of BUAA for Ph.D. Students, and in part by China Scholarship Council under Grant 202106020149.

References

1. Cui, S., Xue, Y., Lv, M., Yao, B., Yu, B.: Cooperative constrained control of autonomous vehicles with nonuniform input quantization. IEEE Trans. Veh. Technol. **71**, 11431–11442 (2022). https://doi.org/10.1109/TVT.2022.3189498
2. Suriyarachchi, N., Mavridis, C., Baras, J.S.: Cooperative multi-lane shock wave detection and dissipation via local communication. In: 2022 30th Mediterranean Conference on Control and Automation (MED), pp. 1080–1086 (2022). https://doi.org/10.1109/MED54222.2022.9837213
3. Cui, S., Cao, F., Yu, B., Yao, B.: Modeling heterogeneous traffic mixing regular, connected, and connected-autonomous vehicles under connected environment. IEEE Trans. Intell. Transp. Syst. **23**, 8579–8594 (2022). https://doi.org/10.1109/TITS.2021.3083658
4. Li, Y., Zhao, L., Gao, K., An, Y., Andric, J.: Revealing driver psychophysiological response to emergency braking in distracted driving based on field experiments. J. Intell. Connect. Veh. **5**(3), 270–282 (2022)
5. Zhu, J., Easa, S., Gao, K.: Merging control strategies of connected and autonomous vehicles at freeway on-ramps: a comprehensive review. J. Intell. Connect. Veh. **5**(2), 99–111 (2022)
6. Gulzar, M., Muhammad, Y., Muhammad, N.: A survey on motion prediction of pedestrians and vehicles for autonomous driving. IEEE Access **9**, 137957–137969 (2021). https://doi.org/10.1109/ACCESS.2021.3118224
7. Polychronopoulos, A., Tsogas, M., Amditis, A.J., Andreone, L.: Sensor fusion for predicting vehicles' path for collision avoidance systems. IEEE Trans. Intell. Transp. Syst. **8**, 549–562 (2007). https://doi.org/10.1109/TITS.2007.903439
8. Xue, Q., Gao, K., Xing, Y., Lu, J., Qu, X.: A context-aware framework for risky driving behavior evaluation based on trajectory data. IEEE Intell. Transp. Syst. Mag. **15**(1), 70–83 (2023). https://doi.org/10.1109/MITS.2021.3120279
9. Brännström, M., Coelingh, E., Sjöberg, J.: Model-based threat assessment for avoiding arbitrary vehicle collisions. IEEE Trans. Intell. Transp. Syst. **11**, 658–669 (2010). https://doi.org/10.1109/TITS.2010.2048314
10. Govea, D.A.V., Fraichard, T., Laugier, C.: Growing hidden Markov models: an incremental tool for learning and predicting human and vehicle motion. Int. J. Robot. Res. **28**, 1486 (2009)
11. Hu, Y., Zhan, W., Tomizuka, M.: Probabilistic prediction of vehicle semantic intention and motion. In: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 307–313 (2018). https://doi.org/10.1109/IVS.2018.8500419
12. Goli, S.A., Far, B.H., Fapojuwo, A.O.: Vehicle trajectory prediction with Gaussian process regression in connected vehicle environment*. In: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 550–555 (2018). https://doi.org/10.1109/IVS.2018.8500614
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
14. Pan, J., Sun, H., Xu, K., Jiang, Y., Xiao, X., Hu, J., Miao, J.: Lane-attention: predicting vehicles' moving trajectories by learning their attention over lanes. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7949–7956 (2020). https://doi.org/10.1109/IROS45743.2020.9341233
15. Hsu, H.-K., Tsai, Y.-H., Mei, X., Lee, K.-H., Nagasaka, N., Prokhorov, D., Yang, M.-H.: Learning to tell brake and turn signals in videos using CNN-LSTM structure. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6 (2017). https://doi.org/10.1109/ITSC.2017.8317782
16. Phillips, D.J., Wheeler, T.A., Kochenderfer, M.J.: Generalizable intention prediction of human drivers at intersections. In: 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 1665–1670. IEEE, Los Angeles, CA, USA (2017). https://doi.org/10.1109/IVS.2017.7995948
17. Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. **36**, 1627–1639 (1964). https://doi.org/10.1021/ac60214a047

18. Ali, Y., Zheng, Z., Haque, M.M.: Modelling lane-changing execution behaviour in a connected environment: a grouped random parameters with heterogeneity-in-means approach. Commun. Transp. Res. **1**, 100009 (2021). https://doi.org/10.1016/j.commtr.2021.100009
19. Liu, Y., Lyu, C., Zhang, Y., Liu, Z., Yu, W., Qu, X.: DeepTSP: Deep traffic state prediction model based on large-scale empirical data. Commun. Transp. Res. **1**, 100012 (2021). https://doi.org/10.1016/j.commtr.2021.100012

Chapter 3
Dynamic Pricing for Mobile Charging Service Considering Electric Vehicles Spatiotemporal Distribution

Baozhen Yao, Heqi Cui, Qian Zhong, Bin Shi, Yongjie Xue, and Shaohua Cui

Abstract As mobile charging service has the advantages of flexible charging and simple operation, it is selected by more and more users of electric vehicles. However, due to the differences in road network density and traffic flow distribution, the uneven distribution of charging demand occurs in different regions. It reduces the service efficiency of mobile charging vehicles during the peak charging demand period, thus affecting the final revenue of operators. In order to solve this problem, this paper proposes a dynamic pricing strategy considering the spatiotemporal distribution of charging demand to induce users to transfer between different regions, which can alleviate the phenomenon that users wait too long during peak demand. In order to realize the city-level operation of mobile charging service, we divide the region into hexagons and make statistics on the charging demand in each region. The established demand updating model can reflect the impact of charging price on users' charging behavior. Finally, we simulate the generation of charging demand in Haidian District, Beijing. According to the demand of each area, a thermodynamic diagram of charging demand is generated.

B. Yao · H. Cui · Q. Zhong · B. Shi
State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, School of Automotive Engineering, Dalian University of Technology, Dalian 116024, PR China

S. Cui (✉)
Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden
e-mail: shaohuac@Chalmers.se; shaoh_cui@buaa.edu.cn

Y. Xue · S. Cui
School of Transportation Science and Engineering, Beihang University, Beijing 100191, PR China

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_3

3.1 Introduction

Electric vehicles (EVs) are considered to be an effective tool to replace petroleum-driven vehicles in the future because of their environmental protection, high efficiency, and low use cost [1–3]. At present, governments are vigorously promoting the popularization and development of the electric vehicle industry. By the end of 2018, the number of EVs in the world had reached 1.2 million. This figure had increased by 50,000 compared with 2017 [4]. EVs need to be driven by batteries, which means that the charging problem of EVs must be solved. At present, there are two main charging methods for EVs. Firstly, drivers use fixed charging piles to charge their electric vehicles. Secondly, the drivers arrive at the battery exchange station to replace the batteries for the electric vehicle, so as to realize the power supplement. Although recharging time can be greatly reduced by replacing batteries, this charging method requires large installation space, and there are still many problems to be solved in related technologies. Therefore, fixed charging pile is still the most widely used charging mode. The problem with the mode is that the charging time is too long, and it will cause a greater burden to the power grid when it is in the peak period of power consumption. In addition, the electric vehicle industry is developing rapidly, the corresponding basic charging facilities are also improving. But there is still a big gap between them. According to the survey of relevant departments, by the end of 2020, the ratio of the number of electric vehicles in China to the number of charging piles is 3 to 1. At the same time, some users of fuel vehicles park irregularly and occupy charging spaces, which also leads to a significant decline in the utilization efficiency of charging piles. Therefore, electric vehicle users urgently need a novel charging mode to alleviate the difficulty of charging during the peak period. Against this background, many companies have begun to vigorously promote mobile charging services [5], and some scholars have also begun to pay great attention to the direction of mobile charging and conducted research [6–9]. Unlike fixed charging posts, mobile charging users can send personal positioning, expected power, expected charging time, and other information to the system center through mobile APP. After receiving information, the system will send instructions to the mobile charging vehicle. Although mobile charging can effectively make up for the shortcomings of traditional charging methods, in the actual operation process, due to the free floating of EVs and the frequent clustering in some hot areas [10], it will lead to the mismatch between the charging demand and the actual charging capacity that can be provided by the region during the peak charging period, that is, the problem of uneven supply and demand. Users in these areas sometimes need to wait for a long time before charging, or even have to give up the mobile charging service because of the long waiting time. Therefore, it is very important for mobile charging operators to formulate intelligent charging strategies and reasonably regulate the charging behavior of users.

Charging price will greatly affect the driving behavior of users and the revenue of operators. It is beneficial to set charging price reasonably to promote the development of EVs. Therefore, researchers attach great importance to the charging price setting of EVs. Li and Ouyang [11] studied from the perspective of charging station, and calculated the pricing range based on real-time energy price, charging station load, etc.. Lu et al. [12] proposed a cost–benefit analysis model on the basis of considering various factors affecting charging pricing. In order to reduce the charging cost as much as possible and meet the charging demand of users, Arif et al. [13] used dynamic pricing strategy and proposed three scheduling algorithms for plug-in vehicles. Based on the valley filling effect of suppliers and user costs, Zou et al. [14] and Hu et al. [15], respectively, designed optimization models for pricing. Shi [16] analyzed the interest relationship between the power system and electric vehicle users with the power demand side management theory and electricity price theory.

These studies are all carried out for fixed charging posts. However, the previous pricing method with a single charging station as the minimum pricing unit cannot be completely migrated to mobile charging services. Because each mobile charging vehicle can be regarded as an independent charging station, if each mobile charging vehicle is priced, it will not only lead to a decline in user satisfaction, but also make the subsequent pricing problem too complex to be reasonably realized. Therefore, this paper proposes a novel pricing mechanism based on the characteristics of mobile charging. Firstly, the mobile charging service area is divided into hexagon areas, then the number of charging demands in each hexagon area is counted. Finally, set a reasonable charging price for the region according to these data. When EVs generate charging demand, users can obtain the charging price of the current region and neighborhood. Figure 3.1 illustrates the user's selection process in these areas. The process that users make choices under the influence of price is the process that the system adjusts the supply and demand relationship of each region through dynamic pricing. Figure 3.1 shows that when users generate charging demand, they will send personal charging information to the cloud server. The system will provide users with optional charging sets and charging prices for each area. Users will select the most suitable service area under the influence of the charging price. This process is also a process in which the system adjusts the supply–demand relationship between regions through dynamic pricing.

The main contributions of this paper are summarized as follows: Firstly, according to the characteristics of mobile charging service, a novel pricing mechanism is proposed to help operators relieve the working pressure of mobile charging vehicles (MCVs) in different time intervals. Secondly, a method for dividing mobile charging service area is proposed. On this basis, reasonably count the charging demand in the region to provide data support for the subsequent regional pricing. Thirdly, considering the impact of charging price setting on user charging behavior, a demand updating model is established to clearly define the relationship between user demand, charging price, and user behavior.

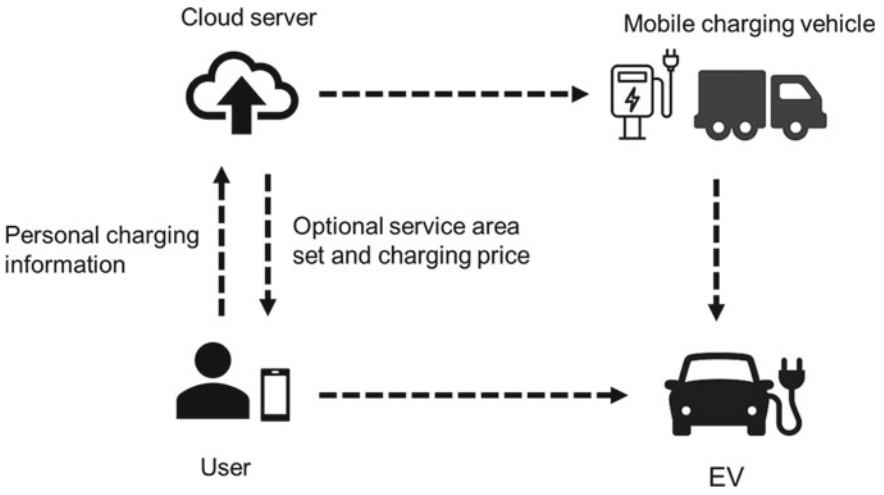


Fig. 3.1 Mobile charging service process

We organize the rest of the paper as follows: In Section 3.2, the division method of mobile charging service area and the advantages of choosing hexagon as the division area shape are introduced. In addition, based on the queue length of each region, the dynamic pricing model of the region is introduced Finally, through the analysis of user behavior, a charging demand update model is established under the influence of charging price. In Section 3.3, taking Haidian District of Beijing as an example, the service area is divided, and a thermal diagram reflecting the distribution of charging demand is designed. In Section 3.4, the research contents of this paper are summarized.

3.2 System Model

In the actual operation process, the mobile charging service demand in each region has a self-balancing mechanism. This is mainly reflected in the longer waiting time required by users when there are more charging demands in the area. The longer the waiting time is, the lower the user satisfaction will be, and the subsequent charging demand rate will also be reduced. Therefore, in order to avoid this problem, a dynamic pricing mechanism is introduced to balance the supply and demand relationship among regions. The advantage of this is that it does not only rely on the self-adjustment of the system, but also helps operators increase service revenue, while ensuring that users have a high degree of satisfaction.

Before introducing a specific pricing mechanism, it is necessary to divide the service areas of mobile charging services. When selecting the graph used to divide the service area, consider that only three kinds of polygons can ensure that the service area can be completely divided without overlapping. They are equilateral triangles, equilateral hexagons, and squares. Compared with the other two kinds of figures, the equilateral hexagon has its unique advantages [17]. The specific advantages are as follows:

(1) Equilateral hexagons have a smaller edge area ratio. This is because compared with the other two shapes, the equilateral hexagon has shorter perimeter under the same unit area, which is beneficial to reduce the deviation caused by the edge effect. At the same time, the statistical accuracy will also be improved when calculating the outflow and inflow of EVs between different regions.
(2) The neighborhood type of equilateral hexagon is more uniform. The region divided according to the square has two kinds of neighbors: the orthogonal neighbors are connected by the side; the square with the opposite corner is connected by only one point. In our research, we need to obtain the transfer of users between different regions, which is unfavorable. All the neighbors of a region divided by a hexagon are connected by adjacent edges. In addition, for an equilateral hexagon region, the distance from it to all its neighbors is equal, which can greatly simplify the cost setting of users moving between different regions.

In practical applications, equilateral triangles are rarely used as the shape of region division because of their large edge area ratio. Although in the orthogonal coordinate system, the square division method is the easiest to achieve, it is the most widely used. However, equilateral hexagon is still widely used in many fields, including geography, computer vision, ecology, and so on, depending on its unique advantages.

Considering that in the actual operation process of mobile charging service, there will be peak time and off-peak time of demand, we will divide the whole day into 24 intervals with 1 hour as a time interval. At the same time, it is assumed that the average charge demand generation rate in each time interval is a fixed value, and the demand rate in different time intervals is variable. Generally, the generation of charging demand in each time interval follows the Poisson process [18]. Let time $t, t = 1, 2, \ldots, T$ represent the T time gap, and the generation rate λ_t of charging demand during the time gap t, so the probability of generating n charging demands during the time gap t is as follows:

$$P\{n\} = \frac{e^{\lambda_t}(\lambda_t)^n}{n!}, n = 0, 1, 2, \ldots \tag{3.1}$$

In the process of charging, the electric quantity to be supplemented for general EVs follows the logarithmic normal distribution [19], the average value of the electric quantity demanded is E, and the standard deviation is E_d. In this paper, in order to simplify the solution process, it is assumed that the charging requirements of all EVs need to be E [20]. Using the relevant knowledge of queuing theory, the queue length

waiting for mobile charging service in each sub-region is

$$Q_{i,t} = \max\{0,\ Q_{i,t-1} + \lambda_{i,t} - V_{i,t} - L_{i,t}\} \tag{3.2}$$

The queue lengths of region i at times $t-1$ and t are $Q_{i,t-1}$ and $Q_{i,t}$, respectively. $V_{i,t}$ indicates the number of MCVs that have not started to work within the time gap. In addition, $L_{i,t}$ represents the number of EVs that have completed recharging within this time interval [21]. When the system collects the queue length in each area of the current time gap, it will judge the supply and demand situation in the area according to the queue length [22], so as to set an appropriate charging price to induce users to charge orderly. For region i, the charging price adjustment scheme is as follows:

$$\widehat{P}_{i,t} = \widehat{P}_{i,0} + e^{\alpha \times Q_{i,t}} - 1 \tag{3.3}$$

$\widehat{P}_{i,0}$ refers to the initial charging pricing of area i, $\widehat{P}_{i,t}$ represents the updated charging price at time t according to the average waiting time of users in region i, and α represents the adjustment proportion of the price according to the length of the queue. In addition, when formulating the charging price, it is also necessary to consider the basic charging cost and the highest acceptable charging price in the market. Therefore, the charging price in this paper should meet the following requirements:

$$P_{min} < \widehat{P}_{i,t}, \widehat{P}_{i,0} < P_{max} \tag{3.4}$$

where P_{max} and P_{min} are the maximum and minimum charging prices allowed by the system during dynamic pricing. When users consider the selection of charging area, they will only consider their current area and its surrounding neighborhood. Figure 3.2 demonstrates the neighborhood that users can select. The reason for this setting is determined by the target users of the mobile charging service; the mobile charging mode is still a supplement to the traditional fixed charging mode. When the owners of EVs choose to use mobile charging, the main reason may be that the residual power of EVs is not enough to move too much distance or users do not want to move too far to get power supplement [23]. In addition, it is the result of a combination of many factors that affect the user's choice of charging area. Because this needs to consider the user's charging cost, the mobile cost between different regions, the time cost, etc. Only when the price difference between different regions is greater than the users' time cost and mobile cost between different regions, users will choose to move to other regions for charging, and the corresponding charging demand rate of each region will also be updated. The influence of charging price on the charging demand rate of each region can be shown in Fig. 3.3. The updated charge demand rate for the region is

$$\lambda_{i,t} = \widehat{\lambda}_{i,t} - \lambda_{i,t}^{out} + \lambda_{i,t}^{in} \tag{3.5}$$

where $\widehat{\lambda}_{i,t}$ represents the estimated charging demand of region i based on historical data statistics, $\lambda_{i,t}$ is the updated final charging demand affected by the price between regions. $\lambda_{i,t}^{out}$ represents the outflow of expected users from region i, and $\lambda_{i,t}^{in}$ represents the demand quantity of expected users flowing into the region from the neighborhood of region i. The value of $\lambda_{i,t}^{out}$ is obtained according to the following formula:

$$\lambda_{i,t}^{out} = \left(\frac{\max\{\left(\widehat{p}_{i,t} - \widehat{p}_{j,t}^{min} - p_0\right), 0\}}{\widehat{p}_{i,t}} \right) \cdot \widehat{\lambda}_{i,t} \tag{3.6}$$

where $\widehat{p}_{j,t}^{min}$ refers to the charging price corresponding to the area i with the lowest charging price in all the neighborhood j of the region and p_0 refers to the driving value of the user's movement between different regions [24]. That is, only when the price difference between regions is greater than this value, the user will have the tendency to transfer, otherwise the user's transfer will not occur. $\widehat{p}_{i,t}$ is solved by the dynamic pricing model. $\lambda_{i,t}^{in}$ is obtained by summing up the outflows of all regions.

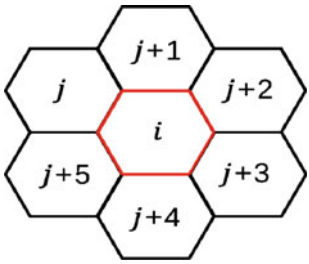


Fig. 3.2 Target area and its neighborhood

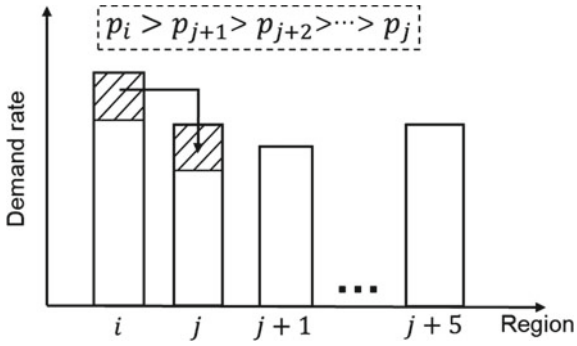


Fig. 3.3 Regional response model based on charging price

3.3 Case Analysis

As the mobile charging mode is a novel charging mode, at present, the relevant data on mobile charging is mainly mastered by a few operators. It is difficult to obtain this part of data, but it is worth noting that the generation of charging demand is often related to the regional traffic flow distribution. This is mainly because the higher the traffic flow, the more likely the charging demand will be generated. Therefore, in this paper, we calculate the traffic flow data in the divided area, and then convert it according to a certain proportion, so as to indirectly reflect the mobile charging demand in this area. We use the car hailing data provided by Didi. These data include GPS data, track data, vehicle speed, and other data of vehicles at different times. The order form contains 4153,586 orders dated November 29, 2016 and December 18 (3 weeks). The longitude of the investigated area is between $116°03^{'} and\ 116°23^{'} E$ and the latitude is between $39°53^{'} and\ 40°09 N$.

We simulate Haidian District of Beijing as the operation area of MCVs, and then divide the area according to the hexagon. Figure 3.4 illustrates the results of the region division. After dividing the regions, we can quickly master the supply and demand of each region by making statistics on the charging demand in each hexagon region. In order to make the distribution of charging demand more intuitive, we use statistical data to generate the thermal diagram of demand distribution. It can be seen from Fig. 3.5 that there is a large demand for charging in the area close to the city center. This is because the traffic network in this area is densely distributed, so the travel probability of EVs is large and the corresponding potential charging demand is also large. Therefore, it can be considered that this sub-region will be the key research area for pricing problems. Due to the scattered distribution of EVs in urban fringe areas, there will be less demand for charging in these areas.

The statistics of charging demand in each region can not only guide operators to reasonably set the number of MCVs in each region, but also help operators to know whether the peak charging demand in each region is within a reasonable range. In the subsequent research, we will also verify whether the dynamic pricing model for the region is in line with the reality based on these data.

3.4 Conclusion

Mobile charging mode can effectively make up for the shortage of traditional fixed charging, but it is worth noting that with the increase of mobile charging demand, the supply and demand relationship between different regions may be unbalanced. This situation will not only affect the service efficiency of the system, but also reduce user satisfaction, and even cause users to give up mobile charging service. Therefore, in order to solve this problem, this paper proposes a dynamic pricing method that adapts to the characteristics of mobile services. By dynamically adjusting the charging price of the region, users can move between different regions. This will improve

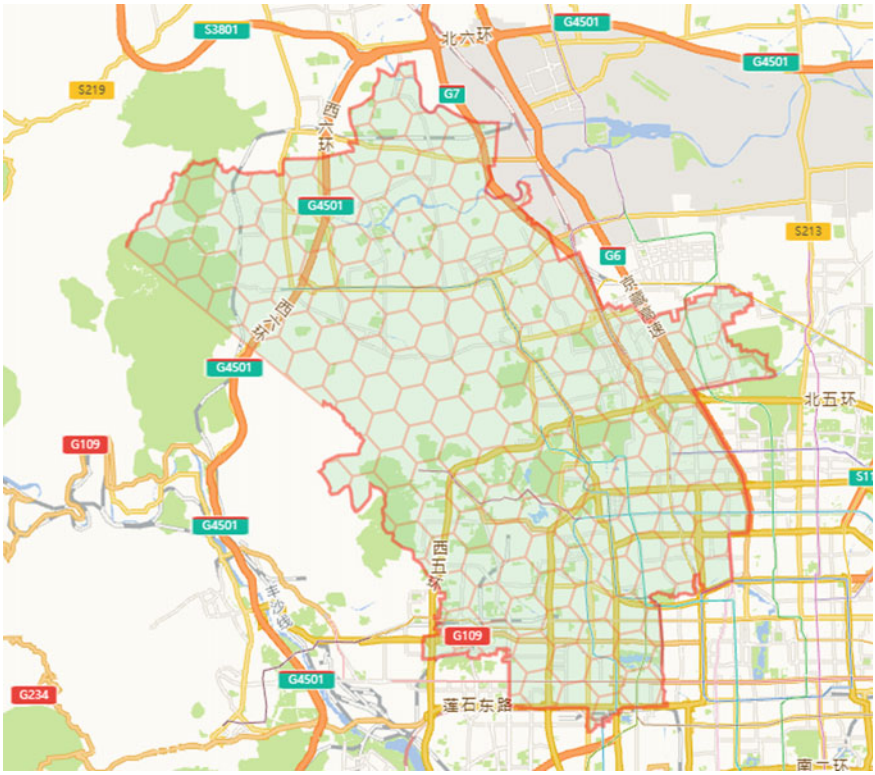


Fig. 3.4 Schematic diagram of regional hexagonal division

the problem that the charging demand is too concentrated during the peak charging period. In addition, this paper also provides a hexagonal division method for mobile charging services. Operators can adjust the prices of different regions according to the demand of the divided regions. In this paper, we take Haidian District of Beijing as an example to divide the service areas, and on this basis, we make statistics on the demand of different regions. The dynamic pricing model and demand updating model are established. They reflect that when there are demand differences and charging price differences between different regions, these differences will affect the final behavior of users. Finally, Haidian District of Beijing is simulated as a designated area for mobile charging service, and the regional division and demand statistics are carried out. These works can help operators to provide data support when setting the number of MCVs in different regions, and can also lay a foundation for subsequent research on pricing issues.

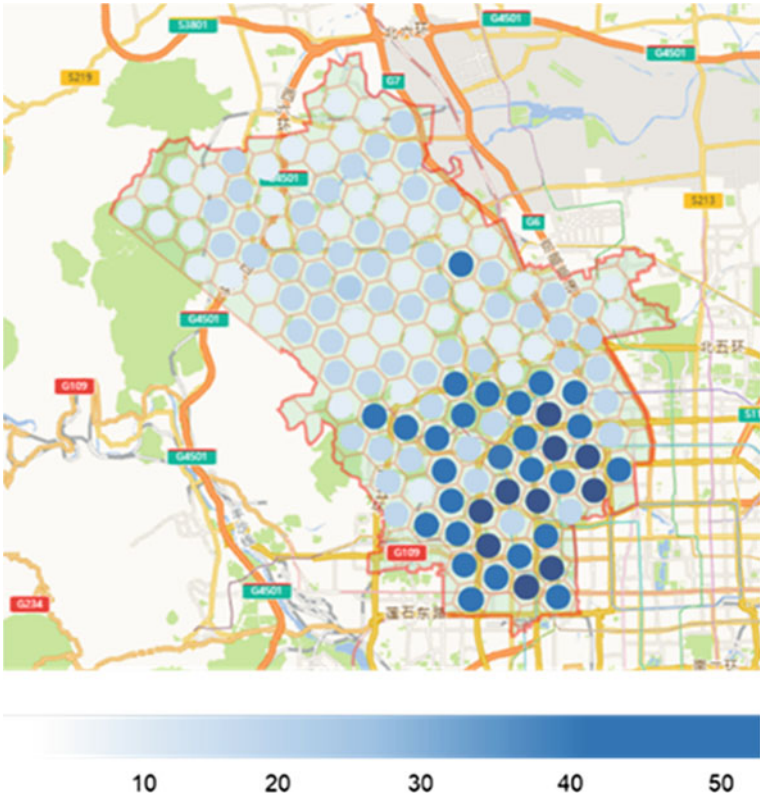


Fig. 3.5 Thermal diagram of charging demand based on data

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 52102393, in part by the AI Center (CHAIR) at Chalmers University of Technology and Swedish Energy Agency, in part by the Academic Excellence Foundation of BUAA for Ph.D. Students, and in part by China Scholarship Council under Grant 202106020149.

References

1. Wirasingha, S.G., Emadi, A.: Plug-in hybrid electric factor. IEEE Vehicle Power & Propulsion Conference **60**(3), 1279–1284 (2011)
2. Tan, J., Wang, L.: A stochastic model for quantifying the impact of PHEVs on a residential distribution grid. IEEE International Conference on Cyber Technology in Automation IEEE, 120–125 (2014).
3. Gao, K., Yang, Y., Zhang, T., Li, A., Qu, X.: Extrapolation-enhanced model for travel decision making: An ensemble machine learning approach considering behavioral theory. Knowl.-Based Syst. **218**, 106882 (2021)

4. Irle, R.: Global EV sales for the 1st half of 2019. URL: http://www.ev-volumes.com/ country/ total- world- plug- in- vehicle- volumes/.
5. Emeç, U., Çatay, B., Bozkaya, B.: An adaptive large neighborhood search for an e-grocery delivery routing problem. Comput. Oper. Res. **69**, 109–125 (2016)
6. Cui, S., Yao, B., Chen, G., Zhu, C.: The multi-mode mobile charging service based on electric vehicle spatiotemporal distribution. Energy **198**, 117302 (2020)
7. Li, B., Page, B.R., Hoffman, J.: Rendezvous planning for multiple AUVs with mobile charging stations in dynamic currents. IEEE Robotics and Automation Letters **4**(2), 1653–1660 (2019)
8. Cui, S., Zhao, H., Zhang, C.: Multiple types of plug-in charging facilities' location-routing problem with time windows for mobile charging vehicles. Sustainability **10**(8), 2855 (2018)
9. Huang, S., He, L., Gu, Y.: Design of a mobile charging service for electric vehicles in an urban environment. IEEE Trans. Intell. Transp. Syst. **16**(2), 787–798 (2014)
10. Cui, S., Ma, X., Zhang, M., Yu, B.: The parallel mobile charging service for free-floating shared electric vehicle clusters. Transportation Research Part E: Logistics and Transportation Review **160**, 102652 (2022)
11. Li, Z., Ouyang, M.: The pricing of charging for electric vehicles in China-Dilemma and solution. Energy **36**(9), 5765–5778 (2011)
12. Lu, K., Liu, S., Niu, X.: Pricing method of electric vehicle charging by using cost-benefit analysis. Journal of Power System and Automation **26**(3), 76–80 (2014)
13. Arif, A.I., Babar, M., Ahamed, T.: Online scheduling of plug-in vehicles in dynamic pricing schemes. Sustainable Energy Grids & Networks **7**, 25–36 (2016)
14. Zou, W., Wu, F., Liu, Z.: Centralized charging strategies of plug-in hybrid electric vehicles under electricity markets based on spot pricing. Automation of Electric Power Systems **35**(14), 62–67 (2011)
15. Hu, Z., Zhan, K., Zhang, H., Song, Y.: Pricing mechanisms design for guiding electric vehicle charging to fill load valley. Appl. Energy **178**, 155–163 (2016)
16. Shi, L.: Design for the electric vehicle charging and discharging price strategy from demand side management perspective. PhD thesis. Chongqing University (2012)
17. Ke, J., Yang, H., Chen, X.: Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. IEEE Trans. Intell. Transp. Syst. **20**(11), 4160–4173 (2019)
18. Alizadeh, M., Scaglione, A., Davies, J.: A scalable stochastic model for the electricity demand of electric and plug-in hybrid vehicles. IEEE Transactions on Smart Grid **5**(2), 848–860 (2014)
19. Ma, T., Mohammed, A.O.: Optimal charging of plug-in electric vehicles for a car-park infrastructure. IEEE Trans. Ind. Appl. **50**(4), 2323–2330 (2014)
20. Gao, K., Yang, Y., Qu, X.: Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. Transp. Res. Part D: Transp. Environ. **97**, 102957 (2021). https://doi.org/10.1016/j.trd.2021.102957
21. Liu, Y., Wang, L., Zeng, Z., Bie, Y.: Optimal charging plan for electric bus considering time-of-day electricity tariff'. Journal of Intelligent and Connected Vehicles **5**(2), 123–137 (2022)
22. Zhang, L., Zeng, Z., Gao, K.: A bi-level optimization framework for charging station design problem considering heterogeneous charging modes. Journal of Intelligent and Connected Vehicles 5(1) (2022).
23. Eliasson, J.: Efficient transport pricing-why, what, and when? Communications in Transportation Research **1**, 100006 (2021)
24. Gao, K., Yang, Y., Qu, X.: Diverging effects of subjective prospect values of uncertain time and money. Communications in Transportation Research **1**, 100007 (2021)

Chapter 4
Guidance Method of Connected Autonomous Vehicles Under Automatic Control Intersections

Lichao Wang, Jiaming Wu, Min Yang, Jiyang Zhang, and Zhiwei Meng

Abstract In this study, a guidance method for connected autonomous vehicles (CAVs) passing through automatic control intersections (ACI) is proposed. The approach favors a new mode of intersection control in an autonomous driving environment. Based on the modeling of automatic control intersections, the functional delineation of the approach lanes was eliminated, and a conflict resolution method for CAVs within the intersections was established considering the objectives of safety and efficiency, and pre-assignment of lanes for CAVs through ACI was realized. Based on the lane pre-assignment results to achieve space–time resource allocation among CAVs, the final passing strategy is determined and back to each CAV. The simulation results show that the method proposed in this study can ensure the safe and efficient passing of CAV through ACI.

4.1 Introduction

As an important means to relieve traffic congestion, reduce traffic pollution, and improve traffic efficiency, intelligent transportation system (ITS) has been widely used in the construction and development of transportation [1]. For the further development of urban transportation, the innovation and applicability of ITSs are crucial [2, 3]. Intersections are an important basic component of urban transportation facilities. With the breakthrough development of technologies such as autonomous

L. Wang · M. Yang
Southeast University, Jiangsu 211102, China

L. Wang · J. Wu (✉)
Chalmers University of Technology, Chalmersplatsen 4, 412 96 Göteborg, Sweden
e-mail: jiaming.wu@chalmers.se

J. Zhang
Design Group, Jiangsu 210014, China

Z. Meng
Jilin University, Changchun 130022, China

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_4

driving and vehicle-to-infrastructure (V2I) cooperation system [4, 5], the way vehicles passing various traffic hubs is expected to undergo a disruptive transformation in an autonomous environment [6, 7].

The transformation is characterized by two main points: one is the elimination of intersection signals, and the other is the elimination of intersection approach lane function delineation[7, 8]. In other words, the basis for whether a vehicle can pass an intersection in an autonomous environment is no longer signal, but relies on the transmission of information between the vehicle to vehicle (V2V), and V2I, so as to obtain the best timing and the best speed for the vehicle to pass the intersection. At the same time, vehicles approaching the intersection approach lane no longer need to change the lane they are currently in and then enter the lane with the designated driving function, i.e. all the approach lanes have the arbitrary turning function. In order to distinguish it from the current intersection control mode, it is thus defined as an automatic control intersection (ACI). Researchers have conducted some relevant studies in this area, and the literature [9] has built a validation environment for traffic management at automated intersections, achieving efficiency improvements and energy savings in intersection control schemes. Literature [10] optimally controls the passing process of a connected autonomous vehicle (CAV) passing a signal-free intersection, ensuring the travel time and safe intervals in the control area. Amir et.al have established an effective cooperative mechanism between connected autonomous vehicles and traffic infrastructure, have constructed a dynamic planning model for signal-free intersections to improve the throughput and increase the efficiency of vehicle passing while ensuring safety [11]. Based on the above, He et.al designed a signal-free automatically controlled intersection with all-direction turn lanes in an automated environment, which has a high passing efficiency compared to signal-controlled intersections [8]. The study by He et al. has been an important inspiration for further ITS. However, in the existing study, He et al. only used the first-in-first-out (FIFO) rule to decide the sequence of vehicles passing the intersection, and this rule may lose its optimal effect during actual operation due to the degree of conflict or the traffic volume in different directions [12, 13].

In an automated driving environment, there are significantly more conflict points within intersections due to the mentioned transformations. In order to improve the above transformation, regulate the order of traffic, reduce vehicle conflicts, and improve the efficiency of passing. It is necessary to propose conflict resolution methods for CAVs within automated controlled intersections to promote the safe operation and to bridge the gaps and deficiencies in the development of existing technologies.

4.2 Method

The technical problem to be solved in this study is to reduce the impact of conflicts inside signal-free intersections and improve the safety and passing efficiency of CAVs passing intersections in an automated driving environment.

4.2.1 The Process of CAV Pass ACI

The whole process of CAV passing the ACI is divided into 3 parts. The first step is to collect the request of each CAV to passing the target intersection, then make the corresponding passing decision according to the passing request of each CAV, and finally return the passing strategy of each CAV.

Where the collection process of each CAV through request is realized at a predefined detection period. Each detection period is continuous so that the continuous control of the intersection can be ensured. The process of developing the passing strategy for each CAV is the core of this study. The objective function is established by considering the time, efficiency and distance of CAV passing the ACI, and the CAV departure lane pre-assignment is realized according to the objective function. According to the pre-assigned lane results to determine whether there is a time–space conflict of each CAV, the CAV departure lane is optimally adjusted, so as to finally determine the passing decision of each CAV, the above specific decision process will be developed in the ***ACI modeling*** section.

Finally, the passing strategy will be back to each CAV, and each CAV will pass the ACI strictly according to the strategy. In particular, the passing strategies of each CAV are solved before the next detection period, the communication delay will be ignored, and each passing strategy can be received and executed by the CAV immediately. The specific operation process is shown in Fig. 4.1.

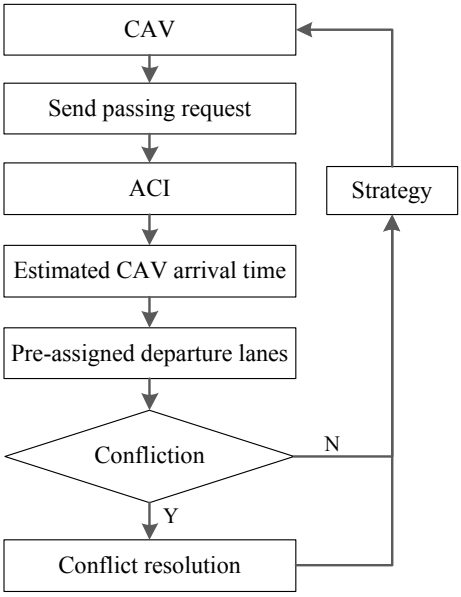


Fig. 4.1 The whole process of conflict resolution

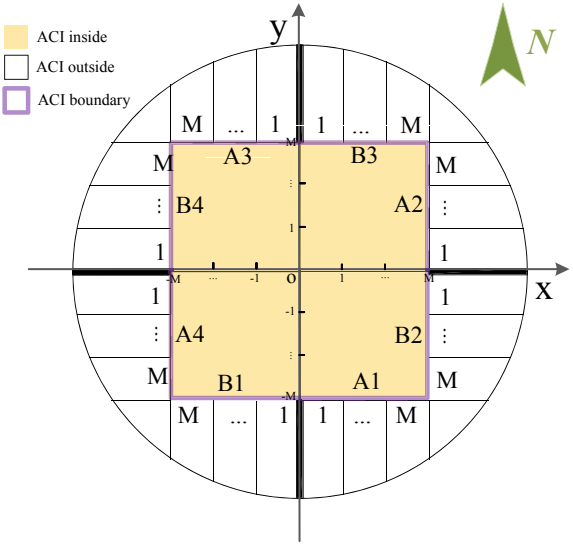


Fig. 4.2 ACI modeling process

4.2.2 ACI Modeling

The ACI needs to be modeled before the conflict resolution model can be constructed. A regular intersection with an equal number of approach and departure lanes is selected as the model base. Define the approach direction of the intersection as A. Let the approach lane of coordinate direction S be A_1, rotated counterclockwise, and the other three approach lanes be A_2, A_3, A_4, respectively. Define the exit direction as B, corresponding to the approach lane number so that the four departure directions are numbered B_1, B_2, B_3, B_4. Take the road centerline as the reference from left to right for the approach lanes numbered $1, 2, \cdots, M$, from right to left for the departure lanes numbered $1, 2, \cdots, N$, $M(1 \leq m \leq M)$ is the number of lanes at approach lane, and $N(1 \leq n \leq N)$ is the number of lanes at departure lane. The ACI modeling process is shown in Fig. 4.2.

4.2.3 Conflict Resolution Process

The trajectory of the CAV between the approach lane and the departure lane is defined as channel (cha) for each CAV's motion process inside the ACI. In each detection period, ACI will collect the passing requests of each CAV and establish the objective function of CAV passing ACI.

$$\min Z = \sum\nolimits_i (\omega_{l_i} + \omega_{t_i} + \omega_{\theta_i}) \tag{4.1}$$

where ω_{l_i} is the channel length resistance of the CAV_i in selecting a certain departure lane; ω_{t_i} is the channel time resistance of the CAV_i in selecting a certain departure lane; and ω_{θ_i} is the steering angle resistance of the vehicle CAV_i in selecting a certain departure lane. The constraints are:

$$\begin{cases} \omega_{l_{i,(Im-On)}} = \frac{l_{i,(Im-On)} - min(l_{i,cha})}{max(l_{i,cha}) - min(l_{i,cha})}, l_{i,(Im-On)} \in \left\{l_{i,cha}\right\} \\ \omega_{t_{i,(Im-On)}} = \frac{t_{i,(Im-On)} - min(t_{i,cha})}{max(t_{i,cha}) - min(t_{i,cha})}, t_{i,(Im-On)} \in \left\{t_{i,cha}\right\} \\ \omega_{\theta_{i,(Im-On)}} = \frac{\theta_{i,(Im-On)} - min(\theta_{i,cha})}{max(\theta_{i,cha})}, \theta_{i,(Im-On)} \in \left\{\theta_{i,cha}\right\} \\ l_{i,(Im-On)} = length(A_{Im} - B_{On}) \\ t_{i,(Im-On)} = \frac{l_{i,(Im-On)}}{v_{i,cha}, (I \neq O, I+O=4 or 6, v_{i,cha}=v_{turn}, else, v_{i,cha}=v_{\text{straight}})} \\ \theta_{i,(Im-On)} = degree(\angle(A_{(4-I)m} A_{Im} B_{On})) \end{cases} \tag{4.2}$$

where I is the number of the approach direction, m is the approach lane number, $1 \leq m \leq M$, O is the number of the departure direction, n is the approach lane number, $1 \leq n \leq N$. $Im - On$ is the channel indicating the approach direction as I, the approach lane as m, the departure direction as O, and the departure lane as n. $l_{i,(Im-On)}$ is the length of the channel for $Im - On$. $t_{i,(Im-On)}$ is the channel $Im - On$ passage time of CAV_i. $\theta_{i,(Im-On)}$ is the steering angle of CAV_i when the channel is $Im - On$.

$l_{i,cha}$ is the length of all selectable channels of CAV_i. $t_{i,cha}$ is the passage time of all selectable channels of CAV_i. $\theta_{i,cha}$ is the steering angle all selectable channels of CAV_i. $v_{i,cha}$ is the travel speed of CAV_i passing the channel, v_{turn} is the travel speed of CAV_i passing the turning channel, and v_{straight} is the travel speed of CAV_i passing the straight channel, where I $\neq$ O. According to the results of the above optimal control problem, the pre-assigned lanes $A_{Im} - B_{On}$ of each CAV passing the ACI can be obtained.

Based on the pre-assigned channels $A_{Im} - B_{On}$ of each CAV_i, the set Y of pre-assigned channel trajectories of all CAV_i is obtained. Based on the starting coordinates $\left(x_{i(A_{Im})}, y_{i(A_{Im})}\right)$ when CAV_i enters the ACI and the ending coordinates $\left(x_{i(B_{On})}, y_{i(B_{On})}\right)$ when it leaves the ACI, the trajectory of CAV_i's channel inside the ACI can be obtained as:

$$\begin{cases} y_i = \frac{y_{i(B_{On})} - y_{i(A_{Im})}}{x_{i(B_{On})} - x_{i(A_{Im})}} x_i + \frac{y_{i(A_{Im})} x_{i(B_{On})} - y_{i(B_{On})} x_{i(A_{Im})}}{x_{i(B_{On})} - x_{i(A_{Im})}}, Turn \\ x_i = n, x_{i(A_{Im})} = x_{i(B_{On})}, Straight(north - south) \\ y_i = n, y_{i(B_{On})} = y_{i(A_{Im})}, Straight(east - west) \\ x_i \in [-M, M], y_i \in [-M, M] \end{cases} \tag{4.3}$$

Then we determine whether there is space-conflict between each CAV by iteratively solving whether there is a common solution between the channel trajectories of each CAV. If there is a space-conflict, the conflict coordinate points $\left(x_{i,conflict}, y_{i^+,conflict}\right)$ need to satisfy the following conditions.

$$\begin{cases} x_{i,conflict} \\ y_{i^+,conflict} \end{cases}, i \neq i^+, i, i^+ \in \mathbb{C}\{1, 2, 3, 4, ...\}, x_{i,conflict}, y_{i^+,conflict} \in Y \quad (4.4)$$

The relationship between the common solution of each CAV channel trajectory and the conflict classification is:

$$\begin{cases} \text{unsolvable}: \text{No conflict} \\ \text{Solvable} \begin{cases} \text{Innumerable solutions}: cha \text{ duplication conflict} \\ \text{Unique solution} \begin{cases} x,\ y \in (-M, M): cha \text{ cross conflict} \\ x = \pm M,\ or,\ y = \pm M: cha \text{ merge conflict} \end{cases} \end{cases} \end{cases} \quad (4.5)$$

Based on the above conflict distance results, the moment when the CAV reaches the beginning of the conflict region is further derived as $t^{start}_{i(i^+),conflict} = t_{i(i^+),I} + \frac{d^{start}_{i(i^+),conflict}}{v_{i(i^+),cha}}$, where $t^{start}_{i(i^+),conflict}$ is the moment when the $\text{CAV}_{i(i^+)}$ reaches the conflict region. $t_{i(i^+),I}$ is the moment when $\text{CAV}_{i(i^+)}$ reaches the ACI boundary. $d^{start}_{i(i^+),conflict}$ is the distance from $\text{CAV}_{i(i^+)}$ to the conflict region. $v_{i(i^+),cha}$ is $\text{CAV}_{i(i^+)}{}'$s speed travels in the cha.

Time-conflict judgment is performed here. If $\left|t^{start}_{i,conflict} - t^{start}_{i^+,conflict}\right| \geq \frac{d_{safe}}{min\left\{v_{i,cha}, v_{i^+,cha}\right\}}$ then it indicates that there is no time conflict in the space-conflict region. If $\left|t^{start}_{i,conflict} - t^{start}_{i^+,conflict}\right| < \frac{d_{safe}}{min\left\{v_{i,cha}, v_{i^+,cha}\right\}}$ then indicates the existence of time-conflict in the space-conflict region where $t^{start}_{i,conflict}$ is the moment when the CAV_i reaches the conflict region. $t^{start}_{i^+,conflict}$ is the moment when the CAV_{i^+} reaches the conflict region. d_{safe} is the safe distance between two CAVs. $min\left\{v_{i,cha}, v_{i^+,cha}\right\}$ is the minimum operating speed of CAV_i and CAV_{i^+} within cha. If there is a conflict between pre-assigned lanes, a departure lane change is made to the$\text{CAV}_{i(i^+)}$. Replace the departure channel of CAV_i and CAV_{i^+} from B_{On} to$B_{On'}$. The exit channel coordinates are converted to$\left(x'_{i(i^+)}, y'_{i(i^+)}\right)_{B_{On'}}$, where$n \neq n'$. Repeat the above conflict judgment process until there is no conflict between CAVs and then the passing strategy can be determined.

4.3 Experimental Analysis

In this study, a regular intersection is selected as the reference model of ACI in the NETLOGO simulation environment, and each road is 6 lanes in both directions, i.e.$M = N = 3$. The width of the lane is 3.5 m. Assuming that all vehicles are CAVs, effective information transfer can be achieved between CAVs and between CAVs and intersections. The CAV sends a pass request to the control center when it is 100 m away from the ACI boundary. The detection period is set to 1 s. Let

$v_{i,cha} = 10$ km/h , $v_{straight} = 20$ km/h, d_{safe}=7 m.The acceleration of each CAV is between $\left[-5m/s^2, 3m/s^2\right]$.

To verify whether the solution speed of the proposed method in this study can be completed before each next detection period. We analyzed the solution capacity of the algorithm at the detection period (1 s), and the results are shown in Fig. 4.3.

The pass strategy proposed in this study is fundamentally a multi-objective dynamic optimization problem. In general, the solution time of multi-objective dynamic optimization problems increases with the complexity of the problem as well as the problem scale [14]. With Fig. 4.3, the solution time is proportional to the number of CAVs. When the number of CAVs is more than 12, the solution time also exceeds the detection period (1 s). The maximum capacity of ACI can be estimated to be 32,000 pcu/h based on the maximum solution capacity during the detection period, but this data is obtained under fully ideal conditions.

Based on the above idealized maximum capacity and detection period, the process of traffic flow passing ACI is simulated. The probability of generating CAV per lane in each approach direction is 1 pcu/5 s. The probability of each CAV going straight is 0.75, the probability of turning left is 0.15, and the probability of turning right is 0.1. Figure 4.4 illustrates the data statistics of all lanes of CAV entering and exiting ACI for a single inlet with a time accuracy of 0.1 s and a simulation time of 6 min.

Each CAV sends a pass request to ACI before passing through ACI, and ACI develops a pass strategy based on the pass request and returns it to each CAV. Each CAV receives the return strategy and starts the corresponding lane selection and speed adjustment. Figure 4.4 shows that the CAV speed was adjusted from higher to lower during the process of passing the ACI, and then speed recovery was achieved

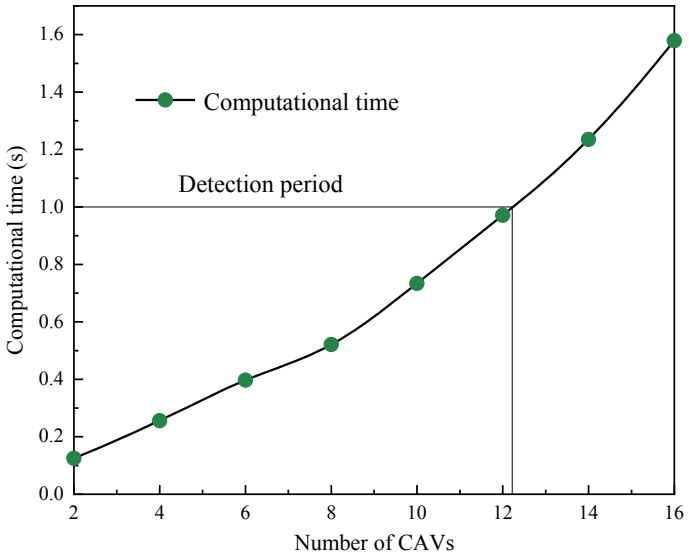


Fig. 4.3 Solving capacity in detection period

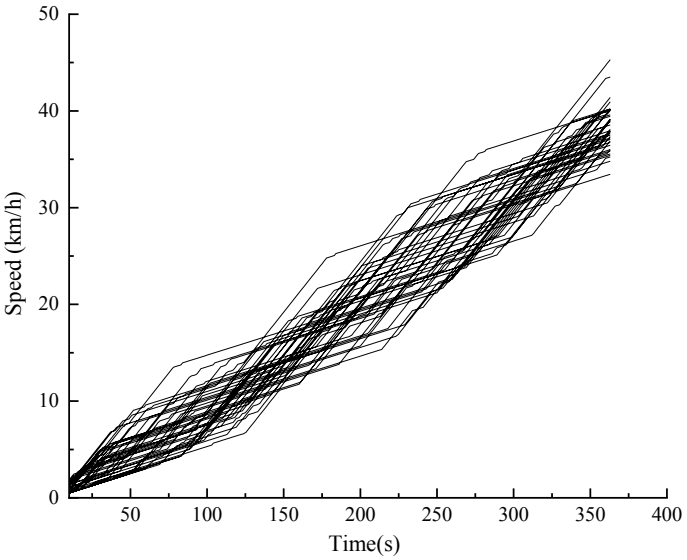


Fig. 4.4 Solving capacity in detection period

after passing the ACI, and passing the ACI at a lower speed also ensured the safe operation of the CAV. The above results show that the ACI ensured that each CAV passed the intersection with as little stopping as possible, indicating that the method proposed in this study can provide an effective coordination strategy for CAVs to pass the ACI. Not only that, this study also improves the passing efficiency of ACI and transforms the control mode of traditional signal intersections.

4.4 Conclusion

This study further deepens the transformation of the existing signal intersection control model by combining the initial ideas of ACI presented in the existing studies. We propose a control strategy for CAV to pass the intersection safely and efficiently, establish the corresponding dynamic optimization problem, analyze the solution time of this control problem, ensure the continuous execution of the control strategy, and finally also prove the effectiveness of the proposed control strategy in this study.

Funding This study is based on research supported by the National Natural Science Foundation of China (Number 52072066), Jiangsu Province Science Fund for Distinguished Young Scholars (Number BK20200014) and ICV-Safe: Testing safety of intelligent connected vehicles in open and mixed road environment (Vinnova 2019-03418). Any errors or omissions remain the sole responsibility of the authors.

References

1. Wang, L., Yang, M., Li, Y., Wang, B., Zhang, J.: Resolution strategies for cooperative vehicle fleets for reducing rear-end collision risks near recurrent freeway bottlenecks. Journal of Intelligent Transportation Systems, 1–19 (2022)
2. Lim, K.L., Whitehead, J., Jia, D., Zheng, Z.: State of data platforms for connected vehicles and infrastructures. Communications in transportation research **1**, 100013 (2021)
3. Wu, J., Ahn, S., Zhou, Y., Liu, P., Qu, X.: The cooperative sorting strategy for connected and automated vehicle platoons. Transportation Research Part C: Emerging Technologies **123**, 102986 (2021)
4. Wu, J., Qu, X.: Intersection control with connected and automated vehicles: a review. Journal of Intelligent and Connected Vehicles **5**(3), 260–269 (2022)
5. Olovsson, T., Svensson, T., Wu, J.: Future connected vehicles: Communications demands, privacy and cyber-security. Communications in Transportation Research **2**, 100056 (2022)
6. Wang, L., Yang, M., Li, Y., Hou, Y.: A model of lane-changing intention induced by deceleration frequency in an automatic driving environment. Physica A **604**, 127905 (2022)
7. Li, B., Zhang, Y.: Fault-tolerant cooperative motion planning of connected and automated vehicles at a signal-free and lane-free intersection. IFAC-PapersOnLine **51**(24), 60–67 (2018)
8. He, Z., Zheng, L., Lu, L., Guan, W.: Erasing lane changes from roads: A design of future road intersections. IEEE Transactions on Intelligent Vehicles **3**(2), 173–184 (2018)
9. Fayazi, S.A., Vahidi, A., Luckow, A.: A Vehicle-in-the-Loop (VIL) verification of an all-autonomous intersection control scheme. Transportation Research Part C: Emerging Technologies **107**, 193–210 (2019)
10. Zhang, Y., Cassandras, C.G.: Decentralized optimal control of connected automated vehicles at signal-free intersections including comfort-constrained turns and safety guarantees. Automatica **109**, 108563 (2019)
11. Mirheli, A., Hajibabai, L., Hajbabaie, A.: Development of a signal-head-free intersection control logic in a fully connected and autonomous vehicle environment. Transportation Research Part C: Emerging Technologies **92**, 412–425 (2018)
12. Zhu, J., Easa, S., Gao, K.: Merging control strategies of connected and autonomous vehicles at freeway on-ramps: a comprehensive review. Journal of Intelligent and Connected Vehicles **5**(2), 99–111 (2022)
13. Gao, K., Yang, Y., Qu, X.: Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. Transp. Res. Part D: Transp. Environ. **97**, 102957 (2021)
14. Wang, J., Gong, S., Peeta, S., Lu, L.: A real-time deployable model predictive control-based cooperative platooning approach for connected and autonomous vehicles. Transportation Research Part B: Methodological **128**, 271–301 (2019)

Chapter 5
Multivariate Sequence Clustering for Driving Preference Classification Based on Wide-Range Trajectory Data

Shuli Wang, Ruo Jia, and Lanfang Zhang

Abstract Accurate driving preferences classification is a crucial component for autonomous connected vehicles in making more safety and more efficient driving decisions. Most existing studies identify drivers' driving preferences based on the historical data of the individual vehicle, and the selected variables are limited to the mechanical motion of the vehicle, which seldomly takes the influence of road traffic conditions and surrounding vehicles into account. This study proposes a driving preferences classification method by multivariate sequence clustering algorithm based on wide-range trajectory data. Based on the specific range of road sections, the selected variables for each trajectory are converted from the time domain to the space domain separately, to capture the dynamic changes of the features along the road area. Multivariate time series clustering combining a weighted Dynamic Time Warping (WDTW) and the *k*-medoids algorithm is used to classify driving preferences into different levels, and a popular internal evaluation metric is employed to determine the optimal cluster result. This study also investigates the heterogeneity of driving behaviors at different driving preference levels. The results show that the proposed method could better recognize drivers' internal driving preferences.

5.1 Introduction

Drivers' driving preferences refer to the behavioral choices of drivers with a great difference under the same traffic condition due to their internal driving styles and driving skills. Accurately understanding and recognizing different driving preferences of surrounding vehicles is essential for autonomous connected vehicles in

S. Wang · L. Zhang (✉)
The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China
e-mail: zlf2276@tongji.edu.cn

S. Wang · R. Jia
Department of Architecture and Civil Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_5

making more safety and more efficient driving decisions [1, 2]. However, the majority of driving preference classifications apply the historical behavioral data of the individual vehicle to identify driving preferences, ignoring the influence of road traffic conditions and surrounding vehicles [3]. Actually, position-relevant traffic factors significantly affect drivers' driving preferences, such as road geometric alignment and fixed-point speed limit [4]. Besides, in real-world traffic scenarios, drivers are diverse and usually make different behavioral choices, even under the same traffic condition [4, 5]. Therefore, analyzing different drivers' driving trajectories under specific road sceneries and traffic flows is necessary to classify drivers' preferences accurately.

This study proposes a driving preferences classification model based on specific road sections. Using wide-range trajectory data which contains all trajectory data that belong to all the vehicles passing the areas covered by Millimeter-wave (MMW) radars, frame-series-based variables (velocity, acceleration, and time headway) for each trajectory within the specific road sections are extracted. Then the selected variables are converted from the time domain to the space domain separately, to capture the dynamic changes of the features along the road area. Multivariate time series clustering combining a weighted Dynamic Time Warping (WDTW) and the *k*-medoids algorithm is used to classify driving preferences for driving safety into different levels. The proposed driving preference classification model could better recognize drivers' internal driving preferences for driving safety.

5.2 Literature Review

There are plenty of works on long-term driving style classification, which require considerably long observations of historical data of the individual vehicle [6, 7]. However, it is widely acknowledged that a driver may vary his/her driving style as the traffic environment changes [8]. Thus, this paper investigates drivers' short-term driving preferences based on short-term observations of specific road sections, which would be more suitable for real-time application. With the development of data acquisition and the increase in feature quantity, researchers tend to employ multivariate sequences classification methods to classify drivers' driving preferences into multiple groups with different driving profiles. Due to the characteristics of high dimensionality and dynamics of multivariate sequences, principal component analysis (PCA) [9], wavelet analysis [10], and dynamic time warping (DTW) [11] are commonly applied to the clustering analysis. Among these, DTW can solve the non-spherical distribution of multivariate sequences data well. Moreover, Artificial Neural Network (ANN) [12] also is adopted to improve the driving style classification performance but suffer from high learning effort.

The variables used to capture drivers' driving preferences are limited to the velocity, acceleration, and mechanical motion of the vehicle [13]. There are no parameters related to traffic conditions that have been taken into account. Actually, traffic conditions can affect drivers' driving preferences significantly. To this

end, this study investigates driving preferences based on all vehicle trajectories of specific road sections. Besides, domain conversion has also been adopted to improve classification accuracy. Some works converted the time domain to the frequency domain to deal with the problem of the variant length of each trajectory in the time domain [14]. Also, some scholars also converted the time domain into the space domain to analyze the influence of position-related traffic factors on driving preferences [15]. Motivated by the above works, this paper uses time-domain conversion on data preprocessing and proposes a driving preferences classification model based on the multivariate sequences clustering method.

