The Spatiotemporal Traffic Accident Risk Analysis in Urban Traffic Network

Chijun Zhang[1,2,5], Jing jin[1,2,5], Qiuyang Huang[3(✉)], Zhanwei Du[2], Zhilu Yuan[4], Shengjun Tang[4], and Yang Liu[1,2,5]

[1] School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun 130117, China
[2] Jilin Province Key Laboratory of Fintech, Jilin University of Finance and Economics, Changchun 130117, China
[3] College of Computer Science and Technology, Jilin University, Changchun 130012, China
`huangqy17@mails.jlu.edu.cn`
[4] Research Institute for Smart Cities and Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, People's Republic of China
[5] Jilin Big Data Research Center for Business, Jilin University of Finance and Economics, Changchun 130117, China

Abstract. Traffic accidents seriously threaten people's lives and property all over the world. Therefore, it is of great significance to human society to have a long-term traffic accidents data with detail temporal and geographic information in a specific space, which can be used for traffic accident hotspots identification to reduce the incidence of traffic accidents. Here, we obtain a one-year dataset of traffic accidents of the city center