5.3 Methodology

In this study, a driving preference classification model is proposed based on the short-term trajectory data of multiple vehicles. Considering position-relevant traffic factors and different trajectory lengths in the time domain, the selected features (speed, acceleration, and time headway) are converted from the time domain to the space domain separately. Then multivariate time-series clustering algorithms are applied to classify drivers' driving preferences into multiple groups (see Fig. 5.1).

5.3.1 Domain Conversion

According to the literature [4], position-relevant traffic factors significantly affect drivers' driving preferences, such as road geometric alignment and fixed-point speed limits. Therefore, the trajectories in a given road section will be applied in the driving preference classification. Besides, each trajectory's length in the time domain is different on account of the travel time of each vehicle through a given road area is different. To this end, the selected variables for each trajectory are converted from the time domain to the space domain separately, to capture the dynamic changes of the features along the road area.

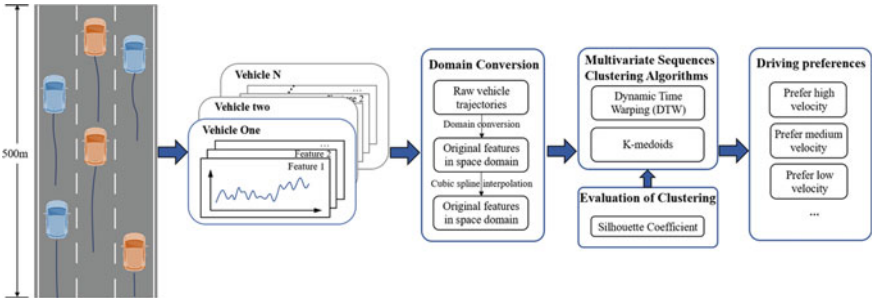


Fig. 5.1 Overall methodology flowchart

In the conversion process, due to the sampling frequency of raw trajectory data in the time domain being time-invariant ($\Delta t = 0.1s$), the sampling scale of these transferred features in the space domain might be deformed. To deal with this problem, the cubic spline interpolation is applied to interpolate the missing data for selected features based on a specific spatial frequency, ensuring that every desired space point along the road has data for each trajectory.

5.3.2 Multivariate Sequences Clustering Algorithms

Based on the selected features for each trajectory in the space domain, the drivers' driving preferences are classified into several groups. Thus, the driving preference classification task could be regarded as a multivariate time-series clustering problem. Multivariate time-series clustering combining a weighted Dynamic Time Warping (WDTW) and the *k*-medoids algorithm is employed.

Weighted Dynamic Time Warping (WDTW)

The WDTW is a distance measurement method to measure the similarity between two given sequence data [16]. A greater WDTW distance suggests a more significant difference between the two sequences. Compared to the DTW, the WDTW incorporates different weights into each child distance $\|(a_i - b_j)\|^2$, where a_i and b_j are two points from two different sequences (A and B). Thus, the distance is computed as

$$d_w(a_i, b_j) = \|w_{|i-j|}(a_i - b_j)\|^2 \tag{5.1}$$

where the weight $w_{|i-j|}$ is related to the phase difference $\Delta p = |i - j|$, the further the two points a_i and b_j are in phase, the larger the value of Δp is, the greater the effect of the weight $w_{|i-j|}$ on the distance $d_w(a_i, b_j)$ is. In the study, the weight value is defined as

$$w_{\Delta p} = w_{\max}/(1 + \exp(-g(\Delta p - m_c))) \tag{5.2}$$

where $w_{\max}$ is a predefined parameter and is often set to be one. The parameter g is a penalty factor that is applied to control the penalization level for the child distance between points with a large phase difference. The parameter m_c is the median value of the sequence.

According to the dynamic programming method, the cumulative cost matrix R can be figured out as follows:

$$R(i, j) = d_w(i, j) + \min\{R(i, j-1), R(i-1, j), R(i-1, j-1)\} \tag{5.3}$$

where $R(0, 0) = 0$, $R(i, 0) = \infty$, and $R(0, j) = \infty$. Therefore, WDTW (A, B) is equal to the cumulative distance $R(n, m)$, where n and m are the lengths of sequences A and B, respectively.

K-medoids method

The k-medoids method is a clustering algorithm similar to the well-known k-means. Different from the k-means uses the mean of the cluster, the k-medoids method applies the medoids of the cluster to update the cluster center, which, thus, is one of the actual members. The criterion function for selecting a medoid of the cluster is that the sum of the distances from the remaining member points in the current cluster to that medoid is minimized. Due to this difference, the k-medoids method is more robust to noise and outliers in the dataset [17].

WDTW distance-based k-medoids method

Given a multivariate sequences (MVS) dataset $S = \{S_1, S_2, \ldots, S_N\}$, where S_i is the ith MVS from the ith vehicle. Each of the S_i consists of m component univariate sequences (CUVS) S_i^p, $p = 1, 2, \ldots, m$, where S_i^p is regarded as the pth feature sequence of the ith vehicle. WDTW is applied to measure the similarity distance between two MVSs S_i and S_j:

$$D_{ij} = \mathrm{WDTW}(S_i, S_j) \tag{5.4}$$

However, MVSs have high-dimensional component sequences, which make it hard to calculate the similarity. Since each dimension of the MVS S_i can be regarded as the CUVS, the similarity between any two MVSs can be calculated as the sum of the similarity of the same CUVS by the WDTW.

$$D_{ij} = \sum_{p=1}^{m} \mathrm{WDTW}(S_i^p, S_j^p) \tag{5.5}$$

From Eq. (5.5), the similarity matrix can be established between any two S_i and S_j.

Then, the k-medoids method is applied to cluster based on the similarity matrices and builds up the correlation matrices for various components and the whole information of multivariate time series.

5.3.3 *Evaluation of Clustering*

Due to the ground-truth data (e.g. class labels) are unavailable for external validation in this study, internal clustering needs to be applied for clustering validation. A popular internal evaluation metric is employed: the Silhouette Coefficient (SC) [18]. Given a clustering result with k clusters, the SC ($s(i)$) for trajectory i is given by:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5.6}$$

where $a(i)$ is the average intra-cluster WDTW distance in the same cluster, and $b(i)$ is the smallest average WDTW distance in the second nearest neighboring cluster. Finally, the average $s(i)$ of all trajectory clusters is the measure of the overall quality of clustering results. The SC is given by:

$$\mathrm{SC} = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{|C_k|}\sum\nolimits_{i=1}^{|C_k|} s(i)\right) \tag{5.7}$$

where $|C_k|$ denotes the number of trajectories belonging to cluster k. The SC values range from -1 to $+1$, where a high value indicates a great cluster configuration.

5.4 Data Preparation

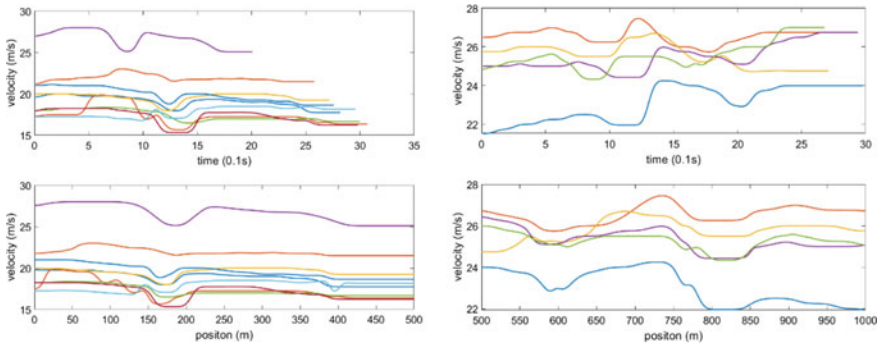
5.4.1 Dataset Description

This study uses the real-world vehicle trajectory data of Hangzhou Xifu Freeway provided by the TJRD TS platform [18]. The data were collected by MMW radars, which are installed on both sides of the road section with the detection set of a two-radar combination. The range for equipment installation is around 1.0 km in length. The dataset contains all frame-series-based data for each track that potentially belong to all the vehicles passing the covered area. The data collection frequency is 10 Hz.

The data recordings with a duration of 15 min corresponding to steady flow are utilized. Thus, 527 trajectory samples of vehicles passing through the study area are extracted to study the driving preferences classification in this work.

5.4.2 Data Preprocessing

Due to the measurement errors of object detection and tracking from MMW radars, there was missing data and some noise in the raw data, such as velocity and acceleration suddenly dropping to zero. To solve these problems, the missing timestamp data were interpolated, and the duplicate timestamps were deleted. The actual velocity and acceleration were repaired and smoothed. The time headway between two consecutive vehicles pair was calculated to estimate the rear-end collision risk. The raw variables (longitudinal velocity, acceleration, and time headway) were finally collected.



(a) Road section one (b) Road section two

Fig. 5.2 Examples of speed profiles in the time and station domains in different road sections

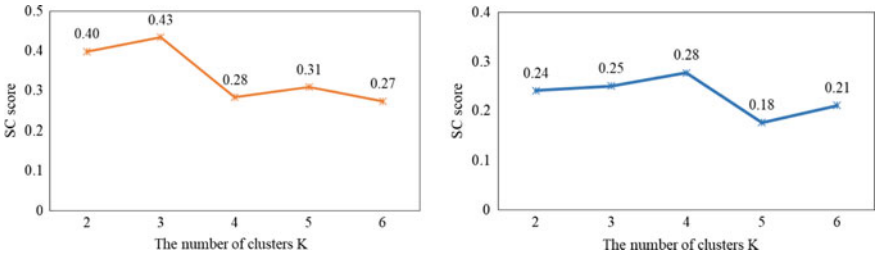
The trajectories were segmented into two sections, 500 m each. For each section, the collected features for each trajectory were then converted from the time domain to the station domain separately according to Sect. 5.3.1. Considering the sampling frequency in the time domain (0.1 s) and travel speed (about 24 m/s), each road section was discretized with a step size of 2 m. For instance, Fig. 5.2 presents several examples of velocity trajectories in the time and station domains. As can be seen from the velocity trajectories in the space domain, the velocity of most vehicles decreases in the specific range of the road section (150–200 and 750–800 m), which is attributed to the speed limit sign at these locations, while these phenomena occur at different time points in the time domain. Finally, the selected features in the space domain for each trajectory were standardized for the multivariate sequence clustering.

5.5 Results and Discussion

5.5.1 Classification Results of Driving Preferences

To find the optimal number of clusters for each road section, different values of K were tested ranging from 2 to 6. According to Fig. 5.3a, for road section one, it is concluded that the optimal K is 3 where SC reaches its maximum. Similarly, for road section two, it is concluded that the optimal K is 4 for clustering.

Table 5.1 shows the summary statistics for each cluster for two road sections, including mean values and standard variances (in brackets) of velocity, acceleration, and time headway. In general, for cluster A, it can be observed that drivers prefer to be more aggressive with greater velocity, and larger variance in velocity. Conversely, for the last cluster, drivers in this driving preference are more cautious with lower



(a) SC for road section one (b) SC for section two

Fig. 5.3 The relationship between the number of clusters and the evaluation metric

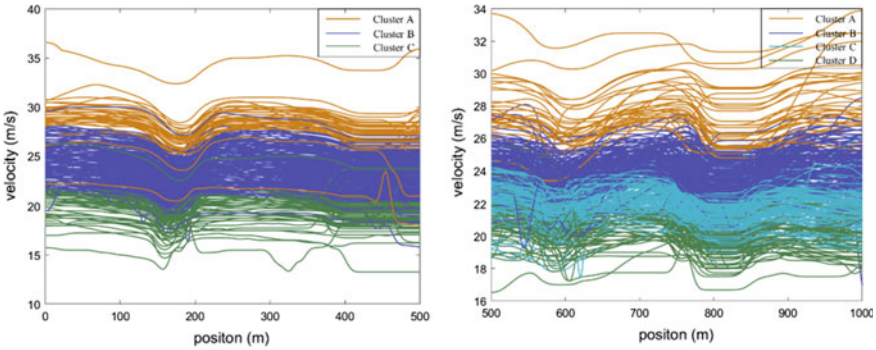
Table 5.1 The clustering results

Clusters	Velocity (m/s)	Acceleration (m/s^2)	Time headway (s)	Driving preferences
Road section one				
A	28.43 (0.92)	−0.06 (0.54)	6.02 (2.18)	Prefer high velocity
B	25.08 (0.71)	−0.07 (0.42)	5.92 (2.52)	Prefer medium velocity
C	19.84 (0.72)	−0.06 (0.40)	5.33 (3.02)	Prefer low velocity
Road section two				
A	26.65 (0.81)	−0.03 (0.54)	6.32 (2.18)	Prefer high velocity
B	24.14 (0.71)	−0.07 (0.42)	5.52 (2.52)	Prefer medium velocity
C	22.52 (0.72)	−0.06 (0.47)	5.73 (3.25)	Prefer slightly low velocity
D	19.87 (0.75)	−0.09 (0.45)	5.43 (2.92)	Prefer low velocity

velocity. According to the significance test, only the velocity values between clusters are significantly different ($p = 0.01$). Thus, driving preferences are defined as the level of preference for speeding driving.

5.5.2 *Characteristics of Different Driving Preferences*

Specifically, for road section one, the drivers are classified into three levels of driving preferences. Cluster C is the most cautious profile, which shows a low velocity that is lower than average traffic velocity. Meanwhile, the drivers' driving preferences are classified into four levels for road section two. Compared to the other three clusters, drivers in Cluster A prefer to travel at a high velocity with a larger variance (see Fig. 5.4).



(a) Road section one (b) Road section two

Fig. 5.4 Velocity trajectories by cluster for different road sections

5.6 Conclusion

In this study, a novel driving preference classification method is proposed to label drivers' different preference levels for speeding driving by classifying short-term trajectory observations within the specific road section. Using wide-range trajectory data which contains all trajectory data that belong to all the vehicles passing the study areas, the main findings could be summarized as follows:

Considering the influence of road conditions, the selected frame-series-based variables (velocity, acceleration, and time headway) for each trajectory within the specific road sections are converted from the time domain to the space domain separately. It could deal with the problem of the variant length of each trajectory in the time domain and capture the dynamic changes of the features along the road area.

The driving preferences are classified into different levels by classifying short-term trajectory observations by multivariate time series clustering combining a WDTW and the *k*-medoids algorithm, which would be more suitable for real-time application.

This work provides a novel method to achieve a great classification of drivers' internal driving preferences excluding the influence of external traffic environments. For future work, the proposed method will be employed on continuous sections along the path to capture the dynamic transitions of driving preferences facing different surrounding environments. It will enable dynamic evaluation of the driving behaviors of surrounding vehicles (especially human-driven vehicles) to enhance the driving safety of autonomous connected vehicles.

Acknowledgements This study was supported by the project "Safety Design Technology of the Multi-Entrances and Exits for Urban Underground Expressway".

References

1. Li, Y., Zhao, L., Gao, K., An, Y., Andric, J.: Revealing driver psychophysiological response to emergency braking in distracted driving based on field experiments. J. Intell. Connect. Veh. **5**(3), 270–282 (2022)
2. Xue, Q., Gao, K., Xing, Y., Lu, J., Qu, X.: A context-aware framework for risky driving behavior evaluation based on trajectory data. IEEE Intell. Transp. Syst. Mag. **15**(1), 70–83 (2023). https://doi.org/10.1109/MITS.2021.3120279
3. Lyu, N., Wang, Y., Wu, C., Peng, L., Thomas, A.F.: Using naturalistic driving data to identify driving style based on longitudinal driving operation conditions. J. Intell. Connect. Veh. **5**(1), 17–35 (2022)
4. Lim, H., Su, W., Mi, C.C.: Distance-based ecological driving scheme using a two-stage hierarchy for long-term optimization and short-term adaptation. IEEE Trans. Veh. Technol. **66**(3), 1940–1949 (2016)
5. Gao, K., Tu, H., Sun, L., Sze, N.N., Song, Z., Shi, H.: Impacts of reduced visibility under hazy weather condition on collision risk and car-following behavior: implications for traffic control and management. Int. J. Sustain. Transp. **14**(8), 635–642 (2020). https://doi.org/10.1080/15568318.2019.1597226
6. Wang, L., Yang, M., Li, Y., Hou, Y.: A model of lane-changing intention induced by deceleration frequency in an automatic driving environment. Phys. A **604**, 127905 (2022)
7. Wang, L., Yang, M., Li, Y., Wang, B., Zhang, J.: Resolution strategies for cooperative vehicle fleets for reducing rear-end collision risks near recurrent freeway bottlenecks. J. Intell. Transp. Syst., 1–19 (2022)
8. de Zepeda, M.V.N., Meng, F., Su, J., Zeng, X.-J., Wang, Q.: Dynamic clustering analysis for driving styles identification. Eng. Appl. Artif. Intell. **97**, 104096 (2021). https://doi.org/10.1016/j.engappai.2020.104096
9. Li, H.: Accurate and efficient classification based on common principal components analysis for multivariate time series. Neurocomputing **171**, 744–753 (2016)
10. D'Urso, P., Maharaj, E.A.: Wavelets-based clustering of multivariate time series. Fuzzy Sets Syst. **193**, 33–61 (2012)
11. Han, T., Peng, Q., Zhu, Z., Shen, Y., Huang, H., Abid, N.N.: A pattern representation of stock time series based on DTW. Phys. A **550**, 124161 (2020). https://doi.org/10.1016/j.physa.2020.124161
12. Zhang, Y., Li, J., Guo, Y., Xu, C., Bao, J., Song, Y.: Vehicle driving behavior recognition based on multi-view convolutional neural network with joint data augmentation. IEEE Trans. Veh. Technol. **68**(5), 4223–4234 (2019). https://doi.org/10.1109/TVT.2019.2903110
13. Constantinescu, Z., Marinoiu, C., Vladoiu, M.: Driving style analysis using data mining techniques. Int. J. Comput. Commun. Control **5**(5), 654–663 (2010)
14. Xue, Q., Lu, J., Gao, K.: Driving style recognition incorporating risk surrogate by support vector machine. In: Smart Transportation Systems 2021, pp. 123–131. Springer, Singapore (2021)
15. Yang, S., Wang, W., Xi, J.: Leveraging human driving preferences to predict vehicle speed. IEEE Trans. Intell. Transp. Syst. (2021)
16. Jeong, Y.S., Jeong, M.K., Omitaomu, O.A.: Weighted dynamic time warping for time series classification. Pattern Recogn. **44**(9), 2231–2240 (2011)
17. Chen, Y., Liu, X., Li, X., Liu, X., Yao, Y., Hu, G., Xu, X., Pei, F.: Delineating urban functional areas with building-level social media data: a dynamic time warping (DTW) distance based k-medoids method. Landsc. Urban Plan. **160**, 48–60 (2017). https://doi.org/10.1016/j.landurbplan.2016.12.001
18. Wang, J., Fu, T.: TJRD TS. https://www.tjrdts.com (2021)

Chapter 6
Optimizing the Deployment of Automated Speed Camera at the Intersections Using GPS Trajectories

Hua Liu, Chuanyun Fu, and Kun Gao

Abstract The economical and rational deployment of automated speed camera is a critical issue for traffic police department to implement speed management efficiently. Based on taxi GPS trajectories collected from Chengdu, 2016, this study optimizes the deployment interval and number of ASCs at the intersections by using K-means clustering and kernel density estimation according to the critical mixed distance halo effect and the delta speed distribution, respectively. Results illustrate that speeding is more likely to happen within the speed limit of 40 km/h rather than 60 km/h. From the whole perspective, with the growing deployment number of ASCs, the upstream distance halo effects gradually increase, while the downstream distance halo effects gradually decrease within the range of about 4500 m. Given that the interaction between two adjacent ASCs, the critical mixed distance halo effect of ASCs is about 215 m and 529 m corresponding to the smaller and larger values of critical delta speed in the northeast direction respectively, and about 315 m and 585 m in the southwest direction. Generally, one ASC should be deployed every 500 m and 600 m within the speed limit of 60 km/h, and every 200 m and 300 m within the speed limit of 40 km/h in the northeast and southwest directions, respectively.

H. Liu
School of Transportation, Southeast University, Jiangsu, China

C. Fu (✉)
School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China
e-mail: fuchuanyun@hit.edu.cn

K. Gao
Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_6

6.1 Introduction

Speeding is a globally serious issue of road traffic safety, which always increases the probability and severity of traffic accidents [1]. In China, the number of traffic accidents and the loss of lives and properties caused by speeding have been among the forefront compared with traffic accidents related to other traffic violations [2]. The proper speed management is crucial to reducing the severity of road traffic accidents and preventing the occurrence of speeding [3].

Existing literature mainly focuses on three aspects. Firstly, the vehicle speeding warning system and intelligent speed adaptation system have been investigated, however, it is still in the stage of experiment [4, 5]. Secondly, traffic control measurements such as adjusting values of speed limit [6], setting warning signs of anti-speeding [7], and releasing advertisements of anti-speeding [8] are adopted. Thirdly, speed enforcements including the traffic police patrol and automated speed camera (ASC) are employed [9–12]. Currently, the intervene measures of speeding in China mainly focus on the ASCs and sometimes combine with other above-mentioned intervention measures. Nevertheless, the occurrence of speeding is still higher [13], which is mainly due to the lack of theoretical basis for deploying ASCs more economically and rationally.

At present, the application of installing ASCs to manage traveling speed of vehicles is usually based on the recorded accident threshold, terrain, visibility, and other environmental conditions by engineers, planners, and designers on the spot. However, the low quality and accuracy issues of traffic accident records [14–16] always disturb the identification of deterrence range of ASC. Fortunately, the GPS trajectory data can be used to analyze the influence of ASCs on the vehicle traveling speed and then ensure the positions of ASCs where the maximum deterrence effect of ASCs can be achieved.

To the best of authors' knowledge, few researchers explore the deployment conditions of ASCs and there is only one research [17] that optimizes the deployment of ASCs based on the prediction models of traffic accidents, however, ignores the deterrence ranges (i.e., distance halo effect) of ASCs. Current researches pay much attention to the distance halo effect of ASCs along the segment rather than at the intersections, which is usually carried out by measuring the traveling speed at multiple locations in the upstream and downstream of ASCs [9, 14, 18]. This method is cumbersome and inaccurate that the distance halo effect of ASCs varies from hundred meters [19] to several thousand meters [11], or even further [10, 20], which can also be solved by GPS trajectory data. Commonly, the deployment of ASCs is involved in the deployment number and interval. To be specific, the traveling speed distribution reflects various driving habits within the range of different speed limits. Additionally, the relative distances of ASCs deployment intervals at the intersections are always shorter than along the segments. Hence, a mixed deterrence effect between two adjacent ASCs should be considered.

In general, there are two knowledge gaps in current researches for optimizing the deployment of ASCs at the intersections. One is that the distance halo effect of

ASCs at the intersections is uncertain. Another gap involves that the ASC deployment only depends on the historical experience and lacks the theoretical basis. Therefore, this study is conducted to optimize the deployment interval and number of ASCs according to the critical mixed distance halo effect and the delta speed distribution at the intersections, respectively.

6.2 Methodology

6.2.1 Data Collecting and Processing

As presented in Fig. 6.1, six intersections located on the Red Star Road in Chengdu, China were selected as the research scope in this study. The coordinates of the start and end points of the selected road are (104.0892, 30.66468) and (104.1144, 30.6985), respectively. There were 891,533 and 1,007,224 trajectory points at six intersections in the northeast and southwest direction from November 5th to 18th in 2016, respectively. Since the obtained speed in the raw dataset was instantaneous speed, the average speed was then calculated to ensure the accuracy of data, followed by map matching and each trip identifying.

Combing the panorama map (2016) from Baidu website with its Developer Platform Coordinated Picker, the relative positions of ASCs at the intersections in the two traveling directions are shown in Table 6.1. Besides, there were two speed limits at six intersections in the two traveling directions, respectively. To be specific, the speed limit is 60 km/h (i.e., in the upstream of fourth ASC and in the downstream of the sixth ASC) and 40 km/h (i.e., from the downstream of the fourth ASC to the upstream of the sixth ASC) in the northeast direction. In the southwest direction, the speed limit in the upstream of the second ASC and in the downstream of the fourth ASC is 60 km/h, and the speed limit is 40 km/h from the downstream of the second ASC to the upstream of the fourth ASC.

6.2.2 Delta Speed Distribution

In order to extract the characteristics of speed distribution to avoid the disturbance of changed and different speed limits at the intersections, the delta speed (ΔV) was proposed, which is defined as the difference between the traveling speed and speed limit.

As presented in Table 6.2, the delta speed distributions in the northeast and southwest directions between on weekdays and weekends are almost consistent. In the northeast traveling direction, the peak value of kernel density estimation on the delta speed of GPS trajectory points within the range of 60 km/h is -18 km/h both on weekdays and weekends. However, this peak value of kernel density estimation is

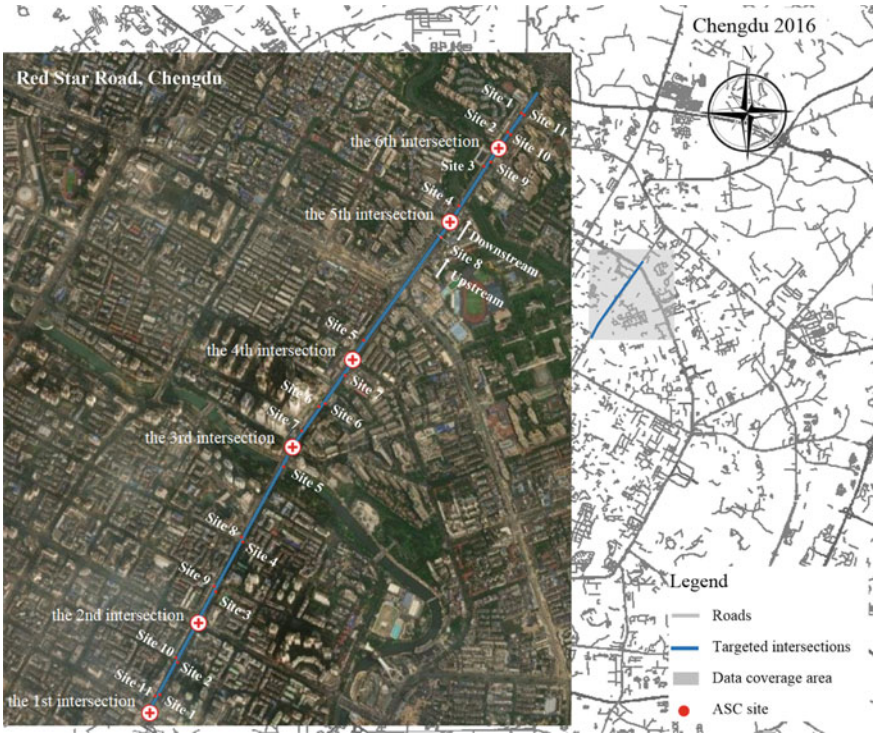


Fig. 6.1 Study area

positive (i.e., 12 km/h on weekdays and 10 km/h on weekends) within the range of 40 km/h. In the southwest traveling direction, the distribution of delta speed is similar with that in the opposite traveling direction.

In general, the probability of speeding is higher within the range of lower speed limit, which is consistent with previous studies [10]. Furthermore, the temporal stability of delta speed distribution is obtained, since the peak values of kernel density estimation on the delta speed of GPS trajectory points are consistent both on weekdays and weekends.

6.2.3 K-Means Clustering

In this study, the method of K-means clustering was employed to quantify the deterrence range of ASC according to the critical delta speed obtained by the clustering center of delta speed.

Firstly, the normalization of original dataset is applied due to the discrepancies of inter-column in different measuring units. After that, the data are divided into k

Table 6.1 Relative position of ASCs at the intersections

In the northeast direction		In the southwest direction	
Order of ASCs	RPFA (m)	Order of ASCs	RPFA (m)
1st intersection	−68	1st ASC	0
1st ASC	0	2nd ASC	185
2nd ASC	134	*1st intersection*	239
2nd intersection	369	3rd ASC	283
3rd ASC	598	4th ASC	500
4th ASC	917	*2nd intersection*	606
5th ASC	1166	5th ASC	1273
3rd intersection	1207	*3rd intersection*	1320
6th ASC	1449	6th ASC	1608
7th ASC	1674	7th ASC	1708
4th intersection	1704	*4th intersection*	1817
8th ASC	2346	8th ASC	2099
5th intersection	2418	9th ASC	2361
9th ASC	2746	*5th intersection*	2655
6th intersection	2785	10th ASC	2877
10th ASC	2839	11th ASC	3031
11th ASC	3024	*6th intersection*	3092

Note The site of the first ASC is considered as the benchmark, and RPFA refers to the relative position of the first ASC

Table 6.2 Distribution of delta speed

Variables		In the northeast direction (km/h)		In the southwest direction (km/h)	
		Weekdays	Weekends	Weekdays	Weekends
Speed limit (km/h)	60	−18	−18	−18	−25
	40	12	10	12	10
	60	−18	−18	−18	−18

clusters by using the elbow method, in which the location of a bend (i.e., the curve of the total within-cluster sum of square errors for each k) in the plot is often selected as an indicator of the optimal number of clusters. The similarity between any two pieces of data is quantified based on Euclidean distance [21].

Given that clustering results, the deterrence ranges of ASC (i.e., distance halo effect) in the upstream and downstream both on weekdays and weekends were then determined according to the positions where the critical delta speed appeared. Moreover, the mixed distance halo effect of ASCs was also further explored, since there

existed an overlapped area formed by the downstream of one ASC and the upstream of its adjacent ASC in the same traveling direction.

For installing ASCs more economically and rationally, the ASCs should be deployed in the case of achieving the maximum deterrence effect of adjacent ASCs. In other words, when the deterrence range of mixed distance halo effect of ASCs is at the critical value of disappearing, this value ($\Delta D_{\text{critical}}$) can be used as the deployment interval of ASCs. It is expressed as:

$$\Delta D_{\text{critical}} = \left|\Delta D_{\text{downstream}} - \Delta D_{\text{upstream}}\right| \tag{6.1}$$

$$\Delta D_{\text{downstream}} = D_{j,\text{downstream}} - D_{j+1,\text{upstream}} \tag{6.2}$$

$$\Delta D_{\text{upstream}} = D_{j+1,\text{upstream}} - D_{j,\text{downstream}} \tag{6.3}$$

where $\Delta D_{\text{critical}}$ denotes the critical value of mixed distance halo effect of two adjacent ASCs; $\Delta D_{\text{downstream}}$ is the difference between the downstream distance halo effect of the jth ASC and the upstream distance halo effect of the $j + 1$th ASC, when the downstream deterrence effect of the jth ASC is higher than the upstream deterrence effect of the $j + 1$th ASC; $\Delta D_{\text{upstream}}$ refers to the difference between the upstream distance halo effect of the $j + 1$th ASC and the downstream distance halo effect of the jth ASC, when the downstream deterrence effect of the jth ASC is lower than the upstream deterrence effect of the $j + 1$th ASC.

On the whole, the deployment interval of ASCs is determined by the critical value of mixed distance halo effect of adjacent ASCs, and the deployment number of ASCs is considered by the distribution of delta speed.

6.3 Results and Discussions

Figures 6.2 and 6.3 present the critical delta speed and its corresponding distance halo effect in the upstream and downstream between on weekdays and weekends in the northeast and southwest directions, respectively. In addition, the mixed distance halo effect with both larger and smaller critical delta speed between one ASC in the downstream, and its adjacent ASC in the upstream is also observed.

Although there exist two critical delta speeds for one ASC, their corresponding distance halo effects are consistent, which also verifies the accuracy of the deterrence range of ASC. From the whole perspective, with the growing deployment number of ASCs, the upstream distance halo effects gradually increase, while the downstream distance halo effects gradually decrease within the range of about 4500 m in this research scope. This is consistent with one study [22] that multiple ASCs are more effective than a single ASC within its research scope of 500 m.

In terms of the interaction between two adjacent ASCs, the mixed distance halo effect should be paid more attention. In the northeast direction, the difference between

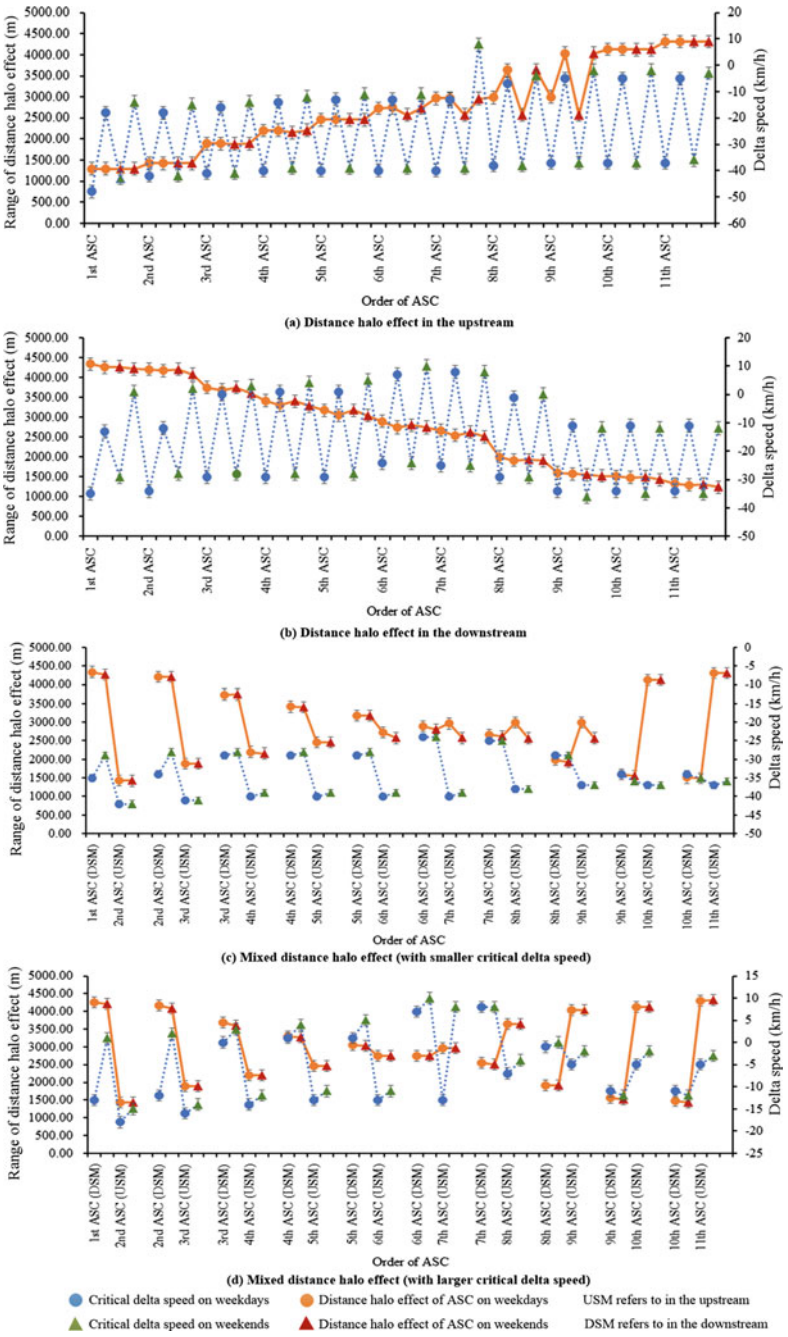


Fig. 6.2 Distance halo effect in the northeast direction

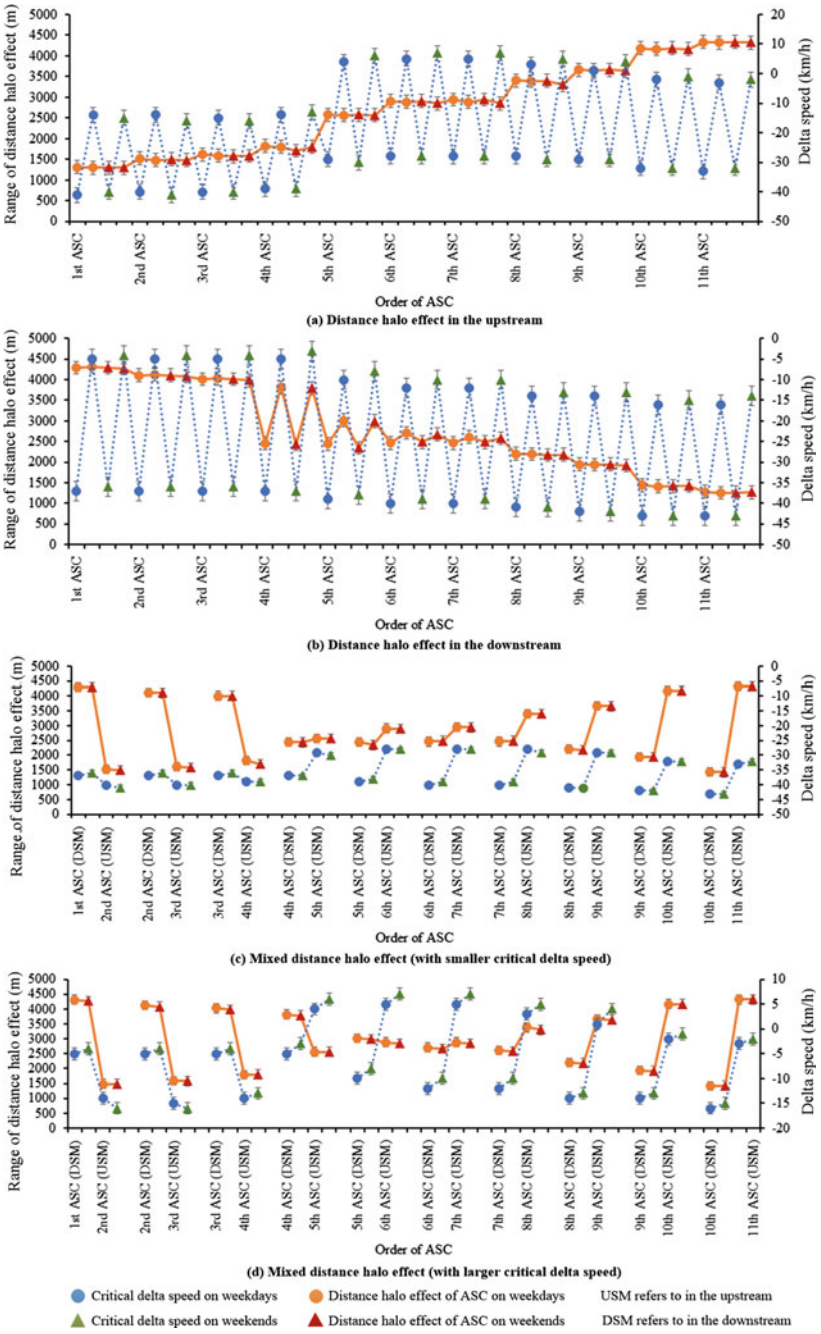


Fig. 6.3 Distance halo effect in the southwest direction

the downstream distance halo effect of one ASC and the upstream distance halo effect of its adjacent ASC gradually decreases until disappearing, and then this difference increases from the sixth and seventh ASCs. In other words, the downstream deterrence effect of one ASC is stronger than the upstream deterrence effect of its adjacent ASC within the range of the upstream of the sixth and seventh ASCs. After that, the opposite trend appears within the range of the downstream of the sixth and seventh ASCs. In the southwest direction, the trend of the difference between the distance halo effect of two adjacent ASCs is similar to that in the northeast direction.

To be specific, the mixed distance halo effects of ASCs are about 1674 m in the northeast direction and 1608–1708 m in the southwest direction, respectively. Though previous studies are not involved in the mixed distance halo effect of ASCs, the values of mixed distance halo effect of ASCs are almost consistent with one study [23] that the range of distance halo effect of ASC is up to 1.5 km from the ASC location. However, the values are much longer than most previous studies where the distance halo effect varies from 300 m to 1.1 km [19, 22, 24–26], since these researches mainly focus on the distance halo effect of ASC in the upstream and downstream, respectively, rather than the mixed distance halo effect of two adjacent ASCs.

In the northeast direction, the critical values of mixed distance halo effect of ASCs are about 215 m and 529 m corresponding to the smaller and larger values of critical delta speed, respectively. In the southwest direction, these values are about 315 m and 585 m, respectively.

As per above-mentioned critical values of mixed distance halo effect of ASCs, the deployment interval of ASCs can be determined. For further identifying the deployment number of ASCs, the distribution of delta speed is also considered. In other words, one ASC should be deployed every 500 m within the speed limit of 60 km/h and every 200 m within the speed limit of 40 km/h in the northeast direction. Similarly, one ASC should be deployed every 600 m within the speed limit of 60 km/h and every 300 m within the speed limit of 40 km/h in the southwest direction.

6.4 Conclusions

The objective of this study is to optimize the deployment of ASCs at the intersections using GPS trajectories more economically and rationally rather than only depending on the historical experience. Combing with the distribution of delta speed within the range of different speed limits and the critical mixed distance halo effect of ASCs, the deployment number and interval of ASCs are then determined.

Results in this study demonstrate that speeding is more likely to happen within the range of lower speed limit. Therefore, more ASCs should be deployed within the speed limit of 40 km/h. In terms of the interaction between two adjacent ASCs, the mixed distance halo effects of ASCs are about 1674 m and 1608–1708 m in the northeast and southwest direction, respectively. The critical mixed distance halo effect of ASCs can be used as the deployment interval of ASCs, which are about 215 m

and 529 m in the northeast direction, and 315 m and 585 m in the southwest direction corresponding to the smaller and larger values of critical delta speed, respectively. In the northeast direction, one ASC should be deployed every 500 m within the speed limit of 60 km/h and every 200 m within the speed limit of 40 km/h. In the southwest direction, one ASC should be deployed every 600 m within the speed limit of 60 km/h and every 300 m within the speed limit of 40 km/h.

However, there are several limitations of this study that deserved to be further explored. Firstly, the characteristic of taxi drivers needs to be further investigated. Secondly, the mixed distance halo effect of ASCs can be extended to encompass a larger number of ASCs, rather than being limited to only two adjacent ones. Finally, other advanced methods for optimizing the deployment of ASCs at the intersections are also deserved to explore in future research.

Acknowledgements This research work was jointly sponsored by the National Natural Science Foundation of China (Grant Number 71801182). We appreciate Eric Jiang from Vandegrift High School to proofread the paper.

References

1. WHO: Managing Speed (2017). https://www.who.int/publications-detail-redirect/managing-speed
2. Fu, C., Zhou, Y., Xu, C., Cui, H.: Spatial analysis of taxi speeding event using GPS trajectory data. In: Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference, pp. 122–127. IEEE, New York (2019)
3. Pilkington, P., Kinra, S.: Effectiveness of speed cameras in preventing road traffic collisions and related casualties: systematic review. BMJ **330**(7487), 331–334 (2005)
4. Van Der Pas, J., Kessels, J., Veroude, B.D.G., Van Wee, B.: Intelligent speed assistance for serious speeders: the results of the Dutch Speedlock trial. Accid. Anal. Prev. **72**, 78–94 (2014)
5. Zhao, G., Wu, C.: Effectiveness and acceptance of the intelligent speeding prediction system (ISPS). Accid. Anal. Prev. **52**, 19–28 (2013)
6. Heydari, S., Miranda-Moreno, L.F., Liping, F.: Speed limit reduction in urban areas: a before-after study using Bayesian generalized mixed linear models. Accid. Anal. Prev. **73**, 252–261 (2014)
7. Wrapson, W., Harré, N., Murrell, P.: Reductions in driver speed using posted feedback of speeding information: social comparison or implied surveillance? Accid. Anal. Prev. **38**(6), 1119–1126 (2006)
8. Plant, B.R.C., Irwin, J.D., Chekaluk, E.: The effects of anti-speeding advertisements on the simulated driving behaviour of young drivers. Accid. Anal. Prev. **100**, 65–74 (2017)
9. Champness, P., Sheehan, M., Folkman, L.M.: Time and distance halo effects of an overtly deployed mobile speed camera. In: Australasian Road Safety Research, Policing and Education Conference, Brisbane (2005)
10. De Pauw, E., Daniels, S., Brijs, T., Hermans, E., Wets, G.: An evaluation of the traffic safety effect of fixed speed cameras. Saf. Sci. **62**, 168–174 (2014)
11. Gao, K., Tu, H., Sun, L., Sze, N.N., Song, Z., Shi, H.: Impacts of reduced visibility un-der hazy weather condition on collision risk and car-following behavior: implications for traffic control and management. Int. J. Sustain. Transp. **14**(8), 635–642 (2020)
12. Fu, C., Liu, H.: Investigating distance halo effect of fixed automated speed camera based on taxi GPS trajectory data. J. Traffic Transp. Eng. (Engl. Ed.) **10**(1), 70–85 (2021)

13. Greaves, S.P., Ellison, A.B.: Personality, risk aversion and speeding: an empirical investigation. Accid. Anal. Prev. **43**(5), 1828–1836 (2011)
14. Christie, S.M., Lyons, R.A., Dunstan, F.D., Jones, S.J.: Are mobile speed cameras effective? A controlled before and after study. Injury Prev. **9**(4), 302–316 (2003)
15. Fu, C., Sayed, T.: Multivariate Bayesian hierarchical Gaussian copula modeling of the non-stationary traffic conflict extremes for crash estimation. Anal. Methods Accid. Res. **29**, 100154 (2020)
16. Fu, C., Sayed, T.: A multivariate method for evaluating safety from conflict extremes in real time. Anal. Methods Accid. Res. **36**, 100244 (2022)
17. Boscoe-Wallace, A.: Optimisation of Speed Camera Locations Using Genetic Algorithm and Pattern Search. A Doctoral Thesis Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University (2017)
18. Retting, R., Ferguson, S., Farmer, C.M.: Reducing red light running through longer yellow signal timing and red light camera enforcement: results of a field investigation. Accid. Anal. Prev. **40**, 327–333 (2008)
19. Xue, Q., Gao, K., Xing, Y., Lu, J., Qu, X.: A context-aware framework for risky driving behavior evaluation based on trajectory data. IEEE Trans. Intell. Transp. Syst. **15**(1), 70–83 (2023)
20. Brackett, M.: Speed control strategies. Saf. J. **9**, 27–39 (1979)
21. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: a K-means clustering algorithm. Appl. Stat. **28**(1), 100–108 (1979)
22. Li, H., Zhu, M., Graham, D.J., Zhang, Y.: Are multiple speed cameras more effective than a single one? Causal analysis of the safety impacts of multiple speed cameras. Accid. Anal. Prev. **139**, 105488 (2020)
23. Bar-Gera, H., Schechtman, E., Musicant, O.: Evaluating the effect of enforcement on speed distributions using probe vehicle data. Transp. Res. F Traffic Psychol. Behav. **46**, 271–283 (2017)
24. Gao, K., Tu, H., Shi, H.: Stage-specific impacts of hazy weather on car following. In: Proceedings of the Institution of Civil Engineers-Transport, vol. 6, pp. 347–359. Thomas Telford Ltd. (2019)
25. Yang, M., Ma, J., Chen, Q., Yang, Y.: Deterrent effect of fixed-site speed enforcement on freeways. In: 16th COTA International Conference of Transportation Professionals, Shanghai (2016)
26. Keenan, D.: Speed cameras-the true effect on behaviour. Traffic Eng. Control **43**, 154–162 (2002)

Chapter 7
Examining the Effect of Speeding Patterns on Speeding-Related Harsh Decelerations for Commercial Drivers with Survival Analysis

Yue Zhou, Chuanyun Fu, Xinguo Jiang, and Haiyue Liu

Abstract Commercial drivers are usually observed with more speeding behaviors than other cohorts. Some of their speeding behaviors occurring under specific conditions may cause more hazards. Hereby, there is a need to verify the hazard caused by different categories of speeding and benefit the anti-speeding countermeasures. This study treats the speeding-related harsh decelerations (SHDs) as a surrogate measure of safety. The speeding distance ratios in different categories (i.e., speeding patterns) and operational attributes are modeled as the risk factors. An Accelerated Failure Time (AFT) model is developed to relate the risk factors to the safe operation hours which means the average duration between the adjacent SHDs. The results indicate that speeding greater than 50% in daytime and speeding greater than 20% on peak hours/at night significantly decrease the safe operation hours on roads with a high-speed limit ($\geq$60 km/h). In addition, it is difficult to keep long safe operation hours when commercial drivers speeding on nighttime roads with a lower speed limit ($\leq$50 km/h). Conversely, speeding less than 20% is not linked to the deterioration of driving safety in all the scenarios.

7.1 Introduction

Commercial drivers (taxis drivers, buses drivers, trucks drivers, etc.) usually have a higher probability of speeding [1, 2], which also raises the likelihood of being involved in accidents. Authorities worldwide have come up with a variety of regulations to prevent commercial drivers from unsafe speeding, including punishment of license point, speed camera, and educational campaigns [3–5]. These efforts have

Y. Zhou · X. Jiang · H. Liu
School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China

C. Fu (✉)
School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China
e-mail: fuchuanyun@hit.edu.cn

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_7

improved the driving safety, but commercial drivers still suffer from extensive crashes caused by speeding in recent years [6]. An important reason could be that some countermeasures are less effective, because they may fail to prevent the speeding of certain patterns that contribute to the risk most [7].

Speeding pattern typically includes the information on where and when the speeding occurs, how serious the speeding is, and relevant attributes of the behavior. Although speeding behaviors of all patterns are viewed as hazardous in common sense, the threat to road safety may vary substantially across the speeding patterns. Several studies have reported the discrepancy among ordinary drivers. For example, Fitzpatrick [8] and Cooper [9] found that not all speeding behaviors were related to crashes. Also, a statistic in 2010 showed that only 6% of fatal crashes were attributed to high-level speeding in France, while 9% and 13% were caused by medium-level and low-level speeding, respectively [10]. Such mixed effects of speeding may also be observed in the group of commercial drivers. Unfortunately, the issue is not sufficiently explored so far. To provide better safety guidelines for commercial drivers, it would be instructive to link the effect of speeding patterns to the actual driving risk.

Evaluating the driving safety of commercial drivers with a reliable crash data is difficult because of the shift work and enforcement avoidance [11, 12]. Fortunately, the use of surrogate measures of safety (SMoS) and the large GPS trajectory data light up a direction for the purpose. SMoS can replace the accident likelihood with highly crash-related indicators, such as conflicts, speed attributes, and harsh driving behaviors (e.g., harsh deceleration, harsh acceleration) [13, 14]. Among the SMoS, harsh deceleration performs as a good indicator of the crash risk, as it implies that drivers are dodging an unexpected incident [15]. This indicator can be measured during a short-term period and further used in a subsequent safety evaluation. In order to detect the harsh decelerations, past studies used GPS trajectories collected by smart phones and vehicular GPS devices [14, 15]. Meanwhile, the large GPS trajectories can help identify the speeding behaviors of commercial drivers in various situations [16]. These superiorities make it easy to simultaneously analyze the speeding attributes and harsh decelerations of a certain driver.

Based on the GPS trajectories of commercial drivers, the current study aims to enrich the knowledge of the relationship between speeding in different patterns and the driving risk reflected by average duration between speeding-related harsh decelerations (SHDs). Our study makes several contributions: (i) the relationship between speeding in different patterns and SHDs is investigated; (ii) the AFT models with different probability distributions are compared in fitting the safe operation hours; and (iii) the study identifies the speeding patterns that are significantly related to the SHDs.

7.2 Data Collection and Analysis

7.2.1 Data Collection

The current study collects the trajectories through vehicular GPS recorders from 3,130 commercial vehicles (i.e., taxis) in Chengdu, China. The GPS recorders are installed in the taxis and able to collect the position, instant speed, and other driving information of the taxi with a frequency of 0.1 Hz. The study period ranges from Nov. 5, 2016, to Nov. 25, 2016 (3 weeks). No festivals or special events are found during the period.

7.2.2 Identification of Speeding Behaviors and SHDs

The study first identifies the speeding behaviors and SHDs with GPS trajectories. A map-matching algorithm is used to correct the GPS positioning error [2, 16] before the identifications. The speeding behaviors are defined as whether the average driving speed of each pairwise GPS point exceeds the posted speed limits. The speeding level is an important norm for law enforcement when penalizing speeders in China. It is classified as Level-1 (i.e., exceeding the speed limit by 3–10%), Level-2 (i.e., exceeding the speed limit by 10–20%), Level-3 (i.e., exceeding the speed limit by 20–50%), and Level-4 (i.e., exceeding the speed limit by 50% or greater).

Harsh deceleration is typically defined as a braking behavior with an extremely large deceleration velocity (represented with a negative sign). The identification of harsh decelerations requires the comparison between the decelerated velocity of the vehicle and a threshold. The threshold varies across the studies due to the traffic and environmental conditions. Generally, past studies adopted a range of threshold between -0.75 and -0.2 g (g is the gravity acceleration i.e., 9.8 m/s^2). We use a minimum acceptance of -0.2 g as the threshold in our cases [17] to avoid any inaccurate measurement. The steps to identify harsh decelerations and SHDs are [15]:

(i) Calculate the average deceleration velocity of each pairwise GPS point by dividing the speed difference between the adjacent points by the corresponding travel time;
(ii) Compare the average deceleration velocity with the threshold of -0.2 g. If it is less than the threshold, then it is defined as a harsh deceleration;
(iii) Examine whether the previous adjacent trajectory of a harsh deceleration is a speeding trajectory (Case 1 in Fig. 7.1); or whether the harsh deceleration itself is recognized as a speeding trajectory (Case 2 in Fig. 7.1). Mark the harsh deceleration as "speeding-related harsh deceleration" (i.e., SHD) if Case 1 or Case 2 is yes; and
(iv) Summarize the attributes of speeding behaviors and record the SHDs.

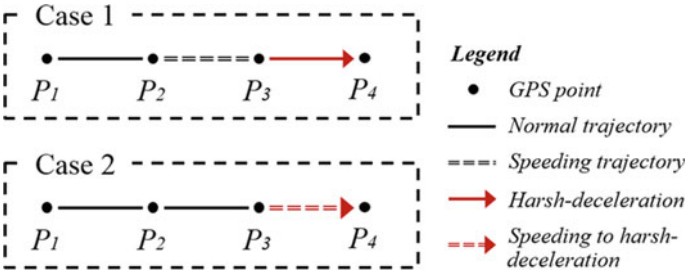


Fig. 7.1 The cases that can be defined as a SHD

The process was conducted for each taxi and we extracted a total of 1,757,680 speeding behaviors and 5,427 SHDs from their trajectories.

7.2.3 Variable Description

The speeding distance ratio is adopted to be the risk factor as it can properly reflect the risk of speeding behaviors [18]. The speeding distance ratio represents the percent of speeding distance to the total driving distance on drivers' routes. The speeding distance ratios are sorted by different speeding patterns and multiplied by 100 to reflect the effect of each 1% change on driving hazard in further models, expressed as

$$R_{(t,s,r)} = \frac{d_{(t,s,r)}}{D_{(t,s)}} \times 100 \tag{7.1}$$

where the footnote (t, s, r) means the speeding pattern with marks t, s, and r, which represents the speeding behaviors are level r, occurring in time of day t and on roads with a speed limit of s. $R_{(t,s,r)}$ is the speeding distance ratio with the speeding pattern (t, s, r), $d_{(t,s,r)}$ is the cumulative speeding distance yielded by the pattern (t, s, r) speeding, and $D_{(t,s)}$ is the total driving distance in time of day t and on roads with a speed limit of s. Each mark can be represented with one of the tags to denote a specific speeding pattern. The details of the speeding pattern are listed in Table 7.1.

The taxi drivers are observed with repeated speeding behaviors and encounters of SHD. Thus, the SHDs can be treated as recurrent consequences of speeding. Our study describes the hazards of speeding by calculating the safe operation hours that means the average duration between the adjacent SHDs. It is expressed as

$$T_i = \frac{H_i}{(m_i + 1)} \tag{7.2}$$

Table 7.1 The description of speeding pattern

Type	Index	Mark	Tag	Explanations
(Time of day, speed limit, speeding level)	Time of day	*t*	*Peak*	Speeding on peak hours (7:00–9:00 & 17:30–19:30)
			Night	Speeding in nighttime (19:30–7:00)
			Day	Speeding in daytime (7:00–17:30)
	Speed limit*	*s*	*Low*	Speeding on roads with speed limits $\leq$ 30 km/h
			Med	Speeding on roads with speed limits of 40 or 50 km/h
			High	Speeding on roads with speed limits $\geq$ 60 km/h
	Speeding level	*r*	*1*	Exceed the speed limits > 3% but $\leq$ 10%
			2	Exceed the speed limits > 10% but $\leq$ 20%
			3	Exceed the speed limits > 20% but $\leq$ 50%
			4	Exceed the speed limits > 50%

Note * The range of speed limits is [20, 100] km/h

where T_i is the safe operation hours for taxi i, H_i is the total operation hours, and m_i is the number of SHD. For those taxis without any SHD, their safe operation hours are equal to the total operation hours. In addition to the speeding distance ratio, several operational attributes of commercial drivers are extracted from the trajectories [1, 2]. Table 7.2 lists the descriptions of the factors.

Table 7.2 The description for operational attributes

Variables	Description	Mean	S.D
RWV	The ratio of vehicle kilometers traveled on weekend among the total vehicle kilometers traveled	0.304	0.167
DNH	Daily nap hours during the operation	1.28	1.56
IPD	Income per 24 working hours (100 Yuan, RMB)	8.903	1.952
LPH	The average frequency of lane changes per 1 working hour	3.105	1.449
LPD	The average frequency of long-trip deliveries (the distance of trip > 10 km) per 24 working hours	3.823	1.922

7.3 Method

Survival analysis models the hazard or survive rate as a function of time and is suitable for evaluating the hazard caused by the recidivism of traffic violations or crashes [19]. The hazard and survival rate can be affected by risk factors. The Cox Proportional Hazard model and AFT model are commonly used to explore the effect of risk factors. Although the Cox model is popular in the semi-parametric structure, the covariates in a Cox model have to meet the assumption of proportional hazard. This assumption is seldom satisfied and therefore restricts the use of the Cox model [20]. Conversely, the AFT model is more flexible, powerful in interpretation, and more robust in estimates than the Cox model. Hereby, the AFT model is used to fit the safe driving hours.

The AFT model fits the individual survival time with a logarithm form and uses a linear relationship to capture the influence of covariates, expressed as

$$\log(T_i) = \boldsymbol{\beta X} + \varepsilon_i \tag{7.5}$$

where T_i is the safe operation hours and ε_i is a random error that needs to be specified with a probability distribution.

If ε_i is assumed to follow an extreme-value density, then it generates an AFT model with exponential or Weibull distribution. If ε_i follows a Gaussian density, then it specifies an AFT model with Lognormal distribution.

7.4 Results

7.4.1 Comparison of Goodness-of-Fit

Table 7.3 presents the model performances. When modeling with exponential and Weibull distributions, the AFT models are inferior to the AFT model with Lognormal distribution in terms of the Log-likelihood and AIC, indicating that Lognormal density is better for fitting the safe operation hours. As such, the AFT model with Lognormal distribution is chosen to provide the factor explanations.

Table 7.3 The goodness-of-fit for different AFT models

Model	Log-likelihood	AIC
AFT-exponential	−2463.40	4950.80
AFT-Weibull	−2457.64	4941.29
AFT-Lognormal	**−2431.61**	**4889.22**

7.4.2 Estimates of the AFT Model

Table 7.4 lists the estimates from the ATF model with Lognormal distribution. The exponent of the coefficients Exp(β) is provided to quantify the effects of covariates on safe operation hours. It is found that all the significant coefficients have a negative sign, signifying that the increase of the factors leads to shorter safe operation hours.

To be specific, a 1% increase in the speeding distance ratios of level-3 and level-4.

speeding occurring on peak hours with a high-speed limit ($R_{(\text{peak, high,3})}$ and $R_{(\text{peak,high,4})}$).

is found to decrease the safe operation hours by 17.9% and 26.7%, respectively. At night, a 1% increase in the speeding distance ratio of level-1 speeding with a low-speed limit reduces the safe operation hours by 5.3%. Similarly, the reductions are 5.3%, 4.4%, 10.6%, and 29.8% respectively when increasing the speeding distance ratio of level-3 and level-4 speeding occurring on nighttime roads with a medium-speed limit.

($R_{(\text{night,med,3})}$ and $R_{(\text{night,med,4})}$) and roads with a high-speed limit ($R_{(\text{night,high,3})}$ and $R_{(\text{night,high,4})}$) by 1%. It also shows that, for each 1% increase in speeding distance ratio, the level-4 speeding in daytime with a high-speed limit ($R_{(\text{day,high,4})}$) could lower the safe operation hours by 42.8%.

The estimates also signify that some operational attributes increase the risk of encountering SHDs. Every 100 RMB increase for the income and 1 more long-trip delivery per 24 working hours can decrease the safe operation hours by 8.8% and 10.2%, respectively. Moreover, 1 more lane-changing maneuver per one working hour will lead to 18.5% lower safe operation hours for taxi drivers.

Table 7.4 The estimates of the AFT model with lognormal distribution

Variables	Estimate	S.D	95% CI	Exp(β)	*p*-value
Constant	8.739	0.029	(8.149, 9.327)	–	<0.001
$R_{(\text{peak,high,3})}$	−0.197	0.084	(−0.361, −0.033)	0.821	0.019
$R_{(\text{peak,high,4})}$	−0.314	0.108	(−0.523, −0.101)	0.733	0.004
$R_{(\text{night,low,4})}$	−0.065	0.022	(−1.088, −0.022)	0.937	0.003
$R_{(\text{night,med,3})}$	−0.054	0.018	(−0.089, −0.019)	0.947	0.003
$R_{(\text{night,med,4})}$	−0.045	0.020	(−0.084, −0.006)	0.956	0.024
$R_{(\text{night,high,3})}$	−0.112	0.041	(−0.192, −0.035)	0.894	0.006
$R_{(\text{night,high,4})}$	−0.354	0.106	(−0.562, −0.148)	0.702	<0.001
$R_{(\text{day,high,4})}$	−0.557	0.111	(−0.775, −0.341)	0.572	<0.001
IPD	−0.092	0.028	(−0.146, −0.038)	0.912	0.001
LPD	−0.107	0.025	(−0.157, −0.057)	0.898	<0.001
LPH	−0.205	0.032	(−0.269, −0.142)	0.815	<0.001
σ	1.871	0.056	(1.765, 1.984)	–	<0.001
No. of subjects	3130				

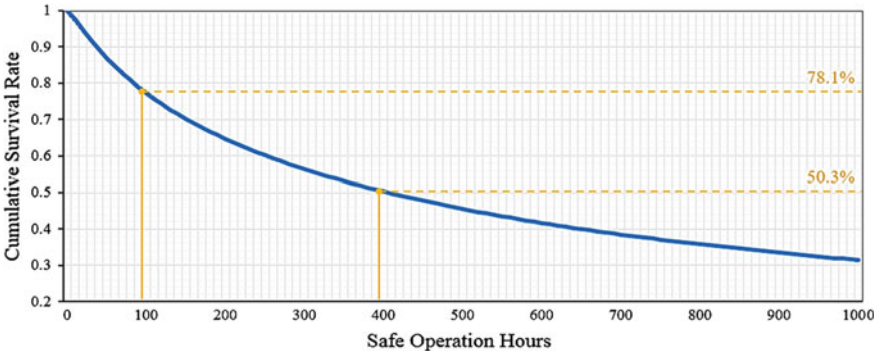


Fig. 7.2 Cumulative survival rate estimated at mean values of the speeding distance ratios

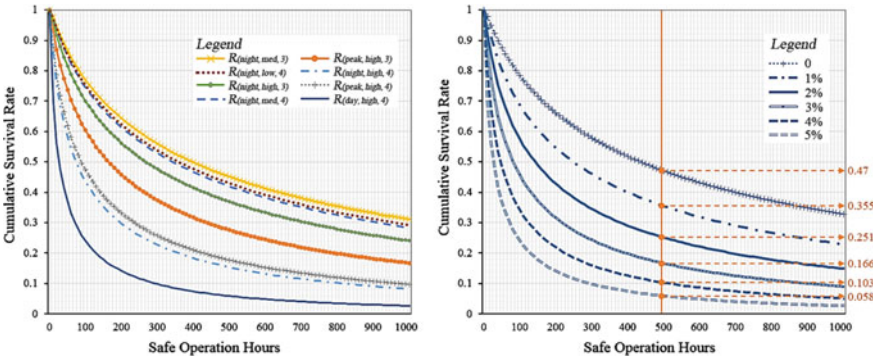
7.4.3 Estimates of the Cumulative Survival Rate

Figure 7.2. illustrates the cumulative survival rates estimated with the mean value of the covariates. The survival rate steeply drops from 1 to 78.1% when the drivers maintain a safe driving to 100 h, and drops more smoothly from 78.1 to 50.3% if the driver could drive safely up to 400 h. This implies that, with the current speeding behaviors, some drivers have larger possibility to encounter SHDs, while others may not drive with increasing hazard if they have already kept long safe operation hours.

Meanwhile, we control the speeding distance ratios in different speeding patterns and compare their effects on the cumulative survival rates. It is achieved by setting all the speeding distance ratios equal to 5% and then estimating the survival rates over time (Fig. 7.3a). Among all the speeding patterns, the survival curve drops faster in the cases of setting $R_{(\text{peak,high,4})}$, $R_{(\text{night,high,4})}$, and $R_{(\text{day,high,4})}$ as 5% each. This result manifests that level-4 speeding with a high-speed limit, especially in the daytime scenario, can cause more hazards than other speeding patterns. Additionally, Fig. 7.3b compares the effects of $R_{(\text{day,high,4})}$ in different magnitudes on the survival rates. With $R_{(\text{day,high,4})}$ increasing from 0 to 5%, the probability of driver keeping a 500 h safe driving decreases from 47% to merely 5.8%. For each 1% change, the most significant drop of the survival rate (-11.5%) occurs when $R_{(\text{day,high,4})}$ increases from 0 to 1%.

7.5 Discussion and Conclusion

Past studies have highlighted the analysis of speeding behavior due to the inherent relationship between speeding and crash. The speeding of commercial drivers should be especially concerned since these drivers generate more speeding records than other cohorts. Meanwhile, the flexible driving styles of these drivers could make the implementation of anti-speeding countermeasures to be improvident and unsustainable. In



(a) Speeding distance ratios equal to 5% (b) $R_{(day,high,4)}$ varies from 0 to 5%

Fig. 7.3 Cumulative survival rate estimated at specific values of the speeding distance ratios

order to guide the management approaches, our study identifies the speeding behaviors and SHDs by GPS trajectories and relates the effect of different speeding patterns on drivers' safe operation hours through the AFT model.

As shown by the model estimates, the speeding no more than 20% does not decrease the safe operation hours statistically. The primary explanation is that commercial drivers typically have better skills in controlling the vehicle and keeping alert to the surrounding environment when driving at a low speed. The professional driving skills can help them avoid the hazards in their routines by slowing down smoother or swerving in advance prior to an unexpected encounter. However, when the drivers speeding severely (>50%) at a high-speed limit (≥60 km/h), the driving speed could bring about an impaired reaction time and deteriorated ability to control the vehicle. Hereby, the drivers could suffer SHDs more frequently. It is also noted that speeding severely at night, even on roads with a speed limit equal to or less than 50 km/h, is the main cause of the SHDs. This is because the drivers may find it hard to notice unexpected encounters in advance with inadequate lighting [21]. Also, drivers in the nighttime are more likely to fatigue which enlarges their perception-reaction time [22].

Another insightful explanation refers to the effect of actual driving speed on road safety. As suggested by Nillson [23], the relationship between the driving speed and accident likelihood can be expressed by a power function. It demonstrates that, even for an equally proportional change on speed, the change with a large initial speed causes more hazards than that with a small initial speed [24]. For instance, increasing speed from 60 to 90 km/h (+50%) is more dangerous than changing from 40 to 60 km/h (+50%). In our cases, committing level-3/level-4 speeding on roads with a medium- or high-speed limit causes a large speed between the driving speed and the speed limits (Table 7.5). Logically, such speeding are more likely to trigger SHDs.

The study also confirms that income, lane changes, and long trips reduce the safe operation hours. Past studies have revealed that commercial drivers with higher

Table 7.5 The speed mean of speeding levels sorted by different speed limits

Speed limit (baseline)		Speed mean of different speeding levels (km/h)			
		Level-1	Level-2	Level-3	Level-4
Low	20 km/h	21.4	23.1	27.6	37.4
	30 km/h	31.9	34.4	39.6	49.8
Medium	40 km/h	42.5	45.8	52.1	71.0
	50 km/h	53.2	57.2	64.2	79.9
High	60 km/h	63.7	68.6	75.9	115.8
	70 km/h	74.3	79.6	92.1	132.6
	80 km/h	84.7	90.7	100.8	164.5
	100 km/h	105.4	113.9	130.5	187.9

revenue tend to commit more severe speeding [1, 2]. Also, they tend to speed in a long trip or implement lane changes to keep a superior speed [2].

In summary, our study shows that only exceeding the speed limit by 50% or greater in the daytime, and 20% or greater during peak hours or at nighttime on roads with a speed limit of no less than 60 km/h increase the possibility of encountering SHDs. Meanwhile, speeding greater than 20% at night with a speed limit of 40–50 km/h or speeding greater than 50% with a speed limit of 20–30 km/h increases the likelihood of encountering SHDs. These findings demonstrate that the hazards are merely related to the extremely large speed, rather than being determined by the speeding level adopted by the current law enforcements. As such, the police could penalize the speeding that really contributes to harsh incidents heavier. If the current laws remain unchanged, we recommend that driving auxiliary system, connected vehicle, and real-time conflict detection [25, 26] could be considered to improve the safety and avoid harsh decelerations of professional drivers. Both measures could extent the perceptibility of the drivers and warning the incoming risky events. Also, factors such as drivers' socio-demographic information, traffic conditions, etc., could be considered in further SHDs modeling.

Acknowledgements This work was jointly supported by the National Natural Science Foundation of China (71801182; 71771191), the Sichuan Provincial Science and Technology Innovation Talents Fund (2019JDRC0023), and the Fundamental Research Funds for the Central Universities (FRFCU5710000111). We also appreciate Eric Jiang from Vandegrift High School to proofread the paper.

References

1. Tseng, C., Yeh, M., Tseng, L., Liu, H., Lee, M.: A Comprehensive analysis of factors leading to speeding offenses among large-truck drivers. Transp. Res. Part F Traffic Psychol. Behav. **38**, 171–181 (2016)

2. Zhou, Y., Jiang, X., Fu, C., Liu. H.: Operational factor analysis of the aggressive taxi speeders using random parameters Bayesian LASSO modeling approach. Accid. Anal. Prev. **157**, 106183 (2021)
3. Watson, B., Siskind, V., Fleiter, J.J., Watson, A., Soole, D.: Assessing specific deterrence effects of increased speeding penalties using four measures of recidivism. Accid. Anal. Prev. **84**, 27–37 (2015)
4. Xue, Q., Gao, K., Xing, Y., Lu, J., Qu, X.: A context-aware framework for risky driving behavior evaluation based on trajectory data. IEEE Intell. Transp. Syst. Mag. **15**(1), 70–83 (2023)
5. Lewis, I., White, K. M., Ho, B., Elliott, B., Watson. B.: Insights into targeting young male drivers with anti-speeding advertising: an application of the step approach to message design and testing (SatMDT). Accid. Anal. Prev. **103**, 129–142 (2017)
6. Wang, Y., Li, L., Prato, C.G.: The Relation between working conditions, aberrant driving behaviour and crash propensity among taxi drivers in China. Accid. Anal. Prev. **126**, 129–142 (2019)
7. Truelove, V., Freeman, J., Szogi, E., Kaye, S., Davey, J., Armstrong. K.: Beyond the threat of legal sanctions: What deters speeding behaviours? Transp. Res. Part F Traffic Psychol. Behav. **50**, 128–136 (2017)
8. Li, Y., Zhao, L., Gao, K., An, Y., Andric, J.: Revealing driver psychophysiological response to emergency braking in distracted driving based on field experiments. J. Intell. Connect. Veh. **5**(3), 270–282 (2022)
9. Cooper, P.J.: The Relationship between speeding behaviour (as measured by violation convictions) and crash involvement. J. Saf. Res. **28**, 83–95 (1997)
10. Viallon, V., Laumon, B.: Fractions of fatal crashes attributable to speeding: evolution for the period 2001–2010 in France. Accid. Anal. Prev. **52**, 250–256 (2013)
11. La, Q.N., Lee, A.H., Meuleners, L.B., Duong, D.V.: Prevalence and factors associated with road traffic crash among taxi drivers in Hanoi. Vietnam. Accid. Anal. Prev. **50**, 451–455 (2013)
12. Shim, J., Park, S.H., Chung, S., Jang, K.: Enforcement avoidance behavior near automated speed enforcement areas in Korean expressways. Accid. Anal. Prev. **80**, 57–66 (2015)
13. Fu, C., Sayed, T.: Multivariate Bayesian hierarchical Gaussian copula modeling of the non-stationary traffic conflict extremes for crash estimation. Accid. Anal. Prev. **29**, 100154 (2021)
14. Lyu, N., Wang, Y., Wu, C., Peng, L., Thomas, A.F.: Using naturalistic driving data to identify driving style based on longitudinal driving operation conditions. J. Intell. Connect. Veh. **5**(1), 17–35 (2022)
15. Zhou, Y., Liu, H., Fu, C.: Investigating contributing factors of hard-braking events on urban road network. In: Proceedings of KES-STS International Symposium, Smart Innovation, Systems and Technologies (2022)
16. Fu, C., Liu, H.: Investigating distance halo effect of fixed automated speed camera based on taxi GPS trajectory data. J. Traffic Transp. Eng. (English Edition) (2019)
17. Botzer, A., Musicant, O., Mama, Y.: Relationship between hazard-perception-test scores and proportion of hard-braking events during on-road driving—An investigation using a range of thresholds for hard-braking. Accid. Anal. Prev. **132**, 105267 (2019)
18. Vlassenroot, S., Broekx, S., De Mol, J., Panis, L.I., Brijs, T., Wets, G.: Driving with intelligent speed adaptation: results of the Belgian ISA-trial. Transp. Res. Part A Policy Prac. **41**(3), 267–279 (2007)
19. Fu, H.: Identifying repeat DUI crash factors using state crash records. Accid. Anal. Prev. **40**, 2037–2042 (2008)
20. Mustefa, Y.A., Chen, D.: Accelerated Failure-Time model with weighted least–squares estimation: application on survival of HIV positives. Arch. Public Health. **79**, 1–9 (2021)
21. Wanvik, P.O.: Effects of road lighting: an analysis based on Dutch accident statistics 1987–2006. Accid. Anal. Prev. **86**, 173–185 (2016)
22. Chu, H.: Risky behaviors of older taxi drivers and suggested requirements for renewing their professional driver's licenses. Transp. Res. Interdiscip. Perspect. **8**, 100272 (2020)
23. Nilsson, G.: Traffic safety dimensions and the power model to describe the effect of speed on safety. Lund Institute of Technology, Lund (2004)

24. Elvik, R.: A re-parameterisation of the power model of the relationship between the speed of traffic and the number of accidents and accident victims. Accid. Anal. Prev. **50**, 854–860 (2013)
25. Fu, C., Sayed, T.: A multivariate method for evaluating safety from conflict extremes in real time. Anal. Methods Accid. Res. **36**, 100244 (2022)
26. Fu, C., Sayed, T.: Identification of adequate sample size for conflict-based crash risk evaluation: an investigation using Bayesian hierarchical extreme value theory models. Anal. Methods Accid. Res. **39**. 100281 (2023)

Chapter 8
Research and Application Analysis of Stepwise Incremental Fine Model for Speeding Behavior

Chuanyun Fu and Jinzhao Liu

Abstract In order to effectively restrict repeated speeding behaviors, this paper first divides the speeding range into three speeding ranges, including low, medium, and high speeding according to the law-enforcement in China. Based on this, we further build the basic framework of a stepwise incremental fine model (SIFM) and a stepped incremental fine model based on the principle of stepped pricing with the minimum total number of speeding times as the objective and the economic cost as the constraint. In addition, the response of drivers to changes in the fine amount was examined using the price elasticity theory. The study also analyzes the application of the stepped incremental fine model using Deyang City as an example. The results show that the stepped incremental fine model can effectively target the speeding drivers and their behavioral characteristics, which then ensures the model could be accepted by most speeding recidivists. The setup of the model is scientific and reasonable.

8.1 Introduction

Speeding behavior is one of the most important factors affecting road traffic safety [1, 2]. In order to restrict speeding behaviors, authorities have implemented a large number of traffic control approaches, but this traffic violation is still prominent to date [3]. One of the major reasons is the failure to constrain speeding recidivists.

Studies have shown that repeated traffic violations (DUI, speeding, etc.) are strongly associated with increased accident rates [4], leading to higher crash severity [5]. Speeding recidivists are more likely to be involved in crashes [6–9]. Raising fines is a common method of restricting speeding behaviors. A related study in Norway found that increasing fines reduced speeding violations [10]. Watson et al. [7] showed that increasing the fine reduced the percentage of recidivism and recidivism frequency of speeding violations, but there was no significant change in the recidivism interval

C. Fu (✉) · J. Liu
School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China
e-mail: fuchuanyun@hit.edu.cn

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_8

and the average number of recidivisms. Liu and Xie [11] found that when the duration of the countermeasure that increasing the cost of the violation increases, the deterrence to the offenders decreases and eventually makes the enforcement slack. Some of the increased fine mechanisms studied cannot effectively target speeding recidivists because they do not reference past speeding counts. Many scholars focus on controlling repetitive speeding behaviors. Polinsky [12] showed that incremental fines can reduce excessive deterrence and compensate for insufficient deterrence. Delhaye [13] found that the optimal fine mechanism depends on the probability of detection of speeding behavior and the strength of the relationship between driver type and speeding record.

The incremental penalty mechanism can be formulated by referring to the ladder pricing theory first proposed by Taylor [14] in the 1970s. Many scholars believe that the mechanism of step pricing can achieve the goal of balanced economic and social development to a certain extent [15, 16]. Ruijs [17] argues that the mechanism of stepped pricing can increase the welfare of low-income people. Also, studies have shown that high prices on the higher step of ladder pricing can effectively restrict uncontrolled overconsumption by consumers in this ladder range category [18]. Some scholars have already established a stepwise incremental penalty model for other traffic violations considering the number of violations. Zeng [19] and Liu [20] modeled stepped incremental penalties for parking violations and motor vehicle red light violations, respectively.

In summary, this paper aims to establish a stepped incremental fine intensity model to restrict speeding behaviors based on the consideration of both the speeding range and frequency of a speeder. The current work could provide support and improvements to the speeding control and traffic management of motor drivers.

8.2 Assumptions and Framework

8.2.1 Model Assumptions

(1) Drivers are rational consumers.
(2) Drivers fully understand the cost of each speeding they commit.
(3) The cycle of the Stepwise Incremental Penalty Model reset is one year.

8.2.2 Basic Framework of the Model

According to the speeding control standards implemented in China, seriously exceeding the posted speed limits could suffer in heavier penalties. As such, the speeding range is divided into low (less than 20%), medium (greater than or equal to 20% but less than 50%), and high (greater than or equal to 50%). The stepwise

incremental fine model for restricting the speeding behaviors is then built based on the classifications of the speeding.

Let the newly implemented ladder fines under each speeding range be divided into m steps according to the number of speeding times, the first ladder speeding times range is $[0, N_1]$, the second ladder speeding times range is $(N_1, N_2]$, and so on, the mth ladder speeding times range is $(N_{m-1}, +\infty)$; P_i denotes the single fine amount under an original speeding range i ($i = 1, 2, 3$, corresponding to the under the low, medium, and high speeding ranges, respectively). The amount of change of the fine amount under each speeding range of each level of the newly implemented ladder fines is $\Delta P = [\Delta P_{i1}, \Delta P_{i2}, \ldots, \Delta P_{ij}, \ldots, \Delta P_{im}]$. Then, the fine amount under each speeding range of each level of the newly implemented ladder fines is $P_a = P_i + \Delta P = [P_i + \Delta P_{i1}, P_i + \Delta P_{i2}, \ldots, P_i + \Delta P_{ij}, \ldots, P_i + \Delta P_{im}]$.

8.3 Stepwise Incremental Fine Model

8.3.1 Determine the Number of Steps and the Number of Speeding in Each Step

Based on the characteristics of drivers' speeding behavior, the lower step should cover the majority of the speeders, while the higher step should target a smaller proportion of speeding recidivists. As the number of steps increases, the proportion of drivers covered by a single step gradually decreases. As such, the proportion of speeders covered in each step is shown in Table 8.1.

Table 8.1 Reference values for the number of steps and coverage rate

Number of steps	Number of speeding intervals in each order			
	First step	Second step	Third step	Fourth step
2	(0, 1]	(1, +∞)	–	–
3	Cover 70 or 80% of speeding drivers	Accumulated coverage of 90 or 95% of speeding drivers	Cumulative coverage of 100% of speeding drivers	–
4	(0, 1]	Accumulatively cover 70 or 80% of speeding drivers	Accumulated coverage of 90% or 95% of speeding drivers	Cumulative coverage of 100% of speeding drivers

8.3.2 Stepwise Incremental Fine Model Based on Driver Response

The fines vary across the frequencies of exceeding the speed limit for a driver. If the driver's total expenditure on speeding is a fixed value, which we could treat it as a "speeding market." Then, the fines should increase as the frequency of exceeding the speed limit decrease as the amount of the fine increases; otherwise, the frequency of exceeding the speed limit should increase as the amount of the fine decreases. The price elasticity can be used to represent the driver's sensitivity to the changes of the penalty and the frequency of speeding, according to the theory of elasticity fine. Based on the fine elasticity of drivers' speeding demand and their response to the implementation of a stepped incremental fine mechanism, a corresponding stepped incremental fine amount optimization model for speeding can be developed. The objective of the model is to minimize the number of speeding in the study area after implementing the model, so the objective function is

$$\min Q_{\mathrm{sa}} = \sum_{i=1}^{3}(w_i Q_{sai}) = \sum_{i=1}^{3} w_i \left[\begin{array}{c} \int_0^{N_1} f(Q_i) Q_{ai1} dQ_i + \int_{N_1}^{N_2} f(Q_i) Q_{ai2} dQ_i \\ + \cdots + \int_{N_{m-1}}^{+\infty} f(Q_i) Q_{aim} dQ_i \end{array} \right] \tag{8.1}$$

The constraints are

$$\left\{ \begin{array}{c} F_{\mathrm{si}} < F_{sai} \\ F_{sai} < (1 + b_0) F_{si} \\ \Delta P_{i(j-1)} < \Delta P_{ij} < \Delta P_{i(j+1)} \\ P_i + \Delta P_{im} < P_{i+1} + \Delta P_{(i+1)1} \\ F_{aij} < (1 + b_j) F_i \end{array} \right. \tag{8.2}$$

where Q_{sa} is the total speeding frequency for all drivers in the speeding range after the implementation of SIFM; w_i is the weight of speeding in the speeding range i; Q_{sai} is the total speeding frequency for drivers in the speeding range i after the implementation of SIFM; Q_i is the speeding frequency for drivers in the speeding range i; Q_{aij} is the speeding frequency for individual drivers exceeded the speed limit under step j of speeding range i after the implementation of SIFM; F_{si} and F_{sai} are total amount of fines for drivers in the speeding range i before and after the implementation of SIFM, respectively; b_0 is the acceptable increase in the amount of fines for all drivers before and after the implementation of SIFM; ΔP_{ij} is the increase in the amount of the fine for speeding under step j of speeding range i after the implementation of SIFM; P_i is the amount of fines for each speeding under speeding range i before the implementation of SIFM; F_{aij} is the amount of fines for each speeding under step j of speeding range i after the implementation of SIFM;

b_j is the growth rate of the acceptable fine amount for the driver in step j before and after the implementation of SIFM; F_i is the total fine amount for the individual driver in speed range i before the implementation of SIFM.

8.4 Case Analysis

This study conducted a case study on the application of the SIFM to speeding in Deyang City, China, to examine the effect of restricting the speeding recidivism. The statistics of the number of speeding violations recorded by off-site enforcement are shown in Table 8.2, which are offered by the Deyang City Traffic Police Detachment in 2017.

8.4.1 Optimize the Number of Fines at Each Level

The acceptance of speeding recidivism with a cost of 100 RMB penalty fine was set as a benchmark. The ratio of the change in the number of times a driver exceeds the speed limit to the change in the fine amount after the fine becomes 200 RMB, 500 RMB, 1000 RMB, and 2000 RMB are counted, and the resulting fine elasticities are assigned weights of 0.5, 0.3, 0.15, and 0.05, respectively. The weighted fine for each driver facing different changes in the fine amount Elasticities is calculated. The weighted fine elasticities of drivers with the same number of actual speeding were averaged to obtain the average fine elasticities of drivers with different numbers of speeding, as shown in Table 8.3.

According to the relationship between consumers' utility and consumption of a good in Stone-Gear theory [21], the principle of maximum utility, and the definition

Table 8.2 The number of vehicles with different speeding times and their frequency

Number of speeding	Vehicle number	Frequency (%)	Cumulative frequency (%)
1	80,230	69.78	69.78
2	20,810	18.10	87.88
3	7401	6.44	94.32
4	3100	2.70	97.01
5	1483	1.29	98.30
6	774	0.67	98.98
7	441	0.38	99.36
8	251	0.22	99.58
9	159	0.14	99.72
10	105	0.09	99.81

Table 8.3 The average value of the driver's fine elasticity under different speeding times

Number of speeding	Average fine elasticity	Number of speeding	Average fine elasticity
1	−0.16117	6	−0.31374
2	−0.19916	7	−0.29721
3	−0.26819	8	−0.20523
4	−0.30533	9	−0.15751
5	−0.33796	10	−0.12832

of the elasticity coefficient of fines, the expression $E = \frac{(1-a)\overline{Q}}{Q} - 1$ for the elasticity coefficient of fines for drivers can be introduced [19]. Because off-site enforcement data do not record driver characteristics that influence speeding behavior, to facilitate subsequent research, the number of speeding Q is used to focus driver characteristics, i.e., so that $E = E(Q)$, the elasticity coefficients of fines for different drivers are reflected by their speeding times. Fitting the average elasticity of fines for drivers with a different number of speeding times gives the following function (8.3).

$$E(Q) = 0.0092Q^2 - 0.0958Q - 0.0659 \cdot R^2 = 0.9242 \tag{8.3}$$

The weight of each speeding range is determined. This study determines the weight of each speeding range according to the severity of the speed corresponding to the speeding range after causing a traffic accident. A study in the United States found a linear relationship between vehicle crash speed and the amount of impact pressure [5]. The relationship between collision damage and collision speed can be expressed by Eq. (8.4), where $\boldsymbol{X}$ is the collision damage; v is the collision speed; a_i and b_i are the two-vehicle stiffness coefficients.

$$X = a_i v - b_i v^2 \tag{8.4}$$

In this paper, the speed range is defined as low, medium, and high gears, and the speed limit value of 60 km/h is taken as the main road in the city, then the low, medium, and high-speed ranges are 60 km/h–72 km/h, 72 km/h–90 km/h and greater than 90 km/h respectively. The collision speed v is taken as the median speed under the three ranges (high range Overspeed is taken as 100 km/h), a_i and b_i are selected as the third level stiffness factor 1.037, and the calculated collision damage is −4288.6, −6478.04, and −9897.34 under the low, medium, and high Overspeed ranges. Therefore, when the low range Overspeed is used as the benchmark with a weight of 1, then $w_1 = 1$, $w_2 = 1.51$, $w_3 = 2.31$.

The probability density function of the number of speeding under each speeding range. When the study area is certain, the number of drivers corresponding to all speeding times under each speeding range is fitted to the proportion of all speeding drivers in that speeding range, then the frequency distribution function of the number of speeding times obtained can be regarded as the probability density distribution

function $f(Q_i)$ of the number of speeding times of drivers under each speeding range, where $i = 1, 2, 3$. The equation of the probability density distribution of the number of speeding under the range of low, medium, and high speeding in Deyang 2017 is (8.5), (8.6), (8.7).

Less than 20% of the low Overspeed range

$$f(Q_1) = 1.154 \times 10^4 \exp\left[-\left(\frac{Q_1 + 13.08}{4.519}\right)^2\right] \cdot R^2 = 0.9966 \tag{8.5}$$

Medium Overspeed range of 20 ~ 50%

$$f(Q_2) = 1915\exp\left[-\left(\frac{Q_2 + 11.28}{4.368}\right)^2\right] \cdot R^2 = 0.9983 \tag{8.6}$$

High Overspeed range of 50% and above

$$f(Q_3) = 1399\exp\left[-\left(\frac{Q_3 + 10.39}{4.143}\right)^2\right] \cdot R^2 = 0.9981 \tag{8.7}$$

Optimization model solving. Referring to the speed limit of 50 km or more than 80 km of road speeding fine standards (Fines of 100 RMB, 150 RMB, and 500 RMB for exceeding the speed limit by 10%–20%, 20%–50%, 50%–70% respectively), according to the text of the low, medium, and high speed range, take $P_1 = 100$, $P_2 = 150$, $P_3 = 500$**.** It is known that the growth rate of per capita consumption level of urban residents in Deyang City in 2018 is 9.0%, and the growth rate of the consumer price index (CPI) is 1.8%. The sum of the capita consumption level and the CPI (10.8%) can be taken as the maximum growth rate of the total amount of fines acceptable to drivers whose speeding times are in the first step of the speeding times range. For the convenience of calculation, the approximation is taken as 10%, that is,$b_1 = 10\%$. The second and third levels of the ladder corresponding to the driver's maximum acceptable total fine growth rate can set $b_2 = 50\%$, $b_3 = 100\%$. The growth rate of the maximum total acceptable fine for all speeding drivers can also be taken as $b_0 = 50\%$. Taking the calibrated parameters into the fine amount optimized model equation yields a stepwise incremental fine amount optimized model for restricting the speeding recidivism, shown below

$$\min Q_{\text{sa}} = \sum_{i=1}^{3} w_i \left[\begin{array}{l} \int_0^{+\infty} f(Q_i) Q_i dQ_i + \int_0^{+\infty} f(Q_i) E \frac{\Delta P_{i1}}{P_i} Q_i dQ_i \\ + \int_1^{+\infty} f(Q_i) E \frac{\Delta P_{i2} - \Delta P_{i1}}{P_i} (Q_i - 1) dQ_i \\ + \int_3^{+\infty} f(Q_i) E \frac{\Delta P_{i3} - \Delta P_{i2}}{P_i} (Q_i - 3) dQ \end{array} \right] \tag{8.8}$$

$$\left\{ \begin{array}{c} F_{\text{si}} < F_{sai} < 1.5 F_{\text{si}} \\ \Delta P_{i(j-1)} < \Delta P_{ij} < \Delta P_{i(j+1)} \\ P_i + \Delta P_{\text{im}} < P_{i+1} + \Delta P_{(i+1)1} \\ F_{ai1} < 1.1 F_i \\ F_{ai2} < 1.5 F_i \\ F_{ai3} < 2 F_i \end{array} \right. \tag{8.9}$$

The model is solved in MATLAB by using the pattern search method. The results of the optimization give the ΔP =[12.055, 31.633, 53.632, 18.078, 54.000, 94.000, 60.285, 174.562, 316.000]. According to the principle of formulating the stepped incremental fines for speeding behavior described in the previous section, the values are rounded so the results are ΔP = [10, 30, 60, 20, 60, 100, 60, 200, 350]. The corresponding stepped fine amounts for each speeding range under each level are ΔP =[110, 130, 160, 170, 210, 250, 560, 700, 850]. An illustration is shown in Fig. 8.1.

8.4.2 Stepwise Incremental Fine Model Effectiveness Analysis

According to the optimized results of the model, the amount of fines for each step under a certain speeding range is significantly increased compared with that before the implementation of the fine policy, as shown in Table 8.4.

The increases in the fine amount of the first speeding ladder under each speeding range are 10, 13.33, and 12%, which can form a "buffer zone" covering about 70% of speeders under each speeding range. The second speeding ladder under each speeding range covers about 95% of speeding drivers cumulatively, and the fine amount is adjusted upward by 30–40% compared with the original values. For the remaining speeders who are on the third speeding ladder and commit speeding more than 3 times in the period, the fines are increased by 60–70% compared to the original values, values, which represent a "heavy penalty zone."

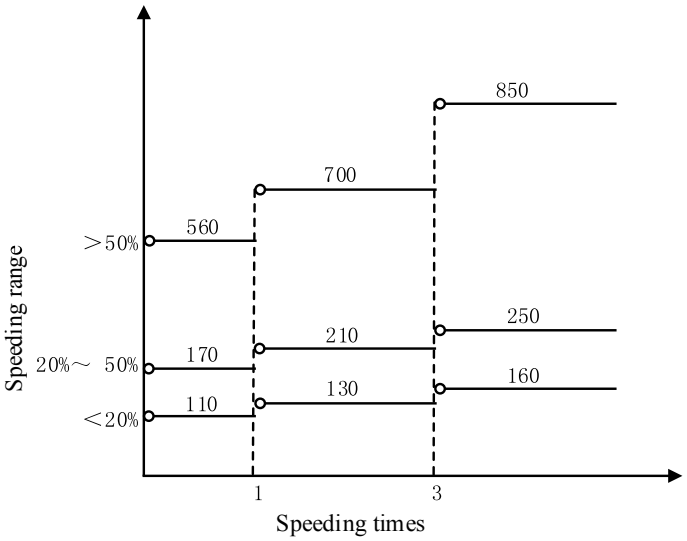


Fig. 8.1 Stepwise incremental fine mechanism for speeding in Deyang

Table 8.4 Changes in the number of fines for each step

Speeding range	Speeding times step	The number of speeding intervals of each step	Original fine amount	Adjusted fine amount	Fine increase range (%)	Cumulative coverage (%)
Low range	Level 1	1	100	110	10.00	69.25
	Level 2	2 ~ 3	100	130	30.00	94.12
	Level 3	>3	100	160	60.00	100.00
Medium range	Level 1	1	150	170	13.33	71.10
	Level 2	2 ~ 3	150	210	40.00	94.77
	Level 3	>3	150	240	60.00	100.00
High range	Level 1	1	500	560	12.00	73.00
	Level 2	2 ~ 3	500	700	40.00	95.87
	Level 3	>3	500	850	70.00	100.00

In addition, the increase in the fine amount for each ladder under the medium and high speeding range is significantly higher than that of the low speeding range, which could lower the possibility that a higher speeding range causes a greater injury severity. It can be seen that the fine amount is set according to the ratio of the speeding recidivism to the total speeding frequency in Deyang City. Hereby, this fine model can effectively restrict the recidivists who are likely to commit repeated speeding behaviors.

The speeding frequency in different ranges after the speeder's response to the model can be obtained after the implementation of the stepped incremental fine

model, as shown in Eq. (8.10). The changes in the speeding frequency in each step are shown in Table 8.5.

$$Q_{saij} = Q_i \int_{N_{j-1}}^{N_j} f(Q_i) Q_{aij} dQ_i \tag{8.10}$$

Table 8.5 illustrates that, under each speed range, the number of speeding decreases as the speeding steps keep rising. The decrease in the total number of speeding times for the first speeding ladder ranges from 1.19 to 1.58% since the speeders with only one speeding record are not sensitive to this penalty ladder. The second ladder has an average decrease rate of about 4%, which is significantly larger than that of the first ladder. The third ladder has a reduced rate of up to 8.57% or greater in the total speeding frequency, which shows that the speeders with more speeding records are more sensitive to the heavier penalty.

On the other hand, the rate of decrease in the number of speeding was significantly higher in the middle and high Overspeed ranges than in the low Overspeed range, especially in the second and third steps of the Overspeed count ladder. As can be seen, the step increment effectively reduces the number of repeat speeding drivers while affirming that drivers have a certain "basic need to exceed the speed limit."

Table 8.5 Changes in the number of speeding times in each step

Speeding range	Speeding times step	The number of speeding intervals of each step	Original speeding times	Number of speeding after response	Rate of change of speeding times (%)
Low range	Level 1	1	58,153	57,462	−1.19
	Level 2	2 ~ 3	47,292	45,604	−3.57
	Level 3	>3	26,619	24,337	−8.57
Medium range	Level 1	1	20,604	20,278	−1.58
	Level 2	2 ~ 3	15,488	14,736	−4.85
	Level 3	>3	8114	7232	−10.87
High range	Level 1	1	1473	1452	−1.42
	Level 2	2 ~ 3	1043	995	−4.65
	Level 3	>3	435	388	−10.76
Total	–	–	179,221	172,485	−3.76

8.5 Conclusion and Future Research Directions

This paper constructs an incremental fine amount optimization model based on the ladder pricing theory that aims to minimize the speeding frequency for the speeding recidivists. The speeding records collected in Deyang City, China, were used to justify the model. Speeding range and frequency were combined to develop the penalty ladders in the model. The results show that the stepped incremental penalty model can reasonably consider the needs of speeding for the speeders and their sensitivity to the penalty. Thus, the model is able to pose a greater deterrence against offenders with more repeat speeding behaviors than those with occasional speeding.

There are limitations in this paper: (1) The paper only evaluates the effect of the stepped incremental penalty model on speeding recidivism theoretically. However, the practical application is not considered to test the effectiveness of the stepped incremental penalty model. (2) The stepwise incremental penalty model does not test the long-term deterrence ability of the speeding recidivists.

In the future, speeding-related outcomes, such as traffic conflicts [22, 23], could be considered for the study of repeat speeding fines.

References

1. Liu, L., Fu, H., Lv, W., Hu, G., Wang, C., Zhang, Z.: Correspondence analysis of traffic accident types and violation types. China Saf. Sci. J. **27**(11), 79–84 (2017). (in Chinese)
2. Fu, C., Liu, H.: Investigating distance halo effect of fixed automated speed camera based on taxi GPS trajectory data. J. Traffic Transp. Eng. (English Edition) 1–30 (2021)
3. Zhang, G., Yau, K.W., Gong, X.: Traffic violations in Guangdong province of China: speeding and drunk driving. Accid. Anal. Prev. **64**, 30–40 (2014)
4. Mesken, J.: Risk factors for traffic unsafety: inventory and selection for research. In: Road Safety Research Foundation SWOV, Leidschendam (2012). (in Dutch)
5. Goldenbeld, C., Reurings, M., Van Norden, Y., Stipdonk, H.: Crash involvement of motor vehicles in relationship to the number and severity of traffic offenses. An exploratory analysis of Dutch traffic offenses and crash data. J. Crash Prev. Injury Control **14**(6), 584–591 (2013)
6. Styles, T., Imberger, K., Cairney, P.: Development of a best practice intervention model for recidivist speeding offenders. In: Australian Road Research Board 2009, BestPI, vol. AP-T134/09. Sydney, Austroads (2009)
7. Watson, B., Watson, A., Siskind, V., Fleiter, J., Soole, D.: Profiling high-range speeding offenders: investigating criminal history, personal characteristics, traffic offences, and crash history. Accid. Anal. Prev. **74**, 87–96 (2015)
8. Gebers, M. A.: Traffic conviction-and accident-record facts. Dept. of Motor Vehicles, Research and Development Section, State of California (1990)
9. Stradlin, S., Meadows, M., Beatty, S.: Characteristics of speeding, violating and thrill-seeking drivers. Int. J. Pediatr. Otorhinolaryngol. **74**(2), 168–172 (2001)
10. Elvik, R., Christensen, P.: Deterrent effect of increasing fixed penalties for traffic offences: Norwegian experience. In: 85th TRB Annual Meeting, Washington (2006)
11. Liu, F., Xie, J.: Game analysis of habitual "violation" and its management. J. Xi'an Univ. Technol. **25**(02), 32–36 (2009). (in Chinese)
12. Polinsky, A.M., Rubinfeld, D.L.: A model of optimal fines for repeat offenders. J. Public Econ. **46**(3), 291–306 (1991)

13. Delhaye, E.: The enforcement of speeding: Should fines be higher for repeated offences? Transp. Plan. Technol. **30**(4), 355–375 (2007)
14. Taylor, L.D.: Demand for electricity: a survey. Bell J. Econ. **6**, 74–110 (1975)
15. Filipovic, S., Tanic, G.: The policy of consumer protecyion in the electricity market. Econ. Ann. **53**, 178–179 (2008)
16. Borenstein, S.: To what electricity price do consumers respond? Residential demand elasticity under increasing-block pricing. Working paper, University of California, Berkeley(2009)
17. Ruijs, A.: Welfare and distribution effects of water pricing policies. Environ. Resour. Econ. (43), 2 (2009)
18. Barta, R.: Stretching "Urban water supplies in colorado-strategies for landscape water conservation". Special Report 13, Colorado Water Resources Research Institute (2004)
19. Zeng, Y.: Research on step punishment intensity model in automated enforcement. Master degree dissertation. Harbin Institute of Technology (2014). (in Chinese)
20. Liu, H.: Research on optimization model of the increasing tiered fine for motorized vehicle running red-light. Master degree dissertation. Harbin Institute of Technology (2016). (in Chinese)
21. Neary, J., Peter, R.C.: Geary's contributions to economic theory. University College Dublin **97**(4), 1–20 (1997)
22. Fu, C., Tarek, S.: A multivariate method for evaluating safety from conflict extremes in real time. Anal. Methods Accid. Res. **36**, 100244 (2022)
23. Fu, C., Tarek, S.: Identification of adequate sample size for conflict-based crash risk evaluation. an investigation using bayesian hierarchical extreme value theory models. Anal. Methods Accid. Res. **39**, 100281 (2023)

Chapter 9
Interaction-Aware Trajectory Prediction for Autonomous Vehicle Based on LSTM-MLP Model

Zhiwei Meng, Jiaming Wu, Sumin Zhang, Rui He, and Bing Ge

Abstract Trajectory prediction is one of the core functions of the autonomous vehicle, it greatly affects the rationality and safety of the decision-making module and the planning module. This is challenging because the motion of the target vehicle is affected by the interactive behavior of its surrounding vehicles. In this paper, we propose the interaction-aware trajectory prediction model for autonomous vehicles based on LSTM-MLP model. The encoder module encoded the history trajectories to extract the dynamic feature of each vehicle in the scenarios by the LSTM model, and then the interaction module captures the interactive feature using the MLP-Max Pooling model. In the end, the decoder module decodes the fusion feature which consists of the dynamic feature of the target vehicle and the interactive feature to output the future trajectory based on the LSTM model. The experiments are carried out on the publicly available NGSIM dataset, and the results show that the proposed model outperforms prior works in terms of RMSE value.

9.1 Introduction

Autonomous driving has been recognized as one of the inevitable development directions of the vehicle industry, and autonomous vehicles will become popularization in the near future [1, 2]. As an important module of autonomous driving technology, trajectory prediction can provide the input information for the decision-making

Z. Meng · S. Zhang · R. He
State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130022, China

Z. Meng · J. Wu
Department of Architecture and Civil Engineering, Chalmers University of Technology, 41296 Goteborg, Sweden

B. Ge (✉)
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
e-mail: gebing1982@163.com

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_9

module and the planning module, and improve driving safety [3–6]. Therefore, the research on trajectory prediction is meaningful.

In this field, there are many challenges, because the future trajectory of the autonomous vehicle is affected by many factors, such as the driving scenarios, and the interaction among vehicles [7, 8]. One of the important challenges is how to consider the interaction between the target vehicle and the surrounding vehicle.

In the past few decades, many researchers have devoted themselves to this challenge [9]. According to the survey, the trajectory prediction methods can be grouped into three different categories. Firstly, the physics-based method is the simplest. It assumes that the future motion of the vehicle only depends on the current motion state, and predicts the future trajectory according to the dynamic model or kinematic model. Besides, it doesn't consider the behavior of the vehicle in the trajectory prediction process. Ammoun et al. [10] introduced a model of trajectory prediction by proposing a dynamic model and more complex equations for the Kalman filtering. In the paper [11], a hierarchical-structured algorithm that fused traffic environment data with the dynamic information of the vehicle was proposed to predict the trajectory of the target vehicle. In a word, the physics-based method has the desired effect under the short-term prediction. But it is not suitable for long-term prediction.

Secondly, the maneuver-based method eliminates the drawback that the physics-based model is unable to use the behavior information. The maneuver-based method firstly will predict the maneuver intention of the vehicle, and then generate the future trajectory. A long-term prediction approach based on a combined trajectory classification and particle filter framework was proposed [12], and the motion prediction was performed based on the highest likelihood behavior. The Hierarchical Mixture of Expert model was proposed through a probability density function and inferring its parameters with previously observed motion patterns to predict the future trajectory [13]. As a result, those methods output the future locations of the vehicle given its maneuver class.

Thirdly, the interaction-based method considers the fact that the future trajectory of the target vehicle is influenced by the surrounding vehicles. It regards the target vehicle and surrounding vehicles as independent entities that influence each other. Social LSTM encoded dynamics of agents and models interaction by sharing information among agents, which is inspired by the recent success of Long-Short Memory networks for different tasks [14]. An encoder-decoder architecture that used convolutional social pooling as an improvement was proposed to output a multi-modal predictive distribution over future trajectories based on maneuver classes [15]. Hasan et al. [16] proposed the pooling strategy into a generative encoder-decoder mode, and then outputs multi-modal distribution of predictions conditioned on a set of semantic maneuvers. Gupta et al. [17] tackled the prediction issue by combining tools from sequence prediction and generative adversarial networks, in his work, the recurrent sequence-to-sequence model with a novel pooling mechanism predicted the future trajectory of the target agent by its historical trajectory. Considering dependencies between the target vehicle and surrounding vehicles, interaction-aware methods can get a more comprehensive understanding of context information.

Inspired by [15, 17], this work proposes the interaction-aware trajectory prediction for autonomous vehicles based on the LSTM-MLP model to improve the accuracy of the future trajectory. Firstly, the proposed model encodes the historical trajectory of all vehicles which are located in the scenario using the LSTM network and then gets the dynamic features. Secondly, the dynamic features are input to the interaction module, and the MLP-Max Pooling structure is used to extract the interactive feature. Thirdly, the dynamic feature of the target vehicle and the interactive feature between the target vehicle and the surrounding vehicles. Finally, the encoder based on LSTM generates the future trajectory by the fusion feature.

The rest of this paper is organized as follows. In Sect. 9.2, we formulate the problem and describe the proposed trajectory prediction model and implementation details. In Sect. 9.3, we first introduce the dataset and evaluation metric and then report the experimental results compared with other models. In Sect. 9.4, we conclude this paper.

9.2 Methodology

9.2.1 Problem Formulation

This work aims to predict the future trajectory of the target vehicle using the historical trajectories of the target vehicle and the surrounding vehicles. The proposed model is shown in Fig. 9.1, it consists of an LSTM encoder, an interaction module based on MLP-Max Pooling structure, and an LSTM decoder. As shown in Fig. 9.1, the gray vehicle represents the target vehicle. On the other hand, the surrounding vehicles are denoted by green vehicles.

This work predicts the future trajectory of 5 s given the historical trajectories of 3 s, and the past trajectory and the future trajectory are shown in the black line segment and the red line segment, respectively.

The input to the model is historical trajectories.

$$H = \left\{h_{v_0}, h_{v_1}, \cdots h_{v_i}, \cdots, h_{v_m}\right\} \tag{9.1}$$

where, H represents the historical trajectories of all vehicles in the scenario, h_{v_i} denotes the historical trajectory of vehicle v_i. Especially, v_0 indicates the target vehicle, while others are surrounding vehicles. Besides, m is the number of surrounding vehicles in the scenario.

$$h_{v_i} = [p_{v_i}^{t-T_h}, p_{v_i}^{t-T_h+1}, \cdots, p_{v_i}^{t_1}, \cdots, p_{v_i}^{t-1}, p_{v_i}^{t}] \tag{9.2}$$

In Eq. (9.2), $p_{v_i}^{t_1}$ represents the trajectory information of vehicle v_i at time $t_1 \in [t - T_h, t]$, as shown in Eq. (9.3). Moreover, T_h is a fixed history time horizon, and time t is a constant.

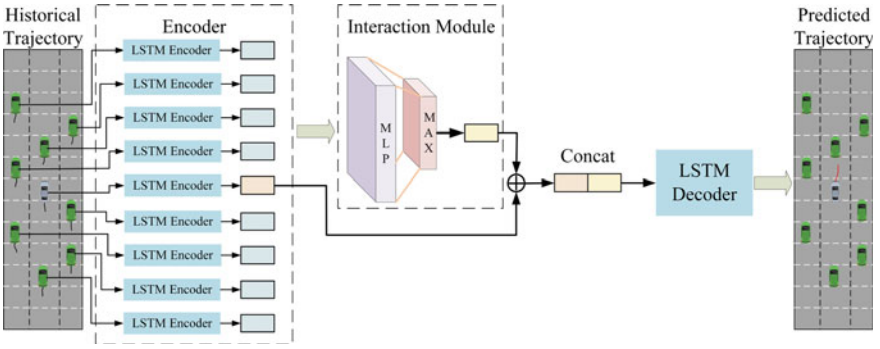


Fig. 9.1 Proposed Model. It consists of three key components: Encoder, Interaction Module, and Decoder. The encoder takes as input the historical trajectories of all vehicles in the scenario and encodes the history as the dynamic feature. The interaction Module takes as input all dynamic features and outputs the interactive feature. The decoder generates the future trajectory of the target vehicle conditioned on the dynamic feature and the interactive feature

$$p_{v_i}^{t_1} = \left(x_{v_i}^{t_1}, y_{v_i}^{t_1}\right) \tag{9.3}$$

where $x_{v_i}^{t_1}$ indicates the coordinate of the X-axis. Similarly, $y_{v_i}^{t_1}$ denotes the coordinate of the Y-axis.

The output of the model is the predicted future trajectory of the target vehicle v_0.

$$F_{v_0} = [f_{v_0}^{t+1}, f_{v_0}^{t+2}, \cdots, f_{v_0}^{t_2}, \cdots, f_{v_0}^{t+T_f}] \tag{9.4}$$

where F_{v_0} represents the predicted future trajectory of the target vehicle. $f_{v_0}^{t_2}$ is the trajectory information of the target vehicle at time $t_2 \in \left[t+1, t+T_f\right]$, it can be described as Eq. (9.5). In addition, T_f is a fixed prediction time horizon.

$$f_{v_0}^{t_2} = \left(x_{v_0}^{t_2}, y_{v_0}^{t_2}\right) \tag{9.5}$$

9.2.2 Model Structure

To build a better interactive relationship between the target vehicle and the surrounding vehicles, this work proposes an LSTM-MLP model, which is designed under the Encoder-Decoder structure. Figure 9.1 shows the proposed model in this paper, and details of the composition module of the proposed model are described below.

LSTM Encoder

Encoder learns the state of each vehicle in the scenario and stores their history of motion using the LSTM network, and the LSTMs used for each vehicle have shared weights.

First, in order to get a fixed length vector a_{v_i}, the history information is embedded by a LeakyRelu non-linearity layer, as shown in Eq. (9.6).

$$a_{v_i} = LeakyRelu\left(Linear\left(h_{v_i}\right)\right) \tag{9.6}$$

where the *LeakyRelu* is the activation function.

Then, the LSTM network takes as input a_{v_i} and outputs the encoded features. And the LSTM weights are shared between all vehicles in the scenario. Finally, those encoded features are used as input to the LeakyRelu non-linearity layer to get the dynamic feature of the vehicle DF_{v_i}. The above process is shown in Eq. (9.7).

$$DF_{v_i} = LeakyRelu\left(Linear\left(LSTM_{enc}\left(a_{v_i}\right)\right)\right) \tag{9.7}$$

Interaction Module

Considering that, in the autonomous driving application scenario, the target vehicle will adjust its trajectory by implicitly reasoning about the motion of its neighbors. Therefore, it is necessary to capture the interdependence among vehicles. Inspired by Social-GAN [7], this work proposed an MLP-Max Pooling structure to solve this issue.

First, the relative positions $pos_{rel}^{t_1}$ between the target vehicle and surrounding vehicles are computed, as shown in Eq. (9.8).

$$pos_{rel}^{t_1} = \left(p_{v_i}^{t_1} - p_{v_0}^{t_1}\right) \tag{9.8}$$

Second, these relative positions are embedded by a Linear layer to get a fixed vector b_{rel}.

$$b_{rel} = Linear\left(pos_{rel}^{t_1}\right) \tag{9.9}$$

Then, the vector b_{rel} is concatenated with each vehicle's encoded features.

$$c_{v_0} = cat\left(b_{rel}, LSTM_{enc}\left(a_{v_i}\right)\right) \tag{9.10}$$

Finally, the concatenate vector c_{v_0} is input to the MLP-Max Pooling model. The MLP process the concatenated vector, then its output is pooled element-wise to compute the interactive vector of the target vehicle using the MaxPooling method. So far, the interactive feature IF_{v_0} is derived.

$$IF_{v_0} = MLP.Max\_Pooing\left(c_{v_0}\right) \tag{9.11}$$

LSTM Decoder

The decoder based on LSTM is used to decode the fusion feature which includes the dynamic feature DF_{v_0} and the interactive feature IF_{v_0}, and it outputs the predicted future trajectory of the target vehicle.

$$F_{v_0} = LeakyRelu\left(Linear\left(LSTM_{dec}\left(\left[DF_{v_0}, IF_{v_0}\right]\right)\right)\right) \tag{9.12}$$

Implementation Details

The proposed model is implemented using Python Programming Language and PyTorch Library [18]. The LSTM networks are used in the encoder module and the decoder module, the encoder LSTM has a 64-dimensional state while the decoder has a 128-dimensional state. For the interaction module, the MLP has a 256-dimensional state. All layers have the same LeakyRelu activation function, and the proposed model is trained with a batch size of 128 for 50 epochs to minimize the weighted root mean square error (RMSE) loss function using Adam.

9.3 Experiment

9.3.1 Dataset

The experiments are carried out on the publicly available vehicle-trajectory dataset NGSIM, which consists of two subsets: US-101 and I-80 [19]. Both subsets were captured at 10 Hz over 45 min, which have three traffic conditions: mild, moderate, and congested. Each trajectory is segmented into 8 s, where we use the first 3 s as the observed history trajectory and the remaining 5 s as the prediction of future trajectory. Additionally, the coordinate of the vehicle in a local coordinate system is provided. We divide the complete dataset into training, validation, and test sets [20].

9.3.2 Evaluation Metric

Root mean square error (MRSE) of the predicted trajectory with regard to the ground truth is used to evaluate the prediction accuracy over the prediction horizon [21]. The RMSE is calculated for each predictive time t_2 as Eq. (9.13).

$$\mathrm{RMSE}_{v_0}^{t_2} = \sqrt{\frac{1}{n}\sum_{z=1}^{n}\left(\left(x_{v_0}^{t_2} - \hat{x}_{v_0}^{t_2}\right)^2 + \left(y_{v_0}^{t_2} - \hat{y}_{v_0}^{t_2}\right)^2\right)} \tag{9.13}$$

where z is the size of the test set, $\left(x_{v_0}^{t_2}, y_{v_0}^{t_2}\right)$ represents the prediction of the future trajectory of the target vehicle v_0 at time t_2, while $\left(\hat{x}_{v_0}^{t_2}, \hat{y}_{v_0}^{t_2}\right)$ denotes the corresponding future ground truth.

9.3.3 Baselines

The performance of this work is compared to the following baselines [22, 23].

- Constant Velocity (CV): This model only uses a constant velocity Kalman filter.
- Vanilla LSTM (V-LSTM): This model uses a single LSTM to encode the historical trajectory of the target vehicle and doesn't consider the interaction between the target vehicle and the surrounding vehicles.
- C-VGMM + VIM: This model proposes a unified framework for surrounding classification and motion prediction based on the variational Gaussian mixture model.
- GAIL-GRU: This model extends Generative Adversarial Imitation Learning to predict the future trajectory of the vehicle.

9.3.4 Results

Comparison results between the different models are reported in Table 9.1, which shows the effectiveness of this work.

It can be seen from Table 9.1 that the proposed model in this work has the best performance, so this work is meaningful. Besides, it is worth noting that the CV model and V-LSTM model have low prediction accuracy compared to others because these models don't consider the dependencies between vehicles. This shows that interaction information is a useful cue for trajectory prediction.

Table 9.1 Root Mean Error (RMSE) for trajectory prediction of different models on the NGSIM dataset

Prediction Horizon (s)	CV	V-LSTM	C-VGMM + VIM	GAIL-GRU	LSTM-MLP (Ours)
1	0.73	0.68	0.66	0.69	**0.59**
2	1.78	1.65	1.56	1.51	**1.29**
3	3.13	2.91	2.75	2.55	**2.15**
4	4.78	4.46	4.24	3.65	**3.24**
5	6.68	6.27	5.99	4.71	**4.58**

9.4 Conclusion

In this work, an interaction-aware trajectory prediction model for autonomous vehicle based on LSTM-MLP is proposed. The proposed model uses the MLP-Max Pooling structure to capture the inter-vehicular interaction among vehicles and employs the encoder-decoder structure to predict the future trajectory of the target vehicle based on LSTM networks. We have demonstrated the effectiveness of the model which predicts the future trajectory considering the interdependencies among vehicles, and our model improves the prediction accuracy compared to some existing methods on the NGSIM dataset. In the future, we will focus on the research that can propose a unified framework to predict the future trajectory of all kinds of agents in the scenarios.

Acknowledgements This work was supported by the Graduate Innovation Fund of Jilin University (No. 2022048) and ICV-Safe: Testing safety of intelligent connected vehicles in open and mixed road environment of VINNOVA (No. Vinnova2019-03418). The first author acknowledged financial support from the China Scholarship Council (CSC No. 202206170069).

References

1. Leon, F., Gavrilescu, M.: A review of tracking and trajectory prediction methods for autonomous driving. Mathematics **9**(6), 660 (2021)
2. Mozaffari, S., Al-Jarrah, O.Y., Dianati, M., Jennings, P., Mouzakitisa A.: Deep learning-based vehicle behaviour prediction for autonomous driving applications: a review. IEEE Trans. Intell. Transp. Syst. **23**(1), 33–47 (2020)
3. Olovsson, T., Svensson, T., Wu, J.: Future connected vehicles: communications demands, privacy and cyber-security. Commun. Transp. Res. **2**, 100056 (2022)
4. Wu, J., Qu, X.: Intersection control with connected and automated vehicles: a review. J. Intell. Connect. Veh. **5**(3), 260–269 (2022)
5. Wang, L., Yang, M., Li, Y., et al.: A model of lane-changing intention induced by deceleration frequency in an automatic driving environment. Physica A **604**, 127905 (2022)
6. Wang, L., Yang, M., Li, Y., et al.: Resolution strategies for cooperative vehicle fleets for reducing rear-end collision risks near recurrent freeway bottlenecks. J. Intell. Transp. Syst. 1–19 (2022)
7. Xue, Q., Gao, K., Xing, Y., Lu, J., Qu, X.: A Context-aware framework for risky driving behavior evaluation based on trajectory data. IEEE Intell. Transp. Syst. Mag. **15**(1), 70–83 (2023)
8. Liu, Y., Lyu, C., Zhang, Y., et al.: DeepTSP: Deep traffic state prediction model based on large-scale empirical data. Commun. Transp. Res. **1**, 100012 (2021)
9. Zhu, Y., Qian, D., Ren, D., et al.: Starnet: Pedestrian trajectory prediction using deep neural network in star topology. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8075–8080 (2019).
10. Nashashibi, F., Ammoun, S.: Real time trajectory prediction for collision risk estimation between vehicles. In: The IEEE 5th International Conference on Intelligent Computer Communication and Processing, pp. 417–422 (2009)
11. Polychronopoulos, A., Tsogas, M., Amditis, A.J., et al.: Sensor fusion for predicting vehicles' path for collision avoidance systems. IEEE Trans. Intell. Transp. Syst. **8**(3), 549–562 (2007)
12. Wohler, C., Hermes, C., Schenk, K., et al.: Long-term vehicle motion prediction. In: The IEEE Intelligent Vehicles Symposium, pp. 652–657 (2009)

13. Wiest, J., Kunz, F., Kressel, U., et al.: Incorporating categorical information for enhanced probabilistic trajectory prediction. In: The 2013 12th International Conference on Machine Learning and Applications, pp. 402–407 (2013)
14. Alahi, A., Goel, K., Ramanathan, V., et al.: Social LSTM: human trajectory prediction in crowded spaces. In: the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 961–971 (2016)
15. Deo, N., Trivedi, M.M.: Convolutional social pooling for vehicle trajectory prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1468–1476 (2018)
16. Zhu, J., Easa, S., Gao, K.: Merging control strategies of connected and autonomous vehicles at freeway on-ramps: a comprehensive review. J. Intell. Connect. Veh. **5**(2), 99–111 (2022)
17. Gupta, A., Johnson, J., Fei-Fei, L., et al.: Social GAN: Socially acceptable trajectories with generative adversarial networks. In: the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2255–2264 (2018)
18. Mo, X., Xing, Y., Lv, C.: Graph and recurrent neural network-based vehicle trajectory prediction for highway driving. In: the IEEE International Intelligent Transportation Systems Conference (ITSC), pp. 1934–1939 (2021)
19. Zhi, Y., Bao, Z., Zhang, S., et al.: BiGRU based online multi-modal driving maneuvers and trajectory prediction. Proc. Inst. Mech. Eng., Part D: J. Automob. Eng. **235**(14), 3431–3441 (2021)
20. Cai, Y., Wang, Z., Wang, H., et al.: Environment-attention network for vehicle trajectory prediction. IEEE Trans. Veh. Technol. **70**(11), 11216–11227 (2021)
21. Mo, X., Xing, Y., Lv, C.: Interaction-aware trajectory prediction of connected vehicles using CNN-LSTM networks. In: The IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, pp. 5057–5062 (2020)
22. Li, X., Ying, X., Chuah, M.C.: GRIP++: Enhanced Graph-based Interaction-aware Trajectory Prediction for Autonomous Driving (2019)
23. Li, X., Ying, X., Chuah, M.C.: Grip: Graph-based interaction-aware trajectory prediction. In: The 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 3960–3966 (2019)

Chapter 10
Speed Profile Optimization for Energy-Saving Operations of Electric Buses

Yajun Liu, Yuting Ji, and Yiming Bie

Abstract Connected and autonomous battery electric buses (EBs) equipped with Vehicle-to-Infrastructure (V2I) communications have great potential in saving energy consumption. However, the trip energy consumption of an EB is mainly affected by its operating speed. How to optimize the speed profile of the EB on the route is a critical issue needing to be solved. In this study, an eco-driving method for one EB that minimizes energy consumption by optimizing the speed profiles of trips between stops is established. In the proposed model, the number of intersections between stops and the different signal timing are considered. The developed model is solved using a genetic algorithm and a case study is conducted with an EB operating on a dedicated bus lane. The results show that the proposed method can flexibly support multiple scenarios of an EB traveling on inter-stop trips and the energy consumption is 0.40–0.43 kWh/km.

10.1 Introduction

The electrification of public transport is increasingly favored by various countries. China has achieved remarkable results in the electrification of public transit vehicles. The number of electric buses (EBs) in China reached 419,500 by the end of 2021, and EB has become the largest bus type. The promotion of EBs in other countries is also increasing, for example, New York City in the United States is expected to replace all 5,700 buses with electric ones by 2040, and the number of EBs in Europe reached more than 2,500 by the end of 2021 [1].

Becoming more automated and connected can make public transportation systems cleaner and more attractive, so it is another research direction [2–4]. Due to its significant advantages, it is foreseeable that connected and autonomous EBs will occupy the future self-driving bus market. However, EBs also have certain drawbacks, such as the frequent need for charging and battery degradation when the battery is

Y. Liu · Y. Ji · Y. Bie (✉)
School of Transportation, Jilin University, ChangchunJilin 130022, China
e-mail: yimingbie@126.com

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_10

overcharged and over-discharged. Therefore, high energy consumption will certainly lead to charging and discharging frequently, and reducing the energy consumption of EBs is essential to reduce the total cost of bus transportation systems. For this reason, various bus eco-driving approaches have been proposed by scholars.

Bus companies and passengers were more concerned about the on-time performance of the public transport system, therefore, eco-driving technology was first applied to private cars [5]. However, as buses become smarter, battery energy consumption can be improved while bus route on-time performance is ensured. References [6–8] proposed certain methods to reduce energy consumption of EBs. However, none of the literature addresses the random residence time fluctuations of EBs, which may render the planned velocity distributions and trajectories inapplicable. Only by knowing the time when EB starts to operate, the speed profile planning can determine the movement state of the bus on road sections and intersections. Considering that the dwell time of EB at downstream stations is also an uncertain variable, we divide the bus route into intervals and optimize the velocity profile of station trips instead of predicting the velocity profile of EB in the whole route or continuous time.

Therefore, this paper optimizes the dynamic speed profile of the inter-stop trip when the vehicle is about to depart from the stop, effectively avoiding the difficult estimation of the downstream intersection signal status due to the random dwell time. In the proposed model, the speed profile of one EB in the inter-stop trip is optimized to exclude the influence of passenger randomness.

10.2 Problem Formulation

10.2.1 Problem Statement

The stop where the studied EB is ready to depart is defined as the target stop, and the trip from the target stop to the next downstream stop is called the target inter-stop trip. The number of intersections in this target inter-stop trip is $N-1(N \geq 1)$. We set the signalized intersections as separation points, so that the target inter-stop trip is divided into N segments. The length of segment $n(1 \leq n \leq N)$ is defined as L_n^{seg} (unit: m).

The target EB is expected to depart from the target stop at time T^{dep} with initial velocity v^{pre} (unit: m/s) and eventually arrive at the downstream stop and decelerate to 0 m/s. The speed profile of the EB is $v(t)$ (unit: m/s), the acceleration is $a(t)$ (unit: m/s^2), and the travel distance from the target stop is $s(t)$ (unit: m), where t is a continuous variable. In addition, we need the green light time windows of signalized intersections to calculate the EB motion status. A total of F green light time windows w_n^g are obtained from segment 1 to segment $N-1$ according to V2I communication, and the lower limit of the fth green light window $w_{n,f}^g(1 \leq f \leq F)$ in segment n is $w_{n,f}^{\min}$ (unit: s) and the upper limit is $w_{n,f}^{\max}$ (unit: s).

10.2.2 Speed Profile Generation Algorithm

Bus Speed Profile Characteristics Analysis

The bus starts from segment 1, travels on the road section first, then may pass through several intersections, and finally arrives at the downstream stop. So inter-stop trip speed profile may also contain acceleration profile, uniform speed profile, and deceleration profile. The EB speed profile generation algorithm aims to organically combine acceleration profile, uniform speed profile and deceleration profile together so that the motion state of the EB can meet the following conditions: (1) the speed profile meets the vehicle's acceleration limit, vehicle, and road speed limit; (2) the bus cannot pass the intersection during the red light; (3) the speed of the bus at the stop is 0 m/s.

Inter-stop Speed Profile Generation Method

In order to portray the "uniformly variable motion—uniform motion—uniformly variable motion" of the vehicle in the segment, this paper divides the speed profile of each segment into 3 phases, the speed profile of phases 1 and 3 is uniform variable speed profile, and the speed profile of phase 2 is uniform speed profile. 3 phases are represented by variables, the duration is $T_{n,p}(T_{f,p} \geq 0)$ (unit: s), and the target speed of each stage is $v_{n,p}^{tar}$ (unit: m/s). For each segment $n(1 \leq n \leq N)$, the speed profile of the EB must be continuous, see Eqs. (10.1)–(10.4).

$$v_{n,p}\left(t_{n,p}\right) = v_{n,p}(0) + \left(v_{n,p}^{tar} - v_{n,p}(0)\right) \cdot t_{n,p} \Big/ T_{n,p},\ p \in \{1,3\},\ 0 \leq t_{n,p} \leq T_{n,p} \tag{10.1}$$

$$v_{n,p}\left(T_{n,p}\right) = v_{n,p+1}(0),\ p \in \{1,2\} \tag{10.2}$$

$$\begin{cases} v_{1,1}(0) = v^{pre} \\ v_{n,3}\left(T_{n,3}\right) = v_{n+1,1}(0) \end{cases} \tag{10.3}$$

$$\sum_{p=1}^{3} \int_{0}^{T_{n,p}} v_{n,p}\left(t_{n,p}\right) = L_n^{seg} \tag{10.4}$$

where $t_{n,p}$ denotes the time difference between the EB running in the segment n and the beginning time of the phase p, s; $v_{n,p}\left(t_{n,p}\right)$ denotes the velocity of the EB at the time $t_{n,p}$, m/s; $\int_0^{T_{n,p}} v_{n,p}\left(t_{n,p}\right)$ is the distance the EB travels in the phase p, m.

If $N > 1$, the end of segment 1 to $N-1$ is the signalized intersection. An EB cannot pass the intersection during the red light when running on, and if it encounters a red light stop then the red light stop duration T_n^r can be calculated according to the signal timing, as is shown in Eqs. (10.5)–(10.7).

$$T_n^r = 0 \quad \text{if } T^{dep} + \sum_{nn=1}^{nn=n-1}\left(T_{nn}^r + \sum_{pp=1}^{pp=3} T_{nn,pp}\right) + \sum_{pp=1}^{pp=3} T_{n,pp} \in w_n^g \tag{10.5}$$

$$\begin{cases} T_n^r = w_{n,f^{arr}+1}^{\min} - T^{dep} - \sum_{nn=1}^{nn=n-1}\left(T_{nn}^r + \sum_{pp=1}^{pp=3} T_{nn,pp}\right) - \sum_{pp=1}^{pp=3} T_{n,pp} \\ v_{f,3}^{tar} = 0 \end{cases} \tag{10.6}$$

$$if\ T dep + \sum_{nn=1}^{nn=n-1}\left(T_{nn}^r + \sum_{pp=1}^{pp=3} T_{nn,pp}\right) - \sum_{pp=1}^{pp=3} T_{n,pp} \notin w_n^g$$

$$w_{n,f^{arr}}^{\max} \le T^{dep} + \sum_{nn=1}^{nn=n-1}\left(T_{nn}^r + \sum_{pp=1}^{pp=3} T_{nn,pp}\right) + \sum_{pp=1}^{pp=3} T_{n,pp} < w_{n,f^{arr}+1}^{\min} \tag{10.7}$$

where nn and pp is the intermediate variable used to assist in the calculation; $\sum_{nn=1}^{nn=n-1}\left(T_{nn}^r + \sum_{pp=1}^{pp=3} T_{nn,pp}\right) + \sum_{pp=1}^{pp=3} T_{n,pp}$ calculates the total time taken by the EB to run from segment 1 to the intersection of the segment n, s; $T^{dep} + \sum_{nn=1}^{nn=n-1}\left(T_{nn}^r + \sum_{pp=1}^{pp=3} T_{nn,pp}\right) + \sum_{pp=1}^{pp=3} T_{n,pp}$ denotes the moment of arrival at the nth intersection, s; T_n^r is the red light stop waiting time for EB, s.

The EB will eventually dwell at a bus stop and wait.

$$v_{N,3}^{tar} = 0 \tag{10.8}$$

According to Eqs. (10.1)–(10.8), the speed profile of the EB on the segment $n(1 \le n \le N)$ is determined by 6 parameters, $v_{n,p}^{tar}$ and $T_{n,p}(p = 1, 2, 3)$, where $v_{n,1}^{tar} = v_{n,2}^{tar}$. By stitching all the speed profile of segments together, the inter-stop trip speed profile can be obtained, and the subsequent objective function calculation can be calculated based on this inter-stop trip speed profile.

10.2.3 Objective Function and Constraints

Objective Function

The vehicle operating economy is usually measured using the total energy consumption of the EBs. The total EBs energy consumption consists of two components, the net energy consumption of the battery driving the wheels and the energy consumption of the battery to meet the needs of the auxiliary devices [11].

According to Newton's second law, the wheels need to do work for EBs to accelerate, decelerate, and overcome gravity, frictional resistance, and air resistance, so the instantaneous power of the wheels is calculated as in Eq. (10.9).

$$p^w(t) = v(t)[\delta ma(t) + mg\sin\theta + C^r mg\cos\theta + \frac{1}{2}\rho^{air}C^d A^{fro}(v(t))^2] \quad (10.9)$$

where $p^w(t)$ is the instantaneous wheel power, W; δ is a factor that considers the inertia of all rotating components in the drivetrain; m is the total weight, kg; θ is the road slope, rad; g is the gravitational acceleration, m/s^2; ρ^{air} is the air mass density, kg/m3; C^r is the rolling resistance parameter; C^d is the aerodynamic drag coefficient; A^{fro} is the top wind area of EB, m^2.

When $p^w(t) < 0$, the EB can convert and store part of the kinetic energy back in the battery (regenerative braking). In addition, taking into account the losses of the battery, the powertrain, and the motor in transferring energy for the wheels to do work, the net power of the battery used to drive the EB is calculated as follows.

$$p^{net}(t) = \begin{cases} p^w(t) \Big/ \left(\eta^{dl}\cdot\eta^{em}\cdot\eta^{bat}\right), & if\ p^w(t) \geq 0 \\ p^w(t)\cdot\eta^{dl}\cdot\eta^{em}\cdot\eta^{bat}\cdot\eta^{reg} & if\ p^w(t < 0) \end{cases} \quad (10.10)$$

where $p^{net}(t)$ is the net power of the battery used to drive the EB wheels, W; η^{dl}, η^{bat} and η^{reg} are powertrain efficiency, electric motor efficiency, battery efficiency, and regenerative braking energy efficiency respectively, %.

The auxiliary devices include both the heating, ventilation, and air conditioning (HVAC) system and other electronics [12]. This paper makes a simplification and considers the energy consumption provided by the bus battery to the auxiliary devices as a known value $p^{aux}(t)$. The total power of the battery is calculated as follows.

$$p^{bat}(t) = p^{net}(t) + p^{aux}(t) \quad (10.11)$$

where $p^{bat}(t)$ is the total power of the battery, W.

The objective function is calculated as follows.

$$Z = \frac{1}{3.6\times 10^6}\int_{T^{dep}}^{T^{dep}+\sum_{n=1}^{n=N}\sum_{p=1}^{p=3}T_{n,p}+\sum_{n=1}^{n=N-1}T_n^r} p^{bat}(t)dt \quad (10.12)$$

where Z is the objective function that represents the total energy consumption of EB operation between stops, kWh.

Constraints

The speed profile of EB should meet the vehicle and road condition restrictions.

$$v_{n,p}^{\min} \leq v_{n,p}^{tar} \leq v_{n,p}^{\max} \quad (10.13)$$

$$a^{\min} \le \left(v_{n,p}^{tar} - v_{n,p}(0)\right) \Big/ T_{n,p} \le a^{\max} \tag{10.14}$$

$$T_{n,p}^{\min} \le T_{n,p} \le T_{n,p}^{\max} \tag{10.15}$$

where $v_{n,p}^{\min}$ and $v_{n,p}^{\max}$ is the minimum and maximum speed, m/s; $a^{\min}$ and $a^{\max}$ is the minimum and maximum acceleration, m/s^2; $T_{n,p}^{\min}$ and $T_{n,p}^{\max}$ is the minimum and maximum duration time, s.

10.2.4 Optimization Model

An optimization model for the eco-driving speed profile was developed as follows.

$$\min Z \tag{10.16}$$

$$\text{s.t.} \begin{cases} v_{n,p}^{\min} \le v_{n,p}^{tar} \le v_{n,p}^{\max} \\ a^{\min} \le \left(v_{n,p}^{tar} - v_{n,p}(0)\right) \Big/ T_{n,p} \le a^{\max} \\ T_{n,p}^{\min} \le T_{n,p} \le T_{n,p}^{\max} \end{cases} \tag{10.17}$$

10.2.5 Solution Algorithm

Since the objective function is a segmented function and the calculation is complex, the exact algorithm such as the Lagrange multiplier method cannot be used to solve the proposed model, and the genetic algorithm (GA) [13] is chosen to solve it with the following steps.

Step 1: Algorithm preparation. The traditional float-encoded GA is applied; the objective function Z itself is selected as the fitness value fvl, and when the fitness value is small, the value of the objective function is smaller and the result is better; the crossover operation and mutation operation are performed according to the basic operation method of float-encoded GA.

Step 2: Population initialization. Random Pop chromosomes are generated and number of the current iteration is recorded as 0. Each chromosome x represents a feasible solution, which is the optimization variable $v_{n,p}^{tar}$, $T_{n,p}(1 \le n \le N)$ in this paper.

Step 3: Record and update the optimal chromosome. Calculate the fitness value fvl_i of the i-th $(1 \le i \le Pop)$ chromosome in the population. If the current iteration is 0, record the minimum fitness value fvl_{best} and the corresponding chromosome x_{best} in this iteration directly; if the current iteration is not 0, compare the minimum

fitness value in this iteration with fvl_{best}, and update the smaller fitness value to fvl_{best}, and the corresponding chromosome to x_{best}.

Step 4: Selection, crossover, and mutation operations. In order to make individuals with smaller fitness values more likely to be selected, the probability of each chromosome being selected is defined as $p_i = \frac{1/fvl_i}{\sum_{i=1}^{Size} 1/fvl_i}$. Reselect Pop chromosomes according to the roulette method. Crossover operation and mutation operation are performed on the chromosomes selected by the roulette method, and the individuals without crossover or mutation are retained until a new population consisting of Pop individuals is generated.

Step 5: Termination judgment. If the termination condition is not satisfied, the number of current iteration will be increased by 1, and step 3, 4, and 5 will be repeated. Termination condition 1: the current number of iterations reaches the maximum iteration times $G_{\max}$. Termination condition 2: the reduction of the average chromosome fitness value of the best chromosome over successive G_{lim} iterations is less than or equal to the set threshold *Toc*. Termination condition 3: The computation time from iteration 0 to the current iteration exceeds the pre-given maximum CPU computation time.

Step 6: The solution is finished. The chromosome x_{best} is used as the optimal solution and the fitness value fvl_{best} is used as the optimal objective function.

10.3 Case Study

10.3.1 Data Collection

In this paper, a bus line operating on a dedicated bus lane in Changchun City, Jilin Province, is selected for the case study. The total length of the bus lane is 5,937 m. The schematic diagram is shown in Fig. 10.1.

An EB in operation is chosen for the case study, and the basic parameters related to energy consumption calculation are shown in Table 10.1.

The speed profile of the EB is optimized for the day 10th August 2022 around 7:00–7:15. The vehicle starts to depart from the first stop at 7:00:03, and the following information is obtained based on V2I communications (Table 10.2).

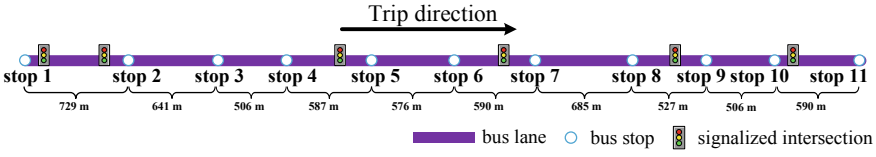


Fig. 10.1 Bus lane diagram

Table 10.1 Basic parameters

Parameter	Value	Parameter	Value
Curb weight	8200 kg	Powertrain efficiency	97%
Size	$8.5 \times 2.5 \times 3.2$ m^3	Electric motor efficiency	95%
Load capacity	60 pax	Battery efficiency	91%
Front area	7.29 m^2	Regenerative braking energy efficiency	60%
Aerodynamic drag coefficient	0.7	Auxiliary devices power	1891 W
Rolling resistance parameter	0.008	Air density	1.293 kg/m^3

Table 10.2 Inter-stop trip information

Trip number	Number of intersections	Dwell time (s)	Cycle length (s)	Green time (s)
1	2	30	47, 62	20, 30
2	0	45	–	–
3	0	12	–	–
4	1	4	68	34
5	0	40	–	–
6	1	11	59	26
7	0	16	–	–
8	1	20	65	21
9	0	16	–	–
10	1	14	74	28

10.3.2 Results

A calculation script is carried out using Python (version: 3.90) to realize the GA, and the computing environment is AMD Ryzen 5 5600 H with Radeon Graphics 3.30 GHz, RAM: 16 GB. The basic parameters are set as $Pop = 100$, $R_{co} = 0.7$, $R_{mu} = 0.01$, $G_{\max} = 100$, $G_{lim} = 30$, and $Toc = 1 \times 10^{-6}$. Taking the first inter-stop trip as an example, the GA optimization process is as shown in Fig. 10.2.

The optimal fitness value of 0.312 kWh is searched around the 82nd iteration with a travel time of 119 s. Due to the existence of two signalized intersections in the first inter-stop trip, the total computing time is 0.32 s. The target speeds for a total of 9 stages in 3 sections are 3.14, 3.14, 8.79, 8.48, 8.48, 6.69, 4.82, 4.82, 0 m/s, and the stage durations are 7.40, 16.95, 7.40, 9.80, 28.03, 20.00, 15.70, 10.00, 3.10 s.

The eco-driving speed profile of the studied EB traveling in the first inter-stop trip is shown in Fig. 10.3. The results of EB in other inter-stop trips are shown in Table 10.3.

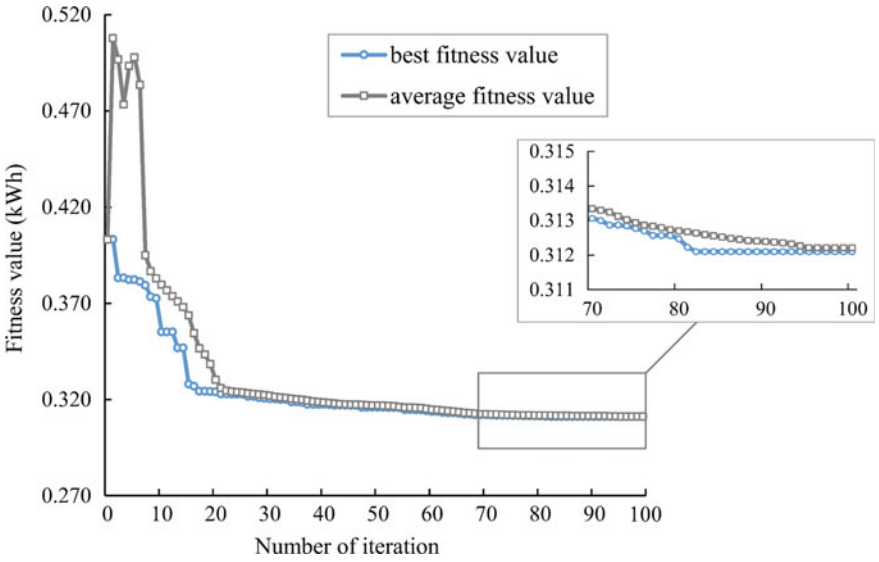


Fig. 10.2 Optimization process

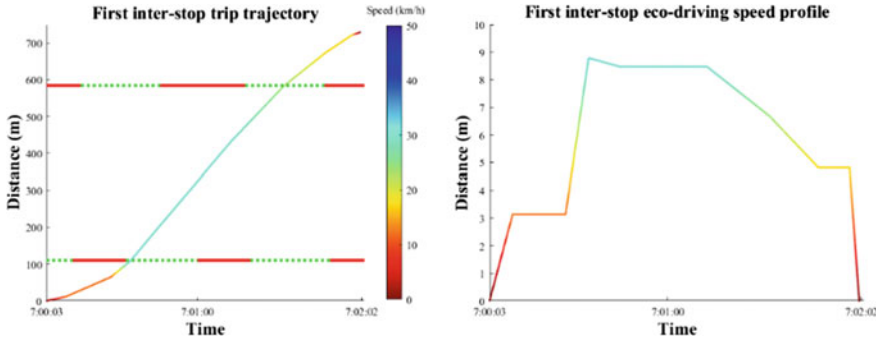


Fig. 10.3 EB eco-driving trip trajectory and speed profile

For the inter-stop trips without intersections, the computing time is relatively shorter, basically maintaining around 0.08 s. For the inter-stop trips with one intersection, the computing time is about 0.14 s. All of the computing times can meet the computational requirement of 0.5 s.

It can be seen from the table that the target speed is around 4–5 m/s and the time duration is around 3.0 s in phase 1 segment 1, and the duration time of phase 2 is longer. The acceleration process consumes more battery energy, so the acceleration should not be too large or too small. In addition, it meets our expectations that the EB keeps a uniform speed for a long duration time of phase 2 to reach low energy consumption.

Table 10.3 Optimization results

Trip No.	Energy consumption (kWh)	Segment No.	Phase duration time (s)			Phase target speed (m/s)		
			1	2	3	1	2	3
2	0.264	1	3.0	120.5	20.0	4.86	4.86	0.00
3	0.212	1	3.0	95.0	20.0	4.75	4.75	0.00
4	0.240	1	3.0	50.3	13.6	5.31	5.31	6.04
		2	16.6	21.0	20.0	4.72	4.72	0.00
5	0.237	1	3.0	100.8	19.8	5.09	5.09	0.00
6	0.239	1	3.0	49.6	14.6	5.35	5.35	5.72
		2	20.0	22.6	20.0	4.20	4.20	0.00
7	0.280	1	3.0	126.4	20.0	4.97	4.97	0.00
8	0.217	1	3.0	31.9	20.0	5.45	5.45	5.33
		2	7.7	31.9	20.0	4.73	4.73	0.00
9	0.212	1	3.2	94.8	20.0	4.75	4.75	0.00
10	0.246	1	3.0	10.0	17.0	3.58	3.58	5.46
		2	12.6	77.3	19.8	4.68	4.68	0.00

As a result, it can be calculated that the energy consumption of EB for the 10 inter-station trips is 0.41–0.43 kWh/km, which reaches a good optimization effect.

10.4 Conclusions

In this paper, an eco-driving speed profile optimization model for EB is established to output the speed profile of the next inter-stop trip, considering the impacts of signalized intersections between adjacent stops.

The method proposed in this paper can adapt to a variety of inter-stop trip scenarios, and even if there are a variety of signalized intersections between stops, this paper can optimize the EB eco-driving speed profile in a shorter time, and exclude the interference of random dwell time. The results of the case study show that the proposed method is effective in saving EB energy consumption.

Acknowledgements This work was supported by the National Natural Science Foundation of China [grant numbers: 52131203 & 52220105001], and Graduate Innovation Fund of Jilin University [grant numbers: 2022206].

References

1. The number of electric buses in Europe reached more than 2,500 by the end of 2021. https://www.mordorintelligence.com/industry-reports/europe-electric-bus-market. Last accessed 13 Apr 2022
2. Bie, Y., Ji, J., Wang, X., et al.: Optimization of electric bus scheduling considering stochastic volatilities in trip travel time and energy consumption. Comput.-Aided Civ. Infrastruct. Eng. **36**(12), 1530–1548 (2021)
3. Ji, J., Bie, Y., Wang, L.: Optimal electric bus fleet scheduling for a route with charging facility sharing. Transp. Res. Part C: Emerg. Technol. **147**, 104010 (2023)
4. The first driverless bus vehicle was piloted on the open road in Shenzhen, China. https://www.sohu.com/a/208309139_467457. Last accessed 26 Apr 2022
5. Zhang, J., Tang, T., Yan, Y., et al.: Eco-driving control for connected and automated electric vehicles at signalized intersections with wireless charging. Appl. Energy **282**, 116215 (2021)
6. Yu, L., Kong, D., Yan, X.: A driving behavior planning and trajectory generation method for autonomous electric bus. Future Internet **10**(6), 51 (2018)
7. Shan, X., Wan, C., Hao, P., et al.: Connected eco-driving for electric buses along signalized arterials with bus stops. IET Intel. Transp. Syst. **00**, 1–13 (2022)
8. Torabi, S., Bellone, M., Wahde, M.: Energy minimization for an electric bus using a genetic algorithm. Eur. Transp. Res. Rev. **12**(1), 1–8 (2020)
9. Gallet, M., Massier, T., Hamacher, T.: Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. Appl. Energy **230**, 344–356 (2018)
10. Nader, A.E., Aboelsood, Z., Hany, E.Z.F.: Novel electric bus energy consumption model based on probabilistic synthetic speed profile integrated with HVAC. IEEE Trans. Intell. Transp. Syst. 1–15 (2020)
11. Bie, Y., Liu, Y., Li, S., et al.: HVAC operation planning for electric bus trips based on chance-constrained programming. Energy **258**, 124807 (2022)

Chapter 11
Control Methods for Energy-Saving Electric Bus Operation

Yuwei Chen and Linhong Wang

Abstract Electric buses (EBs) have the advantages of lower noise, higher driving stability and less emissions, which significantly reduce traffic related air pollutants and bus company operation costs. However, the short driving range is one of its tough problems. This study aims to develop an energy-saving optimization model for a pure electric bus traveling between two continuous stations to reduce energy consumption. Firstly, the state of an electric bus is divided into three steps including time expression, state of motion, and the calculation of energy consumption. Secondly, considering punctuality indicators and comfort of passengers, the optimization model for saving energy is established and solved by Immune Algorithm. Finally, the case study based on a real electric bus route in Meihekou City is conducted to verify the optimization model. The results show that the optimization model can reduce energy consumption effectively no matter whether the traffic light of the intersection is green or red.

11.1 Introduction

Electric buses (EBs) have the advantages of lower noise, higher driving stability and less emissions, which significantly reduce traffic related air pollutants and bus company operation costs [1, 2]. However, driving range of the EB is a very important subject in the current study. Research shows that driving behavior approaching signalized intersections is related to energy consumption. Therefore, energy-saving behavior has a significant effect on improving driving range of EBs.

Current methods for estimating the energy consumption of electric buses fall into three main categories: empirical, dynamics and data-driven methods. The average power consumption of electric buses used in existing studies includes 1.24–2.48 kWh/km [3], 1.20–2.90 kWh/km [4], 1.20 kWh/km [5], 1.50 kWh/km [6]. The dynamics method usually starts from the perspective that the motor provides all the traction power required to operate the vehicle, using dynamics model to calculate the energy consumption [7–9]. The data-driven uses multiple regression or deep

Y. Chen · L. Wang (✉)
School of Transportation, Jilin University, Changchun 130022, Jilin, China
e-mail: wanghonglin0520@126.com

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_11

learning methods to construct an estimation model by multiple influencing factors as input and the energy consumption of the electric bus as output [10–12]. In addition to external factors such as road gradient and temperature, the most important factors are the dynamics of speed and acceleration. This method of achieving energy-saving and emission reduction by improving driving behavior, which can control speed gently, is eco-driving [13–17].

In summary, there are a number of researches on energy consumption and eco-driving. However, most studies on energy-saving focus on area of intersection. As the main of public transport, buses need to take punctuality into account. Therefore, this study proposes control methods for energy-saving electric bus operation considering punctuality and comfort of passengers to minimize energy consumption. Finally, a real EB route in Meihekou City is taken as an example to verify the effectiveness of the control methods for saving energy.

11.2 Methodology

11.2.1 Problem Statement

This section focuses on energy efficient driving strategies for electric buses. The energy-saving strategy only focused on the intersection area lacks the consideration of the correlation between the state of EBs after leaving the upstream station. Therefore, in this section, the research is extended to a bus between two continuous stations.

A bus traveled from its origin to its destination is called a trip. S trips of route H throughout a day are numbered sequentially and denoted by s, $s = 1, 2, 3, \ldots, S$. Each bus station is denoted by n, $n = 1, 2, 3, \ldots, N$. In this section, the section between the upstream station n and the downstream station $n + 1$ is selected as the research section. The schematic diagram of the scenario is shown in Fig. 11.1. Assume that the cycle length is C, with Φ phases. The phase of downstream direction is controlled by phase φ. There are K EBs running in the bus lane within research section, which are represented by k, $k = 1, 2, 3\ldots, K$, from downstream station $n + 1$ to upstream station n. Buses in different numbers can belong to either same or different routes.

In order to achieve the purpose of research, this section takes the minimum energy consumption as the optimization objective, and the speed of each second an EB travels in the future within the research area as the optimization variable. In addition, there are no other vehicles traveling in the front of an EB.

11.2.2 Description of State of Electric Bus

Time expression. We assume that the moment of first trip departs station n *is* T_0, taken T_0 as the starting point of timing, and variable t is defined as the interval

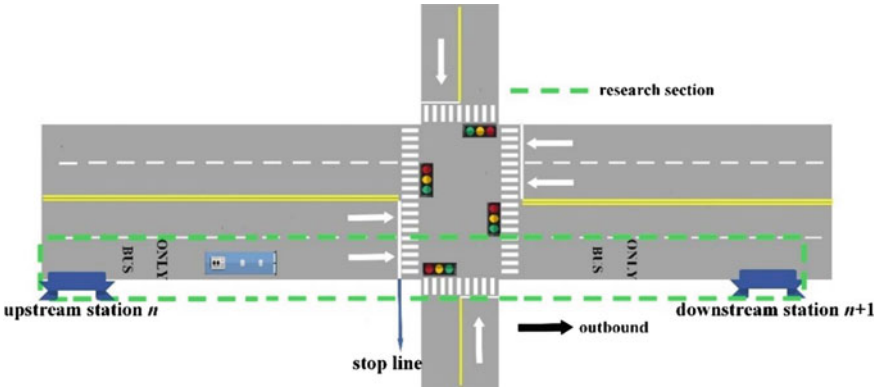


Fig. 11.1 Sketch diagram of the research section

between T and T_0. T is denoted as the real-time moment of EB k.

$$t = T - T_0 \tag{11.1}$$

where T and T_0 are moments. t is time interval, s.

Traffic light of phase φ at moment T_0 is indicated by δ^{T_0}, expressed as Eq. (11.2):

$$\delta^{T_0} = \begin{cases} 0, \text{ red light} \\ 1, \text{ green light or amber light} \end{cases} \tag{11.2}$$

The number of cycle length contained between moment T and T_0 is denoted as the calculation of cycle number, the cycle number corresponding to the phase φ at moment T is expressed as Eq. (11.3):

$$N_0^T = \begin{cases} 0 \text{ , if int}\dfrac{T - T_0 - t_{e0}}{C} < 0 \\ \text{int}\dfrac{T - T_0 - t_{e0}}{C} + 1 \text{ , if int}\dfrac{T - T_0 - t_{e0}}{C} \geq 0 \end{cases} \tag{11.3}$$

where t_{e0} is the remaining duration of T_0 of phase φ, s.

Traffic light of phase φ at moment T is represented by δ^T. $\delta^T = 1$ represents green or amber light, $\delta^T = 0$ represents a red light, which is calculated by Eq. (11.4):

$$\delta^T = \begin{cases} \delta^{T_0} \text{ , if mod}(T - T_0 - t_{e0}, C) + (-1)^{\delta^{T_0}} \cdot \left\{\delta^{T_0} \cdot (C - A) - g_e\right\} > 0 \\ 1 - \delta^{T_0} \text{ , if mod}(T - T_0 - t_{e0}, C) + (-1)^{\delta^{T_0}} \cdot \left\{\delta^{T_0} \cdot (C - A) - g_e\right\} < 0 \end{cases} \tag{11.4}$$

Remaining duration t_T^r is expressed by the difference between the expected ending moment of current phase and moment T.

$$t_T^r = \left\{\left(N_0^T \cdot C + t_{e0}\right) - \left(\delta^T - \delta^{T_0}\right) \cdot \left[\delta^T \cdot (C - A) - g_e\right]\right\} - (T - T_0) \quad (11.5)$$

where $\delta^T - \delta_0$ is the logical variable indicating whether the traffic light at moment T is the same as it at moment T_0.

Calculation of energy consumption. According to the principles of vehicle dynamics, the traction of a vehicle in motion needs to overcome all external resistance, including rolling resistance, air resistance, hill climbing resistance and acceleration resistance.

When EB is in acceleration and even pace, the tractive effort required to run is provided by the energy output of the motor. The calculation of tractive power required to run is expressed as $P_k^T(t)$ in Eq. (11.6):

$$\begin{aligned} P_k^T(t) = {} & \frac{1}{\eta_t \cdot \eta_m \cdot \eta_B} \cdot [m \cdot a_k(t) + mg\cos\theta \cdot \frac{C_r}{1000} \cdot (C_{r1} \cdot v_k(t) \\ & + C_{r2} + \frac{1}{2}\rho \cdot A \cdot C_D \cdot v_k^2(t) + mg\sin\theta \Big] \cdot v_k(t) \end{aligned} \quad (11.6)$$

where η_t is the efficiency of mechanical transmission. η_m is motor efficiency. η_B is the efficiency of a battery system. θ is road gradient. C_r, C_{r1}, C_{r2} are rolling resistance parameters. ρ is the air density, $\mathrm{kg} \cdot \mathrm{m}^{-3}$. A is the windward area of the vehicle, m^2. C_D is the air resistance coefficient. m is the total mass of an EB, kg. $v_k(t)$ is the speed of EB k t seconds after departing station n.

When EB is in a deceleration condition, one part of the reduced energy is used to provide the traction needed to drive the vehicle, one part dissipates as thermal energy, all-electric buses have kinetic energy recovery systems, and the remaining energy is recovered by the wheels into the energy storage device, which is converted into electrical energy to charge the battery. At this point power inputs, energy flows from the wheels to the motor, which is defined as negative. The calculation of braking energy recovery power $P_k^R(t)$ is expressed as Eq. (11.7):

$$\begin{aligned} P_k^R(t) = {} & -\eta_t \cdot \eta_m \cdot \eta_B \cdot \eta_c \cdot [m \cdot a_k(t) + mg\cos\theta \cdot \frac{C_r}{1000}(C_{r1} \cdot \\ & v_k(t) + C_{r2} + \frac{1}{2}\rho \cdot A \cdot C_D \cdot v_k^2(t) + mg\sin\theta \Big] \cdot v_k(t) \end{aligned} \quad (11.7)$$

where η_c is the efficiency of the motor to charge the battery during energy recovery.

The energy consumption of an EB during its journey is expressed in terms of the energy consumed in the acceleration and even pace and the energy recovered in the deceleration.

$$W = \int_0^{t_k^d} \left(P_k^T(t) + P_k^R(t)\right)dt \tag{11.8}$$

where t_k^d denotes the travel time taken by EB k to travel from the upstream stop n to the downstream stop $n+1$, s.

11.3 Objective Function and Constraints

Optimization objectives take three aspects into account including energy consumption, bus punctuality and comfort of passengers. With the objective of minimizing the energy consumption of the electric bus during the driving process, the optimization model is established as Eqs. (11.9)–(11.18).

$$\min Z = W \tag{11.9}$$

where Z is the objective function. W is energy consumption of a bus trip between two stations, kWh.

Punctuality indicators based on the station are constrained by Eqs. (11.10)–(11.12), evaluation indicators based on passenger comfort is constrained by Eq. (11.13), and driving safety is constrained by Eqs. (11.14)–(11.18).

Public service reliability is related to the quality of public transport which is denoted by the ability to provide a stable service over a period of time. The punctuality of a bus between two continuous stations of a route is regarded as punctuality indicator as follows:

$$H_s = T^s_{H,n+1,a} - T^s_{H,n,d} \tag{11.10}$$

$$H_0 = T^s_{0H,n+1,a} - T^s_{0H,n,d} \tag{11.11}$$

$$\mathrm{OTR}(B_s) = \left\{ \frac{H_s - H_0}{H_0} \in [\mu_1, \mu_2] \right\} \tag{11.12}$$

where $T^s_{0H,n,a}$ and $T^s_{H,n,a}$ are the moments when the *sth* trip of route H arrives at station n before and after the optimization, respectively. $T^s_{0H,n,d}$ and $T^s_{H,n,d}$ are the moments when the sth trip of route H leaves station n before and after the optimization, respectively. $\mathrm{OTR}(B_s)$ (on-time rate) is an indicator of the on-time rate of the sth trip based on stations of route H. μ_1 and μ_2 are time range parameters from 90 to 120%.

Jerk, which is calculated by the third order derivative of displacement, is denoted as an evaluation indicator for comfort of passengers. To ensure comfort of passengers, jerk should not be upper than 1.2 m/s^3.

$$j_k(t) \le 1.2 \tag{11.13}$$

where $j_k(t)$ is the jerk of EB k at time t, $m \cdot s^{-3}$.

The moments are discretized in every second which are denoted by T_α ($\alpha = 1, 2, \ldots, T_{n,n+1}^{\max}$), and speed per second is a constant value denoted by decision variables $v_k(T_\alpha)$. An EB should stop at a speed of 0 at the stop line, when a traffic light turns red.

$$|a_{\max}| \leq 2.5 \tag{11.14}$$

$$0 \leq v < v_{\max} \tag{11.15}$$

where $a_{\max}$ is the maximum acceleration speed, $\mathrm{m} \cdot \mathrm{s}^{-2}$.$v_{\max}$ is the maximum speed limit in a bus lane, which takes $40\,\mathrm{km} \cdot \mathrm{h}^{-1}$.

According to $2aX = v_t^2 - v_0^2$, if an EB travels with the maximum speed to the downstream station $n + 1$, it requires at least 25 m to decelerate to the speed of 0. Therefore, in order to ensure that the EB arrives at station $n + 1$ on-time, it starts to decelerate in a_p at monitoring location which is dn meters upstream of intersection i ($dn > 25$).

$$\mathrm{v}\left(T_{dn}^k\right) - a_p \cdot t_p = 0 \tag{11.16}$$

$$v\left(T_{dn}^k\right) \cdot t_p - \frac{a_p \cdot t_p}{2}\left(t_p + 1\right) = X_{n,n+1} - X_{n,dn} \tag{11.17}$$

$$a_p = \frac{v^2\left(T_{dn}^k\right)}{2 \cdot \left(X_{n,n+1} - X_{n,dn} + \frac{v\left(T_{dn}^k\right)}{2}\right)} \tag{11.18}$$

where $\mathrm{v}\left(T_{dn}^k\right)$ is the speed of EB k at monitoring location, $m \cdot s^{-1}$.a_p is arrival deceleration speed, $m \cdot s^{-2}$. t_p is arrival deceleration time, s. $X_{n,dn}$ is the distance between station n and monitoring location, m.

11.4 Solution Algorithm

In Immune Algorithm, antigen represents the objective function needed to be optimized. Antibody represents the feasible solution. In the optimization model of this study, the antigen is the sum of the minimized energy consumption, punctuality index and system comfort. The detailed process of implementing Immune Algorithm is listed as follows:

Step 1: Identify antigen and generate an initial population.

Identify objective function Z as the antigen. The population size is NP and each antibody within the population is represented by $D_i (i = 1, 2, \ldots, NP)$. Value Z of

each antibody represents the affinity function between antibody and antigen which is expressed as $Z_i(i = 1, 2, \ldots, NP)$.

Step 2: Calculate affinity, density and incentive of antibody.

When it comes to algorithms encoding real numbers, we choose Euclidean distance as the most direct way to measure *affinity* between antibodies which is calculated by Eq. (11.19):

$$\text{aff}\left(D_{i,h}, D_{j,h}\right) = \sqrt{\sum_{f=1}^{T_{n,n+1}\max} \left(D_{i,h}^{\alpha} - D_{j,h}^{\alpha}\right)^2} \; (1 \le i, j \le NP) \tag{11.19}$$

where $D_{i,h}$ and $D_{j,h}$ denote the hth generation of antibody i and antibody j respectively. $\text{aff}\left(D_{i,h}, D_{j,h}\right)$ denotes the affinity between the hth generation of antibody i and antibody j. $\text{D}_{i,h}^{\alpha}$ and $D_{j,h}^{\alpha}$ denote the α dimension of the hth generation of antibody i and antibody j respectively.

Antibody density shows the quality of population diversity which is defined by den.

$$\text{den}\left(D_{i,h}\right) = \frac{1}{NP} \cdot \sum_{1}^{NP} S_h\left(D_{i,h}, D_{j,h}\right) \tag{11.20}$$

where $\text{den}\left(D_{i,h}\right)$ represents antibody density within the hth generation. $S_h\left(D_{i,h}, D_{j,h}\right)$ is a binary variable specifying whether the k-th element of antibody i and antibody j are equal to each other, which is expressed in Eq. (11.21).

Antibody similarity is calculated by comparing the magnitude of the antibody affinity and similarity thresholds, as shown in Eq. (11.21).

$$S_h\left(D_{i,h}, D_{j,h}\right) = \begin{cases} 1\,, \text{aff}\left(D_{i,h}, D_{j,h}\right) < \delta_s \\ 0, \text{aff}\left(D_{i,h}, D_{j,h}\right) \ge \delta_s \end{cases} \tag{11.21}$$

where δ_s is the predefined similarity threshold.

As this optimization model is intended to obtain a minimum value, the affinity function $\text{aff}(t)$ is taken as the opposite of the objective function shown in Eqs. (11.22) and (11.23). Antibody incentive sim is used to evaluate the quality of the antibody, taking antibody affinity and density between the antibodies into account. Antibodies with a high affinity and low density generally have a larger incentive.

$$\text{aff}(t) = -Z(t) \tag{11.22}$$

$$\text{sim}\left(D_{i,h}\right) = a \cdot \left(-Z_{i,h}\right) - b \cdot \text{den}\left(D_{i,h}\right) \tag{11.23}$$

where a and b are parameters, usually take $a = 1$ and $b = 2$.

Step 3: Select antibodies based on antibody incentive and clone.

Select the first $\frac{NP}{2}$ antibodies among the population which are sorted by the antibody incentive in descending order and get each of them cloned for the number of Nc1.

Step 4: Mutate and recruit new antibodies.

Mutate $n \cdot \frac{NP}{2}$ antibodies. A non-uniform mutation operator is used, which searches over a large range of population in the first stage and gradually narrows as increasing iterations, dynamically adjusting the search step size for each iteration. Select high quality clones and recombine to form a new population. Each parent antibody and its corresponding cloned antibodies mutated to form a new population. The size of the population remains NP.

Step 5: Repeat iterative process.

The threshold for the number of iterations is set as H_0. When h does not reach the maximum number of iterations, return to step 2 and make $h = h + 1$. When h has reached the maximum number of iterations, output results.

11.5 Case Study

11.5.1 Data Investigation

One EB route 108 in Meihekou City, Jilin Province, China is selected as an example. The interval between Tianyi Market Station and Min'an North Court Station was selected as the research section. Cycle of Min'an-Anju Road signalized intersection with two phases is 88 s. One EB is controlled by the north–south phase, which is the direction of Min'an Road. The effective green time is 60 s in the north–south direction, 25 s in the east–west direction and amber time is 3 s. Travel time, speed of each second, outside temperature, outside humidity, latitude and longitude, altitude, road slope and other information are collected by GPS devices.

11.5.2 Results and Analysis

The results of the speed optimization and energy consumption for the two trips are shown in Table 11.1.

The analysis of velocity profile curve leads to the following conclusions:

(1) Pre-determined traffic light at monitoring location is beneficial to saving energy.

The moment EB reaching intersection can be estimated directly in connected environments. If traffic light is estimated to be green at upstream monitoring location, as shown in Fig. 11.2, the EB can go through the intersection in a smooth speed. In addition, before optimization, speed profile curve fluctuates up and down at the intersection which may be because the EB does not know the remaining time of the

Table 11.1 Comparison of travel time, deceleration time and energy consumption of two trips before and after optimization

Scene	Travel time/s		Deceleration time/s		Energy consumption/(kWh)		Energy-saving rate (%)
	Before	After	Before	After	Before	After	
Green light	48	50	10	15	2.3177	1.4231	38.6
Red light	94	96	20	28	3.2822	1.4507	55.6

green light when it reaches upstream of the intersection, so that it should decelerate its speed to observe and then accelerate immediately. Since the traffic light can be estimated in advance after optimization, speed profile curve becomes smoothly during the intersection.

(2) Increase monitoring distance and extend deceleration time appropriately is beneficial to saving energy.

According to the energy consumption function under deceleration conditions, the longer deceleration time, the higher braking energy recovery. By increasing the monitoring distance, the EB will have enough time to decelerate to the downstream station at a lower deceleration speed. But, increasing the braking time irrationally have a passive impact on travel time and whether the EB can arrive at the downstream station on-time, so that we need to coordinate a balance between deceleration time and the two. As Fig. 11.3 is shown, traffic light is estimated to be red. Before optimization, the deceleration time at the intersection of EB is 7 s and it extends to 11 s after optimization. From the results, longer deceleration time under constraints is beneficial to saving energy.

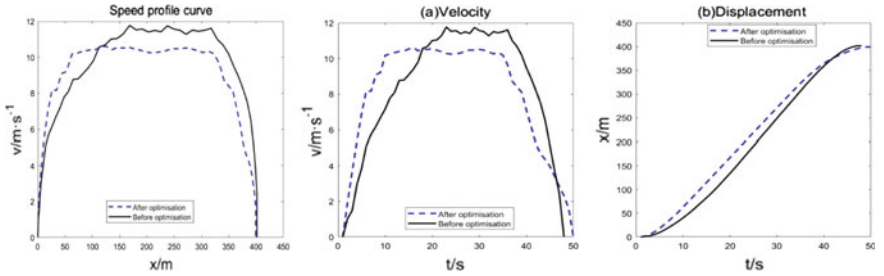


Fig. 11.2 Speed profile curve, velocity–time diagram and displacement–time diagram when traffic light is estimated to be green

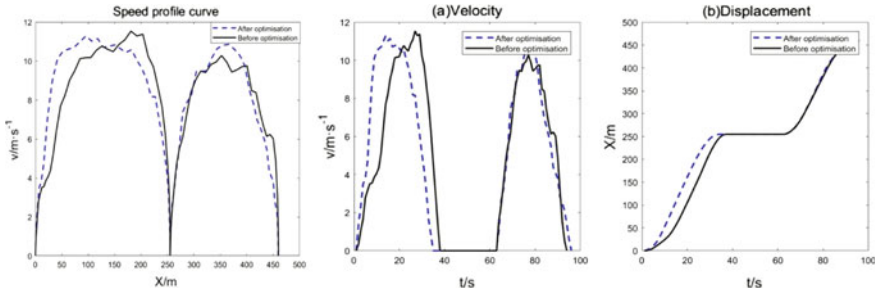


Fig. 11.3 Speed profile curve, velocity–time diagram and displacement–time diagram when traffic light is estimated to be red

11.6 Conclusions

Optimization model for single bus motion of state is developed to minimize the energy consumption between two continuous stations of an EB. Velocity of every second is taken as the optimization variable. Punctuality indicator and the comfort of passengers are regarded as two constraints more than driving safety. The case study based on a real electric bus route in Meihekou City is conducted to verify the optimization model, and the following conclusions can be obtained:

(1) The results of case study show that energy consumption can be reduced by 38.6% when the traffic light is estimated to be green and 55.6% when the traffic light is estimated to be red, respectively, through the optimization model which is verified beneficial to saving energy.
(2) In a connected environment, the optimization model can calculate the planned arrival time at the intersection of an EB by parameters, so that it can judge whether they need to stop or not in advance, or to extend its braking time for saving energy based on the calculation of energy consumption.
(3) More research efforts should be focused on the studies of the optimization for saving energy of taking the impact of EBs in front of them into account.

Acknowledgements This study was supported by Jilin Province technological innovation Program (No. 20230508048RC).

References

1. Bie, Y., Ji, J., Wang, X.: Optimization of electric bus scheduling considering stochastic volatilities in trip travel time and energy consumption. Comput.-Aided Civ. Infrastruct. Eng. **36**(12), 1530–1548 (2021)
2. Bie, Y., Hao, M., Guo, M.: Optimal electric bus scheduling based on the combination of all-stop and short-turning strategies. Sustainability **13**(4), 1827 (2021)

3. Qu, X., Wang, S., Niemeier, D.: On the urban-rural bus transit system with passenger-freight mixed flow. Commun. Transp. Res. **2**, 100054 (2022)
4. He, X., Zhang, S., Ke, W.: Energy consumption and well-to-wheels air pollutant emissions of battery electric buses under complex operating conditions and implications on fleet electrification. J. Clean. Prod. **171**, 714–722 (2018)
5. Li, L., Lo, H.K., Xiao, F.: Mixed bus fleet scheduling under range and refueling constraints. Transp. Res. Part C: Emerg. Technol. **104**, 443–462 (2019)
6. Xylia, M., Leduc, S., Patrizio, P.: Locating charging infrastructure for electric buses in Stockholm. Transp. Res. Part C: Emerg. Technol. **78**, 183–200 (2017)
7. Liu, Y.: Energy efficiency & consumption analysis and system optimization of lithium ion battery bus system and supercapacitor bus system. Beijing Jiaotong University, Beijing (2018)
8. Zhang, L., Zeng, Z., Gao, K.: A bi-level optimization framework for charging station design problem considering heterogeneous charging modes. J. Intell. Connect. Veh. **5**(1), 8–16 (2022)
9. Łebkowski, A.: Studies of energy consumption by a city bus powered by a hybrid energy storage system in variable road conditions. Energies **12**(5), 951–990 (2019)
10. Qiu, J.: A study on power consumption modeling and charging strategy of electric buses. North China Electric Power University, Beijing (2017)
11. Cao, F.: Driving style analysis and energy consumption prediction of new energy buses based on driving data. Huazhong University of Science and Technology, Wuhan (2021)
12. Gao, K., Yang, Y., Qu, X.: Diverging effects of subjective prospect values of uncertain time and money. Commun. Transp. Res. **1**, 100007 (2021)
13. Wei, T.: Research on energy-saving driving behavior and speed optimization method in vehicle networking environment. Changan University (2019)
14. Hu, L., Zhou, D., Huang, J.: Optimal path planning for electric vehicle with consideration of traffic light and energy consumption. Autom. Eng. **43**(5), 641–649 (2021)
15. Yuan, W., Zhang, Y., Wang, H.: Energy-saving driving technique for pure electric buses in intersection. China J. Highw. Transp. **34**(7), 54–66 (2021)
16. Wei, X.: Research on the eco-driving strategy of urban road based on the cooperation vehicle infrastructure system. Jilin University, Changchun (2016)
17. Liu, Y., Wang, L., Zeng, Z., Bie, Y.: Optimal charging plan for electric bus considering time-of-day electricity tariff. J. Intell. Connect. Veh. **5**(2), 123–137 (2022)

Chapter 12
Content and Evaluation of an Enterprise Reference Architecture for Demand-Responsive Public Transport

Mark-Oliver Würtz and Kurt Sandkuhl

Abstract Demand-responsive public transport, such as new mobility solutions (NMS) like car sharing, city bikes, call taxis or e-scooters, have become increasingly popular in big cities. NMS have started to influence how people move around in urban areas, and they are also expected to impact public transport operators and their IT landscape. The paper aims to contribute to an understanding of this impact and uses enterprise architecture models as a basis for visualising and analysing relevant business and IT aspects. More precisely, the focus of the research presented herein lies on developing a reference enterprise architecture for integrating demand-responsive public transport into traditional supply-oriented public transport operators. More precisely, the aim of this paper is to evaluate the initial version of a reference enterprise architecture for demand-responsive public transport. The evaluation is based on expert interviews and focusses on the area of planning and resource utilisation.

12.1 Introduction

Car sharing, urban bicycles, e-scooters, bike sharing, on-demand taxis and on-demand public buses and many other mobility solutions offered in larger cities as pay-as-you-go or subscription models have become increasingly popular and have also started to influence the way people move in short and medium distances in urban areas [1]. These solutions are often categorised as new or innovative mobility solutions (NMS) and have been the subject of research, e.g. in terms of new architectures

M.-O. Würtz (✉) · K. Sandkuhl
Institute of Computer Science, University of Rostock, Rostock, Germany
e-mail: mark-oliver.wuertz@uni-rostock.de

K. Sandkuhl
e-mail: kurt.sandkuhl@uni-rostock.de

M.-O. Würtz
JellyCo GmbH, Stuhr, Germany

K. Sandkuhl
School of Engineering, Jönköping University, Jönköping, Sweden

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_12

[2], business models [3], platforms [4] or acceptance by end users [1]. However, the way in which new mobility solutions affect public transport companies and in particular the IT landscape of these companies has been barely researched so far. Thus, the research presented aims to investigate how demand-responsive NMS can be integrated into public transport companies with the goal of a future reference enterprise architecture.

The initial version of this architecture (see Sect. 12.3) is based on a number of development steps published in previous works: In the first step, a literature analysis surveyed related works, clarified the terminology and confirmed the need for the integration of demand-oriented local public transport into the enterprise architecture of public transport operators. The basis was a case study at Germany's largest public transport company, *Berliner Verkehrsbetriebe* (BVG) [5]. In the second step, the expected impact on the enterprise architecture was investigated based on the example of ticketing. The main result was that public transport companies can only survive in the long term if they are able to integrate NMS platforms into the company and manage them. In order to achieve this goal, public transport companies should know what a future enterprise architecture for mobility platforms would look like; which processes are associated with it and how they can design the interfaces, products and services to be so open that other mobility system providers can join in at any time [6].

The third step was to start the development of a reference enterprise architecture. The analysis of existing works showed that established reference architectures do not sufficiently consider NMS. Therefore, the focus was to investigate (a) the feasibility of an extension of the reference architectures for public transport and (b) the integration capability of NMS using an existing ride-pooling or on-demand ridesharing service, the *BerlKönig* service of the *Berliner Verkehrsbetriebe AöR* (BVG) [7]. Step four further elaborated the enterprise reference architecture [8] that is based on the ITVU core model, a recognised architecture model in the public transport sector [9]. The architecture is summarised in Sect. 12.3.

The aim of this paper is an evaluation of the reference enterprise architecture using the area of planning and resource deployment. Section 12.2 presents the research methodology used before and Sect. 12.4 describes the evaluation and its results. Conclusions and future work are presented in Sect. 12.5.

12.2 Research Approach

The work presented in this paper is part of a research programme that investigates methodological and technological support for the integration of demand-responsive public transport, such as New Mobility Services (NMS), into classic supply-oriented public transport companies. This programme follows the paradigm of design science research (DSR) [10] which consists of the phases problem explication; investigation of business relevance; analysis of root causes and definition of requirements of the design solution; and design and evaluation cycles of the envisioned solution, demonstration, evaluation and communication of results. The envisioned core result of the

DSR research process, called the design artefact in DSR, will be a reference enterprise architecture that specifies how to integrate NMS into the enterprise architectures of public transport companies. "Reference" indicates that this will be a recommendation which applies to the domain of public transport companies and focusses on the general structures and processes but abstracts from the details of implementation.

Our previous work concerned the explication of the problem, business relevance, requirements and creation of the first version of the design artefact, i.e. the reference architecture. The part of our research presented in this paper concerns the evaluation of the initial version of the design artefact in an expert interview and the development of the second version based on the expert's judgement. Thus, the main research question is: From the perspective of an expert in public transport, what utility does the initial reference architecture have and what improvements have to be made?

The research method used to answer this research question is an expert interview including a qualitative content analysis of the expert interview following the procedure recommended by Mayring [11]. As a preparation for the interview, a list of open questions was developed, which covers both the background of the experts interviewed and NMS in general. The interviews were recorded and additionally the interviewer took notes. Mayring's approach includes six steps: step 1 is to decide what material to analyse—in this case obviously the recording of the interview and the notes taken. Step 2 is to explain how the data collection (i.e. in our case the interviews) was prepared and carried out. Step 3 is to explain how the material was transcribed. The material was analysed step by step following the rules of procedure to divide the material into content analytical units. Step 4 concerns the subject-reference of the analysis, i.e. to ensure the connection to the actual subject of the analysis. The subject-reference was implemented by using the subjects mentioned in the research question as categories during the analysis. Step 5 recommends a theory-guided analysis of the data; we used the state-of-the-art theory from previous work. Step 6 defines the analysis technique, which in our case was content summary.

The focus of this publication is on an enterprise reference architecture using the example of "planning and resource deployment" of a public transport company. This artefact was evaluated on the basis of an expert interview according to the architecture approach of Scholz [9] and the expertise of the authors' initial design. The results of this evaluation were considered and implemented in the artefact and comprise the main part of this scientific work.

12.3 Reference Enterprise Architecture for Demand-Responsive Public Transport

Based on the results of [6], this section contains a brief introduction to the reference enterprise architecture for public transport operators that integrate a demand-responsive local passenger transport service (or NMS) into a supply-oriented public transport company. This includes the motivational and strategic elements of the

architecture (Sect. 12.3.1) and selected elements of the business and application architecture (Sect. 12.3.2).

12.3.1 Motivational and Strategic Elements of the Architecture

In this section, the strategic connections that a classic public transport company must fulfil if it wants to integrate demand-responsive transport into its organisation, e.g. the capabilities and control of resources are presented. This strategic view is flanked by the motivation level, which connects and maps requirements, results, end states and changes of the organisation to the responsible stakeholders of the organisation. With this view of the goal architecture, a better understanding of the connection between the strategic level and the motivational level is created.

In Fig. 12.1, the lowest level shows the skills that a modern public transport company must have in order to be able to offer demand-responsive transport services. In addition to the skills that a classic public transport company also needs, such as customer service and IT development, there are skills such as digital platforms and the management of demand-responsive local passenger transport. In addition, there is (nowadays) the ability to manage transport-related environmental protection. In Germany, environmental protection is anchored in the statutes of all public transport companies and must be taken into account.

Figure 12.1 further shows how the capabilities affect the requirements that a company must implement via the control of resources. The implementation of the requirements results in the fulfilment of a final state. This final state changes the organisation, which in turn has to be accepted or even desired by stakeholders. In this context, it should be emphasised that the ability to manage demand-responsive public transport opens up the possibility of serving customers on the last mile. This increases the number of customers and ensures shorter waiting times through better connection options. The better utilisation of means of transport and the increase in the number of customers lead to a reduction of exhaust gases within the area of operation of the public transport company. Customer satisfaction increases due to the improved connections and the environmentally friendly transport service. In addition, turnover increases due to the increased demand. These driving forces contribute to the goals of the respective stakeholders CIO and CFO.

12.3.2 Business and Application Level

This section focusses on the relationship between the processes and the respective roles/actors that a public transport company must have in order to provide driving services. Furthermore, the related elements on the application level are tackled. The

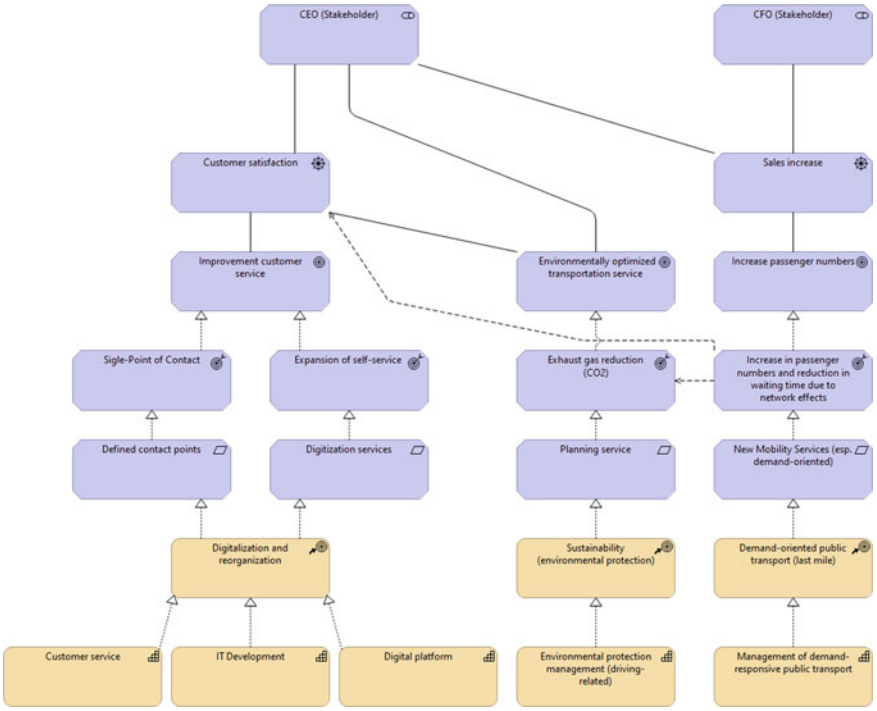


Fig. 12.1 Excerpt from reference enterprise architecture with focus on motivational elements, skills, strategic goals and capabilities

arrangement of the process elements within this architecture artefact follows the usual representation of (value-added) processes, according to which the processes are arranged by input to output from left to right. All further element types of this architectural representation are oriented towards this model. In order to develop a modern architecture that can be adapted to new challenges, two approaches were considered for the implementation of the artefact:

1. Service-oriented architecture approach at the business and application level in order to increase the flexibility of the organisation.
2. Central communication between the components with the help of an enterprise service bus on the application side, which standardises and technically channels the communication.

The reference enterprise architecture is based on the core public transport process shown in Fig. 12.2. The architecture was "thought" based on this process, i.e. the modelling was done along the core process.

The architecture presented here is structurally divided into the executive business roles, services and processes as well as the application services and applications. As presented at the beginning of this section, the EA is divided into the previously

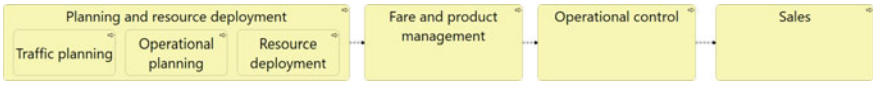


Fig. 12.2 Public transport core process as basis of the reference enterprise architecture

analysed process structure of public transport companies and thus into the following main areas:

- *Planning and use of resources*: contains the planning components for service provision. Here, traffic is combined with operational planning and resource deployment. The goal is to create an integrated, interdependent traffic and operational planning system for efficient service delivery.
- *Fare-/rental-fare collection management*: Here, on the one hand, the supply-oriented planning results are transferred into products and tariffs and, on the other hand, the demand-responsive services are provided with rental prices or rental prices from external service providers are consolidated and made available for billing.
- *Operational management*: In operational management, the services are provided or the provision of services is supported by business services such as traffic dispatching, trip execution or passenger information.
- *Sales*: This area includes the complete travel planning, booking and billing of the fare or rental. The central point of this customer-oriented area is a mobility platform that connects all forms of mobility—whether supply- or demand-responsive—offered by a public transport company.

Data objects have not been considered in this paper and will be subject of future work.

12.4 Evaluation of Reference Enterprise Architecture

The main aim of this paper is to evaluate the developed reference architecture via an expert interview and, if necessary, to improve it. This qualitative interview was conducted with the CIO (Head of IT) of *Bremer Straßenbahn AG* (BSAG). Mr. Jörg Stephan Rings has been working in the public transport sector for 12 years. Before becoming CIO of BSAG in 2018, he worked for 8 years at *Hamburger Hochbahn* in a leading position. Prior to this, he was employed in various industries unrelated to public transport. The evaluation uses the area of "traffic planning" as an example and shows the identified requirements that must be considered for the integration of demand-responsive traffic. The interview consisted of two parts, denoted as [I1] and [I2].

12.4.1 Evaluation of the Motivational and Strategic Elements

In this section, the insights into the strategic relationships in the enterprise reference architecture gained from the expert interview are presented. After an initial presentation of the model including the central goals, the interviewee replied in [I1], minute 53: "Yes. I still know a similar image from the elevated railway". By "elevated railway", the interviewee meant his former employer, the public transport company *Hamburger Hochbahn.*

During the evaluation of the model, the following two main points for an expansion were mentioned:

1. In [I1], minute 56, the interviewee said that customer satisfaction should be extended by **stakeholder satisfaction** so that it also takes into account the satisfaction of other stakeholder groups. Subsequently, the authors decided to list both and to make the connection clear by specifying customer satisfaction as stakeholder satisfaction (arrow from stakeholder to customer satisfaction).
2. In addition, the interviewee pointed out in [I1], minutes 57 and 59 that there is another central goal of the owner or sponsor (in German-speaking countries mostly a municipality or city): In the course of furthering the ecological aspects, **motorised individual transport** is to be reduced or substituted as much as possible.

Both points listed here, taking **stakeholder satisfaction** and the reduction of **motorised individual traffic** into account, are considered in the new version of the architecture. A further planned evaluation will show whether the updated approach is viable.

Figure 12.3 shows the additions made during the evaluation interview as compared to Fig. 12.1.

12.4.2 Evaluation of the Business and Application Level

In this section, the relationships within and between the business and application levels in the area of planning and resource utilisation are critically examined and evaluated. Necessary changes identified by this evaluation are considered, documented and explained. For illustration purposes, we present the initially designed architecture in Fig. 12.4 and the revised version that was the result of the evaluation in Fig. 12.5.

The results of the evaluation shown in Fig. 12.5 are explained in detail in Table 12.1. From here on, all data (especially the red numbers) refer to Fig. 12.5.

The process evaluation of the area "planning and use of resources" concluded with the question of whether the process representation for this area was sufficient ([I1], minutes 90 to 95) on the part of the interviewee.

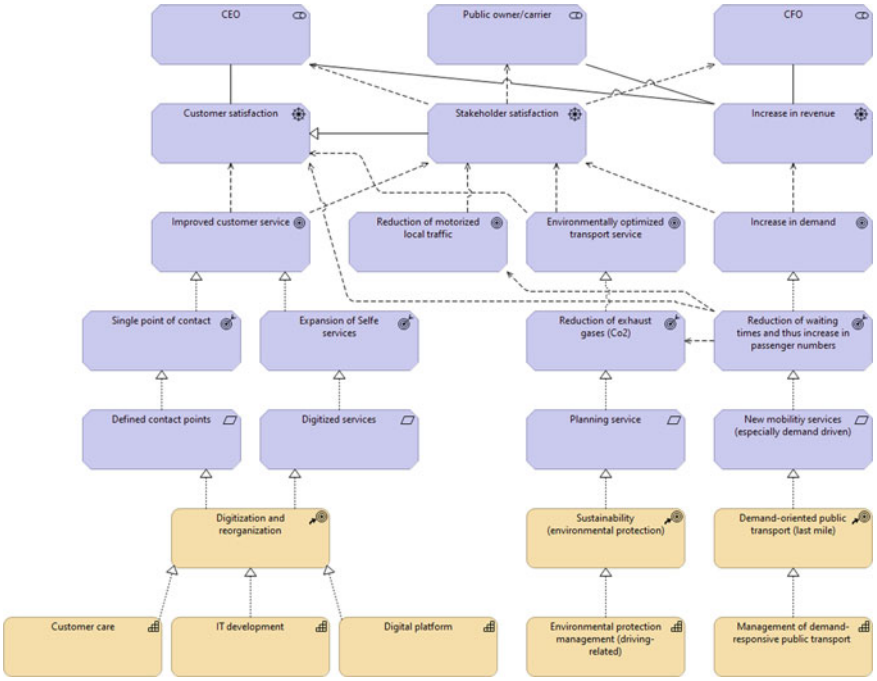


Fig. 12.3 Revised version of the motivational/strategic level, e.g. of Fig. 12.1

The interviewee stated that the

a. **the terms** used are customary in the industry and understandable;
b. **granularity** is sufficient for internal IT understanding and discussions with the specialist department;
c. **completeness** is given and
d. the model is also **easy to understand** for people who are familiar with public transport.

In another interview that took place later [I2] the structure was evaluated further. Here, the Business Services (BS) and Business Role (BR) levels in Fig. 12.5 were considered. The purpose of considering the Business Service is to decouple the role/process relationships via clearly defined services. The considered roles were all attached to the level "economic actor" (BA)—in this case that actor is an employee of the public transport company.

According to the interviewee ([I2] Min. 10), the defined Business Services (BS) are correct and suitable for the model in terms of terminology, granularity and completeness as well as comprehensible for experienced public transport employees. The same applies to the correspondingly modelled business roles in this area [I2].

Finally, the technical service (TS) and application levels were evaluated. The application level is decoupled from the processes with the help of technical services.

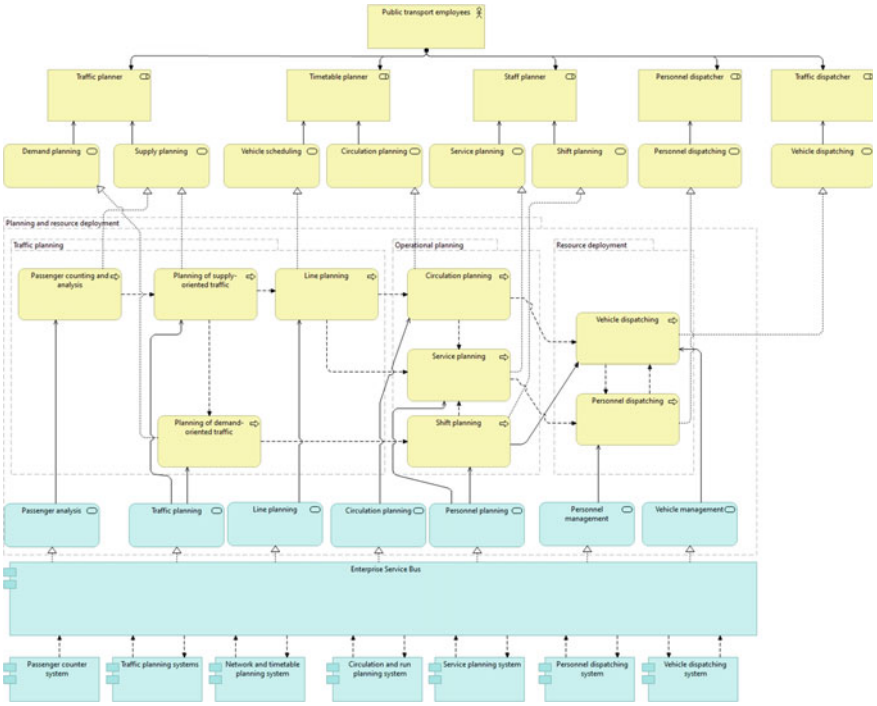


Fig. 12.4 Excerpt from business architecture (initial version)

The technical services access the specialised applications via an Enterprise Service Bus (ESB). This data hub should enable companies to develop new services quickly and easily in the future, as all data is available via this ESB application and may be "orchestrated" via a central application. The interviewee was convinced by the concept of introducing a data hub, as *Hamburger Hochbahn* is pursuing this approach as a universal service bus in the long term [I2]. The technical services were finally evaluated in the context of the interview, and it was noted (only) under TS1 that the term should be changed from "vehicle management" to "vehicle dispatching" for better understanding.

The applications communicate with each other via the ESB application. This ensures that all interfaces that are loaded/unloaded from/to specialist applications, such as the passenger counting and analysis system, are available to other specialist applications via the central ESB application.

The "Planning and Resource Deployment" area includes seven specialised applications that are supposed to exchange data with each other via the ESB application. It should be mentioned that the AP1 (passenger count and analysis system) application usually consists of more than one application or can consist of more than one application. The following application systems are listed as examples: passenger counting

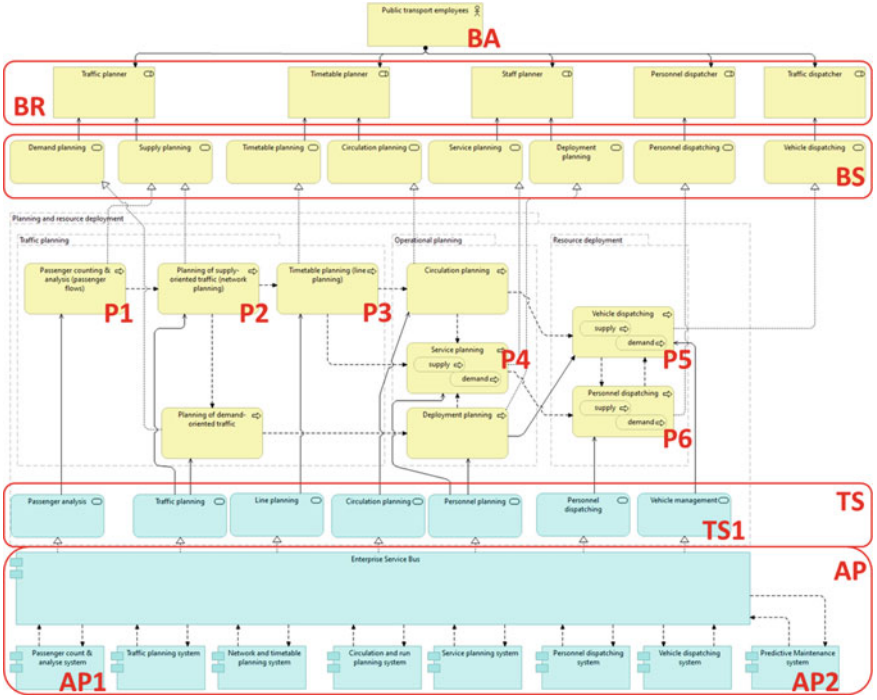


Fig. 12.5 Revised version of business architecture, e.g. of Fig. 12.4

system (in the buses and trams and as a background system), passenger survey and statistical extrapolation systems, urban planning systems.

Furthermore, the interviewee suggested in [I2] Min. 26:30 to introduce another specialised application for the maintenance of the fleet of the transport company. This predictive maintenance system (AP2) could transfer the vehicle data generated in the vehicles via the vehicle computers (onboard unit) and evaluate it, e.g. with the help of a digital twin, and thus calculate and implement optimised maintenance.

At the end of the second interview [I2], the interviewer asked the interviewee the following questions to conclude the basic evaluation:

Question 1 in [I2], min. 69: "What is your overall impression of the enterprise architecture presented?". Answer 1: "Honestly? Very good, I have always dreamed of having something like that. We even discussed it once in the VDV[1] in our work group 2 years ago. Wouldn't it be great if we had such a blueprint of what our entire world would look like? The unanimous tenor was, there is no such thing".

Question 2 in [I2], Min. 70: "Would such an enterprise architecture make your work as CIO easier?". Answer 2: "Yes, of course, the first impression has already

[1] VDV: The Association of German Transport Companies (VDV) is an industry association for public transport with more than 640 members.

Table 12.1 Explanation of evaluation results

#	[Interview]/ Minute	Explanation
P1	[I1], from 41	"Passenger flows" should be added in brackets to the name of the process "passenger count and analysis system" to make the viewer of the model understand that this is not (only) pure passenger count data, but also passenger survey results, evaluated mobile phone data and – considered very important by the interviewee - urban planning data that accurately represents current as well as future transport requirements.
P2	[I1]/from 50 [I1]/between 55 and 63	The interviewee specified the need to expand the terminology of the process with "network planning" in brackets to represent the common technical term. The interviewee outlined the customary procedure in the supply-oriented planning process, with the stages network planning, line planning and rotation planning (from a rough overview to fine details) and demand-responsive planning, which is a separate process downstream of network planning. In the context of the interview, the interviewee thought of many new services that could be developed on the basis of the information gained. For example, it would be possible to offer targeted non-transport-related products at selected high-traffic locations (account points).
P3	[I1]/63	The interviewee suggested renaming the item "line planning" to "timetable planning (line planning)" for better understanding (in the public transport sector, as many transport companies (mostly smaller ones) use the term "timetable planning" rather than "line planning".
P4	[I1]/73	Due to the different requirements, the duty scheduling must be differentiated between the demand-responsive duty scheduling and the supply-oriented rotation scheduling in terms of processes. For this purpose, corresponding sub-processes (demand/supply) were added to the model.
P5	[I1]/36	As with D, vehicle dispatching must be differentiated according to demand-responsive and supply-oriented, as both sub-processes have different characteristics. Supply-oriented vehicle dispatching mainly revolves around central fleet management, which is provided at the depots of the public transport company. Demand-responsive vehicle dispatching, e.g. with sharing service providers, revolves around the service personnel who return the distributed vehicles - mostly at night - to specific locations (e.g. mobility hubs).
P6	[I1]/36 and 65	Staff scheduling also differs, see the reasoning in E. According to the interviewee, there is a significant difference between stationary and mobile planning of staff assignments.

answered that", … "for us this is actually a help when you want to either expand or perhaps migrate".

Question 3 in [I2], Min. 71: "Do you consider the Enterprise Architecture for Public Transport to be scientifically relevant?". Answer 3: … "whether I think an enterprise architecture is scientifically relevant? Yes, immediately. PT of course I do, yes. Because, I can benefit from it".

12.5 Conclusions and Future Work

From the results of the first evaluation presented here, which were obtained on the basis of an expert interview, it can be seen that the initially derived structure, which is based on a purely supply-oriented architectural approach according to Scholz [9] and the experience of the authors, can be expanded in a targeted manner. On the basis of these results, it can be seen which architecture modules must be expanded or adapted for demand-responsive traffic.

The interviewee confirmed the usefulness of such a reference enterprise architecture for his work as CIO. This "green field" view of a coherent enterprise architecture helps with the expansion and migration of new demand-responsive processes, services and technical building blocks. The question posed in the introduction, how new mobility solutions (NMS) affect the EA of public transport organisations, can be answered after the first evaluation with the statement that the integration of NMS (demand-responsive transport) has an impact on all areas of an organisation. Not only do processes have to be extended, but capabilities, roles and business services on the one hand and technical services, application and data areas on the other hand are also affected. The modular architecture leaves open which areas will be provided within the public transport company and which will be provided externally, e.g. by a service provider.

In future work, the second area of the public transport core process, which comprises tariff and product management, operational management and sales of an enterprise reference architecture, will be presented. Based on this, the authors will conduct a further evaluation with the help of an expert interview and present the results in a final publication. It is planned to conduct the second evaluation interview with the Munich Public Transport Company (MVG). This means that the architecture was developed with the help of the largest public transport company in Germany, the *Berliner Verkehrsbetriebe* (BVG), and evaluated by a small- and a medium-sized public transport company (Bremen, Munich).

References

1. Kamargianni, M., Li, W., Matyas, M., Schäfer, A.: A critical review of new mobility services for urban transport. Transp. Res. Procedia **14**, 3294–3303 (2016)
2. Pflügler, C., Schreieck, M., Hernandez, G., Wiesche, M., Krcmar, H.: A concept for the architecture of an open platform for modular mobility services in the smart city. Transp. Res. Procedia **19**, 199–206 (2016). https://doi.org/10.1016/j.trpro.2016.12.080
3. Hildebrandt, B., Hanelt, A., Piccinini, E., Kolbe, L., Nierobisch, T.: The value of IS in business model innovation for sustainable mobility services-the case of carsharing. 1008–1022 (2015)
4. Marchetta, P., Natale, E., Pescapé, A., Salvi, A., Santini, S.: A map-based platform for smart mobility services. 19–24 (2015)
5. Würtz, M.-O., Sandkuhl, K.: Impact of new mobility services on enterprise architectures: case study and problem definition. In: International Conference on Business Information Systems, pp. 45–56 (2020)

6. Würtz, M.-O., Sandkuhl, K.: Neue Mobilitätsdienste und ihre Auswirkungen auf Unternehmensarchitekturen. In: INFORMATIK 2020 (2021)
7. Würtz, M.-O., Sandkuhl, K.: Towards a reference architecture for demand-oriented public transportation services. IEEE Xplore. https://ieeexplore.ieee.org/document/9626226 (2021)
8. University Rostock: A target enterprise architecture approach for new mobility services in demand-responsive public transport (2022)
9. Scholz, G.: IT systems in public transport: information technology for transport operators and authorities. dpunkt.Verlag, Heidelberg. http://gbv.eblib.com/patron/FullRecord.aspx?p=4658847 (2016)
10. Johannesson, P., Perjons, E.: An Introduction to Design Science, 1st edn. Springer, Cham, Heidelberg (2014)
11. Mayring, P.: Qualitative content analysis. Forum Qual. Sozialforsch. **1**(2) (2000). https://doi.org/10.17169/fqs-1.2.1089

Chapter 13
Prospects of the Activity-Based Modelling Approach: A Review of Sweden's Transport Model-SAMPERS

Omkar Parishwad and Ruo Jia

Abstract The rapid changes in global development scenarios, such as technological advancements, lifestyle decisions and climate change, call for updated transport models to test micro-level policy decisions. This paper explores the advances in activity-based transport modelling in simulating travel demand in urban scenarios, focusing on Sweden's National Transport model. Sampers is used for impact analysis, investment calculations for traffic simulations, transport policy implementation evaluations, and accessibility and impact analysis of extensive changes in land use and transport systems in cities and regions of Sweden. This research systematically compares individual components, sub-models, and algorithms and discusses integrations with cutting-edge agent-based models. Furthermore, recent research and projects for Sampers are investigated, highlighting its advantages over current models, potential gaps and limitations, and long-term development prospects. The study concludes by cross-referencing Sampers' global developments and regional needs to assess its long-term development prospects.

13.1 Introduction

The advancements in behavioural research have highlighted the strengths of an Activity-based modelling approach, such as the integrity of its sub-models, reduced interdependence between the four stages, and has resulted in a paradigm shift in travel behaviour analysis. Despite these advantages, the inability of Activity Based Model's (ABM's) to reflect such realism foreshadows the obvious need to improve

O. Parishwad (✉) · R. Jia
Department of Architecture and Civil Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden
e-mail: omkarp@chalmers.se

R. Jia
e-mail: ruoj@chalmers.se

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_13

upon the present transportation models. Moreover, researchers and transport planners are increasingly relying on ABM, the third-generation models for simulating transport demands based on the simulated behaviour for activities.

This research aims to identify recent advances in ABM for its individual components and integrated approaches, particularly agent-based models. The advantages and limitations of these advancements are analysed systematically to determine their potential applications in urban scenarios.

The recent advances [1] over various integrated approaches in ABM for the utility maximization based econometric models and the rule-based computational process models, are discussed here. We seek to note integrations of major advancements in ABM's for their potential advantages and limitations over apt applications in a systematic manner.

Overall, this paper aims to contribute to the ongoing debate on the prospects of the activity-based modelling approach by providing a comprehensive analysis of recent advancements in ABMs, their potential advantages and limitations, and their integration with cutting-edge agent-based models. The specific contribution of this research is the development recommendations for the evolved [2] Sampers.

13.2 Theoretical Background

Activity-based models focus on individual activity schedules, which include activity choice, sequence, mode and route choice as well as time dynamics, in sync with the transport network and its attributes. The model's output being simulation of scenarios and forecasting variations in trip chains for activities, modes of transportation, routes, and other constraints.

13.2.1 ABM Components and Applications

Activity-based models (ABMs) are travel demand models that simulate the entire process of individual decision-making regarding their daily activities and travel choices. ABMs typically consist of three components: activity generation, activity scheduling, and mobility assignment. The activity generation component determines an individual's desire to participate in different activities and is often modelled using heuristic, hazard-based, or micro-behavioural models. The Agent-based Dynamic Activity Planning and Travel Scheduling (ADAPTS) model, an American computational process model for the Chicago region, uses concurrent competing hazard models to determine whether to generate a new out-of-home activity of a specific type.

The activity scheduling component determines each activity's start time, duration, and location while considering all constraints. This component is the most researched of the three and often uses probability-based models to determine these factors.

The Canadian model for the Greater Toronto region, Travel Activity Scheduler for Household Agents (TASHA), generates activity agendas and their attributes based on empirical distributions. While the complex rule-based ABMs, A Learning-Based Transportation Oriented Simulation System (ALBATROSS) of the Netherlands uses decision heuristics. The DaySim ABM model, which uses a disaggregate dynamic network traffic assignment tool TRANSIMS router developed for Sacramento and Jacksonville, USA, is an application of the micro-behavioural model.

The mobility assignment component determines the mode choice and vehicle allocation based on the previously determined activity and scheduling choices. This component often uses heuristic, hazard-based, or micro-behavioural models. Agent-based models (ABMs) are also commonly used to represent human behaviour in ABMs, as they can produce complex travel behaviours using simple heuristics. Recent research on the integration of activity-based and agent-based models has shown that these models can forecast traffic in greater detail. For example, the MATSim (Multi-Agent Transport Simulation) model has been enhanced [3] to include network, zone, traffic counts, road pricing, trip length distribution, route assignment algorithms, and behaviour modification programs (flexible scheduling, ridesharing), and so on. MATSim, the stand-alone agent-based modelling framework, has also been shown to produce traffic flows that are more accurate than other models, such as Canadian EMME (multimodal equilibrium model).

Aside from MATSim and ALBATROSS, other notable integrations include FEATHERS, developed initially as a computational process model for Flanders, Belgium. The FEATHERS model has also been successfully applied in new contexts. Also included is TRANSIMS, a disaggregate dynamic network assignment tool integrated with multiple ABMs such as DaySim, ADAPTS, and others via sequential integration. SimMobility is another recent multiscale integrated agent-based simulation platform that includes a learning day-by-day module. POLARIS, which integrates dynamic simulation of travel demand, network supply, and network operations to solve the difficulty of integrating dynamic traffic assignment and disaggregate demand models, is also available. Gains in computational efficiency and performance enable previously distinct aspects of the urban system to be included in planning models.

While ABMs offer many benefits, including a more comprehensive understanding of travel behaviour and the ability to forecast traffic in greater detail, they also have limitations. For example, agent-based models have high computational complexity and require well-defined conditions and constraints. Additionally, there needs to be a streamlined process for calibrating and imputing model parameters, which can result in non-reproducibility and difficulties in understanding agent interactions with other agents and environmental parameters.

13.2.2 Overview of Sweden's National Transport Model

Sampers, the national transport model system developed by the Swedish Trans- port Administration (Trafikverket [4]), is a powerful tool for cross-modal passenger transport analysis. It forecasts future traffic volumes for different scenarios, simulations by varying infrastructure availability, and parameters such as GDP, fuel price, employment, population growth, to be contrasted in a socio-economic constraint. Sampers model is integrated with its socio-economic model (Samkalk) for comparison and investigative analysis, using its calculation module divided into matrix program, route analysis, effect models, and financial program.

The model comprises several travel forecasting modules, including the frequency choice model, destination choice model, mode of transportation choice model, and route choice model, linked in a nested logit model estimated using data from national travel habit surveys (RIKS-RVU/RES). A generic overview of the recently proposed Sampers in comparison to the model systems can be acquired from Fig. 13.1 below. The logit models compute probability distributions for how an individual or group of individuals chooses between diverse alternatives. Furthermore, the logit model computes log sums, which measure the combined benefit generated by the model's selectable options. The log of the denominator of this logit choice probability, the activity-based accessibility measure (logsum), gives the expected utility from a set of alternatives and links different choices, as in nested logit models. This metric is ideal for project evaluation because it expresses the consumer's benefits from all travel alternatives.

The route choice model comprises modules built into the EMME software toolkit and accounts for congestion as a factor that controls route choice in the algorithm

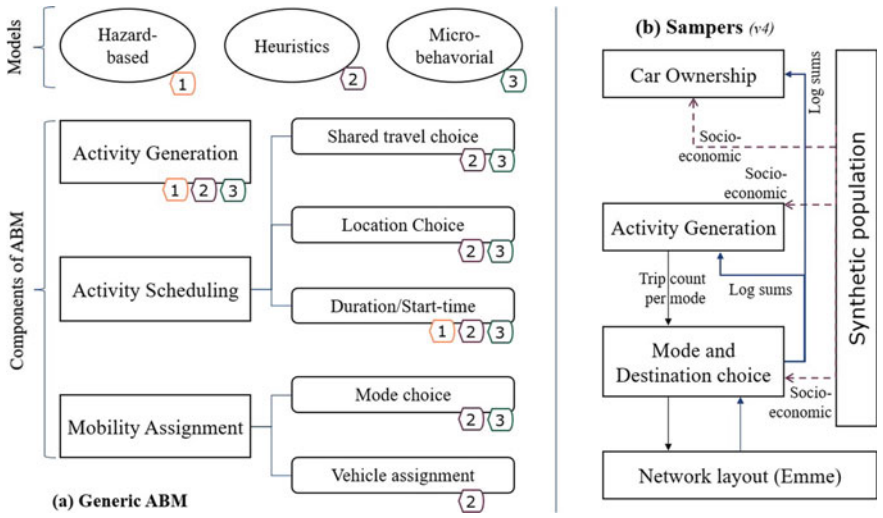


Fig. 13.1 Overview structure of model and sampers

for car traffic and is vital for the travel times obtained in route choice simulations. To achieve equilibrium between supply (travel times and costs) and demand, the car route choice model must iterate between the first three steps of the Sampers model (frequency, destination, and mode choice) and the route choice model.

The socio-economic module determines the socio-economic benefit of one investigation alternative over another, using data on travel by various modes of transportation, travel times, traffic volumes, and traffic revenue on various public transportation routes as input. The model used travel times, travel costs, greenhouse gas emissions, and accident values to calculate benefits.

Sampers is used primarily for impact analysis and investment calculations for traffic simulations, transport policy implementation evaluations, as well as accessibility and impact analysis of extensive changes in land use and transport systems in cities and regions [5]. It can also be used for micro-modelling scenarios that investigate intersections for minor variations in traffic control and travel times, meso-modelling constraints that investigate the effects of institutional conditions on traffic counts, interactions between transport systems and land use, and impact studies of IT-based or relevant scenarios on travel demand.

While the model has several sub-models, including a module for connecting journeys, it is used sparingly due to high uncertainty and limitations in handling travel between Sweden and abroad. Several methods for managing trips need to be calculated in the demand model, including static arrays and quota adjustments for existing travellers. Overall, Sampers is a valuable tool for transportation planning and analysis in Sweden, but certain innovations in the model are worth exploring to realise the limitations and further recommendations.

13.3 Sampers Innovation Projects

Transportation systems are continuously evolving with technological innovations such as electric vehicles, connected and collaborative intelligent transport systems, and autonomous vehicle technology, along with lifestyle-altering events such as the recent pandemic and sustainability issues related to climate change. As a response to these changes, several recent industries and university-led research projects analyse scenarios and work towards introducing relevant changes in the existing Sampers model. In this section, we will discuss some ongoing and completed projects that focus on improving Sampers for planning prospects in Sweden.

13.3.1 IHOP Project: Building Dynamic Network Models for Next-Generation Activity-Based Models

IHOP aimed to create a dynamic and disaggregate person/network simulation system. This project framework aims to develop a model for car traffic, to handle queues with back-blocking, the interaction between vehicles at junctions, weaving stretches, and other scenarios to make socioeconomic calculations more comparable and consistent. With the vision of a traffic system with connected and collaborative autonomous vehicles, the project evolved to develop an overall control model to optimize the behaviour of the traffic system. A natural choice for network deployment of demand matrices constructed from activity-based models was to integrate dynamic network deployment methods in Sampers.

IHOP2 [6] connected the Regent travel demand model and the TransModeler network assignment package via a new agent-based interface layer based on the MATSim transport simulation toolkit. The main goal of this effort was to demonstrate that such a coupling, with frequent status updates in the system between all vehicles and immediate updates of the route options, is possible. The interface layer could link the mesoscopic network model (TransModeler) to the activity-based demand model (Regent) by defining agents routed through the network. The agents could also communicate with each other in real-time, providing the ability to perform various collaborative maneuvers such as lane changing and car following. The results showed that the coupled models could accurately represent travel behaviour, with its interface between the demand model and mesoscopic, dynamic network layout, a shift from estimation to implementation in its forecasting system.

IHOP3 focused on ensuring an economically consistent analysis of the simulated travel behaviour in such a system. To achieve this, IHOP3 developed a simulation method for economically consistent integration of Sampers and MATSim. This resulted in the specification of a joint, fully dynamic, and person- centric utility function in both Sampers and MATSim. Given the different time and travel demand resolutions of the two models, different utility functions were used in Sampers and MATSim, resulting in different models of travel experience and different cost–benefit analysis results. However, the integrated model developed in IHOP3 was able to overcome these discrepancies and provide a more accurate and consistent estimation of travel behaviour in the network.

The IHOP project successfully improved the Sampers model by creating a more dynamic and disaggregate person/network simulation system. With the development of a simulation method for economically consistent integration of Sampers and MATSim, the project achieved a shift from estimation to implementation in forecasting systems.

13.3.2 Simulations for Long-Distance Trips: Vehicle Fleet Electrification

This research [7], initiated in 2018, aimed at reducing the carbon footprint based on technological advances in electric drivetrain and battery technology that allow them to travel long distances independent of public charging infrastructure during most suburban and commuting trips. This effort was part of a larger project to electrify the vehicle fleet in Sweden. The impact of large-scale vehicle electrification on long-distance trips was assessed by combining a Swedish agent-based long-distance transport model with a detailed energy consumption and battery charging model. This new approach developed a microscopic transport model using agent-based simulations to assess different system parameters' impact on total energy consumption.

The goal was to derive the charging infrastructure needs or the overall system cost for all electromobility-related technologies. By simulating different scenarios, the impact of different system configurations on energy consumption and charging infrastructure requirements were assessed, providing valuable insights into the potential benefits of vehicle electrification for long-distance travel.

13.3.3 Modelling and Analysis Module for EVs, ITS and Autonomous Vehicles

Significant research is being conducted on the future scenarios and impacts of self-driving vehicles in academic institutions and industry projects. Swedish research institutes are developing AV-induced transport system simulations [8] to better understand the impacts of autonomous vehicles on the country's transportation system. Meanwhile, other studies are focused on identifying research gaps and performing cost analysis modelling for AV freight operations, as well as introducing comprehensive impact assessment tools [9] to assess the overall impact of AVs on the transportation system.

One key aspect of this research is its connection to the Swedish transport model, Sampers, which is used to review AVs in relation to system assessment. This project conducts analyses to identify uncertainties and gaps in current modelling tools and the freight transport module and derive socioeconomic calculation results by identifying research needs for effective relationships. The project evaluates the calculation system's future development needs concerning behavioural, cost and emission changes to function in a future with a high proportion of connected, self-driving vehicle systems in the Swedish transport system. Time values, travel costs for passenger traffic, and transport costs for goods transport must all be estimated and updated.

Another vital consideration in this research is the need to develop appropriate algorithms for the possibility of using the exact automated vehicle for multiple modes of transport by members of the same family without the need for a driver or for the

vehicle to park itself where parking costs are reasonable. These efforts aim to identify the charging infrastructure needs and overall system costs for all electromobility-related technologies, as well as assess the impact of large-scale vehicle electrification on long-distance trips. Overall, this research aimed to understand better self-driving vehicles' impacts on Sweden's transportation system and develop effective resource allocation strategies to support the growth of this technology.

13.3.4 Synthetic Sweden Mobility (SySMo): Agent-Based Models

The Synthetic Sweden Mobility (SySMo) project at Chalmers is an ongoing research effort to address the shortfalls in the Swedish Transport model- Sampers. This comprehensive effort to create a new agent-based integrated model of Sampers envisioned a decision support system framework based on a combination of several computing tools and techniques in synthetic information systems and large-scale agent-based simulations.

The SySMo model employs a stochastic approach combined with Neural Networks, a machine learning technique to generate a synthetic population and behaviourally realistic daily activity-travel schedules for each agent. This new approach allows for creating a synthetic population representative of the actual population and can be used to generate activity-travel schedules for each agent. The current model provides a valuable planning and visualization tool for illustrating Swedish population mobility patterns [10].

Using SySMo, we can better evaluate mobility and transport scenarios in Sweden, which can help identify areas for improvement and inform policy decisions. The model can also be used to simulate different future scenarios, such as the adoption of AVs, and assess their potential impacts on the transportation system.

13.4 Limitations and Recommendations for Sampers

In this era of the IT revolution, new data resources for travel demand analysis must be analysed and validated for real-time transport modelling. The challenge remains in extracting accurate activity detail from big data and integrating them for travel demand models. Various algorithms- genetic, evolutionary, fuzzy, and techniques such as advanced text and data mining, natural language processing, and machine learning/neural networks- must be tested to evolve the models further.

Sampers has its limitations; for example, the impact of new infrastructure on buildings and housing is not modelled. Where people live and work is included as a given condition in the forecast, produced by Statistics Norway. Sampers cannot model travel chains in which individuals select modes of transportation based on the

activities to be performed during the day. If you drive to work, shopping trips are separate from work trips; instead, each trip is modelled separately.

Changed preferences and values cannot yet be modelled over Sampers. If, for example, people's attitudes towards driving change or cycling are perceived as more attractive due to increased health and environmental awareness in the future, this is not captured in the model. The model also does not capture remote working scenarios, which would reduce the need for work or business trips.

13.4.1 Long-Term Development Recommendations

Integration with emerging technologies: As emerging technologies such as connected and automated vehicles continue to evolve, it is essential to integrate them into Sampers to predict travel behaviour and demand accurately. This integration could involve developing agent-based models that simulate the behaviour of these vehicles and their interactions with other vehicles and the transportation system. Sampers can aim to integrate Land Use Transport Interaction (LUTI) models for policy implementation, such as the 'Transport Infrastructure Land-use Interaction Simulation model' (TIGRIS XL) developed since the 1990s, to better address the impact of new infrastructure on buildings and housing.

Incorporating spatial and temporal constraints: Investigating evolved spatiotemporal concepts like the Time–space prism modelling can capture the spatial and temporal constraints that people use to construct the patterns of their trips and activities. Integrating such modelling techniques into Sampers could improve the model's accuracy in predicting travel behaviour by capturing the constraints that affect travel choices.

Addressing equity concerns: Sampers could be further developed to address equity concerns in transportation planning. This could involve analyzing the model's outputs for fairness and ensuring that the model is inclusive of all demographic groups.

Improved data collection: Sampers could be updated to incorporate more detailed, accurate, and real-time data sources as technologies evolve. Research projects concerning integrating real-time mobile, GPS tracking, crowd-sourced, and social media data sources to understand travel behaviour better are ongoing.

Enhancing policy analysis capabilities: Sampers could be further developed to provide policymakers with a more detailed analysis of the potential impacts of policy decisions. This could involve adding a more detailed analysis of policy decisions' environmental, social, and economic impacts.

13.5 Conclusions

Sampers remains a critical tool for transportation planning and decision-making in Sweden. Its mechanisms are based on observable behaviour. Measures examined today for advancements in self-driving vehicles, information systems, and electric vehicles might have a lengthy development time, so the advantages of the measures are frequently computed over forty years. However, technological advancements have resulted in significant changes that might render present estimates exceedingly questionable, if not meaningless.

The Swedish Transport Administration must develop an approach for how evaluation methodology and forecasting tools should be developed to reasonably consider technological development in the transport planning process, considering both the opportunities and problems that development may entail. Furthermore, addressing its limitations and incorporating emerging technologies and modelling techniques can provide valuable insights for policymakers and ultimately lead to more sustainable and equitable transportation systems.

References

1. Tajaddini, A., Rose, G., Kockelman, K., Vu, H.: Recent progress in activity-based travel demand modeling: rising data and applicability. In: models and technologies for smart, sustainable and safe transportation systems. IntechOpen, (2020). https://doi.org/10.5772/intechopen.87681
2. Petersson, K., Engelson, L., Selling, E.: Sampers 3.4. 4 Användarhandledning. Trafikverket, (2020)
3. Delhoum, Y., Belaroussi, R., Dupin, F., Zargayouna, M.: Multi-agent activity-based simulation of a future neighborhood. In: Agents and Multi-agent systems: technologies and applications 2021. Springer, pp. 501–510 (2021). https://doi.org/10.1007/978-981-16-2994-5_42
4. Trafikverket.: Sampers. Accessed 10 Jan. 2022. https://bransch.trafikverket.se/tjanster/system-och-verktyg/Prognos-och-analysverktyg/Sampers/
5. Kristoffersson, I., Berglund, S., Samuelsson, S., Almström, P., Algers, S.: Sampers4: Skattning av regionala efterfr. (2018). http://urn.kb.se/resolve?urn=urn:nbn:se:trafikverket:diva-5502
6. Canella, O., Flotteröd, G., Johnsson, D., Kristoffersson, I., Larek, P., Thelin, J.: Flexible coupling of disaggregate travel demand models and network simulation packages ("IHOP2")
7. Marquez-Fernandez, F., Bischoff, J., Domingues–Olavarrıa, G., Alakula, M.: Assessment of future EV charging infrastructure scenarios for long-distance transport in Sweden. IEEE Trans. Transp. Electrification., 8, 615–626 (2021)
8. Pernest, A., Kristoffersson, I.: Effects of driverless vehicles: Comparing simulations to get a broader picture. Eur. J. Transp. Infrastruct. Research. 1–23 (2019)
9. Gao, K., Yang, Y., Qu, X.: Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. Transp. Res. Part D: Transp. Environment. 97, 102957 (2021)
10. Tozluoğlu, Ç., Dhamal, S., Yeh, S., Sprei, F., Liao, Y., Marathe, M., Barrett, C.L., Dubhashi, D.: A synthetic population of Sweden: datasets of agents, households, and activity-travel patterns. Data in Brief, **48**(2352–3409), 109–209 (2023)

Chapter 14
A Novel Longitudinal Control Strategy of Connected Automated Vehicle in Heterogeneous Traffic Flow and String Stability Analysis

Li Genze, Zhang Lanfang, and Wang Shuli

Abstract To cope with the randomness derived from the human driving in heterogeneous traffic consists of human-driving vehicles and connected automated vehicles (CAVS), a longitudinal car-following control strategy of CAV is proposed based on the original Intelligent Driver Model (IDM) model and model predictive control (MPC) structure. The string stability of heterogeneous platoon is verified by head-to-tail string stability criteria. Results indicate the strategy proposed can reflect the relationship between the speed and string stability and prove the adaptability to different traffic conditions.

14.1 Introduction

In recent years, the concept of automated vehicles (AV) has been gradually accepted and developed by the public since their first introduction at DARPA Grand Challenge [1]. However, many automated vehicles today only use on-board sensors to perceive the environment, which may not be robust to various driving conditions. In particular, as on-board sensors are only able to obtain current information about the immediate neighborhood, these automated vehicles often have difficulties anticipating the motion of surrounding vehicles reliably [2]. The introduction of connected automated vehicle (CAV) technology, along with substantial advances in sensing, computation, recognition, and control technologies, has enabled the development of AVs that can

L. Genze (✉) · Z. Lanfang
College of Transportation Engineering, Tongji University, Shanghai, China
e-mail: Ligenze123@tongji.edu.cn

W. Shuli
Department of Architecture and Civil Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden
e-mail: shuli@chalmers.se

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_14

accurately sense their local environment, detect and classify objects, interpret the changes in the surrounding driving environment of the system, and perform complex maneuvers, while ensuring safe and efficient operations [3].

The development of CAV from the application to the current traffic flow to the complete popularization will inevitably go through a complex research progress. Future road traffic will be composed of CAV and human-driving vehicles (HV), resulting in a heterogeneous traffic flow state. In order to avoid the uncertainty to the current traffic system resulting from the heterogeneous traffic flow, dedicated lane for CAV is believed as one of the potentially effective solutions to eliminate the driving pattern differences between CAVs and HVs. Nevertheless, setting CAV dedicated lane will reduce the number of lanes for accommodating other regular vehicles. If not set properly, it will lead to a great waste of road resource and cause severe congestion in the traffic flow, decreasing the overall throughput of the road. Especially when the market penetration rate of CAVs reaches a certain level, it is difficult to avoid the heterogeneity of driving behavior between CAVs and HVs only by allocating lanes. For the purpose of ensuring the efficient, stable and safe vehicle operation state and completing the basic car-following driving task, it is necessary to find a CAV longitudinal control strategy suitable for mixed traffic.

Compared with HVs, CAVs often perform corresponding driving tasks following relatively fixed behavior patterns without obvious volatility according to presented driving strategies. Taking car following as an example, in pursuit of safer and more efficient traffic flow state, CAVs often adopt more aggressive driving strategies, such as maintaining a shorter following headway, distance, etc. and choose shorter reaction time when following the preceding vehicle. Obviously, the deterministic control strategies cannot capture uncertainties and stochasticity that are inevitable in human car-following behavior. The difference between CAV and HV driving strategies may influence the driving behavior of CAV, and even interfere with the CAV's judgment and decision-making when the preceding vehicle encounters an emergency due to behavioral uncertainty to a certain extent. When the disturbance occurs, the longitudinal control strategy unsuitable for heterogeneous traffic flow will lead to the formation and development of oscillations. In this paper, we attempt to propose a CAV longitudinal control strategy suitable for the situation where the preceding car is HV considering the randomness of human driving behavior to keep the traffic flow stable. The strategy proposed adopts MPC as the control framework and calculates the reference trajectory according to the improved IDM model which is also proposed in this paper considering the preceding HV randomness. The strategy string stability is also verified to prove the adaptability to different traffic conditions for the sake of future application.

14.2 Literature

Numerous microscopic car-following models have been proposed to imitate driving pattern of manual driving system. These models provide us with structures to work on car-following strategy in mixed traffic environment and identify the extents of potential paradigm shifts.

Although, for a long time, it has generally been believed that the effect of stochastic factors is marginal in traffic dynamics, recent studies demonstrated that behavioral stochasticity is associated with a number of features observed in real world traffic, such as long oscillation periods and concave and stochastic growth patterns of oscillation magnitudes [4]. Because classical car-following models cannot well explain these features, a number of stochastic car-following models have been developed in the hope of characterizing the real vehicle operating state and accurately reflecting the propagation and evolution of traffic oscillation. For example, Kim T, Zhang HM [5] found traffic waves propagate in a stochastic manner and their speeds are determined by the relative differences between gap times and the reaction times of drivers. Wagner [6] proved that randomness of speed and time headway can be attributed to an internal stochasticity of the driver itself by model-based analysis. Treiber proposed a minimal general model simultaneously considering string instability, external white acceleration noise and indifference regions implemented by action points [7]. The PATH laboratory at the University of California, Berkeley has conducted long-term research on the ACC/CACC vehicle following model. The proposed model is simple and has been verified by real vehicle tests. In addition, some scholars have modeled ACC/CACC car-following behavior from the perspective of optimal control, with the methods including Proportional–Derivative (PD), Model Predictive Control (MPC), Artificial Intelligence (AI), etc. [8].

14.3 Control Design

A platoon with m + 1 vehicles, as shown in Fig. 14.1, is discussed. The main control objective is to regulate the subject vehicle's longitudinal motion to follow its predecessor with a constant time gap. The inter-vehicle spacing, the speed of the subject vehicle and the predecessor, and the acceleration of the subject vehicle are assumed to be detected by on-board sensors, including radar, speedometer, and accelerometer. The acceleration and control of the predecessor are collected by the V2V communication system. The communication topology is assumed to be predecessor-follower (PF) topology, which is a distributed communication structure that is more resilient to communication failure compared with centralized communication topology.

The longitudinal vehicle dynamics model, which is assumed to be the same for all the vehicles, is first formulated as follows:

$$\dot{a}_i = \frac{u_i - a_i}{\tau} \tag{14.1}$$

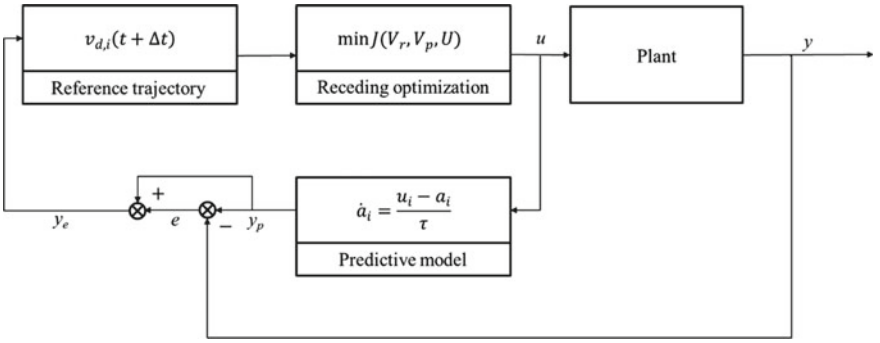


Fig. 14.1 Structure of SAIDM-MPC

where a_i is the acceleration of vehicle i, u_i to be interpreted as desired acceleration, and τ a time constant representing engine dynamic, set to be 0.1 s. In the remainder of this section, the MPC- and original CACC-based control strategies are introduced. The systems above are based on the same spacing policy.

14.3.1 Spacing Policy

In the widely used CAV with a constant time gap (CTG) policy [9], the prevailing definition of time gap denotes the time it takes the subject vehicle to approach its predecessor at the current speed. The desired spacing $d_{r,i}$ at time t for this CACC family is formulated as

$$d_{r,i}(t) = r_i + v_i(t) \times g \tag{14.2}$$

where i is the index of the subject vehicle, r is the standstill distance, v_i is the speed of the subject vehicle, and g is the desired time gap. The value of g is set to be 1.1 s for the purpose of safety, which is a suitable value for commercial auto-driving system. Under this formulation, the spacing error is

$$e_i(t) = d_i(t) - d_{r,i}(t) = p_{i-1}(t) - p_i(t) - v_i(t) \times g - L_i - r_i \tag{14.3}$$

where i-1 is the index of the preceding vehicle, d_i is actual spacing, p_{i-1} and p_i are longitudinal rear-bumper positions of the preceding vehicle and the subject vehicle, respectively, and L is the vehicle length.

The IDM takes the maximum acceleration a into account as well as the comfort deceleration b, where the spacing policy differs from the CTG, as shown below:

$$d_{r,i}(t) = r_i + v_i(t) \times g + \frac{v_i \Delta v}{2\sqrt{ab}} \tag{14.4}$$

where Δv is the velocity difference between two adjacent vehicles. The intelligence in the desired distance decision leads to the wide applicability in various traffic conditions of IDM.

14.3.2 Design of SAIDM-MPC Control Law

As mentioned above, IDM model determines the desired spacing from the preceding car based on the fixed time headway, which is consistent with the widely used spacing policy CTG. In addition, IDM is suitable for various traffic conditions due to the consideration of traffic flow and ergonomics. In the vehicle-connected environment, the acceleration of preceding car can be acquired in real time and becomes an indispensable input for the decision of desired acceleration. In consequence, a_{i-1} is introduced as the variable of original IDM, as shown below:

$$u_i = a\left[1 - \left(\frac{v_i}{v_0}\right)^{\delta} - \left(\frac{d_{r,i}(t)}{d_i}\right)^2\right] + \beta a_{i-1} \tag{14.5}$$

where v_0 is the free traffic flow speed and δ, β are the parameters needed to be calibrated. In heterogeneous traffic, the driving behavior stochasticity of HV preceding vehicle may disturb the following stability of CAV platoon, and even cause bottleneck-free congestion (ghost jam) under certain traffic volume probably. A generic stochastic equation can reflect the influence of stochasticity on the following speed v_i, including a deterministic part and a diffusion term:

$$dv_i = f(v_i, d_i, \Delta v_i)dt + g(v_i, d_i, \eta)dt \tag{14.6}$$

where f(.)f(.) denotes a deterministic car-following model as formula (8) in this paper while g(.)g(.) denotes the stochastic source which depends on the state of the considered vehicle and the stochastic process η. Field data of vehicle operation (NGSIM) demonstrate that the vehicle speed noise follows normal distribution. Winer process is adopted modeling the random deviations from the optimal speed of the vehicle. Moreover, to enhance the non-negativity of speed, the model below is proposed to model the CAV-following behavior in heterogeneous traffic, called as SAIDM:

$$dv_i = \left\{a\left[1 - \left(\frac{v_i}{v_0}\right)^{\delta} - \left(\frac{d_{r,i}(t)}{d_i}\right)^2\right] + \beta a_{i-1}\right\}dt + \sigma\sqrt{v_i}dW_n \tag{14.7}$$

where σ is the constant parameter describing the noise strength. A standard genetic algorithm is used for calibrating the parameters δ, β, σ. Trajectory data provided by NGSIM is selected for the calibration. The rest of the parameters in the model above is derived from the previous research and the specific control acquirement

Table 14.1 Parameters in SAIDM

Symbol	Description	Unit	Value
r_i	Standstill distance	M	1.5
G	Safety time headway	S	1.1
A	Max. acceleration	$m.s^{-2}$	1.4
B	Comfort acceleration	$m.s^{-2}$	2
v_0	Free traffic flow speed	$Km.h^{-1}$	120
δ	/	/	15.36
β	/	/	0.67
σ	Dissipation coefficient	$\sqrt{m}.s^{-1}$	0.88

(considering the security, comfort level, vehicle dynamic performance, etc.). Table 14.1 shows the parameters in SAIDM.

MPC scheme is adopted to achieve the control objectives, that is, keeping constant and safe time headway with preceding car and driving as fast as possible. MPC does not acquire that the vehicle dynamic model is highly accurate due to the error compensation, which simplifies the modeling and calculation process. By receding horizon control, MPC can control the car in real time to deal with complex traffic conditions. Owing to customizing the reference trajectory, the optimized path can be smooth to comfort the passengers working or relaxing in the car. According to the SAIDM, the desired speed of the vehicle following can be calculated for satisfying the control demand. To construct the reference trajectory, SAIDM is discretized using an explicit Euler–Maruyama scheme:

$$\mathrm{v}_{d,i}(t+\mathrm{v}t) = \mathrm{v}_i(t) + \left\{ a\left[1 - \left(\frac{\mathrm{v}_i(t)}{\mathrm{v}_0}\right)^{\delta} - \left(\frac{d_{r,i}(t)}{d_i(t)}\right)^2\right] + \beta a_{i-1}(t)\right\} \Delta t + \sigma\sqrt{\mathrm{v}_i(t)}\sqrt{\mathrm{v}t}\,\omega_n \tag{14.8}$$

where Δt represents the time step, $\omega_n \sim N(0, 1)$. The desired speed $v_{d,i}(t + k\Delta t)$ every time step k during the predictive horizon is used to construct the reference trajectory. The desired acceleration (system output) $u_i(\mathrm{t})$ is the solution to the following optimization function:

$$\min J(V_r, V_p, U)$$

$$J\left(V_r, V_p, U\right) = \sum_{x=t}^{t+T} \beta_1 (v_{r,i}(x) - v_{p,i}(x))^2 + \beta_2 u_i^2(x) + \tag{14.9}$$

where $\mathrm{V_r V_r}$ is the desired speed sequence, V_p is the predictive speed sequence, equalling to $[v_{p,i}(t + \Delta t), v_{p,i}(t + 2\Delta t) \ldots v_{p,i}(t + T)]^T$, U is desired acceleration sequence, and $\beta_1, \beta_2, \beta_3$ are constant coefficients. The first term denotes the cost due

Table 14.2 Control parameters in SAIDM-MPC

Description	Unit	Value
Sample time	S	0.1
Prediction horizon	S	10
Control horizon	S	2
β_1	/	0.78
β_2	/	0.45
β_3	/	0.23

to deviation from the desired speed, the second term is the cost due to acceleration or deceleration. The last term represents a cost due to a short time gap with the preceding vehicle. Specifically, the function $P(.)$ denotes penalty when the time headway reduces from the desired value while following the preceding vehicle, and the function is defined as

$$P(.) = e^{-\alpha(g-g_i(x))} \tag{14.10}$$

where $g_i = (p_{i-1}(t) - p_i(t))/v_i$ is the actual time headway and α is a constant. Some constraints are defined considering the safety-comfort and rules-regulation relevant to a road network, such as

$$\begin{cases} -a_{min} \le u_i(x) \le a_{max} \\ p_{i-1}(x) - p_i(x) \ge r_i + v_i(x) \times g + \frac{v_i(x)\Delta v}{2\sqrt{ab}} \end{cases} \tag{14.11}$$

$$\text{For } x = t,\ \ t + \Delta t, \ldots,\ t + T - \Delta t$$

As the design of the vehicle dynamic model, reference trajectory, objective's function, and constraints, the structure of the SAIDM-MPC is shown in Fig. 14.1.

The parameters of the controller are derived through system tuning as shown in Table 14.2. To evaluate the performance of SAIDM-MPC, ordinary ACC and CACC are selected as comparative items and introduced then.

Ordinary ACC and CACC are selected as comparative items to reflect the string stability, respectively.

14.4 Analysis of String Stability

How a perturbation at the lead vehicle is initially amplified across the following upstream vehicles. Vehicle flow is string stable if speed fluctuations are smoothed when they propagate upstream. From the perspective of the heterogeneous vehicle, platoon consists of HDV and CAV, the definition of head-to-tail string stability in this

paper is based on transfer function theory. According to this definition, the HDV-CAV platoon is head-to-tail string stable if the speed fluctuations of the head can be attenuated by the tail CAV vehicle [10]. To achieve this, the following constraint should be satisfied:

$$|G(s)| = \left| \frac{\tilde{V}_{tail}(s)}{\tilde{V}_{head}(s)} \right| < 1 \tag{14.12}$$

where s is the Laplace operator, $\tilde{V}_{head}(s)$ denotes the Laplace transform of speed fluctuation of the head HDV, while $\tilde{V}_{tail}(s)$ represents the equivalent of the tail CAV vehicle, the $G(s)$ is the transfer function of speed fluctuations propagating through the mixed platoon, and $|\cdot|$ stands for the amplitude of transfer function in frequency domain. Equation (14.12) can guarantee speed and space stability simultaneously.

The speed fluctuation of the head HDV propagates upstream along the HDV-CAV platoon before it reaches the tail CAV. During this period, the transfer function of the speed fluctuation can be calculated using the car-following model mentioned above. For the HDV-CAV-following pairs, the ordinary CACC will degrade to the ACC as a result of the lack of wireless communication not considering the acceleration of preceding car. Taking the Laplace transform of position perturbation linearization with zero initial conditions, the transfer function $G_{H-C}(s)$ of speed fluctuations between lead HDV and CAV adjacent can be calculated as follows:

$$G_{H-C}(s) = \frac{\tilde{V}_i(s)}{\tilde{V}_{i-1}(s)} = \frac{f_i^d + f_i^{\Delta v} s}{s^2 + s\left(f_i^{\Delta v} - f_i^v\right) + f_i^d} \tag{14.13}$$

where $\tilde{V}_i(s)$ is the Laplace transform of speed fluctuation of the CAV, while $\tilde{V}_{i-1}(s)$ denotes that of its preceding lead HDV $_{\text{i-1}}$. Similarly, the transfer function $G_{C-C}(s)$ of speed fluctuations between adjacent CAVs can be derived as follows:

$$G_{C-C}(s) = \frac{\tilde{V}_i(s)}{\tilde{V}_{i-1}(s)} = \frac{f_i^d + f_i^{\Delta v} s + s^2 f_i^a}{s^2 + s\left(f_i^{\Delta v} - f_i^v\right) + f_i^d} \tag{14.14}$$

where $\tilde{V}_i(s)$, $\tilde{V}_{i-1}(s)$ represent the Laplace transform of speed fluctuation of the ego and preceding CAV, respectively. The transfer function $G(s)$ can be calculated according to Eq. (14.12) as shown below:

$$G(s) = G_{H-C}(s)\left[G_{C-C}(s)\right]^{m-1} = \frac{\left(f_i^d + f_i^{\Delta v} s\right)\left(f_i^d + f_i^{\Delta v} s + s^2 f_i^a\right)^{m-1}}{\left[s^2 + s\left(f_i^{\Delta v} - f_i^v\right) + f_i^d\right]^m} \tag{14.15}$$

As mentioned before, the HDV-CAV platoon can be head-to-tail string stable when it satisfies the constraint in Eq. (14.12). By substituting Eqs. (14.12)–(14.15), the generalized string criterion of the HDV-CAV platoon can be obtained as follows:

$$|G(j\omega)| = \left| \frac{\left(f_i^d + f_i^{\Delta v} j\omega\right)\left(f_i^d + f_i^{\Delta v} j\omega - \omega^2 f_i^a\right)^{m-1}}{\left[j\omega\left(f_i^{\Delta v} - f_i^v\right) + f_i^d - \omega^2\right]^m} \right| < 1 \tag{14.16}$$

where j is imaginary unit, ω is system frequency domain, and $s = j\omega(\omega \geq 0)$. Equation (14.16) is suitable for the string stability of the original CACC and ACC not considering the stochastic driving behavior fluctuation of lead HDV. When it comes to SAIDM, formula (14.7) is transformed by linearization as follows:

$$\begin{aligned} d\dot{y}_i(t) = &\left[f_i^d(y_{i-1}(t) - y_i(t)) + f_i^v \dot{y}_i(t) + f_i^{\Delta v}(\dot{y}_i(t) - \dot{y}_{i-1}(t)) + f_i^a \ddot{y}_{i-1}(t)\right]dt \\ &+ g_i^v \dot{y}_i(t) dW \end{aligned} \tag{14.17}$$

where $g_i(.) = \sigma\sqrt{v}$ derived from formula (14.10), while g_i^v denotes the partial derivatives of $g_i(.)$ with respect to v. Both sides of Eq. (14.17) are divided by dt:

$$\begin{aligned} \frac{\dot{y}_i(t)}{dt} = &f_i^d(y_{i-1}(t) - y_i(t)) + f_i^v \dot{y}_i(t) + f_i^{\Delta v}(\dot{y}_i(t) - \dot{y}_{i-1}(t)) + f_i^a \ddot{y}_{i-1}(t) \\ &+ g_i^v \dot{y}_i(t) \frac{dW_n}{dt} \end{aligned} \tag{14.18}$$

The Wiener process is not derivable. For computation simplification, the formal derivative of Wiener process dW_n/dt regarded as white Gaussian noise is used to calculate the transfer function:

$$\begin{aligned} \ddot{y}_i(t) = &f_i^d(y_{i-1}(t) - y_i(t)) + f_i^v \dot{y}_i(t) + f_i^{\Delta v}(\dot{y}_i(t) - \dot{y}_{i-1}(t)) + f_i^a \dot{y}_{i-1}(t) \\ &+ g_i^v \dot{y}_i(t) \varepsilon_n(t) \end{aligned} \tag{14.19}$$

where $\varepsilon_n(t)$ denotes the white Gaussian noise, obeying $N(0, \sigma_0^2)$. Similar to the $G(s)$ deduction method about the original ACC and CACC, the $G(s)$ about SAIDM can be shown as follows:

$$G(s) = \frac{\left(f_i^d + f_i^{\Delta v} s\right)\left(f_i^d + f_i^{\Delta v} s + s^2 f_i^a\right)^{m-1}}{\left[s^2 + s\left(f_i^{\Delta v} - f_i^v - g_i^v S_i(\omega)\right) + f_i^d\right]^m} \tag{14.20}$$

where $S_i(\omega)$ denotes the noise power spectral density, used for representing the characteristics of $\varepsilon_n(t)$ in the frequency domain. Known that $S_i(\omega)$ is equal to its variance when the mean of $\varepsilon_n(t)$ is 0, the generalized string criterion of heterogeneous platoon about SAIDM can be obtained as follows:

$$|G(j\omega)| = \left| \frac{\left(f_i^d + f_i^{\Delta v} j\omega\right)\left(f_i^d + f_i^{\Delta v} j\omega - \omega^2 f_i^a\right)^{m-1}}{\left[j\omega\left(f_i^{\Delta v} - f_i^v - \sigma_0^2 g_i^v\right) + f_i^d - \omega^2\right]^m} \right| < 1 \tag{14.21}$$

14.5 Results of String Stability

Based on the string criterion of original ACC, CACC, and SAIDM described before, respectively, Eq. (14.13), (14.14) and (14.15) are substituted into Eq. (25) (30) to calculate the stability region of related variables. Known from the equation above, the string stability of heterogeneous HDV-CAV platoon is determined as the equilibrium speed v and desired time gap g with the same platoon size. The stability region at different v ~ (0, 33 m.s^{-1}) and different g~(0, 5 s) can be calculated applying the software Python, as shown in Fig. 14.2.

In Fig. 14.2, different colors indicate whether the model satisfies the string stability criterion. Purple region represents the platoon can keep string stable at corresponding speed and desired time gap, while green region represents unstable conditions. The extent of the stable region does not vary much with different platoon sizes, which indicates that the model chosen can keep the platoon string stable within a certain platoon size scope. Since the string stability criterion of ACC and CACC do not reflect the relationship between the speed and string stability, the model application robustness cannot be guaranteed. As far as ACC and CACC are concerned, the platoon

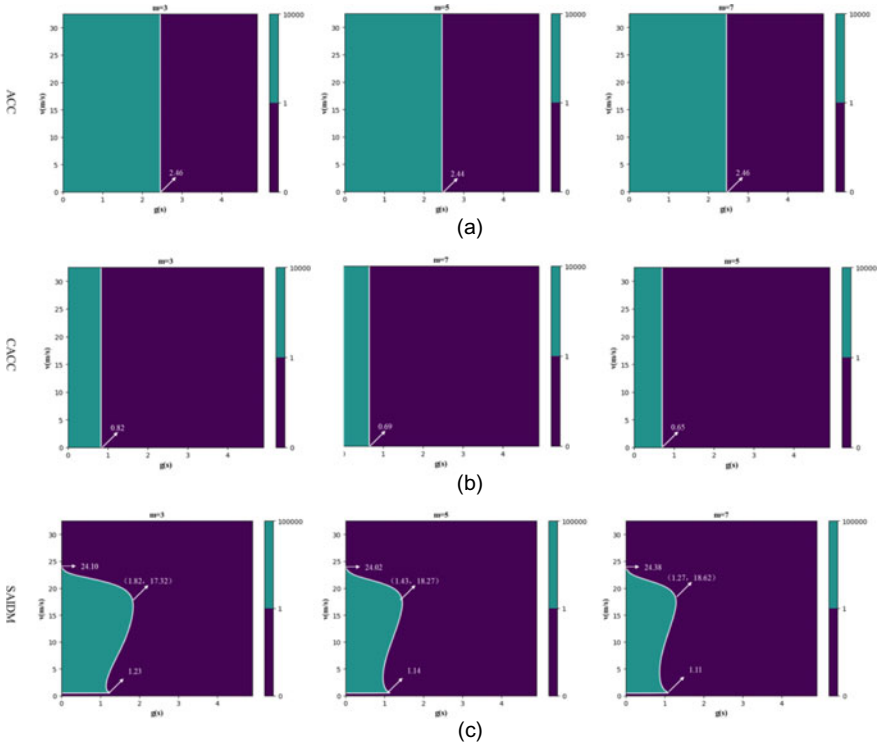


Fig. 14.2 String stability region of HDV-CAV platoon (**a**: ACC; **b**: CACC; **c**: SAIDM)

can keep string stable when g is greater than about 2.5 and 0.7 s, respectively. At short time headway, CACC outperforms ACC in terms of string stability, helping to ensure the traffic efficiency. For SAIDM, the string stable area decreases with the platoon size increasing, which indicates CAV formation driving is feasible within a certain platoon size range under the control law based on SAIDM. When the desired time gap is greater than 1.82 s, platoon can maintain stability at any value of speed in regular traffic speed range, which proves the adaptability of SAIDM to different traffic conditions. Also, string stability results support for the selection of following time headway in the CAV longitudinal control.

14.6 Conclusion and Future Work

(1) Due to the wireless communication, the acceleration of preceding car introduced to the traditional IDM model, as well as the speed fluctuation derived from the human driving randomness, helps to the following-car's speed and acceleration decision.
(2) The stability analysis (local and string stability) proved that the SAIDM-MPC can guarantee the driving stability within normal traffic speed range and safe time headway (above the TTC threshold), reflecting the capacity to cope with emergencies happened and dissipate the fluctuation in traffic.

For further research, traffic stability should be verified and the performance of SAIDM-MPC in terms of driving safety and traffic efficiency should be tested by simulation.

References

1. Thrun, S., Montemerlo, M., et al.: Stanley: The robot that won the DARPA Grand Challenge. J. Field. Robot. **23**(9), 661–692 (2006)
2. Baek, M., Mun, J., et al.: Driving environment perception based on the fusion of vehicular wireless communications and automotive remote sensors. Sensors **21**(5), 1228–1244 (2021)
3. Yang, H., Zhao, J., Wang, M.: Optimal control of automated left-turn platoon at contraflow left-turn lane intersections. J. Int. Con. Vehicles **5**(3), 206–214 (2022)
4. Wang, Y.J., Li, X.P., et al.: Stability analysis of stochastic linear car-following models. Transp. Sci. **54**(1), 274–291 (2020)
5. Kim, T., Zhang, H.M.: Interrelations of Reaction Time, Driver Sensitivity, and Time Headway in Congested Traffic. Transp. Res. Rec. **2249**, 52–61 (2011)
6. Wagner, P.: Analyzing fluctuations in car-following. Trans. Res. Part B: Method. **46**(10), 1384–1392 (2013)
7. Treiber, M., Kesting, A.: The intelligent driver model with stochasticity: new insights into traffic flow oscillations. Trans. Res. Part B: Method. **117**, 617–627 (2019)
8. Atam, E., Helsen, L.: Ground-coupled heat pumps: Part 1-Literature review and research challenges in modeling and optimal control. Renew. Sustain. Energy Rev. **54**, 1653–1667 (2016)

9. Jerison, D.C., Levine, L.: Mixing Time And Eigenvalues Of The Abelian Sandpile Markov Chain. Trans. Am. Math. Soc. **372**(12), 8307–8345 (2019)
10. Zhang, T., Zou, Y., et al.: A cruise control method for connected vehicle systems considering side vehicles merging behaviour. IEEE ACCESS **7**, 6922–6936 (2019)
11. Zhu, J., Easa, S., Gao, K.: Merging control strategies of connected and autonomous vehicles at freeway on-ramps: a comprehensive review. J. Int Conn Veh **5**(2), 99–111 (2022)

Chapter 15
Dynamical Classification to Improve the Selection of the Driver-Cargo Transportation Duo for a Trucking Company

Ivanhoe Arias, Saul Lazcano, L. Dávila-Nicanor, and Maricela Quintana

Abstract The transfer of merchandise and products has played a fundamental role during this time of the pandemic, many countries have a high percentage of land transport routes by road and roads dedicated to the delivery of products. It is a great concern to establish the best conditions for human nature and cargo transportation to avoid accidents.

The design of a system through a dynamic classifier has been developed to establish the best selection in the driver-cargo transportation duo, with an analytical approach based on Bayes' Theorem that is supported by a software architecture through design patterns at runtime from the perspective of object-oriented, to optimize memory resources and improve the efficiency of the classification process.

The conditions evaluated and the selected variables are the product of an exhaustive analysis between related works and archival data evidence from a local company.

I. Arias (✉) · S. Lazcano · L. Dávila-Nicanor · M. Quintana
Centro Universitario UAEM Valle de México, Laboratorio de calidad de Software. Boulevard Universitario s/n, Predio San Javier, Atizapán de Zaragoza, Estado de México, México
e-mail: iariass002@alumno.uaemex.mx

S. Lazcano
e-mail: slazcanos@uaemex.mx

L. Dávila-Nicanor
e-mail: ldavilan@uaemex.mx

M. Quintana
e-mail: mquintanal@uaemex.mx

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_15

15.1 Introduction

Traffic accidents represent a serious problem that entails various consequences ranging from economic losses to injuries and deaths. According to the Technical Secretariat of the National Council for Accident Prevention (STCONAPRA), in Mexico, there is an average of 16,500 annual loss of lives and economic losses of more than 150 billion pesos, associated with accidents transit, which represents around 1.7% of the Gross Domestic Product (GDP) [1]. In the years 1997 and 2020, there have been 17,792,838 traffic accidents involving pedestrians, cyclists, and vehicles in general, of which 2,574,516 accidents are associated with cargo transportation (pickup trucks, trucks, and tractor-trailers) [2]. The Ministry of Infrastructure, Communications and Transport (SICT) has among its functions the prevention of accidents in federal public transport, through various tools such as training courses, information campaigns and the issuance of Official Mexican Standards (NOM) focused on verifying various elements, such as the adequate mechanical condition of the vehicles, the state of physical and mental health of drivers, to name a few [3]. Analyzing the main causes that cause accidents in cargo transportation, INEGI groups them into three major factors, Analyzing the main causes that cause accidents in cargo transportation, INEGI groups them into three major factors: accidents due to the driver 80%, accidents due to external agents (weather, road conditions) 13% and accidents due to mechanical problems of the vehicle 7% [2].

External factors such as road conditions, adverse weather conditions and lack of proper road signage, among others, are characterized by being highly changeable and uncontrollable. However, the human and mechanical factors can be subject to strict controls that allow, at a given moment, to reduce the rate of road accidents.

In this context, the objective of the present work is the design of a classification system that allows to improve the way in which the transport companies carry out the assignment of the driver and the cargo transportation to attend a particular service, taking as starting points the driver health and the physical and mechanical conditions of cargo transportation, looking for the most suitable duo to achieve a reduction in the rate of accidents in cargo transportation vehicles. The data of a trucking company (that requests anonymity) is taken as a case study, for the construction of the system.

This work is organized as follows: Sect. 15.2 describes the related work, in Sect. 15.3 an analysis of the data provided by the local company is carried out to accurately determine the variables to be considered for the characterization and development of the classification system. Section 15.4 presents the architecture proposed for the design of the classifier and finally in Sect. 15.5 the results and conclusions are shown, as well as future work.

15.2 Related Work

In improving the allocation of trips in cargo transportation, according to [4], the evaluation and prediction of human behavior conditions in normal and extreme conditions are adequate, developing a functional psychodynamic method to study the decision-making mechanisms of transport workers. From this perspective, research on accidents in cargo transportation according to [5], points to the human factor as the main responsible for causing accidents. Among the causes, is the lack of reliability of the driver in the combination of the Human Factor—Transport—Environmental System is indicated. In the analysis of the factors that intervene in the driver's driving behavior [6], proposes a fatigue monitoring questionnaire, conducting surveys of experts, where researchers predict positive effects in monitoring fatigue, such as the reduction of accidents, but do not deny the possible dangers due to the adaptation of the behavior of drivers, stating that it is particularly important to develop a positive attitude towards driving without fatigue, to improve the individual responsibility of drivers. In the work [7], a proposal was made to classify the selection of the driver using automatic analysis of the curriculum vitae (CV) through the WEKA (Waikato Environment for Knowledge Analysis) software. The created system allows receiving the trained WEKA classifier and applying it to analyze the new data in an automatic mode. This proposal makes it possible to give an initial estimate of each record in the incoming data and to establish elements to a parameterized selection of the human being driver analysis by [8] considering driver distraction, fatigue, the driver's degree of familiarity with the road, and the presence of alcohol in the blood, as the most significant factors that contribute to road transport accidents, taking into account, the type of vehicle, its speed, to predict the probability of road accidents on the road in curved sections. From another perspective [9], A route planning model was developed with the minimum distance of each, the objective is to obtain the most optimal for each trip, considering the travel distance of the vehicle, features of the road, such as the longitudinal slope, the flat curve, the uniformity of the pavement and the damage of the pavement to prevent risks on the road. Alonso et al. [10], propose an expert system capable of taking decisions to avoid collisions in urban traffic or minimize damage, due to inattention or some other unexpected situation, said rule-based expert system is capable of detecting a potentially dangerous situation and acting accordingly on the vehicle brake, considering the maximum legal speed in urban traffic conditions, the braking decelerations of the vehicle behind and the vehicle in front, the transverse and longitudinal forces of the inclination of the road, speed and distance.

According to data shown in Table 15.1, the most used variables for this type of analysis are: fatigue or tiredness, age (in drivers), condition or symptom, speed, size, and load (in vehicles). For the classification of adequate options between drivers and available vehicles, the design of a system with a dynamic classifier has been proposed. In this case, a classification function based on Bayes' Theorem is established, supported by a dynamic evaluation of objects through design patterns at runtime. Thus, an architecture is established that at runtime can generate objects to

Table 15.1 Variables by approaches used in related works

Work related	Approach (Transportation and human behaviors)	Variables used
Zhang et al. [11]	• Transport route • Physical characteristics of the transport • Driver efficiency, (Reaction times)	• Transport distance and load • Degree of mechanical modernization • Traffic information publication delay time and employee control • Command level coefficient
Gholamhosseinian & Seitz [12]	• Physical–mechanical conditions of transport • Physical characteristics of the transport • Transport security • Road conditions	• Physical and kinematic characteristics of the vehicles • Speed, Acceleration/deceleration • Weight and wheelbase • Classes, quantity, and shape of the vehicle, that is to say: height, width and length
Marinov et al. [4]	• Psychophysical health conditions of the driver	• Personal traits and conscience of a person; intellect, character, way of thinking, abilities, motives, and various behavioral reactions • Discipline, perseverance, will, interest, morality, and reactions speed in dangerous situations
Singla et al. [13]	• Psychophysical health conditions of the driver	• Functionality of the Nephron (anatomical and physiological unit of the kidney) • Diastolic and systolic blood pressure • Body mass index, age, and blood sugar • Smoke

evaluate an unlimited set of combinations to locate the best proposal. This scheme optimizes memory and time resources, improving efficiency in locating the better options.

15.3 Characterization of Human Behavior and Mechanical Transport Status Variables in the Local Environment

To estimate the values in terms of probability, we have worked with source documents from transport companies that have supported the investigation, under the restriction of anonymity. The local Company provided an information bank that it has managed for two years and eight months, where the record of a total of 46,490 transport trips made was kept. In this information, the statisticians estimate that 12% of the trips suffered accidents, according to the SICT, INEGI, the related work and the practice of real data, the causes that are being considered are divided into three categories: Variables due to causes that were attributed to the driver, Variables that were attributed

Table 15.2 Percentage of accidents based on the nature of the cause in database

Factor	Percentage (%)
Accidents due to the driver	77
Accidents due to external agents (weather, road conditions)	16
Accidents due to mechanical problems of the vehicle	7

Table 15.3 Causes of accident attributed to the driver

Variables	Percentage (%)
Excessive speed	42.20
Invaded lane	16.03
Imprudence	12.08
I don't keep distance	9.54
Did not give way	6.58
Wrong turn	4.86
Dozing	3.78
Improper passing	1.92
Drunk state	1.09
Badly parked	0.95
No respect stops sign and traffic light	0.72
Glare	0.18
Under the influence of drugs	0.07

to the vehicle and Variables that were attributed to the road. Table 15.2 shows the percentage of accidents according to the main variables considered. According to the related work, they are: transportation [8, 11, 12], Driver [4, 7, 13] and Road [9, 12]. It is possible to observe in Table 15.3, that a high percentage corresponds to the driver.

Based on the information from the aforementioned company, the variables that affected the occurrence of accidents have been analyzed. According to Table 15.3, a percentage is observed that is related to the driver, in this case, Excessive speed, Invaded Lane and Imprudence are the variables with the highest value, with a limit of 10%. In this same way, the physical part of the vehicle and the road were analyzed.

Table 15.4 shows the variables attributed to the vehicle that could cause an accident. The most frequent variables are: Tires, mechanical failures and Brakes.

In relation to the analysis of the Road in Table 15.5 we have the highest weighting for the variables: slippery road, wet road.

To obtain the risk function, Weka's a priori association algorithm [7] is used, and the variables to be considered (Tables 15.3 and 15.4) are selected taking into account the related works [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Table 15.6 shows the relationships found in the information analyzed.

Table 15.4 Causes of accident attributed to the vehicle

Variables	Percentage (%)
Tires	47.82
Mechanical failure	24.27
Brakes	11.89
Electrical fault	5.58
Poorly secured cargo, overweight and excess dimensions	3.89
Direction	2.43
Suspension	1.94
Axis	1.21
Engine and transmission	0.97

Table 15.5 Causes of accident attributed to the route

Variables	Percentage (%)
Wet	47.09
Slippery	35.35
Objects in the Path	5.93
Irruption of cattle	5.6
Damage	3.73
Lack of signs	2.3

Table 15.6 Associations between the analyzed variables

Accident association		Percentage (%)
Driver and road	Excessive speed, Wet	36
Driver	Imprudence, Invaded Lane	20
Driver and vehicle	Excessive speed, Tires	9
Driver and vehicle	Invaded lane, Imprudence, Tires	9
Driver	Excessive speed, did not give way	8
Driver and vehicle	Imprudence, Invaded Lane, Tires, I don't keep distance, Mechanical failure	4
Driver and road	Invaded lane, Objects in the way	3
Driver and road	Excessive speed, Lack of signs	2
Driver	Invaded lane, Dozing	2
Driver and vehicle	Mechanical failure, Dozing	2
Driver	Imprudence, Excessive speed	2
Driver and road	Dozing, Objects in the way	2
Driver and vehicle	Invaded lane, Mechanical failure, Suspension	1

Table 15.7 Relation of variables in the database and those found in the related work

Main variables	Database variables	Related work variables
Driver	Excessive speed	Speed [8, 12, 14]
	Invaded lane	Motion [14]
	Imprudence	Fatigue [6, 8]
Vehicle	Tires	Adhesion coefficient [6]
	Mechanical failure	Physical and kinematic characteristics of the vehicles [12, 15]
	Brakes	No data
Road	Slippery road	No data
	Wet road	No data

In Table 15.7, the relation of main variable, variables found in the database and related work are shown. Between them are: speed, fatigue, and physical and kinematic characteristics of the vehicles.

15.4 Proposed Architecture Software System

The main goal of the proposed system is to design a classification scheme at runtime from the perspective of the object-oriented paradigm. The medical information coming from the driver's health diagnosis and the mechanical inspection of the vehicles are used to generate an object set that, through the rules of Bayes' Theorem, analyzes and selects the better driver-transport pair. The information registered by the driver is according to local norm NOM-087-SCT-2-2017 [1], in which driving times and breaks are established during the trip, and it is loaded through the system interfaces. As shown in the use case diagram in Fig. 15.1, five user views have been determined: the doctor, the driver, the logistics manager, the mechanic, and the knowledge system. The driver fills out his hours-of-service log, which is part of the doctor's information to diagnose the driver's physical health. The mechanic's view will evaluate the physical–mechanical conditions of the motor transport by NOM-068-SCT-2–2014 [16]. The view that corresponds to the logistics manager can plan trips, according to the suggestions or options that the knowledge system presents. Additionally, it will also be possible to monitor, and control scheduled trips. The knowledge system will have a classification function for the selection of the better options, this is in the driver-transport duo relationship supported by the Wrapper design pattern. This will allow the classification and evaluation at runtime, with the only limitation being the size of the memory where the system is executed.

The model-view-controller (MVC) design pattern has been considered in Fig. 15.2 in the design of software architecture. The inputs to the Model are the information of the medical diagnosis and the technical opinion of the transport that exists in the database. Setting the probabilities of the related variables (excessive speed, invaded

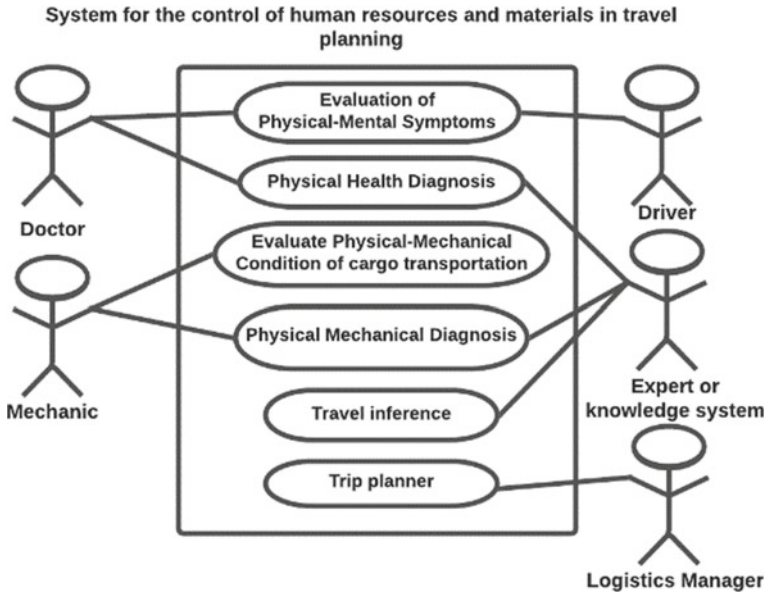


Fig. 15.1 Use case of the proposed system

lane; imprudence, tires, mechanical failure), networks of class nodes are projected, and the classification function operates according to Bayes' Theorem, generating dynamic objects and an evaluation at runtime. The Wrapper design pattern generates as many class nodes as the number of possibilities that exists, according to the information in the database. The Controller establishes the communication between the View and the Model, in this way the information from the service log, the medical diagnosis, the technical opinion, and the data of the scheduled trips and new trips are sent to the model to update the information in the database. The scheduler displays the best options based on what the knowledge system estimates. Through the sight, the interfaces that will be presented are the form that corresponds to the service hours log that will be filled out by the driver, the forms that correspond to the medical diagnosis, and the technical opinion, filled out by the doctor and the mechanic. In the view of the logistics manager, the classification with the best options for new trips and scheduled trips is presented.

Figure 15.3 shows some of the system interfaces, in the first view (a) the login interface is presented, and the access data is requested. According to each view, the system will present the corresponding set of options; in the case of the driver, the service log will be filled in, which is presented in the second interface of the table (b). The third user interface (c) corresponds to the mechanical diagnosis and in the fourth view (d), the scheduler trip is presented to the logistics administrator so that he can select the trips according to the recommendations that the knowledge system provides.

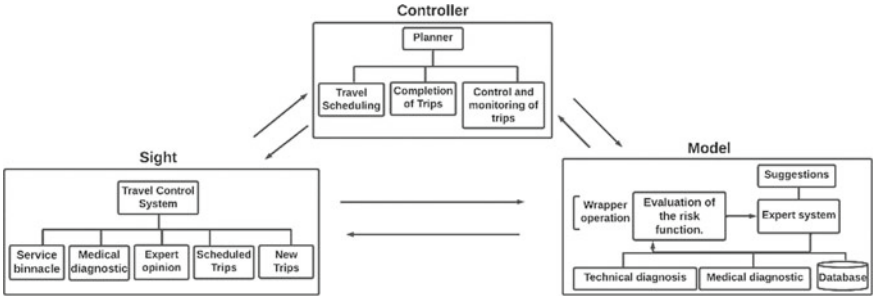
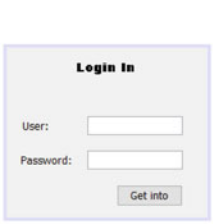
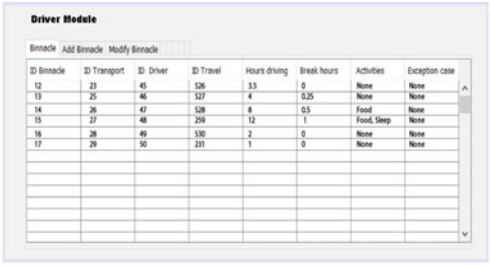


Fig. 15.2 Model-View-Controller Architecture of the Proposed System

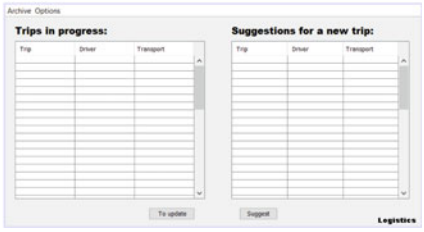


a) Login interface



b) The service log interface

c) Mechanical diagnosis interface



d) Recommendations that the knowledge system provides

Fig. 15.3 System interfaces

15.5 Conclusions and Future Work

In the process of allocating resources to cargo trips, there is a combination of factors that can become a risk of accidents. According to the studies carried out in this paper, the variables involved obey two fundamental aspects: conditions of human nature and cargo transportation conditions. In the development of this work, the design of a system through a dynamic classifier has been proposed, with an analytical approach based on Bayes' Theorem that is supported by a software architecture through design patterns at runtime from the perspective of object-oriented. This

scheme operates with the MVC and Wrapper design patterns, the latter allows optimization of memory resources, having as its only limitation the characteristics of the memory size and the speed of the processor where it is executed. In this way, the classification function analyzes viable conditions between available drivers and vehicles, thus establishing a software architecture that at runtime generates objects that evaluate a set of combinations in the selection of the best driver-cargo transportation duo. According to the studies and the results obtained, it is estimated that as future work it is appropriate to include a third aspect of conditions related to the road, such as lack of signs, wet pavement or floor, objects, or damages on the road, to enrich the classification scheme.

The proposed system aims to reduce the margin of risk in the assignment of the duo driver and truck. Part of the future work will consist of measuring the accident rate that the company has once the system is implemented, in order to quantify in a practical way, the reduction in risk levels and its total economic impact.

Acknowledgements This work was supported by the Quality on Software Laboratory of University Center UAEM Valley of México of the Autonomous University of Mexico State.

This work was partially supported by the PNPC-CONACyT 2021-23, Master´s thesis on Master of Degree Postgraduate of the Autonomous University of México State.

References

1. Diario Oficial de la Federación NOM-087-SCT-2–2017, Retrieved October 2021, from: https://www.dof.gob.mx/nota_detalle.php?codigo=5529381&fecha=28/06/2018
2. INEGI. (06 de Mayo de 2022). Land traffic accidents in urban and suburban areas. Retrieved August 2021, from INEGI: https://www.inegi.org.mx/sistemas/olap/proyectos/bd/continuas/transporte/accidentes.asp
3. Gobierno de México. (16 de septiembre de 2020). Secretaría de Infraestructura, Comunicaciones y Transportes. Retrieved August 2021, from http://www.sct.gob.mx/transporte-y-medicina-preventiva/medicina-preventiva/quienes-somos/mision-y-vision/
4. Marinov, M., Asaul, A., Skorokhodov, D., Malygina, E.: Development of methods for professional and axiological certification of transport experts' activities. Transp Res Procedia **50**, 436–443 (2020). https://doi.org/10.1016/j.trpro.2020.10.052
5. Bęczkowska, S. A., Grabarek, I.: The importance o0f the human factor in safety for the transport of dangerous goods. Int. J. Environ. Res. Public Health 18(14) (2021). https://doi.org/10.3390/ijerph18147525
6. Karrer, K., Rötting, M.: Early evaluation of driver fatigue monitoring systems in expert surveys (2007)
7. Osipovs, P.: Classification tree applying for automated CV filtering in transport company. Proc. Comput. Sci. **149**, 406–414 (2019). https://doi.org/10.1016/j.procs.2019.01.155
8. Cheng, G., Cheng, R., Pei, Y., & Xu, L.: Probability of roadside accidents for curved sections on highways. Math. Probl. Eng. (2020). https://doi.org/10.1155/2020/9656434
9. Ren, C., Wang, X., Gao, G., & Li, J.: Urban regional logistics distribution path planning considering road characteristics. Discret. Dyn. Nat. Soc. (2020).https://doi.org/10.1155/2020/2413459
10. Alonso, L., Oria, J. P., Fernández, M., & Rodríguez, C.: Car crash prevention expert system in urban traffic based on ultrasounds (n.d.)

11. Zhang, G. S., Shen, X. Y., Hua, J., Zhao, J. W., & Liu, H. X.: System dynamics modelling for dynamic emergency response to accidents involving transport of dangerous goods by road. J. Adv. Transp. (2021).https://doi.org/10.1155/2021/2474784
12. Gholamhosseinian, A., Seitz, J.: Vehicle classification in intelligent transport systems: an overview, methods and software perspective. IEEE Open J. Intell. Transp. Syst. **2**, 173–194 (2021). https://doi.org/10.1109/OJITS.2021.3096756
13. Singla, J., Kaur, B., Prashar, D., Jha, S., Joshi, G. P., Park, K., Tariq, U., & Seo, C.: A novel fuzzy logic-based medical expert system for diagnosis of chronic kidney disease. Mob. Inf. Syst. (2020).https://doi.org/10.1155/2020/8887627
14. Albusac, J., Vallejo, D., Castro-, J.J., Gzlez-Morcillo, C.: An ex-pert fuzzy system for improving safety on pedestrian crossings by means of visual feedback. Control. Eng. Pract. **75**, 38–54 (2018). https://doi.org/10.1016/j.conengprac.2018.03.008
15. Shafi, U., Safi, A., Shahid, A. R., Ziauddin, S., & Saleem, M. Q.: Vehicle remote health monitoring and prognostic maintenance system. J. Adv. Transp. (2018).https://doi.org/10.1155/2018/8061514
16. Diario Oficial de la Federación NOM-068-SCT-2-2014, Retrieved October 2021, from: http://www.dof.gob.mx/nota_detalle.php?codigo=5378850&fecha=19/01/2015

Chapter 16
Analysis of Driver Navigation Software Use Experience Based on Structural Equation Model

Wenhua Xu, Wenyi Wang, and Weiwei Qi

Abstract To explore the key factors affecting drivers' experience of using navigation software, 306 valid questionnaires about basic personal attributes, perceptual experience, interactive experience, and emotional experience were collected. Applying R language to construct and analyze the structural equation model (SEM) of drivers' navigation software use experience, which includes 5 latent variables and 21 observed variables. The final model was obtained by eliminating the correlations with insignificant significance levels based on the initial model, which was constructed by combining SEM and user experience (UX). The model demonstrated the relationships and corresponding parameters among the variables characterizing the navigation software drivers' experience of using the software. The results showed that the factors can be divided into 5 categories: prompt acceptability, reaction operability, perception of attention, perception of recognition and perception of use. Perception of recognition, reaction operability and prompt acceptability positively influenced drivers' perception of use. Prompt acceptability positively influenced the perception of recognition and reaction operability. Perception of attention positively influenced perception of recognition, reaction operability, and prompt acceptability. The model not only helped to explain the influencing factors of using navigation software, but also helped to suggest voice prompting requirements, which can further improve drivers' navigation software use experience.

W. Xu
Guangzhou Northring Intelligent Transport Technology Co., Ltd, Guangzhou 510030, China

W. Wang · W. Qi (✉)
School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China
e-mail: ctwwqi@scut.edu.cn

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_16

16.1 Introduction

In recent years, navigation software has developed rapidly as an important tool for route navigation and travel planning. Generally, navigation improves the efficiency of users in unfamiliar environments, and it also helps in solving road congestion. However, using navigation systems results in some negative effects on driving behavior to some extent while driving vehicles [1–4]. Some scholars have done research on the effects of navigation software on driving behavior through real-vehicle road tests [5] and simulated driving tests [2, 6]. Bian and Yang et al. studied the effect of prompt wording of navigation on drivers' driving efficiency and stability [7] and the effect on drivers' overall behavior [8].

Several scholars have studied navigation software from the perspective of cognitive load. Michael et al. found that increased cognitive task loading decreased warning compliance behavior [9]. Wang et al. argued that in-vehicle information systems, including navigation, mostly occupy four resource channels: visual, auditory, motor and cognitive, causing distraction [10]. Sun et al. proposed a load model formed by three load factors: time occupancy, level of information processing, and task setting transformation [11]. Some scholars have also focused on the design aspects of navigational voice prompts. Li et al. pointed out that navigation voice prompts should be simple and clear [12]. And in the study of the organization of information at the speech interface, Tan et al. stated that the information should not exceed the amount that a person can remember in a short term, with a maximum of 5–7 items [13]. This told us that the content of the voice prompts and the amount of information prompts contained may have some influence on the drivers. David et al. found that drivers held different attitudes towards each navigation voice, including their preferences and trust in their use, which further affected driving behavior [14].

There are many studies adopting the technology acceptance model (TAM) [15] to explore the user acceptance of particular systems, including navigation products. On in-vehicle navigation systems, some studies have revealed the effects of user perceptions, attitudes and usage intentions of navigation systems [16–18]. On mobile navigation applications, Yang explored the factors that affect drivers' usage intention based on the TAM integrated by adding three new constructs (drivers' sense of direction, navigation application affinity and distraction perception) [19].

However, there are few studies on the design elements of navigation software and their impact from the perspective of drivers' subjective experience of using navigation software. In this paper, an evaluation model of drivers' navigation software use experience was constructed in order to clarify the design elements of it.

16.2 Structural Equation Modeling of Driving Navigation Software Use Experience Evaluation

16.2.1 User Experience

User experience (UX) is all about how users feel before, during and after using a product or system, originating from the user's internal and physical state, which comes from the user's previous experiences, attitudes, skills, abilities, personality and the context of use [20]. Park evaluated UX with home appliances based on the usability, affect, and user value principles [21]. Zhu et al. explored the UX of geriatric medical APP by establishing a structural equation model between design elements, instinct, behavior, reflection level and overall user satisfaction [22]. This study took navigation software as the object, combined its characteristics to propose UX evaluation elements, and constructed indicators for analysis.

16.2.2 Basic Structure of the Structural Equation

Invoking the structural equation model (SEM), a validation model, many of the influencing factors studied in this paper are related to driver psychology (latent variables) that cannot be directly measured by explicit indicators and require explicit variables to analyze the relationship of them. It allows for the existence of multiple dependent variables and accepts errors in both the independent and dependent variables.

The SEM is formulated as follows.

$$\eta = B\eta + \Gamma_1\xi + \Gamma_2 I + \zeta \tag{16.1}$$

In the above equation: B: The matrix of coefficients between the endogenous latent variables. η: Endogenous latent variables. ξ: Exogenous latent variables. I: Dominant variables. Γ_1: Matrix of coefficients of exogenous latent variables on endogenous latent variables. Γ_2: Matrix of coefficients of explicit variables on endogenous latent variables. ζ: Residual vectors.

16.2.3 Structural Equation Model Construction

Based on UX and SEM, an evaluation of products can be established so that suggestions can be proposed to enhance product satisfaction. For example, Yang proposed to construct a SEM to evaluate the purchase intention of Alipay financial products using product utility, cognitive experience, emotional experience, and association experience [23]. Hu et al. proposed to analyze the SEM of the elder's satisfaction for intelligent healthcare products using six factors such as usefulness, usability, and

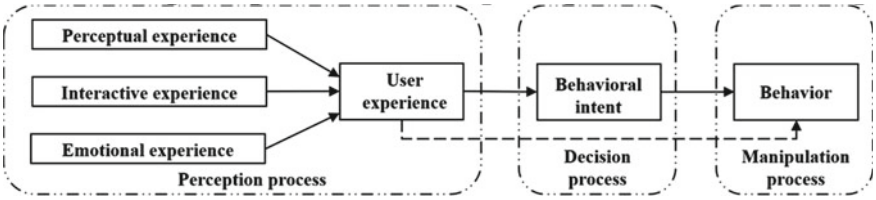


Fig. 16.1 Theoretical structure of the model

accessibility [24]. Many scholars have introduced SEM into the analysis of product application design factors. Similarly, this paper aimed to resort to UX and SEM to identify design factors and their importance in navigation software. Moreover, this paper adopted the driver "perception-decision-manipulation" behavior model proposed by Feng et al. [25] to construct the theoretical structure as in Fig. 16.1. It can be seen that the experience of navigation software affects driving behavior, then improving experience ratings can facilitate decisions about driving behavior. The UX of navigation software is influenced by its own design factors, so the driver's experience was explored from the perspective of design elements. The design factors were explored in terms of the audio-visual perception, reaction perception and emotion perception.

Based on the theoretical model, this paper hypothesized the influencing factors of navigation perception process and the relationships between them. The influencing factors are all psychological factors, which are difficult to measure directly. For this reason, we designed explicit variables for each potential factor. Perception of recognition (RG): Focused on driver's attitudes toward navigation software. Reaction operability (RA): Driver's reaction to navigation prompts and reaction space perceptions, observing by their subjective perception of the prompt timing, space and content. Prompt acceptability (AC): Driver's acceptance of the perceptions of prompt content, voice and speed. Perception of attention (AT): Driver's perception of the ability of the navigation prompts to draw his or her attention. Perception of use (UP): Driver's overall perception of using the voice navigation software. The latent and explicit variables are shown below in Fig. 16.2.

16.2.4 Driver Navigation Software Usage Experience Model

Based on the proposed theoretical model and each latent and explicit variable, a SEM of driver navigation software use experience was established, as Fig. 16.3. Hypothesis proposed, H1: RG positively affects UP. H2:RA positively affects UP. H3: RA positively affects RG. H4: AC positively affects UP. H5: AC positively affects UP. H6: AC positively affects RA. H7: AT positively affects RG. H8: AT positively affects AC. H9: AT positively affects RA.

Latent variable	Observed variable
Perception of recognition (RG)	Navigation software is a good aid. (RG1) Navigation software is beneficial to driving. (RG2) Navigation software is reassuring. (RG3) I trust navigation. (RG4)
Reaction operability (RA)	Navigation software gives me enough time to react and operate. (RA1) Navigation software does not prompt too late. (RA2) The frequency of navigation software prompts is not too high. (RA3) The prompt timing is well integrated with the road traffic signs. (RA4) I quickly understand the prompts every time. (RA5) I am able to react quickly to each navigation prompt. (RA6)
Prompt acceptability (AC)	Navigation software prompts is concise. (AC1) The amount of information in the prompts is satisfied. (AC2) Default voice of the navigation software sounds comfortable. (AC3) The speed of the prompts is not too slow. (AC4) The speed of the prompts is not too fast. (AC5)
Perception of attention (AT)	Default voice of the prompts can easily draw my attention. (AT1) The prompts voice still easily draws my attention after listening for a long time. (AT2) I am never surprised by the voice prompts for navigation. (AT3)
Perception of use (UP)	I have no problem driving to unfamiliar destinations with the navigation.(UP1) I have no problem driving on the highway with the navigation software. (UP2) I have no problem using the navigation software continuously for an hour.(UP3)

Fig. 16.2 Correspondence and numbering of latent and explicit variables

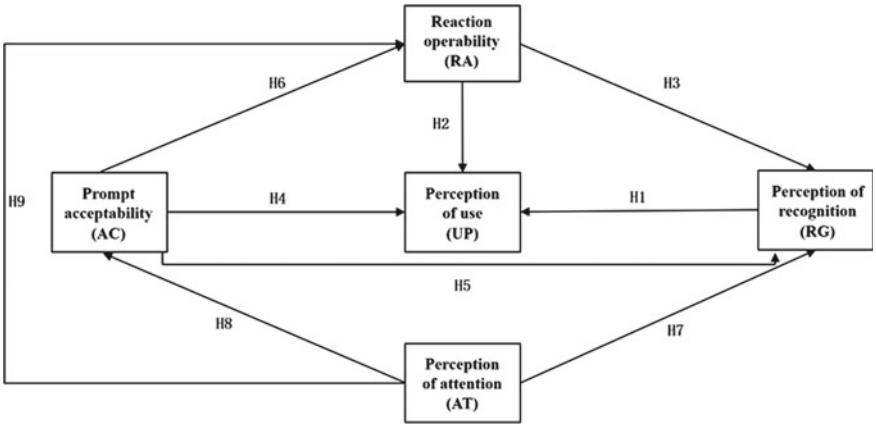


Fig. 16.3 Model hypothesis of driver navigation software use experience

16.3 Questionnaire Survey

16.3.1 Survey Time and Number of Questionnaires

The survey focused on drivers who had used navigation software. There are 329 questionnaires collected, and after checking the logic of the answers, 306 questionnaires are valid, with the effective rate of 93.01%. Among the personal attributes of the drivers surveyed in the questionnaire, there is a relatively even gender with 51.31%

Table 16.1 Latent variables reliability test results

Latent variable	Observed variable	Cronbach's alpha	
		Each	Overall
Perception of recognition (RG)	RG1, RG2, RG3, RG4	0.847	0.935
Reaction operability (RA)	RA1, RA2, RA3, RA4, RA5, RA6	0.897	
Prompt acceptability (AC)	AC1, AC2, AC3, AC4, AC5	0.884	
Perception of attention (AT)	AT1, AT2, AT3	0.803	
Perception of use (UP)	UP1, UP2, UP3	0.873	

being male. The respondents who "occasionally use", "often use" and "always use" navigation occupy 86.93%. Therefore, the data of questionnaire has its significance for the follow-up study.

16.3.2 Reliability and Validity Analysis of the Data

16.3.2.1 Reliability Analysis of Latent Variables

The reliability analysis aimed to measure the reliability of the scale data, and was judged using the Cronbach's alpha. Reliability test results are shown in Table 16.1. The overall reliability coefficient of 0.935 (>0.9) and each variable's reliability higher than 0.8 indicate that the quality of the study data reliability is very high. Data can be used for further analysis.

16.3.2.2 Latent Variable Validity Analysis

Validity analysis is primarily a measure of the soundness of quantitative data through factor analysis or exploratory factor analysis. KMO is often used as a measure of validity and the closer the value is to 1, the more suitable it is for factor analysis. Validity analysis results are shown in Table 16.2. The overall KMO value of 0.89 close to 0.9 indicates high validity and is well-suited for factor analysis. Although the KOM of UP do not reach 0.7, it is acceptable to continue the analysis. And the remaining latent variables with KMO over 0.7 indicates that it is well-suited for factor analysis.

Table 16.2 KMO value and Bartlett spherical test of evaluation scale

Latent variable	Observed variable	KMO	Bartlett's Test of Sphericity		
			chi–squared approximation	*df*	*p*-Value
RG	RG1, RG2, RG3, RG4	0.78	568.896	6	0.000
RA	RA1, RA2, RA3, RA4, RA5, RA6	0.85	1184.273	15	0.000
AC	AC1, AC2, AC3, AC4, AC5	0.79	946.162	10	0.000
AT	AT1, AT2, AT3	0.72	313.433	3	0.000
UP	UP1, UP2, UP3	0.67	537.747	3	0.000

Table 16.3 Fit test indicators of initial structural equation model

Indicator	Adaptation value	Value	Adapted or not
χ^2/df	<3	2.98	Yes
GFI	>0.9	0.87	No
CFI	>0.9	0.93	Yes
NFI	>0.9	0.90	Yes
NNFI/TLI	>0.9	0.91	Yes
IFI	>0.9	0.93	Yes
RMSEA	<0.08	0.08	Yes

16.4 Model Validation Analysis

16.4.1 Model Assumptions

In this paper, nine hypotheses were proposed as described in 2.4. Apply R language to analyze the relationship between each latent and observed variable, calculate and test the proposed model. The results of the model fit test are shown in Table 16.3. 6 of the 7 indicators meet the criteria, and the rest, GFI, gets close to 0.9. Integrated considering, the model was acceptable and can be further revised.

16.4.2 Exclusion of Correlations with Insignificant Significance Levels

The initial model path coefficients passed the test with a significance level of <0.05 for all factors except for the relationship between RA and RG, and AT and RG. And

the coefficients of path relationships between latent variables and between latent and explicit variables were appropriate. When correcting, deleting one path relationship also affects the others, so they should be deleted one by one. The path relations of RA and RG with worse correlation were chosen to be deleted first.

The results of the modified model and its fitness indicators were obtained, and compared with the initial model, the change of fitness was not obvious, the χ^2/df decreased by 0.02, and the change of other indicators was less than 0.01. The modified model was still basically fit, though the GFI (0.87) failed to meet the standard (0.9). Although the Modification Index (MI) of several paths between observed variables was higher than 10, this paper did not consider adding paths according to MI values to correct the model, considering that there was an unexplained part among the observed variables themselves. The p-values of paths are all <0.05 which mean passed the test. (as Table 16.4.) Therefore, output the standardized results and shown in Fig. 16.4.

16.4.3 Analysis of Model Results

(1) **Relationship between RG and UP**

The standardized path coefficient is 0.24, meaning that the higher the driver's recognition of navigation software, the more satisfied he or she is with the use of it.

H1 holds, suggesting that the RG positively affects UP.

(2) **Relationship between reaction operability RA, UP and RG**

The standardized path coefficient between RA and UP is 0.29, indicating that the more satisfactory is RA, the more satisfactory is UP. The unstandardized path coefficient between RA and RG is only 0.09, which is poorly correlated, so the path between them is removed.

H2 holds, suggesting that RA positively affects UP. H3 does not hold.

(3) **The relationship between AC and RA, UP, and RG**

The standardized path coefficients between AC, RA, UP, and RG are 0.30, 0.21, and 0.33, respectively, which shows that the higher the AC satisfaction, the more satisfied the driver is with the RA, UP and RG.

H4 holds, suggesting that AC positively affects UP. H5 holds, suggesting that AC positively affects RG. H6 holds, suggesting that AC positively affects RA.

(4) **Relationship between AT and RG, RA, and AC**

The standardized path coefficients between AT and RG, RA, and AC are 0.26, 0.61, and 0.77, respectively. The more satisfied a driver is with AT, the more satisfied he or she is with RG, RA and AC.

H7 holds, suggesting that AT positively affects RG. H8 holds, suggesting that AT positively affects RA. H9 holds, suggesting that AT positively affects AC.

Table 16.4 Modified model unstandardized path coefficients and significance levels

Path	Estimate	S.E	z-value	P	Whether to pass the test
UP ← RG	0.289	0.075	3.881	0.000	Yes
UP ← RA	0.339	0.107	3.164	0.002	Yes
UP ← AC	0.228	0.106	2.159	0.031	Yes
RG ← AC	0.291	0.093	3.128	0.002	Yes
RG ← AT	0.213	0.085	2.519	0.012	Yes
RA ← AC	0.276	0.078	3.526	0.000	Yes
RA ← AT	0.522	0.082	6.327	0.000	Yes
AC ← AT	0.705	0.066	10.729	0.000	Yes
RG1 ← RG	1.000	–	–	–	Yes
RG2 ← RG	1.471	0.109	13.541	0.000	Yes
RG3 ← RG	1.401	0.106	13.180	0.000	Yes
RG4 ← RG	0.999	0.102	9.772	0.000	Yes
UP1 ← UP	1.000	–	–	–	Yes
UP2 ← UP	1.141	0.079	14.417	0.000	Yes
UP3 ← UP	1.229	0.083	14.749	0.000	Yes
RA1 ← RA	1.000	–	–	–	Yes
RA2 ← RA	1.396	0.114	12.278	0.000	Yes
RA3 ← RA	1.505	0.136	11.075	0.000	Yes
RA4 ← RA	1.460	0.130	11.196	0.000	Yes
RA5 ← RA	1.627	0.154	10.542	0.000	Yes
RA6 ← RA	1.591	0.146	10.865	0.000	Yes
AC1 ← AC	1.000	–	–	–	Yes
AC2 ← AC	1.058	0.056	18.800	0.000	Yes
AC3 ← AC	0.839	0.078	10.743	0.000	Yes
AC4 ← AC	0.854	0.075	11.416	0.000	Yes
AC5 ← AC	0.997	0.084	11.922	0.000	Yes
AT1 ← AT	1.000	–	–	–	Yes
AT2 ← AT	1.429	0.109	13.124	0.000	Yes
AT3 ← AT	1.378	0.105	13.151	0.000	Yes
UP1 ↔ UP3	−0.071	0.014	−4.921	0.000	Yes
UP1 ↔ AC2	0.021	0.007	2.885	0.004	Yes
UP1 ↔ RA1	0.052	0.013	3.980	0.000	Yes
RG4 ↔ UP1	0.061	0.014	4.399	0.000	Yes
RA1 ↔ RA2	0.066	0.019	3.551	0.000	Yes
RA5 ↔ RA6	0.231	0.029	8.002	0.000	Yes
RA5 ↔ AT2	0.061	0.019	3.153	0.002	Yes

(continued)

Table 16.4 (continued)

Path	Estimate	S.E	z-value	P	Whether to pass the test
AC1 ↔ AC2	0.040	0.012	3.310	0.001	Yes
RG1 ↔ AC2	−0.030	0.008	−3.608	0.000	Yes
AC3 ↔ AT1	0.111	0.014	7.831	0.000	Yes
AC3 ↔ AT3	0.087	0.017	4965	0.000	Yes
AC4 ↔ AC5	0.111	0.015	7.496	0.000	Yes

16.4.4 The Total Effect of Each Factor on Navigation Software Usage Experience

The direct effect and indirect effect between each influencing factor were further obtained, and the total effect was obtained cumulatively. The results of the effect of each variable in the above SEM of drivers' navigation software use experience are shown in the following Table 16.5.

16.5 Conclusion

To study the factors influencing the satisfaction of drivers' navigation software use experience, we established a SEM combining with UX, and clarified the relationship between drivers' prompt acceptability (AC), reaction operability (RA), perception of attention (AT), perception of recognition (RG) and perception of use (UP). RG, RA and AC positively affect drivers' UP. AC positively influences RG and RA. AT positively influences RG, RA and AC.

16.6 Discussion

The findings of the model can be applied to the optimisation of voice prompt design elements in navigation software to improve the driver's navigation software experience, which can have a positive impact on driver behavior. And it is a preliminary study of the SEM of navigation software use experience, and multi-group perspective research can be carried out to make a remedy for the deficiency in further studies.

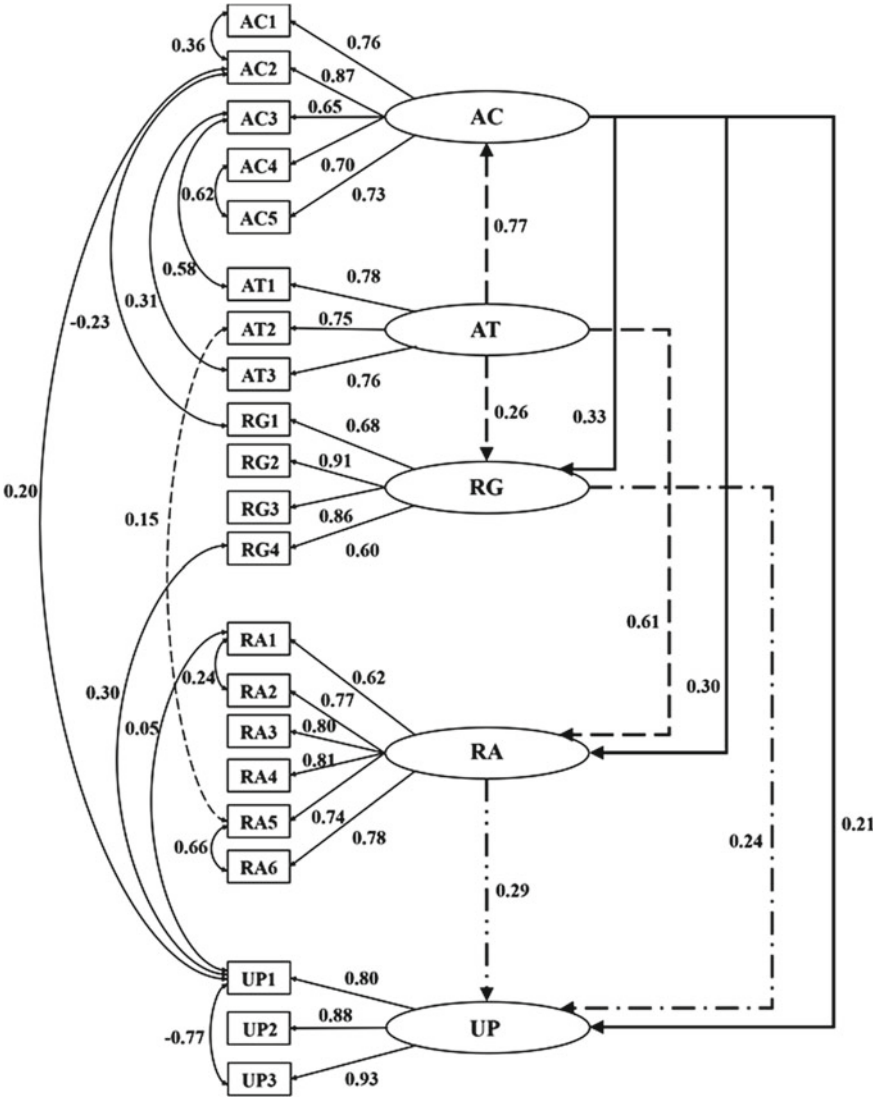


Fig. 16.4 Model's standardized correction model (Explicit variables error term not shown)

Table 16.5 Total effect of factors on drivers' experience of using navigation software

Variables	Direct effect	Indirect effect	Total effect
Perception of recognition	0.216	0.000	0.216
Reaction operability	0.129	0.000	0.129
Prompt acceptability	0.200	0.133	0.333
Perception of attention	0.000	0.394	0.394

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 52072131), the Key Research Projects of Universities in Guangdong Province (No. 2019KZDXM009), the Natural Science Foundation of Guangdong Province (No. 2023A1515010039).

References

1. Yared, T., Patterson, P., All, E.S.A.: Are safety and performance affected by navigation system display size, environmental illumination, and gender when driving in both urban and rural areas? Accid. Anal. Prev. **142**, 105585 (2020)
2. Xue, Q., Gao, K., Xing, Y., Lu, J., Qu, X.: A context-aware framework for risky driving behavior evaluation based on trajectory data. IEEE Intell. Transp. Syst. Mag. **15**(1), 70–83 (2023)
3. Hamish, A., Merat, N.: Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. Transport. Res. F: Traffic Psychol. Behav. **8**(2), 79–96 (2005)
4. Patrick, S., Mustapha, M., Nicol, J., et al.: The effects of modern navigation on driver distraction. Proceedings of the Human Factors and Ergonomics Society Annual Meeting **48**(18), 2022–2026 (2004)
5. Zhu, T., Wu, L., Lu, Q.: The effect of vehicle information on driver blink characteristics and load patterns. J. Southwest Jiaotong Univ. **50**(03), 504–510 (2015)
6. Tian, W.: Research on the influence of mobile phone navigation on driving behavior based on simulated driving. Tsinghua University, (2016)
7. Yang, L., Bian, Y., Zhao, X., et al.: The influence of the complexity of the wording of navigation announcements on driving behaviour. J. South China Univ. Technol. (Natural Science Edition) **49**(03), 139–148 (2021)
8. Bian, Y., Yu, J., Zhao, X., et al.: A comprehensive study of the influence of navigation announcement wording on driving behavior. J. South China Univ. Technol. (Natural Science Edition) **48**(11), 30–37 (2020)
9. Michael, S.W., Mary, O.U.: Effects of concurrent cognitive task loading on warning compliance behavior. Proceedings of the Human Factors and Ergonomics Society Annual Meeting **43**(6), 525–529 (1999)
10. Wang, Y., Zhang, W., Wu, S.: A review of research on in-vehicle information systems and driver safety. Sci. Technol. Rev. **27**(13), 105–110 (2009)
11. Sun, T., Xu, Y.: Theories and key models of cognitive load. Psychol. Res. **5**(02), 93–96 (2012)
12. Li, R., Liu, Y.: Application and problems of voice prompts in vehicle navigation electronic maps. Electron Technol. **20**(05), 79–83 (2013)

13. Tan, K., Tang, Y., Lv, Y.: A study of the information architecture of speech interfaces. China Science and Technology Information **11**, 118–120 (2012)
14. Large, R.D., Burnett, E.G.: The effect of different navigation voices on trust and attention while using in-vehicle navigation systems. J. Safety Res. **49**, 69–75 (2014)
15. Davis, F.D.: User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int. J. Man Mach. Stud. **38**(3), 475–487 (1993)
16. Park, E., Kim, K.J.: Driver acceptance of car navigation systems: integration of locational accuracy, processing speed, and service and display quality with technology acceptance model. Pers. Ubiquit. Comput. **18**(3), 503–513 (2014)
17. Chen, C., Chen, P.: Applying the TAM to travelers' usage intentions of GPS devices. Expert Syst. Appl. **38**(5), 6217–6221 (2011)
18. Xu, C., Wang, W., Chen, J., et al.: Analyzing travelers' intention to accept travel information. Transp. Res. Rec.: J. Transp. Res. Board **2156**(1), 93–100 (2010)
19. Yang, L., Bian, Y., Zhao, X., et al.: Drivers' acceptance of mobile navigation applications: an extended technology acceptance model considering drivers' sense of direction, navigation application affinity and distraction perception. Int. J. Hum Comput Stud. **145**, 102507 (2021)
20. Yu, N., Huang, Y.: Important factors affecting user experience design and satisfaction of a mobile health App-A case study of daily Yoga App. Int. J. Environ. Res. Public Health **17**(19), 6967 (2020)
21. Park, J., Han, H.S., Park, J., et al.: Development of a web-based user experience evaluation system for home appliances. Int. J. Ind. Ergon. **67**, 216–228 (2018)
22. Gao, K., Yang, Y., Qu, X.: Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. Transp. Res. Part D: Transp. Environ. **97**, 102957 (2021)
23. Yang, H. M.: A study on the influence of user experience on Harbin citizens' willingness to purchase Alipay financial products. Harbin University of Commerce (2022)
24. Hu, S., Jia, Q., Zhang, L., et al.: Intelligent escort product design for wellness based on structural equation modelling. J. Mach. Des. **38**(7), 110–117 (2021)
25. Feng, S., Hwang, Q., Zhang, Y., et al.: "Perception-Decision-Manipulation" behavioural model for drivers. J. Transp. Syst. Eng. Inf. Technol. **21**(1), 41–47 (2021)

Chapter 17
Design of a Highway Traffic Safety Assessment System Based on Crash Data Mining and Modeling

Luo Li, Shuolei Qin, and Weiwei Qi

Abstract With the advent of the era of traffic information technology, traffic data appears to be extremely large, and the traditional way of managing highway crashes has become difficult to adapt to the many implications of the complicated data. In this paper, we design a highway traffic safety assessment system based on crash data mining and modeling. By constructing a visualized highway traffic crash data analysis and mining platform, which is used as a basis to gather traffic crash analysis solutions and build a think tank, a safety assessment system is formed that integrates traffic crash information visualization, crash data analysis and mining technology and comprehensive solutions for crash risk prediction. The system provides crash data pre-processing solutions, crash pattern mining solutions, factor analysis solutions and crash risk prediction solutions. Based on the above solution, the basic framework and seven main functional modules of the highway traffic safety assessment system are designed and implemented. The system achieves systematic and automated data processing, pattern mining, factor analysis and risk prediction, and aims to make highway traffic safety management more efficient.

17.1 Introduction

In the field of big data in transportation, there is an interactive development relationship between big data and transportation models. Due to the multi-source and multi-dimensional character of big data, information fusion technology is indispensable for applications in traffic system optimization, integrated assessment, data-driven modeling and traffic safety management. Existing highway management systems are able to capture not only traffic flow data, but also specific traffic crash data.

L. Li
Guangzhou Northring Intelligent Transport Technology Co., Ltd., Guangzhou 510030, China

S. Qin · W. Qi (✉)
School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China
e-mail: ctwwqi@scut.edu.cn

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_17

There are many experts and scholars who have researched and analyzed this data and dedicated themselves to exploring crash factors. However, a systematic traffic safety impact assessment platform has still not been well developed. Therefore, how to develop a highway traffic safety assessment platform based on crash data mining and modelling from the spatial and temporal distribution patterns and influencing factors of highway traffic crashes is the research focus of this paper.

In many studies, various models have been used to model the factors influencing crashes based on historical traffic crash data to enable safety analysis. Among them, the most used are logit models and probit models. For example, Milton et al. [1] used a random coefficient logit model for modeling the distribution of injuries caused by road traffic crashes for crash data in Washington State. Satoshi et al. [2] used an ordered probit model to investigate the risk of serious crashes in winter conditions.

The accuracy of predictions of crash consequences is a critical issue for decision makers. However, statistical modeling techniques have inherent assumptions that are often difficult to hold on complex crash severity data [3, 4]. A consequence of using these methods is that many statistical models of road crash severity fail to achieve predictions with greater accuracy when tested using samples outside of the training set [5]. In addition, problems associated with the data, such as missing values, overdispersion, and heterogeneity, may reduce the credibility of the model assumptions [4]. According to several studies, machine learning techniques tend to show better predictive performance than statistical methods [6]. However, there are still many data quality issues associated with traffic crash databases. The problem of unbalanced data categories is particularly prominent [7, 8].

In most parts of the world, serious crashes are relatively rare in comparison to minor crashes [9]. The imbalance in the data therefore poses a considerable challenge to the statistical methods used for data analysis [8, 10]. And often the minority category in unbalanced data is the one that is more important and of greater interest than the majority category. The class imbalance problem can be overcome in two directions, namely at the algorithmic level and at the data level [11, 12]. At the data level, association rule mining, an important data mining method, focuses on finding association rules implied in a dataset. The earliest concepts related to association rule mining were introduced in 1993 by R. Agrawal, Swami and Imielinskih et al. Among them, the Apriori algorithm is one of the most classical algorithms in the field of association rules. There are many studies to improve and optimize the Apriori algorithm [13].

Bayesian networks (BNs) can perform probabilistic inference on uncertainty problems and are a model for studying classification learning problems. Zong [14] showed that the Bayesian network model had a better fit than regression when performing the crash severity analysis. Yahaya et al. [15] applied three data balancing methods comparatively for the imbalance of crash data in Ghana. And based on this, different Bayesian network models were investigated to explore the factors affecting fatal injury crashes. Song et al. [16] developed a Bayesian network model based on traffic crash data from Kunshan. The model test results showed that the model performed well in terms of accuracy and AUC metrics. Alkheder et al. [17] used support

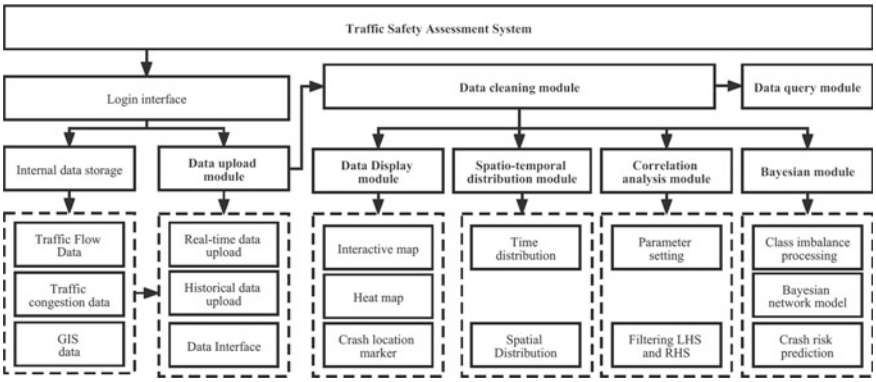


Fig. 17.1 System functional structure diagram

vector machines, decision trees, and Bayesian network models to conduct a comparative analysis of the factors influencing crash severity. In terms of performance, the Bayesian network model had the highest prediction accuracy.

17.2 System Functional and Architecture Design

17.2.1 System Architecture Design

The overall functionality of the Highway Traffic Safety Assessment System is implemented by first acquiring raw highway crash data through manual uploads or application programming interfaces. The raw data enters the data cleaning module and is cleaned into display data and training data. The display data is transferred to the data presentation module, the data query module and the spatio-temporal distribution module, where it is made available to the user in the form of maps, tables and statistical charts. The training data is transferred to the association analysis module and the Bayesian network module for modeling and prediction. The modules are relatively independent and interlinked within the Shiny framework. The overall structure of the system is shown in Fig. 17.1.

17.2.2 System Architecture Design

The architecture of the whole system is divided into four main layers: front-end UI layer, display layer, business layer and data layer.

In the front-end UI layer, the basic structure of the web page is built using Shiny calls to the Bootstrop framework, including the grid system, link styles, backgrounds,

etc. Within the basic structure, various web components, such as drop-down menus, navigation bars, progress bars, paging, thumbnails, dialogs, media objects, etc., form the overall layout of a web page. Effect factors of route deviation.

In the display layer, Shiny's responsive interactions, such as reactive and render* functions, help the front-end UI layer to interact with the business layer. Meanwhile, the system's map presentation will be done in two ways: static maps are drawn using the system's internally stored vector map data; dynamic maps use the leaflet(a JavaScript library for interactive maps) to call the application programming interface to publish map services.

On the Server side, the business layer receives data from the data layer and implements the core business functions of the system such as crash data presentation, crash data query, spatio-temporal pattern analysis, association analysis, category resampling and Bayesian networks.

The data layer can read and write data from the crash database, geodatabase and road database on the one hand, and can also obtain data through manual data uploads or interfaces from other systems on the other. The data is cleaned in the data layer and finally transmitted to the business layer. The architecture is shown in Fig. 17.2

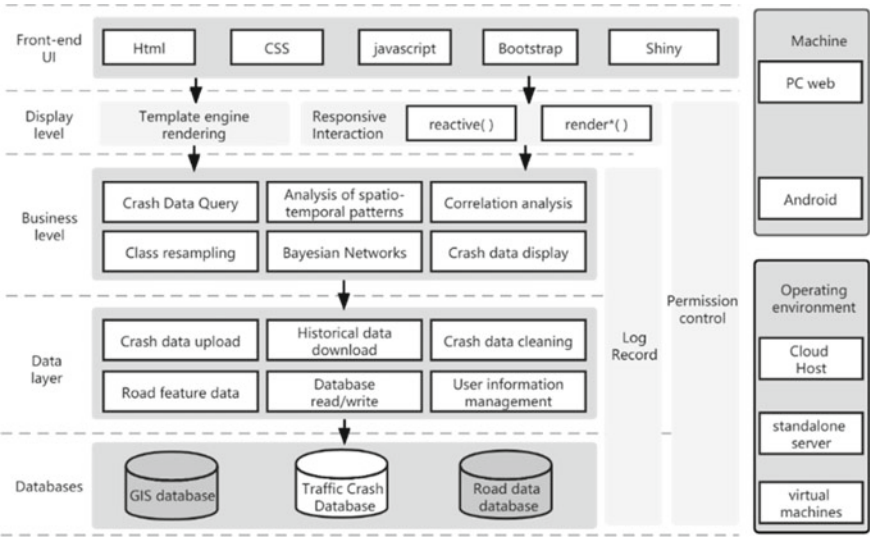


Fig. 17.2 System architecture diagram

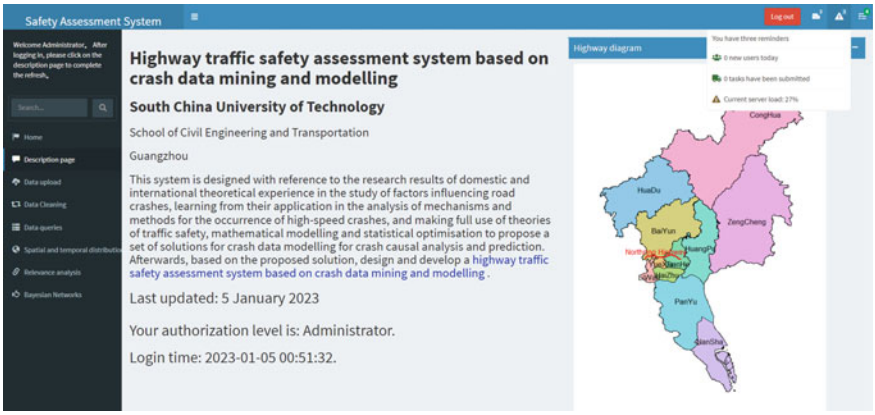


Fig. 17.3 Schematic diagram of the system

17.3 System Implementation

17.3.1 Basic System Framework

When the user opens the system, a blank page is first loaded and a login box is displayed in the center of the page to form the login screen. At the same time, the system will be preloaded with packages, functions and data. When the login is successful, it will unlock the overall UI interface and access the system, as shown in Fig. 17.3.

17.3.2 Data Presentation Module

The Data Display module is loaded in the main interface page. The data display module provides multiple types of dynamic crash maps for the user to explore freely. The user selects how many recent crashes they want to see by using the number slider. When the user hovers the slider over a number for 500 ms, the Server side filters the recent crash data to match the requirement. This data is used to create an interactive map and presented in a data table. In the interactive map, the recent crashes are marked at their location. When the user clicks on the marker, the information on the location of the crash, the type of crash, the cause of the crash, the date of occurrence, the direction of occurrence, the line type, the number of casualties and the number of facilities damaged is displayed.

In the crash heat map sub-page, the user selects the properties of the crash to be viewed from the 8 conditional filtering widgets on the UI side of the page and clicks on the "Start Plotting" button, the Server side filters the data according to the conditions selected by the user. At this stage, the data is point data, the Server side

further converts the point data into density data and then uses interpolation to convert the density data into raster data. The raster data is then plotted on a map to create a heat map. The heat map shows the distribution of crashes and the user can have a direct impression of the geographical distribution of crashes by looking at the heat map.

17.3.3 Data Upload Module

On the data upload page, after the user has logged in, the Server side activates the data upload interface with the historical database for use, while the Server side searches for the time of the latest data in the database and displays it on the UI side for the user's reference. If the user chooses to upload new data, the data uploaded on the UI side is transferred to the Server side through the data upload interface and becomes the raw data to be used. The data upload interface also recognizes the type of data uploaded by the user. If the user uploads data in a format other than the one prompted by the UI, the data upload interface will indicate that the upload has failed.

17.3.4 Data Cleaning Module and Query Module

On the crash data cleaning page, the user selects what to clean using a series of radio boxes. Once the cleaning is complete, the data table for the display data and the data table for the training data are displayed at the bottom of the page.

The data cleaning process is a combination of the raw data to be cleaned and other auxiliary data. For the original crash data information table, the Server side first processes the crash time, converting the inconsistently formatted crash time into a uniform time format. The Server side then extracts the corresponding day of the week and season information from the time of occurrence and divides the time period to which the crash belongs from the moment of occurrence.

After the crash time cleaning is completed, the Server side extracts the road alignment information table from the database and merges it with the crash data table using the chainage as the key. Next, the Server side connects the traffic flow information table, the traffic obstruction information table and the crash data information table with the time of crash as the key. Once the connection is completed, the traffic flow information, blocking information and the proportion of large vehicles will be added to the crash data table.

Once the above operation is completed, the Server side will use the three attributes in the crash data table, namely the number of fatalities, injuries and damage to facilities, to classify the crash type. The crash types are classified as property damage, injury and fatality.

After classifying the crash type, the Server side extracts the traffic sign and merge areas information table from the database and merges it with the crash data table,

using the chainage as the key. Finally, the Server side adds WGS1984 coordinates to the crash data table using the road coordinates table. At this point, the raw data is cleaned to display data. Further, the table headers of the display data are modified and the continuous variables are discretized and recoded to form the training data.

17.3.5 Temporal and Spatial Distribution Module

The temporal and spatial distribution module plots the temporal and spatial distribution of crashes based on the cleaned display data transferred from the data cleaning module.

In the spatial distribution sub-page, the user selects the attributes of the data they wish to view by using the conditional filtering field. Once the selection is complete, the user clicks the "Start Plotting" button, at which point the Server side reads the filtering criteria selected by the user and filters the crash data set from the displayed data according to the criteria. At the same time, the Server side extracts the geographic information system(GIS) data from the database and plots the crash data in the UI side in the form of point and bar charts.

In the time distribution sub-page, after the user sets a date range in the date selection box on the UI side, the Server side receives this range and filters the crash data within this time range from the displayed data. Afterwards, the Server side uses the filtered data to draw the year, season, month, week and hour distribution of crashes and display them on the UI side.

17.3.6 Association Analysis Module

This module contains solutions for the analysis of factors influencing traffic crashes. Before the data can be analyzed for association, the data set must be recoded into transactional data to suit the model requirements. Transactional data requires each row to specify a single instance as the transformed table contains only zeros and ones, the transformed data is stored as a sparse matrix in order to reduce the memory pressure during computation. The sparse matrix takes the form of a triplet in the system, i.e. the non-zero elements of the matrix are stored as (row, col, data). Further, the factors influencing the occurrence of highway crashes are analyzed using the A priori algorithm [18] using traffic crash data that has been pre-processed.

After using the Apriori algorithm, the modeling results are obtained. The user can continue to filter the antecedent items or consequent items of interest on the UI side and Server will filter the rules of interest to the user and output them as a table on the UI side.

17.3.7 Bayesian Network Module

The Bayesian network module is divided into a resampling function and a Bayesian network modeling function.

Typically, highway crash data exhibits a significant imbalance in the classification of crash types, i.e. the number of casualty crashes is much greater than the number of property damage crashes. The "resampling" function uses undersampling, oversampling and manual data synthesis techniques to deal with the imbalance in crash types in the crash data.

When the user enters the Bayesian Network page, the Server side will calculate the raw imbalance of the data table to be sampled and output it to the UI side. Users can choose between different resampling methods. Once selected, the Server side responds by processing the traffic crash type variable in the raw imbalance data using random undersampling, random oversampling or synthetic minority over-sampling technique, depending on the processing method selected by the user.

Random undersampling gives a balanced set of instances by randomly eliminating a majority of the class samples. Random oversampling obtains a balanced set of instances by randomly replicating a minority class sample. The synthetic minority over-sampling technique method uses the generation of artificial data rather than repeating the original observations to resolve imbalances. Assume that there are W minority samples in the training set, each represented by a vector $X = \{x_1, x_2, ..., x_n\}$. In the synthetic minority over-sampling technique, based on these W minority class samples, a new sample NW is synthesized manually (N is a positive integer). This results in the number of minority samples being similar to the number of majority samples. The main steps of the synthetic minority over-sampling technique are as follows [19]:

In the first step, the k nearest neighbors of the minority class sample x_i are calculated, using the Euclidean distance metric. The k nearest neighbors are labeled as ${x_i}^{near}$, near $\in \{1, \ldots, k\}$.

In the second step, a new sample ${x_i}^1$ is synthesized by randomly generating $\theta \in (0, 1)$ and randomly selecting a sample ${x_i}^{near}$, near $\in \{1, \ldots, k\}$ from the k nearest neighbors of x_i.

$$ {x_i}^1 = x_i + \theta({x_i}^{nn} - x_i)\# \tag{17.1} $$

In the third step, the second step is repeated N times for each minority class sample, resulting in the synthesis of NW new samples.

After the resampling is completed to get the balanced data, on the one hand, the Server will read the balance of the balanced data and output the information to the UI side, and on the other hand, the balanced data will be used for the Bayesian network model building.

After the user has completed the resampling, the Server transfers the balanced data to the Bayesian network modeling function and uses the balanced data set to construct

a Bayesian network model. The Bayesian network represents the probabilistic dependencies between a given set of random variables $\boldsymbol{X} = \{X_1, X_2, \ldots, X_p\}$ succinctly as $\mathcal{G} = (\boldsymbol{V}, A)$, where each node $v_i \in \boldsymbol{V}$ corresponds to a random variable X_i. At the same time, the Bayesian network contains a global probability distribution with a parameter Θ.

For a given crash sample data D, structure learning of a Bayesian network model is first performed to find a network structure that best matches it. Using a scoring-based search algorithm, the model is considered as a topology that expresses the joint probability distribution of variables. The structure with the maximum posterior probability is the optimal structure. The algorithm is a combinatorial optimization process, with the core points being the scoring function and the search algorithm. The process is defined as:

$$OM = (\mathcal{G}, \Omega, \mathbb{F})\# \tag{17.2}$$

In the above equation, $\mathcal{G}$ is the search space, containing the various network structures that may be formed between all the variables of the sample data D. Ω is the set of constraints, such as a most basic constraint that all nodes should form a directed acyclic graph. $\mathbb{F}$ is the scoring function, and the extreme value point is the optimal structure of the network.

First, the user needs to perform Bayesian network model structure learning. During the structure learning process, three scoring functions are supported to score the degree of fit of the structure to the data, after which a search algorithm is used to find the higher scoring local optimal solution or global optimal solution to obtain multiple better model structures [16].

The first scoring function is the AIC scoring function, which is expressed as shown below:

$$f_{AIC}(G, D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} log \frac{m_{ijk}}{m_{ij}} - \sum_{i=1}^{n} (r_i - 1) q_i \# \tag{17.3}$$

The second scoring function is the BIC scoring function with the expression:

$$f_{BIC}(G, D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} log \frac{m_{ijk}}{m_{ij}} - \frac{logm}{2} \sum_{i=1}^{n} (r_i - 1) q_i \# \tag{17.4}$$

The third scoring function is the BD scoring function, whose expression is:

$$f_{BD}(G \mid D) = log(P(G)) + \sum_{i=1}^{n} \sum_{j=1}^{q_i} \left[log\left(\frac{\Gamma(\alpha_{ij^*})}{\Gamma(\alpha_{ij^*} + m_{ij^*})} \right) + \sum_{k=1}^{r_i} log\left(\frac{\Gamma(\alpha_{ijk} + m_{ijk})}{\Gamma(\alpha_{ijk})} \right) \right] \# \tag{17.5}$$

α_{ijk} denotes the hyperparameters taken in the dirichlet distribution, $\alpha_{ij^*} = \sum_{k=1}^{r_i} \alpha_{ijk}$. m_{ijk} denotes the number of samples for which variable X_i takes the k-th value while $pa(X_i)$ takes the j-th value, $m_{ij^*} = \sum_{k=1}^{r_i} m_{ijk}$.

Based on the scoring of the structure using the scoring function, the hill-climbing method is used to find the solution with the higher score. The hill-climbing method determines whether or not to select the operation by continuously performing local operations of adding and subtracting edges as well as deleting edges during the search and depending on whether or not the score changes.

Once the model structure is selected, the user clicks on the Start Training button in Model Learning and the Server side learns the parameters of the model using balanced data and a Bayesian network structure. Based on the selected model structure G, the maximum likelihood estimation is used to learn the parameters of the model. If G has n variables $X = \{x_1, x_2, \ldots, x_n\}$, where variable x_i has r_i values, and its parent node $pa(x_i)$ has a total of q_i combinations of values, and if x_i has no parent node, then $q_i = 1$. In this case, the parameters of the demanded G are $\theta = \left\{\theta_{ijk} \mid i = 1, \ldots, n; j = 1, \ldots, q_i; k = 1, \ldots, r_i\right\}$. The logarithm of the likelihood function is:

$$l(\theta \mid D) = logL(\theta \mid D) = log\prod_{l=1}^{m} P(d_l \mid \theta) = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} m_{ijk} log\theta_{ijk}\# \tag{17.6}$$

where m_{ijk} is the sample size at $x_i = k$ and $pa(x_i) = j$. Using the Lagrange multiplier method for the above equation, the estimates of the parameters are obtained.

After the model parameters are successfully learned, the user selects the existing evidence on the UI side and clicks the Show Results button. the Server side inputs the events into the model, gets the probability of the model output and displays it on the UI side. At the same time, the Server adds the road segment variable to the conditions selected by the user. The whole road section is divided into smaller sections in 1 km units, and the probabilities of these sections are predicted from the model in turn, and the predicted images are plotted and the results are output in conjunction with the GIS data.

17.4 Conclusion

In this paper, a highway traffic safety assessment system based on traffic crash data mining and modeling is investigated and developed. The statistical functions of the system, such as data cleaning, spatio-temporal pattern analysis, resampling, correlation analysis and Bayesian networks, are mainly implemented in the R language, and the back-end and front-end are built by Shiny with the Bootstrap framework developed on HTML, CSS and JavaScript. The system is able to achieve the functions of highway traffic crash causation analysis and risk prediction.

The system is built on the basis of the existing highway management system, combined with the traffic crash information visualization, crash data analysis and mining technology and crash risk prediction integrated solutions proposed in this paper. It can highlight the characteristics of traffic crash data analysis under information technology conditions and carry out traffic safety statistical analysis. The model system it builds can comprehensively assess factors affecting road safety for urban and regional traffic safety management.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 52072131), the Key Research Projects of Universities in Guangdong Province (No. 2019KZDXM009), the Natural Science Foundation of Guangdong Province (No. 2023A1515010039).

References

1. Milton, J.C., Shankar, V.N., Mannering, F.L.: Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accid. Anal. Prev. **40**(1), 260–266 (2008)
2. Hyodo, S., Hasegawa, K.: Factors affecting analysis of the severity of accidents in cold and snowy areas using the ordered probit model. Asian Transp. Stud. **7**, 100035 (2021)
3. Ma, Z., Mei, G., Cuomo, S.: An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. Accid. Anal. Prev. **160**, 106322 (2021)
4. Savolainen, P.T., Mannering, F.L., Lord, D., et al.: The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accid. Anal. Prev. **43**(5), 1666–1676 (2011)
5. Yahaya, M., Fan, W., Fu, C., et al.: A machine-learning method for improving crash injury severity analysis: a case study of work zone crashes in Cairo, Egypt. Int. J. Inj. Control. Saf. Promot. **27**(3), 266–275 (2020)
6. Abdel-Aty, M.A., Abdelwahab, H.T.: Predicting injury severity levels in traffic crashes: a modeling comparison. J. Transp. Eng. **130**(2), 204–210 (2004)
7. Mujalli, R.O., López, G., Garach, L.: Bayes classifiers for imbalanced traffic accidents datasets. Accid. Anal. Prev. **88**, 37–51 (2016)
8. Vilaça, M., Macedo, E., Coelho, M.C.: A rare event modelling approach to assess injury severity risk of vulnerable road users. Safety **5**(2), 29 (2019)
9. Montella, A., Aria, M., D'Ambrosio, A., et al.: Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accid. Anal. Prev. **49**, 58–72 (2012)
10. Leevy, J.L., Khoshgoftaar, T.M., Bauder, R.A., et al.: A survey on addressing high-class imbalance in big data. J. Big Data **5**(1), 1–30 (2018)
11. Xu, Z., Shen, D., Nie, T., et al.: A hybrid sampling algorithm combining M-SMOTE and ENN based on random forest for medical imbalanced data. J. Biomed. Inform. **107**, 103465 (2020)
12. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
13. Raj, S., Ramesh, D., Sethi, K.K.: A Spark-based Apriori algorithm with reduced shuffle overhead. J. Supercomput. **77**(1), 133–151 (2021)
14. Zong, F., Xu, H., Zhang, H.: Prediction for traffic accident severity: comparing the Bayesian network and regression models. Math. Probl. Eng. 475194 (2013)
15. Yahaya, M., Guo, R., Fan, W., et al.: Bayesian networks for imbalance data to investigate the contributing factors to fatal injury crashes on the Ghanaian highways. Accid. Anal. Prev. **150**, 105936 (2021)

16. Song, Y., Kou, S., Wang, C.: Modeling crash severity by considering risk indicators of driver and roadway: a Bayesian network approach. J. Safety Res. **76**, 64–72 (2021)
17. Alkheder, S., AlRukaibi, F., Aiash, A.: Risk analysis of traffic accidents' severities: an application of three data mining models. ISA Trans. **106**, 213–220 (2020)
18. Borgelt, C., Kruse, R.: Induction of association rules: Apriori implementation. In: Compstat: Proceedings in Computational Statistics, pp. 395–400. Physica-Verlag HD (2002)
19. Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C.: Safe-level-smote: safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. In: Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp. 475–482. Springer (2009)

Chapter 18
Visualization Method of Urban Motor Vehicle Trajectory Based on License Plate Recognition Data

Minggui Xu, Bin Rao, Yue Li, and Weiwei Qi

Abstract With the advent of the era of big data, efficient transportation plan decision-making and management are inseparable from the support of traffic big data. The license plate recognition data of road intersection video monitoring has the advantages of mature collection methods, low additional cost, high data accuracy and wide application range. Based on the license plate recognition data of a city traffic electronic police system, this paper firstly presents the sources of license plate recognition data and analyzes the data structure and features; then based on the MySQL database, the grouped character stitching method is used to construct a vehicle trajectory database, which contains the spatial trajectory and time trajectory of each vehicle in one day; finally, the map matching is combined with the GIS system to realize the visualization of the vehicle trajectory. The travel trajectory of motor vehicles in the region is accurately extracted, and the travel expectation line is completely presented. The results show that there are 115,466 motor vehicles traveling in the city in one day, and the trajectory map composed of all vehicles shows the expectations of urban motor vehicle travel, indicating that the motor vehicle trajectory visualization method in this paper has good operability and practicality.

18.1 Introduction

With the rapid development of intelligent transportation systems, traffic monitoring systems are also becoming more and more mature. Traffic surveillance cameras are everywhere on roads and intersections, and the license plate data of vehicles passing on the road can be collected by video detection technology and license plate recognition technology. These data contain a lot of traffic information, but it

M. Xu
Guangzhou Conghua Zhujiang Real Estate Development Co., Ltd., Guangzhou 510990, China

B. Rao · Y. Li · W. Qi (✉)
School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China
e-mail: ctwwqi@scut.edu.cn

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_18

is difficult to visualize the trajectory of vehicles through these data. Therefore, how to construct vehicle trajectory data and visualize the trajectory processing through license plate recognition data has become a very important issue.

Currently for the acquisition of vehicle trajectory routes generally use GPS devices installed on the vehicles. The GPS data on the vehicle contains the real-time positioning information of the vehicle, i.e., the space–time information of the vehicle, which can get the trajectory route of each vehicle [1, 2]. However, it is difficult to obtain GPS data on ordinary vehicles, and the GPS data upload ports of each vehicle are not uniform and the data formats are not the same, which makes it very difficult to obtain vehicle trajectory data through GPS data. Nowadays, the city's traffic monitoring system has been very perfect, and video monitoring equipment has been installed at most intersections, which can't only realize traffic violation capture, but also real-time traffic monitoring and real-time understanding of the traffic operation condition. With the continuous improvement of the recognition accuracy of the license plate recognition system, the data accuracy and completeness have been greatly improved [3, 4].

In terms of travel pattern research, Inagaki et al. [5] collected trajectory data of rented bicycle users and compared the travel behavior of electric bicycles with that of ordinary bicycles. Qin et al. [6] obtained trajectory data based on cell phone communication log data and performed cluster analysis on the trajectory data for discovering hotspot paths in the road network. Dong et al. [7] used mobile communication location data for statistical analysis of residents' trips. Bing et al. [8] analyzed the spatial and temporal characteristics of small car tourist trips and the flow patterns of tourists' trips based on a large number of license plate data of small car trips. Xiao et al. [3] analyzed vehicle travel behavior and summarized travel patterns based on private car trajectory data. It can be directly applied to destination prediction, POI recommendation and route planning, etc. Tang et al. [9] proposed a hybrid model for traffic flow prediction, which captures spatial correlations through the volume transition matrix of trajectory estimation obtained from license plate recognition data and the network weight matrix quantified from different detectors and introduces attention mechanism and genetic algorithm to determine and optimize the time dependence to achieve high accuracy and stable traffic flow prediction. Mo [10] proposed a hybrid framework for estimating dynamic origin–destination (OD) demand that fully exploits the information available in license plate recognition (LPR) data. Zhong et al. [11] proposed a new temporal representation that integrates temporal and spatial dimensions by mapping time to a temporal profile in color space to visualize urban vehicle travel paths.

The above research has analyzed urban traffic from multiple perspectives through traffic big data, which plays an important role in the research and development of urban traffic. Based on the license plate recognition data collected by the traffic electronic police system, this paper combines MYSQL database and GIS to realize the construction and visualization of vehicle trajectory data and dig the urban travel pattern, which can provide a reliable data basis for the traffic control measures of urban traffic management departments.

18.2 License Plate Recognition Data Characteristics

18.2.1 License Plate Data Source

The license plate recognition data is derived from the actual data of a city's traffic monitoring system. By installing video surveillance equipment on signal-controlled intersections, together with a traffic electronic police system, the traffic conditions at all intersections are monitored, as well as illegal capture [12]. It is able to obtain license plate data of vehicles passing in each direction at each intersection, including information such as license plate number, license plate type, capture time, capture location, affiliation, travel direction and vehicle type. The license plate recognition data collected at all intersections in a day is about 1.3 million, and the data is collected 24 h a day. So the amount of data collected in a year is as high as 470 million. The results obtained from processing and analyzing such a huge amount of data can accurately reflect the traffic conditions of the city, and provide accurate data support for traffic planning, management and construction of the traffic management department [13].

This paper focuses on selected correlation data from Sunday, March 28 and Monday, March 29, 2021. The data covers 34 signalized intersections with a total of over 2.65 million data records, containing almost all vehicle movements over the two days.

18.2.2 License Plate Data Structure

The license plate recognition data acquired in this paper records the data of 34 intersections in a city on Sunday, March 28 and Monday, March 29, 2021. The data is collected 24 h a day without interruption, and the data volume is 1.3–1.5 million items a day. The raw plate recognition data table contains the following attribute fields: "License plate number", "License plate type", "Capture time", "Capture location", "Affiliation", "Travel direction", and "Vehicle type". Each original record contains seven fields, and the plate recognition data is shown in Table 18.1. The data time period contains information for 24 h all day on weekends and weekdays.

In the table, "License plate number" is the encrypted vehicle license plate information; "License plate type" is the type of vehicle license plate, indicating that the vehicle is a pure electric new energy car; "Capture time " is "2021–03-28 00:00:01", which means the data was sampled at 0:0:1 on March 28, 2021; "Capture Location" and "Affiliation" represents the location of the intersection where the vehicle was captured; "Travel direction" is "West->East" which means that the vehicle was traveling in a west to east direction; The "Vehicle Type" is "52", which means that the vehicle type is car.

Table 18.1 Structure of vehicle license plate data

Field	Data
License plate number	084,674
License plate type	Pure electric new energy vehicle
Capture time	2021-03–28 00:00:01
Capture location	***
Affiliation	***
Travel direction	West->East
Vehicle type	52

18.3 Vehicle Trajectory Database Construction

18.3.1 License Plate Data Pre-processing

The license plate recognition data is collected through the license plate recognition system. According to the previous section license plate data structure and source introduction, the amount of data recognized by the system tends to be full data. But there is a certain amount of data noise in the originally collected license plate recognition data [14]. In order to ensure the accuracy of the subsequent trajectory construction, it is necessary to carry out data pre-processing before the trajectory construction.

First of all, after preliminary research and analysis of the data, it was found that the original license plate recognition data had the following two problems: missing data and data redundancy. For example, if the content of the "license plate number" field in the license plate data is "unlicensed car", it is judged to be missing data; Redundant data refers to the appearance of consecutive data for the same vehicle at an intersection, which is manifested by the presence of multiple identical data in the fields of "License plate number", "Capture time" and "Affiliation". With SQL database processing, missing data is deleted directly and redundant data is deleted from duplicates, leaving only one valid piece of data.

In order to facilitate the construction of the vehicle trajectory database and the visualization of the vehicle trajectory, after the initial cleaning of the original plate recognition data, the 34 intersections need to be numbered and the time converted to a uniform format, starting at 0:00:00 am of the day and converting the time to seconds.

The data after data cleaning, intersection numbering and time conversion is shown in Table 18.2.

Table 18.2 License plate data after pre-processing

License plate number	Location	Time
008,802	30	0
013,273	13	1
069,285	29	8538
032,832	3	86,399
……	……	……

18.3.2 Vehicle Trajectory Database Construction

This paper adopts the method of grouping character splicing based on MYSQL database to build the vehicle trajectory database, taking each intersection as a travel trajectory point. First of all, the license plate number field is grouped to filter out all vehicles, and then the "location" and "time" fields are stitched according to the ascending order of time to establish the vehicle trajectory data table, so as to build a vehicle trajectory database [15]. The steps of MYSQL-based grouped character splicing trajectory database construction method are as follows.

(1) Group according to the license plate number.
(2) Arrange each group of license plate numbers in ascending chronological order after grouping.
(3) Splice the location and time field values of each group of license plate numbers separately, separated by ",".
(4) Save the result as a new trajectory table.

The result obtained after the run is shown in Fig. 18.1, which is the vehicle trajectory data of March 28th, in which there are 115,466 vehicles. "Location trajectory" is the spatial trajectory points that the vehicles pass through in a day, and the "time trajectory " is the time trajectory at the location of the intersection corresponding to the "location trajectory ".

For example, the location trajectory of the vehicle with the license plate number "009,734" is "19,33,2" and the time trajectory is "5702,5754,5823", which means that the vehicle traveled from intersection 19 to intersection 33 to intersection 2, passing through each intersection at 5702 s (1:35:02), 5754 s (1:35:54) and 5823 s (1:37:03) respectively.

18.4 Visualization of Vehicle Travel Trajectories

Using vehicle trajectory data, combined with geographic information systems, QGIS software is used to match the trajectory points passed by the vehicle with the map, so that the trajectory route of all vehicles can be displayed on the map, i.e., to visualize the trajectory, which is convenient for traffic management departments to view the operation of vehicles in the traffic road network from a macro perspective [16].

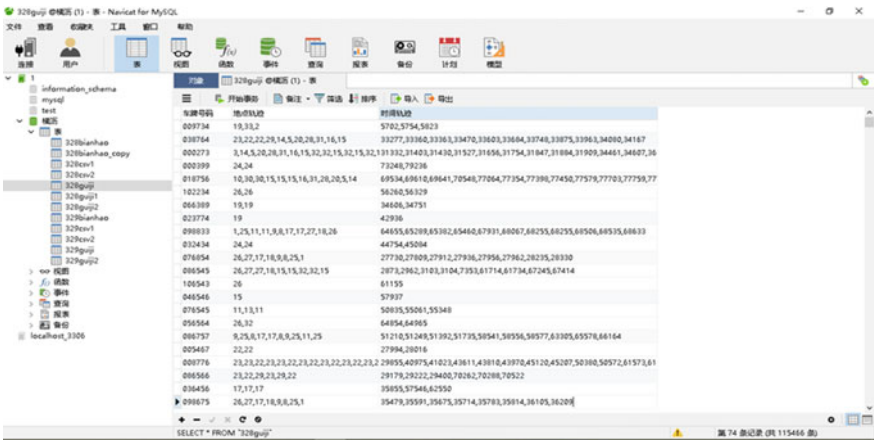
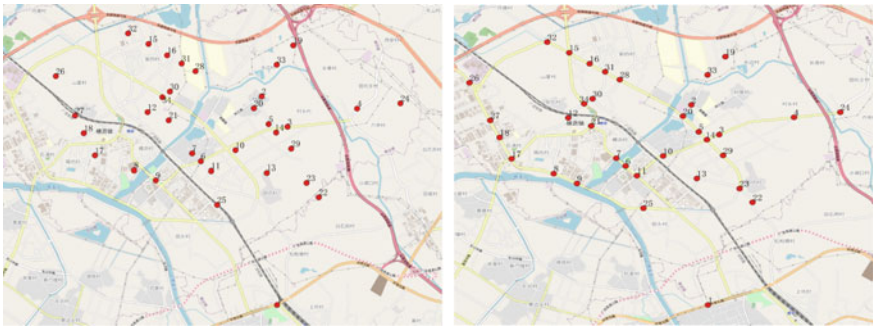


Fig. 18.1 Vehicle trajectory database based on MYSQL

18.4.1 Vehicle Travel Trajectory Point Matching

To match the trajectory data with the map, we first need to determine the latitude and longitude coordinates of all the intersections in the road network. This paper uses the coordinate extraction system of Baidu Maps to query the latitude and longitude coordinates of all the intersections. Firstly, the names of the intersections are matched with their locations in the actual map. Then the location points are identified from Baidu Maps and the latitude and longitude are extracted and recorded. Once the intersection latitude and longitude coordinates were determined, the longitude and latitude coordinates field value was changed to "POINT (latitude and longitude coordinates)" as it needed to be matched with the map in QGIS.

After obtaining the complete intersection latitude and longitude coordinates, the trajectory point data was imported into QGIS, as shown in Fig. 18.2a. It was found that the intersection location did not match the map and there was some deviation because the intersection latitude and longitude coordinates were in the BD09 coordinate system (Baidu coordinate system) while the map in QGIS was in the WGS coordinate system (Earth coordinate system), so there was some deviation. Through the GeoHey plugin in the QGIS toolbox, the intersection latitude and longitude coordinates were converted from the Baidu coordinate system to the Earth coordinate system, and after successful conversion, as shown in Fig. 18.2b, it can be seen that the intersection location successfully matched with the map road network intersection.



(a) (b)

Fig. 18.2 Trajectory points match the map

18.4.2 Vehicle Travel Trajectory Route Matching

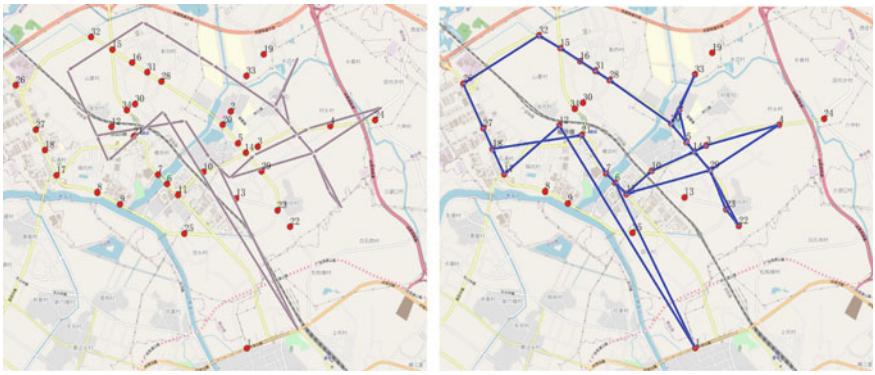
The trajectory line is the trajectory route of each vehicle, realizing the trajectory route to match with the map and displaying on the map the specific intersection through which each vehicle is traveling, in order to facilitate a macro view of the city's traffic operation.

The longitude and latitude coordinates corresponding to each intersection are combined with the license plate recognition data to establish a trajectory coordinate data table, as shown in Table 18.3.

Import the obtained vehicle coordinate trajectory from MYSQL database into QGIS software, and set the detection shape as linear. After successful import, as shown in Fig. 18.3a, it can be seen that there is an offset between the trajectory and the map position, also due to the deviation of the Baidu coordinate system and the Earth coordinate system, but the coordinates of the layer could not be successfully converted. This is because the layer is not a point layer, but a line layer, where the combined shapes are judged to be non-normal geometry and cannot be converted, so the layer needs to be processed.

Table 18.3 Vehicle trajectory coordinates

License plate number	Trajectory coordinates
048,546	linestring(113.965917 23.01862,113.951869 23.023915,113.960932 23.02074)
026,973	linestring(113.997719 23.047977,113.993883 23.04375,113.9904 23.036846)
082,326	linestring(113.978736 23.020628,113.991574 23.02024,113.978736 23.020628)
061,753	linestring(113.942846 23.041268,113.959527 23.050609)
035,890	linestring(114.02244 23.035319,114.02244 23.035319)
……	……

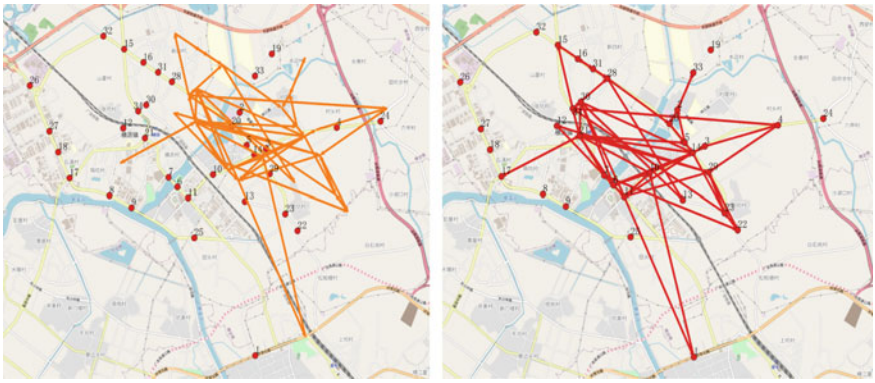


(a) Trajectory before coordinate conversion (b) Trajectory after coordinates conversion

Fig. 18.3 Route of a car driving trajectory on March 28

Using the "Check validity" tool in the QGIS toolbox, Three new layers will be created after checking the validity of the layer's graphics. They are "Output invalid", "Output error" and "Output Valid". Look at each layer separately and find that there are no graphics on the "Invalid Output" layer, which means that no trajectory is invalid. The "output error" layer is the point position of each intersection, indicating that the point coordinates do not form a graph. The "output valid" layer will be converted from BD09 to WGS coordinate system, as shown in Fig. 18.3b.

Figures 18.3 and 18.4 show the trajectory routes of the vehicle with license plate number "057,204" before and after the conversion on 28 March and 29 March respectively.



(a) Trajectory before coordinate conversion (b) Trajectory after coordinates conversion

Fig. 18.4 Route of a car driving trajectory on March 29

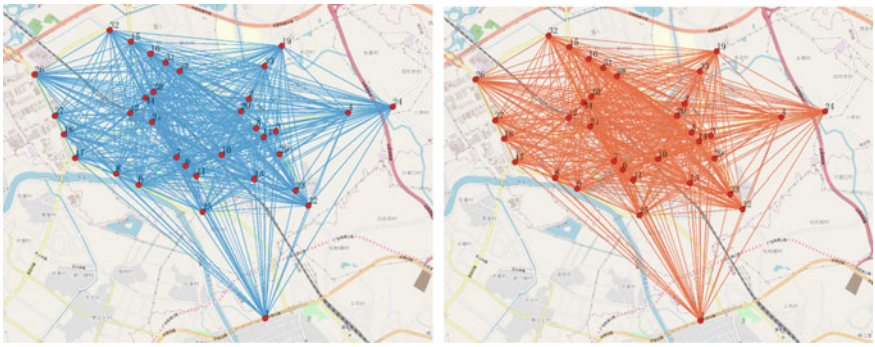
18.4.3 Vehicle Travel Desire Line Generation

Traditional vehicle travel OD survey is a wide-ranging, very large workload, requiring many units and departments to collaborate with each other and work together to complete. Before conducting the OD survey, survey forms need to be designed and printed, and the principle of form design is to meet the requirements of the survey, but also to be concise, so that the respondents can easily fill out or answer. Commonly used vehicle OD survey methods are shown in Table 18.4. Traditional OD survey parties are also limited by the choice of sampling rate, for cities with a population of more than 1 million, the sampling rate is generally about 4%. Sampling methods include simple random sampling, stratified sampling, equidistant sampling, and whole-group sampling. After completing the OD survey, the results need to be collated and checked for accuracy, and the relative error of the survey results is kept within 5% to meet the requirements, and finally multiplied by the amplification factor to expand to the full sample.

Based on the vehicle trajectory database, import all vehicle trajectories into the QGIS software, as shown in Fig. 18.5a and b are the trajectory maps of all vehicles on March 28 and March 29 respectively, which means that the city motor vehicle travel desire line map was formed. The motor vehicle trip desire line shows the distribution of trips between intersections within the city, and can directly obtain the traffic generation results in the four-stage method of traffic planning, which is more efficient than the traditional manual sampling survey. Besides, the traffic distribution results obtained are relatively more perfect and accurate.

Table 18.4 Commonly used vehicle OD survey methods

Survey method	Content
Distribution form survey method	The survey form will be sent to motorists, implemented by the vehicle management system to each person, filled out and recycled, and needs to be filled out before the mobilization and explanation work
Roadside inquiry method	Investigation stations are set up on the road to stop vehicles and question drivers. Since the investigation process requires stopping vehicles, it has a large impact on normal road traffic
Registration of vehicle license plate method	A number of survey stations were set up on the road network, and each station noted down the last 3–4 digits of the vehicles that passed through the station, as well as the time of passage, and finally, the summary was checked
Vehicle tethering method	Attach the tag to both sides of the windshield or door handle of the vehicle at the parking place, with the tag indicating the place of issuing the tag and the time of driving out, and unhook the tag at other parking places and note the place and time of driving in
Postcard survey method	Postcards with survey items will be mailed or sent to drivers, who will fill them out and collect them, with a recovery rate of no less than 20% to be valid



(a) March 28 (b) March 29

Fig. 18.5 Vehicle travel desire line

18.5 Conclusion

As one of the most important means of travel for residents, information on the travel routes of urban motor vehicles can reflect to a certain extent the situation of urban traffic travel.

This paper firstly presents the sources of license plate recognition data and analyzes the data structure and features. Then the vehicle trajectory database was then constructed based on the MYSQL database using the grouping character stitching method, and the trajectory data was combined with the GIS system to visualize the vehicle travel trajectory. The main research findings are as follows.

(1) Based on the MYSQL database, a vehicle travel trajectory database was constructed using the grouping character stitching method. The trajectory database contains the spatial and temporal trajectories of 115,466 motor vehicles traveling in a day, which can reflect the daily motor vehicle travel rate of the city to a certain extent.
(2) The trajectory data combined with GIS enables the visualization of the travel trajectory of motor vehicles, not only to view the travel trajectory of each vehicle, but also to form a travel desire line map of all motor vehicles with the city intersection as the OD, providing a more complete traffic distribution result and providing a more accurate and complete data to support urban traffic planning.

In the process of traditional traffic survey, often only intersection flow data can be obtained, but not the specific driving path of each vehicle. The vehicle travel trajectory database constructed through license plate recognition data can provide more detailed vehicle travel data, realize the visualization of vehicle travel trajectories, and generate vehicle travel desire lines, which can provide traffic management departments with more intuitive urban vehicle operation and provide data support for traffic management, control and planning schemes.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 52072131), the Key Research Projects of Universities in Guangdong Province (No. 2019KZDXM009), the Natural Science Foundation of Guangdong Province (No. 2022A1515010123).

References

1. Quddus, M., Washington, S.: Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. Transp. Res. Part C Emerg. Technol. **55**, 328–339 (2015)
2. Fan, J., Fu, C., Stewart, K., Zhang, L.: Using big GPS trajectory data analytics for vehicle miles traveled estimation. Transp. Res. Part C Emerg. Technol. **103**, 298–307 (2019)
3. Xiao, Z., Xu, S., Li, T., et al.: On extracting regular travel behavior of private cars based on trajectory data analysis. IEEE Trans. Veh. Technol. **69**(12), 14537–14549 (2020)
4. Jin, X., Tang, R., Liu, L., Wu, J.: Vehicle license plate recognition for fog-haze environments. IET Image Process. **15**(6), 1273–1284 (2021)
5. Gao, K., Tu, H., Sun, L., Sze, N.N., Song, Z., Shi, H.: Impacts of reduced visibility under hazy weather condition on collision risk and car-following behavior: Implications for traffic control and management. Int. J. Sustain. Transp. **14**(8), 635–642 (2020)
6. Xue, Q., Gao, K., Xing, Y., Lu, J., Qu, X.: A context-aware framework for risky driving behavior evaluation based on trajectory data. IEEE Intell. Transp. Syst. Mag. **15**(1), 70–83 (2023)
7. Li, Y., Zhao, L., Gao, K., An, Y., Andric, J.: Revealing driver psychophysiological response to emergency braking in distracted driving based on field experiments. J. Intell. Connect. Veh. **5**(3), 270–282 (2022)
8. Bing, H., Bo, K., Ling, Y., et al.: Discovering the graph-based flow patterns of car tourists using license plate data: a case study in Shenzhen, China. J. Adv. Transp. 1–15 (2020)
9. Tang, J., Zeng, J., Wang, Y., Yuan, H., Liu, F., Huang, H.: Traffic flow prediction on urban road network based on license plate recognition data: combining attention-LSTM with Genetic Algorithm. Transportmetrica (Abingdon, Oxfordshire, UK) **17**(4), 1217–1243 (2021)
10. Mo, B., Li, R., Dai, J.: Estimating dynamic origin-destination demand: A hybrid framework using license plate recognition data. Computer-Aided Civ. Infrastruct. Eng. **35**(7), 734–752 (2020)
11. Zhong, Y., Song, W., Kim, C., Hwang, C.: Coupling path visualization and its application in preventing electromagnetic interference. IEEE Trans. Electromagn. Compat. **62**(4), 1485–1492 (2020)
12. Lin, C., Chuang, C., Lin, H.: Edge-AI-based real-time automated license plate recognition system. Appl. Sci. (Basel) 12(3) (2022)
13. Gao, T., Liu, Z., Lian, S., Yue, S., Zhang, J.: Crossing road monitoring system based on adaptive decision for illegal situation. Appl. Soft Comput. **11**(7), 4399–4412 (2011)
14. Wang, W., Tu, J.: Research on license plate recognition algorithms based on deep learning in complex environment. IEEE Access **8**, 91661–91675 (2020)
15. Qi, X., Ji, Y., Li, W., Zhang, S.: Vehicle trajectory reconstruction on urban traffic network using automatic license plate recognition data. IEEE Access **9**, 49110–49120 (2021)
16. Gan, N., Zhang, M., Zhou, B., Chai, T., Wu, X., Bian, Y.: Spatio-temporal heuristic method: a trajectory planning for automatic parking considering obstacle behavior. J. Intell. Connect. Veh. **5**(3), 177–187 (2022)

Chapter 19
A Two-Stage Teaching Philosophy for Postgraduate Students

Kai Wang, Ying Yang, Yue Zhang, and Xiaobo Qu

Abstract In this study, we develop a two-stage structured teaching philosophy that is able to cater to the needs of postgraduate students with different expectations about the learning outcomes. At the postgraduate level, different students have distinct expectations about their future careers. Some may want to pursue their career as an engineer, with little interest in research and development. Others may have different views and would like to pursue their career as a researcher or an academic in the future. Therefore, it is necessary to take into account the distinctions among students' expectations about courses. The proposed teaching philosophy divides the course sessions into two stages. In the first stage, fundamental and common knowledge bases for the courses are delivered to all students, which ensures the students receive the necessary and basic knowledge that is required for both industrial and academic pathways. Afterward, the second stage leverages the flipped classroom model to let the students choose their learning and course materials with different emphases as per their own expectations and interests. Customized learning and teaching materials are prepared for students who prefer the industrial pathway and students who show more predilections for the industrial pathway. We will use a master course about transportation engineering to empirically test this teaching philosophy and evaluate its performance, including a comparison with the conventional teaching process. The results demonstrate that the new structure is well received by students and is much beneficial for improving students' subjective evaluations of the courses and performances in learning.

K. Wang · X. Qu
School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China

Y. Yang (✉)
Unit of Urban Mobility and Behavior Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden
e-mail: yying@chalmers.se

Y. Zhang
Department of Education and Special Education, University of Gothenburg, 41296 Gothenburg, Sweden

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_19

19.1 Introduction

In the curriculum development plan for master's students, students will be equipped with the required knowledge base to be engineers in relevant industries or an academic in the fields of research [1, 2]. It has been recognized that there are significant distinctions between these two groups of students. Although both pathways do have some similar knowledge bases, they are very different at the next stage [3]. Academic researchers are inclined to discover new knowledge about the physical and natural world or human society that are not unrevealed by the human yet. In other words, researchers focus more on creating new knowledge with comparatively fewer considerations of practical applications and constraints. Researchers need to ask a scientific question, formulate a hypothesis, and test that hypothesis based on experiments or deductive methods. These essential skills for scientific studies ask for excellent theoretical background and abstract thinking, which should be well-trained during education. Therefore, learning knowledge and developing capacities requested by scientific research are more beneficial for students who show a predilection to the research pathway.

On the contrary, engineers tend to make the best use of the existing and learned knowledge to solve practical problems, focusing on promoting our daily lives such as reducing cost, improving efficiency and safety, and so on. Engineers tend to provide more attention to practical situations and constraints of specific problems. This aim is to apply mature technology, methods, and strategies to successfully solve a particular problem. Hence, technical expertise and practical knowledge about existing solutions and mature techniques in a professional field are more crucial for industrial applications as compared to abstract thinking and theoretical deductions. For example, skillfully using practical software or solution tools are more important for students who prefer the industry pathway as these skills and knowledge are core competence for industrial jobs. Students in the research pathway look at 5–10 years' time, and target general science, while the industry looks at the current situation. In summary, there are non-negligible differences in the education needs for research and industrial pathways [4].

After years of bachelor's study, Master's students have already made initial plans for their future career pathways. On account of the divergent career plans or expectations, different students are supposed to show quite different interests and focus during their studies at the master level [5, 6]. As per empirical observations, students who prefer the research pathway generally have stronger desires to develop his/her academic and research ability in terms of theoretical backgrounds and research skills. As a consequence, these students present stronger motivations for absorbing more knowledge concerning academic research and may seek the opportunity to know the emerging research topics and expertise, which is part of preparations to pursue a Ph.D. or other research careers in the future. However, Master courses generally do not mandatorily include clear and enough channels to provide research ability training for Master students as these are not standard requirements for all master students. In this regard, Master students tending to pursue the research pathway may

receive less training about research ability than they expect in Master courses, which potentially results in negative impacts on students' self-development and passions. Nonetheless, if much content about academic research is included in courses, these would increase the difficulty on account of diversity in students' abilities and backgrounds. More importantly, students who prefer the industry pathway are not very interested in a lot of theoretical deduction and research content, and may not be satisfied with the courses because they expect more practical expertise and technical skills to solve practical problems.

In short, it is challenging to balance the expectations about course contents from students with different interests and career plans as the required focuses and contents from the two types of students to some extent, conflict with each other. However, diversity among students is a rather common phenomenon or situation that teachers should well deal with. As a result, it is of great utmost to develop a course structure that is able to establish the nexus among learning objectives, students' expectations as well as distinctions between the two pathways. More specially, flexible teaching philosophy or framework that accommodates different learning objectives and expectations, is mandatory to realize better master education outcomes that are appropriate for students choosing both research and industrial pathways.

This above-mentioned situation is particularly evident for transport engineering students at the Master's level. The transport discipline is under revolutionary changes in the past few years, mainly due to the advent of connected, automated, electric vehicles and shared mobility [7–9]. As a result, the master courses in transport engineering constantly introduce a bit more about these emerging technologies and advances happening in the area. As per our empirical observations, this progress of research and development immediately attract the attention of some students, who are very passionate about it. They show much interest to learn more about the knowledge and theories behind these new technologies. A few students even expressed their interest in pursuing a Ph.D. in this area. There is actually a fundamental paradigm change in the discipline resulting from emerging technologies. To achieve a sustainable and diverse education and support research development, it is very important to not only train our master's students to be outstanding engineers in our master's program, but also prepare themselves with the required fundamental basis to be an excellent potential researcher. As a consequence, the teaching philosophy also needs to be adjusted to accommodate the different objectives and expectations of different master students. This aims to endow the course structure with flexibility and comprehensiveness to satisfy diverse students' expectations and needs.

In this study, we develop a two-stage structured teaching philosophy that is able to cater to the needs of postgraduate students with different expectations about the learning outcomes. At the postgraduate level, different students have distinct expectations about their future careers. Some may want to pursue their career as an engineer with little interest in research and development, while others may have different views and would like to pursue their career as a researcher or an academic in the future. Thus, it is necessary to take the distinctions among students' expectations into consideration. The previous course structure does not differentiate the distinct

views of our master's students, while a two-stage structured teaching philosophy is employed.

The proposed teaching philosophy divides the course sessions into two stages. In the first stage, fundamental and common knowledge bases for the courses are delivered to all students. The first stage ensures the students receive the necessary and basic knowledge that is required for both industrial and academic pathways. Afterward, the second stage leverages the flipping learning to make the students choose their learning and course materials with different emphases as per their own expectations and interests. More specially, two customized learning and teaching materials are prepared in the course. One is designed for students who prefer the industrial pathway in the future, with more academic and theoretical content. Another is for students who show more predilections for the industrial pathway, with more professional tools and knowledge concerning practical solutions. We use a master course about transportation engineering to empirically test this teaching philosophy and evaluate its performance, including a comparison with the conventional teaching process. We further compare the performances of both structures using student evaluations and students' average marks. The results demonstrate that the new structure is well received by our students.

The rest of this report is organized as follows. Section 19.2 introduces the two-stage course structure and associated assessment plan. We use a course in transport engineering to conduct an empirical comparison of students' feedback and performances using our proposed course structure and the conventional course structure. In Sect. 19.3, the results of the before and after comparisons are presented, including both qualitative and quantitative measures. Section 19.4 concludes this study and proposes the way forward.

19.2 Methods

On the one hand, some contents and knowledge are standard bases in a course, and thus indispensable for all students, no matter what pathways the students are interested in and plan to pursue. These contents are essential and should be certainly taught in the course. On the other hand, as above-mentioned, different students may also have some divergent expectations for a course due to their preferred career pathways in the future. If the same teaching process in terms of learning contents, material, and evaluations throughout a course is used for all the students, their distinct expectations will be ignored, which would reduce students' passions and proactive learning motivations in the course. Consequently, these may have negative impacts on students' acquired knowledge and their future development. Hence, it is necessary to use a flexible teaching paradigm to take care of different needs.

Taking into account the two obligations of a course, we propose a two-stage teaching philosophy. The basic idea is to include two different instructional approaches, including student-centered teaching and the flipped classroom model, by integrating their mutually complementary merits. Both approaches are conducted

based on the constructivism theory in educational research. Constructivism theory, one of the most important learning theories in pedagogy, believes that people acquire knowledge through constructing meaning from their experiences and interactions with the world [10, 11]. Employing a student-centered teaching philosophy entails taking students as the main role of the classroom and teachers as the active coaching roles. The flipped classroom model is based on the learning process explained by constructivism theory which highlighted the importance of creating a learning model to encourage students to think and interact more with the learning content.

In the first stage, student-centered teaching is used to deliver the fundamental and mandatory content to all students using the same learning materials, lectures, and discussions. Teachers provide basic introductions and explanations about theories and contents in the courses, divide students into several small groups, assign homework, and exercise in groups to understand the present theories in courses through discussions with peers and shared experience from solving problems. Teachers instruct by giving feedback on the homework and exercises to enhance students' understanding and correct students' mistakes. The teacher teaches and delivers the main contents during courses via lectures, discussions during the course, and tutorials. Because all students need this fundamental knowledge, the student-centered teaching approach which is more impressive is used for these contents. However, fundamental knowledge is hard to satisfy different students' demands. As stated in the introduction, two types of students have various learning ambitions. Some students enjoy the deduction process and calculation of a theory. However, others may be not interested in theoretical perspectives. These students may lose their passion and interest in proactively participating in the courses. To address this, we utilize the flipped classroom model in the second stage to consider different expectations of students who prefer different career pathways in the future.

The flipped classroom model adopts flexible teaching and learning approach in which instructional modes of student learning, in-class activities, and out-class homework are flipped [12, 13]. Differing from the conventional teaching paradigm, the flipped classroom model moves reading and learning materials to pre-course periods and makes the best of the in-class time to let students conduct deeper discussions and potential practice in the class with assistance from teachers.

As the flipped classroom model moves the material reading to pre-class periods, teachers can prepare two different sets of learning materials in advance, concentrating on different aspects of research and industrial pathways. Students can choose their own preferred materials for learning as per their own expectations. For the materials of the research pathway, more content about emerging technology and theoretical knowledge will be covered. In contrast, more content about technical tools and practical solutions will be included in the materials for students choosing the industry pathway. In this way, different students' learning expectations and interests can be satisfied. Moreover, the flipped classroom model uses a sustainable way and has essential advantages to encourage students' active learning, which is very important. The ultimate goal of teaching is to let the students acquire the corresponding knowledge and develop self-learning abilities as per their own career interests. In students' future development, the ability to active learning or self-learning may be

more important than the previously obtained knowledge. It is worth mentioning that the teacher has to prepare more material and in-class activities for different groups due to flipped classroom model. These, to some extent, increase teachers' pedagogical burdens and workloads. Nonetheless, the prepared materials are sustainable and can be conveniently reused in future courses.

We conduct an empirical before- and after-analysis based on a field experiment to test the effectiveness and recognition of the proposed two-stage teaching structure and compare its performances with the traditional teaching structure. The experiment is conducted at a Swedish university, taking advantage of a master's course in transportation engineering. In 2018, we used a traditional "teacher-centered" teaching structure as most engineering courses did. The teaching materials are exactly the same for both pathways, with around 40% fundamental knowledge, 40% practical project-based materials, and 20% research components. For the assessment, students will be required to answer all questions in the final exam. In 2020, we used the proposed two-stage teaching structure for the same course. However, 40% of fundamental knowledge remained unchanged (same as the contents in 2018) and was delivered to students in the first stage. For the rest 60% of the sessions, we use flipping classes and prepared structured teaching materials with some pre-reading requirements (both practice targeting engineers and research targeting Ph.D. studies) before attending the lecture. Students can choose their own preferred materials for self-learning before the course. In the flipped classroom model, half of the students study research articles and the other half of the students learn practical reports. The teachers comment and help the class to have a good understanding of both practice and research. It is expected that students will be more self-motived in reading/exploring the pathway that their future career path might be associated with. It should be noted that a clear syllabus is critical for the success of course modification [14, 15]. Further, the teachers need to prepare good materials for research papers for the research pathway and practical project reports for the industry pathway. Simultaneously, the evaluations of students' learning performances for different students should be distinguished as they focused on different types of learning materials. For the assessment, two compulsory aspects are evaluated. The first aspect is to reflect the student's learning progress about fundamental knowledge in the first stage. The second aspect concentrates more on the learning contents in the second stage. We design four selective problems, in which two problems are designed for the industry pathway, and the other two are customized for the research pathway. Students need to select two out of the four problems to answer in the evaluation as per their choice of learning materials in the second stage.

19.3 Results

To evaluate the students' feedback towards the proposed two-stage teaching structure, we employ both qualitative and quantitative measures [16]. The qualitative measures are executed via the form of student interviews. For 2018, we randomly selected

ten student representatives and discussed their subjective evaluations concerning the courses. In 2020, we also randomly select ten student representatives and discuss the courses with them to identify the difference between the two teaching strategies. We design a few questions for these two groups of students to answer, including what goes right, what goes wrong, and how we can improve in different aspects of the course. We employ the five-point Likert Scales to measure students' evaluations on six dimensions, including prerequisites, learning outcomes, learning, literature assessment, and workload. At the same time, the overall impression of all aspects of the courses is surveyed.

Based on the collected data, we would like to test whether students believe this experiment is in alignment with their career development plan. The results are summarized in Table 19.1. The results demonstrate that there is a noticeable improvement in the investigated items in 2020 as compared to those in 2018. The students' evaluation of prerequisites increases by about 5.5%. The scores about learning outcomes and learning improve by 15.3% and 23.7%, respectively. These imply that the proposed teaching structure is more beneficial for students' learning from the students' perspectives as compared to the conventional teaching process. Additionally, a notable increment in the student's evaluation of assessment (48%) is observed after implementing the proposed teaching strategy. This indicates that more flexible assessment processes for different students in the proposed strategy are welcomed by the students. The proposed teaching structure lets students answer questions about the aspects (i.e., contents for research or industry pathways) that they prefer to learn, and thus motivate the students' self-interest to learn and explore relevant knowledge, which consequently leads to students' higher evaluation about "assessment" in the course.

As for the overall course score, the students' evaluation improves by 40.2% in 2020, which is considerable. This hints that the proposed teaching structure is indeed beneficial for increasing students' satisfaction with the course contents as compared to the conventional teaching approach and is well accepted by students. However, the students' evaluations of workloads have no differences over the two years. Besides students' subjective feedback, we also evaluate the students' objective study outcomes using the students' exam grades. The results are presented in Table 19.2. The results demonstrate that the average exam grade improves from 73.3 to 76.7 (4.6% increase), indicating the positive effect of the proposed teaching structure on students' learning outcomes and exam performances. Better performance in exams

Table 19.1 Students' evaluations of different aspects of the course in different years

Year	Prerequisites	Learning outcomes	Learning	Literature	Assessment	Workload	Overall
2018	3.81	3.4	2.91	3.92	3.04	3.23	2.81
2020	4.02	3.92	3.6	4.3	4.5	3.22	3.94
Improvement (%)	5.5	15.3	23.7	9.7	48.0	−0.3	40.2

Table 19.2 Students' grades: comparison between the two years (average marks and deviation)

Year	Student' average grades	Deviation
2018	73.3	10.6
2020	76.7	8.4
Improvement	4.6%	−2.2

can be well explained by the self-determination theory. Self-determination theory argues that human beings have an innate psychological need for self-motivation and personality integration [17]. Allowing students to choose what materials they want to learn is a kind of autonomy-supportive teaching style, which is proven to be helpful in increasing students' autonomous motivation [18]. At the same time, autonomous motivation as one type of intrinsic motivation in self-determination theory can help students create positive learning outcomes in academics [19]. However, due to the large deviation among students, the improvement is not significant in statistics.

One of the features of the proposed two-stage teaching structure is that we leverage flipping teaching to satisfy different students' learning expectations and interests on account of their preferred career pathways. Nonetheless, there are still some challenges in using flipped classes in real practice that are worth noting. The first one is that the quality of pre-course materials should be secured at the cost of increasing teachers' preparation workloads. In the flipped classes, the pre-course materials are very important, as discussed before. On account that it is a pre-course, the teachers could provide immediate responses to the question by students when the students view and learn the materials. Therefore, the materials should be well-prepared, taking the special needs of different students into consideration. More specifically, the pre-course materials and videos should be more detailed and comprehensive as compared to normal face-to-face courses. This may take more time for the teacher to prepare, but it is worthy as the materials could be reused in a sustainable way for future courses.

The second problem of a flipped class is ensuring every student spends enough time preparing before courses. The process is kind of dependent on the students' self-learning before the courses. Some passive students may be negative in preparing the courses, and then they may perform very badly in the courses as there is no studying surveillance like conventional in-class teaching. This results in a concern that teachers have to ensure students read the pre-course materials responsibly and understand the contents mainly by themselves. Unfortunately, there is no efficient and explicit way to guarantee that students will adequately prepare the materials. Even though it is the student's choice to show responsibility for the course or not, teachers always want to ensure every student learns from the course and gets the credit for their degree correctly. One potential solution may check the quality of preparation in the course through some quizzes. If a student is found to show irresponsibility in the preparation, the teachers could provide customized reminders and warnings.

19.4 Conclusion

In this study, we propose a two-stage structured teaching philosophy that is able to cater to the needs of students with different expectations about the learning outcomes. This intends to help students not only to be equipped with the most fundamental knowledge about the discipline but also to learn materials that are in alignment with their future career plans. The first stage teaches all students the same fundamental and common knowledge bases about a course. Afterward, the second stage leverages flipped learning and provides two options for the students. For students who have a clear idea about their future, they can pick up the most valuable components from this flipped learning for their future career path, while maintaining a satisfactory understanding of the other pathway. Also, students who are still in the stage of preparing for their future will have a clearer picture of what the two pathways look like, and they will select an appropriate pathway based on their interests and talent.

We use a before and after-analysis in a transport engineering course to test the effectiveness and recognition of the proposed teaching strategy. The results show that there are improvements in students' subjective evaluations of the course and their objective performance in learning knowledge and exams after implementing the proposed strategy as compared to the conventional teaching process. The empirical results indicate positive effects of the two-stage structured teaching philosophy on students' subjective satisfaction, learning outcomes, and performances.

References

1. Sander, P., Stevenson, K., King, M., Coates, D.: University students' expectations of teaching. Stud. High. Educ. **25**(3), 309–323 (2000)
2. Hassel, S., Ridout, N.: An investigation of first-year students' and lecturers' expectations of university education. Front. Psychol. **8**, 2218 (2018)
3. Stinebrickner, T., Stinebrickner, R.: Learning about academic ability and the college dropout decision. J. Law Econ. **30**(4), 707–748 (2012)
4. Williams, B., Alias, M.: Strategic pathways to engineering education research: case study of a top-down initiative. In: Proceedings of the Research in Engineering Education Symposium (2011)
5. Gerwel Proches, C.N., Chelin, N., Rouvrais, S.: Think first job! Preferences and expectations of engineering students in a French "Grande Ecole". Eur. J. Eng. Educ. **43**(2), 309–325 (2018)
6. Zeivots, S., Schuck, S.: Needs and expectations of a new learning space: research students' perspectives. Aust. J. Educ. Technol. **34**(6) (2018)
7. Gao, K., Yang, Y., Li, A., Li, J., Yu, B.: Quantifying economic benefits from free-floating bike-sharing systems: a trip-level inference approach and city-scale analysis. Transp. Res. Part A: Policy Pract. **144**, 89–103 (2021)
8. Qu, X., Yu, Y., Zhou, M., Lin, C.-T., Wang, X.: Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach. Appl. Energy **257**, 114030 (2020)
9. Zhou, M., Yu, Y., Qu, X.: Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. **21**(1), 433–443 (2019)

10. Schunk, D.H.: Learning Theories an Educational Perspective, 6th edn. Pearson (2012)
11. Bada, S.O., Olusegun, S.: Constructivism learning theory: a paradigm for teaching and learning. J. Res. Method Educ. **5**(6), 66–70 (2015)
12. Vaughan, M.: Flipping the learning: an investigation into the use of the flipped classroom model in an introductory teaching course. Educ. Res. Perspect. **41**, 25–41 (2014)
13. Hung, H.-T.: Flipping the classroom for English language learners to foster active learning. Comput. Assist. Lang. Learn. **28**(1), 81–96 (2015)
14. Armstrong, P.: The CDIO syllabus: learning outcomes for engineering education. In: Rethinking Engineering Education, pp. 45–76. Springer (2007)
15. Leifler, O., Dahlin, J.-E.: Curriculum integration of sustainability in engineering education—A national study of programme director perspectives. Int. J. Sustain. Higher Educ. (2020)
16. Donovan, J., Mader, C.E., Shinsky, J.: Constructive student feedback: online versus traditional course evaluations. J. Interact. Online Learn. **9**(3), 283–296 (2010)
17. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**(1), 68 (2000)
18. Williams, G.C., Deci, E.L.: Internalization of biopsychosocial values by medical students: a test of self-determination theory. J. Pers. Soc. Psychol. **70**, 767–779 (1996)
19. Guay, F.: Applying self-determination theory to education: Regulations types, psychological needs, and autonomy supporting behaviors. Can. J. Sch. Psychol. **37**(1), 75–92 (2022)

Chapter 20
A Network-Wide Traffic Speed Estimation Model with Gaussian Process Inference

Chen Qiu and Ruo Jia

Abstract Accurate urban road traffic speed analysis and prediction are important for the application of intelligent transportation systems. However, the limited and inefficient traffic state monitoring infrastructure installed on urban roads makes it difficult to monitor the traffic state of an entire network. Moreover, the complex characteristics of urban road networks may lead to difficulties for traditional statistical and traffic flow models in dealing with this type of complex relationship. Therefore, this study proposes a network-wide traffic speed estimation model with full spatial and temporal coverage and selects floating vehicle trajectory data in an actual road network for experiments. The results show that the proposed model can accurately estimate the full spatiotemporal traffic state of a traffic network with only partial data input. This method can be effectively applied to urban road state estimation and can provide a scientific basis for traffic management departments to formulate congestion mitigation strategies.

20.1 Introduction

Traffic congestion has become a common problem in urban megacities and inherently affects the efficiency of the residents' travel and logistics. It is important to accurately estimate and predict traffic conditions to effectively avoid and relieve traffic congestion.

In the past decade, due to the rapid development of new technologies such as artificial intelligence, big data, and the Internet of Things, various terminals and electronic field equipment have generated massive amounts of data, which have penetrated various business fields of the transportation industry to become an important production factor [1–3]. Traditional data processing technologies cannot satisfy

C. Qiu
Zhejiang Institute of Communications Co., Ltd., Hangzhou 310 000, China

R. Jia (✉)
Chalmers University of Technology, 41296 Gothenburg, Sweden
e-mail: ruoj@chalmers.se

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_20

the real-time processing requirements of large-scale data nor mine the important values contained in the data. Data-driven technologies, such as artificial intelligence and big data, as new engines for scientific and technological development, are still in the early stages of exploration and application. As such, they are not expected to bring about major changes in transportation system simulation and deduction, prediction, and decision-making in the near future [4–6]. In particular, under the impetus of traffic big data, predicting traffic flow and road traffic congestion status has become a research focus for scholars in the field of transportation in recent years [7]. Traffic congestion prediction is based on the extraction of prior knowledge from relevant historical data and other environmental factors to estimate future traffic flow states.

Urban road networks require real-time traffic state monitoring to achieve effective urban traffic control and management. However, traffic state monitoring infrastructure (such as traffic detectors) installed on urban roads is limited and inefficient, which makes it difficult to monitor the traffic state of the entire network [8]. The analysis of the urban road traffic state is mainly divided into methods based on statistics and methods based on traffic flow theory. However, most only focus on inferring the road state of a single road segment, so it is difficult to estimate the traffic state over the entire time and space [9].

Williams et al. [10] used the exponential smoothing method to model traffic flow data to predict traffic flow on urban roads. Based on the periodicity of traffic flow evolution, Ding et al. [11] proposed an autoregressive comprehensive moving average (STARIMA) model based on time and space to predict the short-term traffic flow change of the city over the next five minutes.

Urban road networks exhibit complex characteristics. For example, owing to the limitations and differences in lanes and channelization, very complex network topology relationships may exist between very short sections [12]. Traditional statistical and traffic flow models are difficult to deal with in complex relationship modeling. However, owing to the limitations of traffic state data, it is necessary to infer the traffic states of spatially separated locations from the observed road traffic states. Therefore, data-driven Bayesian inference that can describe random processes and traffic state propagation is very important, and more complex statistical models are needed. Urban road traffic network systems have complex nonlinear structures and are restricted by various environmental variables such as weather and traffic accidents [13]. Therefore, traditional data-driven prediction algorithms are unsuitable for real traffic scenarios. As a mainstream data-driven prediction method, the artificial neural network prediction method has attracted much attention in the field of traffic flow prediction. Generally, the artificial neural network method does not need to build a complex physical model, but only needs to use rich historical data to build the key characteristics of traffic flow so that the model can effectively learn and predict the short-term traffic flow in the future. Kumar et al. [14] considered the key characteristics of traffic flow, such as traffic flow, speed, density, and time, as input variables and then used an artificial neural network (ANN) model for training to predict short-term traffic flow. In 2015, Ma et al. [15] introduced the deep learning

theory to predict large-scale traffic congestion. Zhang et al. [16] designed an end-to-end structure of ST-Resnet with unique attributes to process spatiotemporal traffic flow data. The model adopts the residual network architecture with the attributes of simulating the time proximity and periodicity of traffic flow and integrates the residual network and convolution network to predict the traffic flow in each area of the road.

In summary, traditional traffic state estimation methods based on statistical models ignore the spatial attributes of urban road networks. Traffic flow theory research, represented by basic traffic flow maps and macro basic maps, ignores the time variability and randomness of traffic networks, and it is difficult to estimate the traffic state of the entire road network under the condition of spatiotemporal sparse data. Therefore, there is an urgent need for an urban road network traffic estimation method that can reasonably model the topological characteristics of the traffic network and the propagation and inference of the traffic state with the aim of estimating the traffic state covered by the entire time and space of the traffic network under partial data input.

20.2 Problem Statements

For the entire space–time traffic state estimation problem, the model needs to consider the static topology of road network information and obtain the macro-spatial information of the urban road network to better estimate the relationship of traffic state propagation between road segments. In this study, a Gaussian process Bayesian structure was constructed. The correlation between road segments is characterized by conditional Bayesian conditional probability, and graph information is introduced into the kernel function of the Gaussian process through the graph Laplace operator to better study the propagation of traffic state between road segments and road segments within the network.

20.2.1 Gaussian Process

The Gaussian process is an effective method for data-driven modeling [17]. The Gaussian process model is suitable for analyzing problems with uncertainty. It can effectively approximate any continuous function and encode it reasonably in combination with the prior information. For example, by selecting different covariance functions, different degrees of difference or specific functions can be coded, such as periodicity and symmetry.

Assuming X^n is the data set, for random functions $f : X \to \mathbb{R}$, a Gaussian process is defined as $f \sim GP(\mu, k)$, where $\mu(\cdot)$ represents the mean value of the function, and $k(\cdot, \cdot)$ represents the kernel function. For the points in the dataset $\mathbf{x} \in X^n$, the random vector $f(\mathbf{x})$ is a multivariate Gaussian vector with mean vector $\boldsymbol{\mu} = \mu(\mathbf{x})$

and covariance matrix $\mathbf{K_{xx}} = k(\mathrm{x}, \mathrm{x})$. Without loss of generality, we assume that the prior mean μ is zero.

20.2.2 Variational Inference

Although prior and posterior Gaussian processes have great flexibility, they have two main defects. First, if the likelihood function of the kernel function is non-Gaussian, the posterior Gaussian process cannot be calculated analytically. Second, the computational complexity of the Gaussian process is very high, reaching $O(N^3)$, where N is the number of training data points, making the model unsuitable for large datasets.

To solve these problems, modern variational inference introduces a set of induced points $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_M]^T$ by constructing set M, in which $\mathbf{z}_m \in \mathbb{R}^{D\times 1}$ provides an effective solution. The induced point $\mathbf{u} = [f(\mathbf{z}_1), \ldots, f(\mathbf{z}_M)]^T$ is a variational parameter, which is a set of subsets of random variables of a subset for Gaussian function $f(\mathbf{x})$. By assuming that $m(\mathbf{x})$ is zero, the Gaussian process under this condition can be written as

$$f(\mathbf{x})|\mathbf{u} \sim GP\ominus(\mathbf{k}_{zx}^T\mathbf{K}_{zz}^{-1}\mathbf{u}, k_\theta(\mathbf{x}, \mathbf{x}) - \mathbf{k}_{zx}^T\mathbf{K}_{zz}^{-1}\mathbf{k}_{zx}) \tag{20.1}$$

where $\mathbf{k}_{zx} = [k_\theta(\mathbf{z}_1, \mathbf{x}), \ldots, k_\theta(\mathbf{z}_M, \mathbf{x})]$, $[\mathbf{K}_{zz}]_{ij} = k_\theta(\mathbf{z}_i, \mathbf{z}_j)$.

It can be deduced that $p(\mathbf{u}) = \mathcal{N}(0, \mathbf{K}_{zz})$, and the variational posterior distribution $q(\mathbf{u})$ of the induction point $\mathbf{u}$ obeys the Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$. According to the theory of variational inference, the evidence lower bound (ELBO) of the induction point is

$$\mathcal{L}(\theta, \mathbf{Z}, \mathbf{m}, \mathbf{S}) = \sum_{n=1}^{N} \mathbb{E}_{q(f(\mathbf{x}_n))}[\log p(y_n|f(\mathbf{x}_n))] - \mathrm{KL}[q(\mathbf{u})||p(\mathbf{u})] \tag{20.2}$$

Therefore, the variational distribution $q(f(\mathbf{x}_n))$ can be easily calculated by solving ELBO.

20.3 Case Study

This section evaluates the proposed full spatiotemporal estimation algorithm and compares it with the mainstream deep learning space–time estimation models in practical application scenarios, including GCN [18], DCNN [19], and DeepWalk [20].

20.3.1 Data Description

Experimental data were obtained from the PEMs Bay dataset in California, USA. It is the urban road network traffic data collected by the California Department of Transportation and includes 325 nodes and 2369 road segments. The time span was six months (January 1, 2017 to May 31, 2017). The time interval of the data was 5 min, and the average speed of traffic flow at each node was recorded. This experiment obtains and matches the Open Street Map road network according to the node data, establishes the adjacency matrix and calculates the Laplacian. In the model calculation and result analysis, 150 and 175 nodes were randomly selected as the test and verification sets, respectively. The experiment was repeated 10 times to calculate the mean value.

20.3.2 Model Prediction Results and Analysis

To evaluate the performance of the traffic flow speed prediction algorithm in this study, percentage error (MAPE) of Graph Gaussian Process (GGP) was used as the evaluation index. The smaller the MAPE value is, the higher the prediction accuracy, and the stronger the feature expression ability of the model.

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\frac{\left|y_i - \hat{y}_i\right|}{y_i} \times 100\%, \tag{20.3}$$

where y_i and $\hat{y}_i$ represent the actual vehicle speed and predicted vehicle speed, respectively, and m is the dimension of the dataset.

As shown in Table 20.1, in terms of road state estimation performance, the Gaussian process considering the graph information proposed in this study is more effective and accurate than other estimation models. For different kernel functions, the GGP model considering the Laplace operator has the best performance, and the MAPE reaches 15.3%; it can better estimate the traffic situation over time and space. However, only through GGP, the model whose kernel functions are RBF and Matérn has a general effect and fails to achieve the effect of the GCN.

Figure 20.2 shows the predicted mean value of the entire time and space vehicle speed in the entire road network, and Fig. 20.3 shows the predicted standard deviation of the entire time and space vehicle speed. Noticeably, the model proposed in this study can effectively capture global dependence. Moreover, the Bayesian conditional probability can represent the correlation between road segments, which allows the model to estimate the distribution of traffic states. Meanwhile, it can also be seen from Fig. 20.3 that for areas with long node distances, the impact of road network topology information is small, resulting in a higher estimation of the standard deviation by the model.

Table 20.1 Model performance

Model	MAPE (%)
GGP-RBF	21.2
GGP-Matérn	19.6
GGP-Laplacian	15.3
GCN	18.3
DCNN	23.1
DeepWalk	33.8

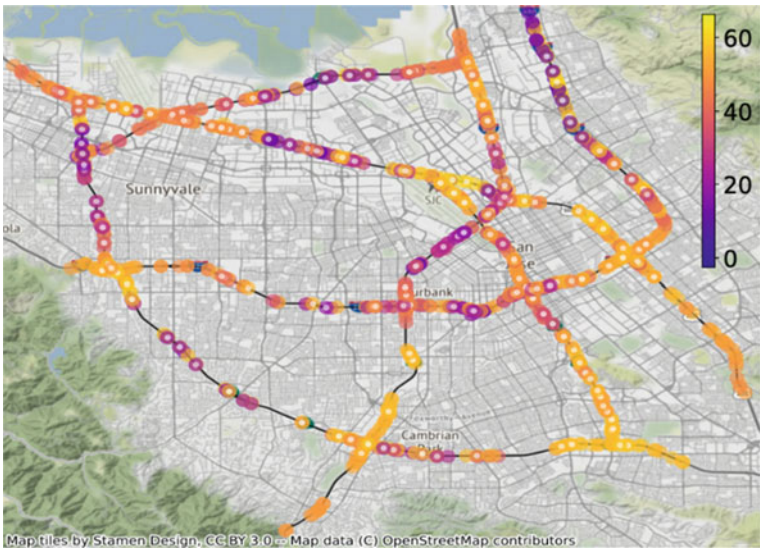


Fig. 20.2 Estimated mean value of the whole time and space vehicle speed

20.4 Conclusion

For the entire space–time traffic state estimation problem, the static topology of road network information must be considered to estimate the comprehensive traffic state as much as possible through limited data input. In this study, a Gaussian process based on a Bayesian structure was constructed. The correlation between road segments is characterized by conditional Bayesian conditional probability, and graph information is introduced into the kernel function of the Gaussian process through the graph Laplace operator to better study the transfer of traffic state between road segments and road segments within the network. The analysis of PEMs traffic data revealed that the predicted value of the model proposed in this study was relatively consistent with the actual road traffic flow state. This method was superior to the GCN, DCNN, and DeepWalk methods. This is an effective traffic state estimation and prediction model.

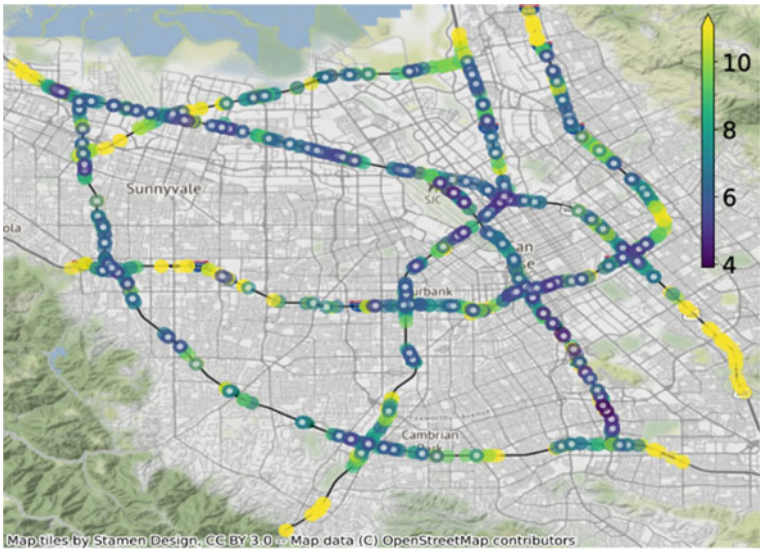


Fig. 20.3 Estimated standard deviation of the whole time and space vehicle speed

This study only analyzed the traffic state and propagation law in the network from the perspective of average travel speed. However, to determine the relationship between network nodes and edges, this study constructs a Laplace operator only by distance, lacks the analysis and modeling of the characteristics of the actual road speed limit and intersection signal control, and relies excessively on training data, resulting in a poor generalization effect. In future research, by incorporating traditional traffic flow theory into the existing approach, we will explore an accurate deduction system of complex traffic systems integrating data-driven and model-driven methods to realize efficient engineering applications.

References

1. Li, Y., Zhao, L., Gao, K., An, Y., Andric, J.: Revealing driver psychophysiological response to emergency braking in distracted driving based on field experiments. J. Intell. Connect. Vehic. **5**(3), 270–282 (2022)
2. Gao, K., Yang, Y., Qu, X.: Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. Transp. Res. Part D: Transp. Environ. **97**, 102957 (2021)
3. Ma, Z., Zhang, P.: Individual mobility prediction review: data, problem, method and application. Multimodal Transp. 100002 (2022).https://doi.org/10.1016/j.multra.2022.100002
4. Liu, Z., Wang, Z., Cheng, Q., Yin, R., Wang, M.: Estimation of urban network capacity with second-best constraints for multimodal transport systems. Transp. Res. Part B: Methodol. **152**, 276–294 (2021)
5. Liu, Y., Wu, F., Lyu, C., Li, S., Ye, J., Qu, X.: Deep dispatching: a deep reinforcement learning approach for vehicle dispatching on online ride-hailing platform. Transp. Res. Part E: Logist. Transp. Rev. **161**, 102694 (2022)

6. Qiu, J., Huang, K., Hawkins, J.: The taxi sharing practices: matching, routing and pricing methods. Multimodal Transp. 100003 (2022)https://doi.org/10.1016/j.multra.2022.100003
7. Cheng, Q., Liu, Z., Lin, Y., Zhou, X.S.: An s-shaped three-parameter (S3) traffic stream model with consistent car following relationship. Transp. Res. Part B: Methodol. **153**, 246–271 (2021)
8. Liu, Z., Liu, Z., Fu, X.: Dynamic origin-destination flow prediction using spatial-temporal graph convolution network with mobile phone data. IEEE Intell. Transp. Syst. Mag. PP(99), 2–15 (2021)
9. Yang, L., Ruo, J., Xue, Xie, et al.: A two-stage destination prediction framework of shared bicycles based on geographical position recommendation. IEEE Intell. Transp. Syst. Mag. **11**(1), 42–47 (2018)
10. Williams, B.M., Durvasula, P.K., Brown, D.E.: Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. Transp. Res. Rec. **1644**(1), 132–141 (1998)
11. Gao, K., Yang, Y., Zhang, T., Li, A., Qu, X.: Extrapolation-enhanced model for travel decision making: An ensemble machine learning approach considering behavioral theory. Knowl.-Based Syst. **218**, 106882 (2021)
12. Huang, D., Liu, Z., Liu, P., Chen, J.: Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. Transp. Res. Part E: Logist. Transp. Rev. **96**, 1–19 (2016)
13. Huang, D., Xing, J., Liu, Z., An, Q.: A multi-stage stochastic optimization approach to the stop-skipping and bus lane reservation schemes. Transportmetrica A: Transp. Sci. **17**(4), 1272–1304 (2021)
14. Kumar, K., Parida, M., Katiyar, V.: Short term traffic flow prediction for a non urban highway using artificial neural network. Procedia Soc. Behav. Sci. **104**, 755–764 (2013)
15. Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y.: Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. Part C: Emerg. Technol. **54**, 187–197 (2015)
16. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Thirty-first AAAI Conference on Artificial Intelligence (2017)
17. Williams, C.K., Rasmussen, C.E.: Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA (2006)
18. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks (2016). arXiv preprint arXiv:1609.02907
19. Jia, R., Chamoun, R., Wallenbring, A., Advand, M., Yu, S., Liu, Y., Gao, K.: A spatio-temporal deep learning model for short-term bike-sharing demand prediction. Electron. Res. Archiv. **31**, 1031–1047 (2023)
20. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710 (2014)

Chapter 21
Rule-based Recommendation System for Traffic Congestion Measures

Yasmine Amor, Lilia Rejeb, Nabil Sahli, Wassim Trojet, Ghaleb Hoblos, and Lamjed Ben Said

Abstract Traffic congestion has become a serious concern in both developed and developing countries. Increasing demand for urban transport has led to plenty of issues including longer travel times, higher fuel consumption and greater vehicular crash rates and therefore to a deterioration in the quality of life. On grounds of the wide range of problems that traffic congestion can cause, the study of traffic congestion measures and their implementation is a crucial step that should be considered in analyzing traffic. However, these measures might vary per country. They are context-sensitive. Therefore, the purpose of this study is to develop a recommendation system able to generate the congestion measures in accordance with the context under study. The goal of this research is to assist researchers and traffic operators to choose the most suitable congestion measures to the studied area.

21.1 Introduction

Traffic congestion has been getting worse throughout many countries of the world representing an undeniable threat to the quality of urban life. Traffic congestion mainly manifests as a gradual slowing down of traffic speed, which increases travel time, fuel consumption, and environmental pollution when compared to a continuous traffic flow. That's why, the study of traffic congestion is growing in importance

Y. Amor (✉) · L. Rejeb · L. Ben Said
Institut Supérieur de Gestion de Tunis, SMART Lab, Université de Tunis, 41 Avenue de la Liberte Bouchoucha, Bardo 2000 Tunis, Tunisie
e-mail: yasmine.amor@isg.u-tunis.tn

N. Sahli
Computer Science Department, German University of Technology (GUtech), Athaibah P.O. Box 1816, Muscat, Sultanate of Oman

W. Trojet
Higher College of Technology, Abu Dhabi, UAE

Y. Amor · G. Hoblos
IRSEEM, Technopole du Madrillet, Av. Galilee, Saint-Etienne du Rouvray, 76800 Normandie, France

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9_21

within the field of Intelligent Transportation Systems (ITSs). To assess the severity of traffic congestion, several congestion measures have been developed. We classify them into five categories: Speed, Travel Time, Delay, Level of Services, and Congestion Indices. They depend on various factors and might vary per country. There is no unique universal method of measuring road traffic conditions. In different countries, even in different states of a country, different measures can be used [1]. Congestion measures are indeed context-sensitive. In this paper, we use the term "context" to refer to the set of circumstances that result into a traffic congestion. Examples of circumstances include "heavy rain," "bad road infrastructure," or "accident". Consequently, choosing the most appropriate measure to apply in a road traffic analysis can be challenging. In fact, deciding on the congestion measure is required in a preliminary stage when studying traffic. To the best of our knowledge, previous studies related to traffic congestion relief have not established the relationship between the studied contexts and the congestion measures. Therefore, our aim in this study is to find out the relation between these two different axes. In light of this, we developed a recommendation system that assists researchers and traffic operators in deciding which measure is the most suitable to the context they are dealing with.

The paper is organized as follows: Sect. 21.2 presents a brief literature review of the existing congestion measures and a representation of the most common contexts. Section 21.3 describes the rule-based recommendation system and its functioning. Section 21.4 shows some given results. Finally, Sect. 21.4 presents the conclusions derived from this study and some perspectives.

21.2 State of the Art

Several measures have been developed to assess congestion. We propose to classify them into five main categories: Speed, Travel Time, Delay, Level of Services, and Congestion indices. Some studies used the Speed as congestion measure [2–7]. They considered diverse contexts including the time of the day, the infrastructure's condition, the climate, the presence of bottlenecks, etc. Malecki et al. [8] used both Speed and Travel Time measures. The authors conducted their research in Szczecin, Poland, considering in their context on-street car parking. They studied the behaviour of various drivers in the process of on-street parking and focused on three major behaviors: patience, perceptiveness, and the usage of indicators like turn signals. They developed a multi-agent system and simulated traffic using Cellular Automata (CA). Experiments' final results indicated that impatient drivers exacerbate congestion with a 19% increase in Travel Times. The Speed measure was also used by Boni et al. [2] who studied the effects of severe weather and driver behavior on traffic congestion. They dealt with the context of a severe rainstorm, which is very frequent in urban areas across many countries. Rainstorms lead to road waterlogging and poor visibility, which influences driving behavior and results in traffic congestion. These conditions also lead to potential accidents that will in their turn cause congestion.

Vehicles' speeds were calculated under various water depths and visibility conditions. Another study was conducted by Elleuch et al. [9] who used as well Speed and Travel Time to estimate traffic congestion. They developed an Intelligent Traffic Congestion Prediction System (ITCPS) based on a Neural Network. In addition to Speed measures, Travel Time is also a persistent measure that has been widely used [8, 10–14]. Farrag et al. [10] developed a micro-simulation model studying the local area's network and the driver behaviours. They involved three sub-models: a car-following model, a lane-changing model, and a gap-acceptance model. The authors treated the congestion throughout peak morning and evening hours every day. Vlieger et al. [11] were concerned by a context with a bad infrastructure considering the driver's behavior. They studied aggressive versus normal driver's behaviors based on fuel consumption, and deduced that this latter increased with up to 40% for aggressive driving compared to normal driving. Zhibin et al. [13] used reinforcement learning in Variable Speed Limit (VSL) control strategies in order to reduce Travel Time. Two scenarios were evaluated for both stable and fluctuating traffic demands at freeway bottlenecks. Another study was performed by Zhang et al. [14] who examined the early and evening rush hours, while considering bottlenecks congestion. The authors employed the Delay and the Level of Services measures in addition to the Travel Time. A number of other studies used the Delay as congestion measure [3, 5, 6, 15]. For example, Hamzah et al. [15] studied the context of car accidents that present a primary cause of traffic congestion. They investigated the human behavior and its impact on this context. The final results and analysis helped them in developing new traffic rules and policies, in order to prevent accidents, and improve roads safety. Other studies used the Level of Services as congestion measure [16, 17]. For instance, Ulfi et al. [16] studied various types of aggressive driving behaviors mainly improper speed, inattentiveness, display of hostility, impatience, and disobedience of traffic signals. Final results showed that driver's impatience is the behavior that must be addressed in order to alleviate traffic congestion in Bandung, Indonesia. Congestion was measured by the value of volume-capacity ratio, based on the data from Dishub Kota Bandung and from real time GPS tracking. Ichsan et al. [17] analyzed the level of service and the policy directions of Majene Axis Road of Campalagian Market in Polewali, Mandar. The result revealed that the primary cause of traffic jams is irregular parking.

To sum up, several traffic congestion measures were used in the literature. Each with its own benefits and drawbacks. However, the use of congestion measures may vary from one study to another. Due to the wide variety of congestion measures, researchers may struggle to decide which ones are the most suitable for their case study. Upon reviewing the studies mentioned in this section, it becomes evident that the selection of the measure is strongly linked to the addressed context. The contexts themselves are made up of a variety of parameters such as the weather conditions, the driver's behaviors, the roads infrastructure, the hour of the day, etc. To the best of our knowledge, most existing systems addressed particular contexts. None has come up with a generic framework that may work for any traffic context. Our aim is thus to develop a recommendation system to help choose the most relevant measures for

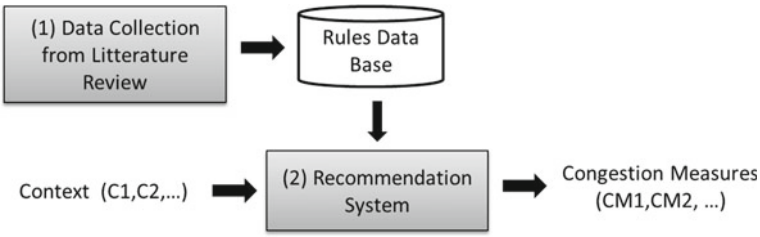


Fig. 21.1 Inputs and outputs of the proposed recommendation system

each particular context. In the next section, we present our recommendation system and its functioning.

21.3 Proposed Approach

We propose a rule-based system that relies on the similarity of contexts in order to make recommendations. Figure 21.1 illustrates the general architecture of our rule-based recommendation system. In fact, two distinct phases are represented here. The first one is the data collection phase which is used to build the basis of our set of rules. The second one concerns the development of the recommendation system itself. This latter takes the context that we want to study as input, analyzes it and computes the similarities with the existing information in the data base and then generates the possible congestion measures that best fit the input context. To describe this input context, and after analyzing the different circumstances from the literature review, we categorize traffic contexts based on nine parameters (bottlenecks, hour of the day, day of the week, roads infrastructure, weather, work zones, incidents, Special events, and drivers behavior). Consequently, each context will be represented in our system by a weighted sum of these nine parameters. It is worth mentioning here that some context parameters may cause more harm than others and are then considered as more important. Besides, a context parameter may have more impact on traffic congestion in one country than in another country. For example, in countries with poor infrastructure, the parameter "Road Infrastructure" will be considered as the most important. In other countries characterized by bad weather conditions (snowy weather for example), the weather parameter will be considered as the most important one. This makes it impossible to come up with a universal model, applicable in any situation. However, we can determine the most commonly used congestion measures, which is the aim of our recommendation system.

In the following, we present the data collection process, show the functioning of our recommendation system and discuss the obtained results.

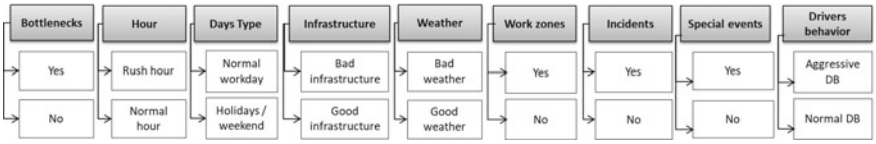


Fig. 21.2 Context representation

21.3.1 Data Collection

The data base holds the domain knowledge necessary for the problem-solving. In our study, data were gathered from the literature review that we conducted. We developed a form for context representation that provides a detailed and precise depiction of environments.

Figure 21.2 shows the considered parameters in our study. This representation helps providing a relatively accurate perception of the studied environments, capable of combining different context parameters that may be present at the same time. We also take into consideration the weight of each context parameter in order to get closer to reality and to focus on the most pertinent parameters to the studied area.

Actually, we have 9 context parameters and the final context is the combination of all of them. Context parameters are represented by $(C_1, \ldots, C_9)$ where

C_1: The presence of bottlenecks
C_2: The hour of the day
C_3: The day of the week
C_4: The roads infrastructure
C_5: The the weather conditions
C_6: The presence of work zones
C_7: The presence of incidents
C_8: The presence of Special events
C_9: The drivers behavior

A given context represented by all these context parameters is given as input to the system. This latter analyzes it and gives in return the congestion measures that can be used, with different weightings, where the measure having the highest weight is the most appropriate one. The output of the system is presented by $(CM_1, \ldots, CM_5)$ where

CM_1: Speed
CM_2: Travel Time
CM_3: Delay
CM_4: Level of Services
CM_5: Congestion Indices

In a rule-based system, the knowledge is represented as a set of rules. Each rule specifies a relation or recommendation and is represented by an IF (condition) THEN (action) form. The rule is activated and the action part is performed as soon as the condition clause of the rule is fulfilled.

As shown in Fig. 21.3, our system's rules are represented by the following form:

IF Condition (Context 1 AND Context 2 AND . . . AND Context n) THEN Action (Congestion Measure1 OR . . . OR Congestion Measure n)

The condition is the combination of the considered context parameters and the action

Fig. 21.3 Condition—action rule representation

is the possible congestion measures.

The next section is dedicated to the functioning of the system.

21.3.2 System Reasoning and Context Analysis

Our system takes as input the considered contexts, which should be entered by a system operator. This latter has to assign a different weighting to the various context parameters (depicted in Fig. 21.2) depending on their relative importance. The system captures the input and processes it. The principle is to link the rules that are defined in the knowledge base with the input context. To do so, we calculate the similarity between this Input Context and the different contexts of our rules. In particular, we use Cosine similarity measure. It calculates the similarity between two vectors of an inner product space. Cosine similarity is measured by the cosine of the angle between these two vectors and determines whether the vectors are pointing to the same direction. We opt for the use of cosine similarity because even if two data objects are far apart according to the Euclidean distance, still there could be a slight angle between them.

The first vector is related to the new input context and the second one is related to the context of the considered rule from our data base.

We present below the equation of the cosine similarity measure.

$$SimCos(V1, V2) = V1 * V2/(||V1|| * ||V2||) \tag{21.1}$$

where V1 and V2 are the two vectors.

V1 = (C_{11},C_{12}, . . . ,C_{1n}), where C_{11}, C_{12}, . . ., C_{1n} are the contexts related to V1. V2 = (C_{21},C_{22}, . . ., C_{2n}), where C_{21},C_{22}, . . ., C_{2n} are the contexts related to V2.

The scalar product of the vectors V1 and V2 is given by (2).

$$V1 * V2 = C_{11} * C_{21} + C_{12} * C_{22} + \cdots + C_{1n} * C_{2n} \tag{21.2}$$

The norm of the vector V1 is

$$||V1|| = SQRT(C_{11} * C_{11} + C_{12} * C_{12} + \cdots + C_{1n} * C_{1n}) \tag{21.3}$$

The norm of the vector V2 is

$$||V2|| = SQRT(C_{21} * C_{21} + C_{22} * C_{22} + \cdots + C_{2n} * C_{2n}) \tag{21.4}$$

After calculating the similarities between the input context and the different contexts of our rule base, we consider all the rules for which the similarity is greater than 0.7. We chose this threshold in order to ensure the greatest number of measures with a meaningful similarity. For each of these selected rules (named winning rules), we attribute scores to the different congestion measures. The score attributed to each measure is the average of the measures existing in the winning rules. These scores are calculated in order to determine the most used congestion measures for the given input context.

21.3.3 Proposed Algorithm

The selection of the most appropriate congestion measures for a given context is given by Algorithm 1. In our study, the final context is composed of 9 specific context parameters. Thus, if we have two possibilities for each context, the total number of combinations is 2^9. Therefore, we have 512 possible contexts. On the other hand, for the congestion measures, one or more measure could be used, which makes a total of 2^5-1 combinations.

Algorithm 1 Traffic congestion measures selection Algorithm

Input: New context C($C_1, \ldots, C_9$) + Weightings of the contexts W($W_1, \ldots, W_9$)
Output: The appropriate congestion measures CM($CM_1, \ldots, CM_5$) + Their scores
Begin
$Input_C \leftarrow C_1 * W_1 + \cdots + C_9 * W_9$
V1 $\leftarrow$ (C_{1_1},$C_{1_2}, \ldots, C_{1_n}$)
Initiate (Speed, TT, Delay, LoS, CI)
$WinningRules \leftarrow 0$
FOR i $\leftarrow$ 1 à n /* n is the number of rules in our base */
$V2_i \leftarrow (C_{2_1}, C_{2_2}, \ldots, C_{2_n})$
$CM2_i \leftarrow (CM_1, CM_2, \ldots, CM_5)$
$Similarity \leftarrow$ SimCos (V1, $V2_i$)
IF ($Similarity > 0.7$)
$WinningRules \leftarrow WinningRules + 1$
UpdateScore (Speed , TT , Delay , LoS, CI)
End IF
End FOR
AvgScore (Speed , TT , Delay , LoS, CI , $WinningRules$)
End

After receiving a new context as input, the system generates a first vector V1 related to that context. After that, we initiate all the measures scores to 0. We browse our rules data base, creating for each rule a vector V2 to be compared with vector V1. We calculate the similarity between the two vectors and for each similarity value higher than 0.7, We consider the congestion measures related to the winning rule. Following that, we update the scores of our measures by incrementing their values if they are existing in the winning rule. After that, we calculate the average scores of the congestion measures according to the number of the winning rules. The next section presents some results.

21.4 Experiments and Results

We used Java to implement our system. We created an interface that allows the users to select the parameters they want to take into account and assign each one a rating according to its importance. The process begins once the user enters its context. Table 21.1 depicts the tested scenarios. Nine different scenarios were considered. For each scenario, we assign a high weight to a specific context parameter that we want to focus on.

For the first scenario, we considered *bottlenecks* as our major concern. We attributed the highest weight to *bottlenecks*' context parameter. We also take into account the *hour of the day* (rush hour) and the *type of the day* (normal work day) but with lower weights comparing to the *bottlenecks*. For the other remaining parameters, we considered a good *weather*, a good *road's infrastructure*, a normal *drivers behavior* and an absence of any *incidents*, *work zones*, or *special events*. The generated output includes Speed and Delay as the most adequate measures with a score of 0.37 each. Travel Time could also be used in this context. This can be explained by the fact that bottlenecks are defined by localized sections that experience reduced speeds and

Table 21.1 Scenarios and results

Scenario	Systems input									Systems output				
	Contexts									Congestion measures				
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	CM_1	CM_2	CM_3	CM_4	CM_5
SC 1	0.7	0.15	0.15	0	0	0	0	0	0	0.37	0.12	0.37	0.12	0.02
SC 2	0	0.1	0.2	0.7	0	0	0	0	0	0.41	0.41	0.04	0.14	0.0
SC 3	0	0.1	0.2	0	0.7	0	0	0	0	0.52	0.09	0.34	0.05	0.0
SC 4	0	0.1	0.2	0	0	0	0.7	0	0	0.22	0.22	0.50	0.06	0.0
SC 5	0	0.1	0.2	0	0	0	0	0	0.7	0.39	0.14	0.28	0.14	0.05
SC 6	0.25	0.25	0.25	0.25	0	0	0	0	0	0.38	0.26	0.19	0.1	0.07
SC 7	0	0	0	0	0.2	0.2	0.2	0.2	0.2	0.39	0.12	0.39	0.0	0.1
SC 8	0	0	0.2	0	0	0	0.4	0	0.4	0.3	0.18	0.51	0.0	0.01
SC 9	0	0.3	0.3	0.4	0	0	0	0	0	0.47	0.39	0.12	0.01	0.01

inherent delays due to a recurring operational influence or a non-recurring impacting event.

For the second scenario, we identified a *bad-road infrastructure* as the main focus. A *rush hour* and a *normal work day* were also considered. In this case, the most relevant congestion measures are Speed and Travel Time.

The third scenario is related to a bad *weather* condition. The highest weight is thus attributed to *weather*. Considering a bad weather, the highest score was given to the Speed measure. In fact, due to bad weather such as high wind gust, heavy rains, or snow, both vehicles' speed and volume can be affected [20].

In scenario 4, we paid attention to the *incident* context. Results showed that the most adequate measure is the Delay with a score of 0.5, followed by the Speed and the Travel time with a score of 0.22.

Scenario 5 considered a high weight for the *drivers behavior*. The Speed was the most suitable congestion measure for this context. This can be explained by the related behaviors such as inappropriate speed, acceleration, and deceleration behaviors. In fact, when congestion occurs, drivers assess the situation and adjust their speed accordingly. The Travel Time and Delay measures can also be used.

Indeed, congestion can be classified as recurring or non-recurring. Recurring congestion occurs every day, due to the excessive number of vehicles on roads during rush hours. However, non-recurring congestion is caused by irregular events such as bad weather, work zones, incidents, or special events. We thus identified two other scenarios (6 and 7) in order to test the recurring and the non-recurring contexts, respectively. For scenario 6, we considered all the recurring contexts and none of the non-recurring ones. We associated equal weights for all of them. Result showed that when we deal with recurring congestion, the most used measures are the Speed and The Travel Time.

For scenario 7, we took into account every non-recurring context and none of the recurring contexts. For this case, the Speed and the Delay are the most adequate measures. For the two last scenarios, we simulated the studies of Hamzah et al. [17] and Elleuch et al. [9]. Knowing the measures that they used, our aim was to test the performance of our system. For scenario 8, the output of the system is 0.51 for the Delay measure and 0.3 for the Speed measure. The Travel Time may be used as well, but with a low score. For scenario 9, the system recommends to use the Speed measure in a first place, then the Travel Time in a second place. As discussed in the section "State of the art," these were the measures that the authors used in their studies. Another observation drawn from the presented results is that the Congestion Indices are the less used measures. In [18], this measure has also a very low overall score as deemed to be inaccurate and insensitive to significant changes in road traffic.

21.5 Conclusion and Perspectives

In this paper, we proposed a recommendation system for the optimal traffic congestion measures capable of addressing various contexts. Our main concern was to find out the most relevant congestion measures for each new situation, based on a

rule base. Actually, by taking into account a weighted combination of nine distinct context parameters, we are able to handle a large number of possible contexts.

The weighting of the parameters is required in order to fit in every country's context. As a consequence, the user can focus on the parameters that are most important to him, despite the fact that their importance varies greatly from one country to another. In future works, more detailed context parameters can be addressed like traffic heterogeneity, inadequate traffic controllers, poor lane discipline, breakdowns, collisions, etc. The final aim is to enhance the performance of Intelligent Transportation Systems by choosing the most adequate congestion measures. The effectiveness of our paper is to assist researchers working on traffic congestion management. As recommending the most effective congestion measure depends on the accuracy and reliability of the data used to identify the context, it is important to use the most appropriate sources to collect these data. Our future plan involves expanding the system by providing recommendations on the optimal data sources for each context, in addition to the traffic congestion measures.

Acknowledgements This work was partially supported by The Research Council of Oman [grant number BFP/RGP/ICT/22/327]

References

1. He, F., Yan, X., Liu, Y., Ma, L.: A traffic congestion assessment method for urban road networks based on speed performance index. Procedia Eng. **137**, 425–433 (2016)
2. Su, B., Huang, H., Li, Y.: Integrated simulation method for waterlogging and traffic congestion under urban rainstorms. Nat. Hazards **81**(1), 23–40 (2016)
3. Lu, H., Zhu, Y., Shi, K., Lv, Y., Shi, P., Niu, Z.: Using adverse weather data in social media to assist with city-level traffic situation awareness and alerting. Appl. Sci. **8**(7), 1193 (2018)
4. Sun, S., Chen, J., Sun, J.: Traffic congestion prediction based on GPS trajectory data. IJDSN **15**(5), 1550 (2019)
5. Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X.: Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors **17**(7), 1501 (2017)
6. Wan, Q., Peng, G., Li, Z., Inomata, F.H.T.: Spatiotemporal trajectory characteristic analysis for traffic state transition prediction near expressway merge bottleneck. TR-C **117**, 102682 (2020)
7. Wegerle, D., Kerner, B.S., Schreckenberg, M., Klenov, S.L.: Prediction of moving bottleneck through the use of probe vehicles: a simulation approach in the framework of three-phase traffic theory. J. Intell. Transp. Syst. **24**(6), 598–616 (2020)
8. Malecki, K.: A computer simulation of traffic flow with on-street parking and drivers' behavior based on cellular automata and a multi-agent system. J. Comput. Sci. **28**, 32–42 (2018)
9. Elleuch, W., Wali, A., Alimi, A.M.: Neural congestion prediction system for trip modelling in heterogeneous spatio-temporal patterns. Int. J. Syst. Sci. **51**(8), 1373–1391 (2020)
10. Farrag, S.G., El-Hansali, M.Y., Yasar, A.-U.-H., Shakshuki, E.M., Malik, H.: A microsimulation-based analysis for driving behaviour modelling on a congested expressway. JAIHC **11**, 5857–5874 (2020)
11. De Vlieger, I., De Keukeleere, D., Kretzschmar, J.: Environmental effects of driving behaviour and congestion related to passenger cars. Atmos. Environ. **34**(27), 4649–4655 (2000)
12. Chen, M., Yu, X., Liu, Y.: PCNN: deep convolutional networks for short-term traffic congestion prediction. IEEE Trans. Intell. Transp. Syst. **19**(11), 3550–3559 (2018)

13. Li, Z., Liu, P., Xu, C., Duan, H., Wang, W.: Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks. EEE Trans. Intell. Transp. Syst. **18**(11), 3204–3217 (2017)
14. Zhang, X., Yang, H., Huang, H.-J., Zhang, H.M.: Integrated scheduling of daily work activities and morning-evening commutes with bottleneck congestion. Transp. Res. Part A Policy Pract. **39**(1), 41–60 (2005)
15. Al Najada, H., Mahgoub, I.: Big vehicular traffic data mining: towards accident and congestion prevention. In: IWCMC, IEEE, pp. 256–261 (2016)
16. Mutia, U., Siallagan, M., Putro, U.S., Marsetyawan, M., Alamanda, D.: Investigating aggressive driving behavior in reducing traffic congestion on Bandung city. IJTM **17**(2), 151–165 (2018)
17. Hamzah, M.I., Jinca, M.Y., Hamzah, B.: Traffic jams on majene axis road of campalagian market in polewali mandar regency, west sulawesi, indonesia. Traffic **5**(7), 27–34 (2016)
18. WorldBank: Cairo Traffic Congestion Study: Final Report. World Bank (2013)
19. He, F., Yan, X., Liu, Y., Ma, L.: A traffic congestion assessment method for urban road networks based on speed performance index. Procedia Eng. **137**, 425–433 (2016)
20. Afrin, T., Yodo, N.: A survey of road traffic congestion measures towards a sustainable and resilient transportation system. Sustainability **12**, 4660 (2020)

Author Index

Y. Bie et al. (eds.), *Smart Transportation Systems 2023*, Smart Innovation, Systems and Technologies 356, https://doi.org/10.1007/978-981-99-3284-9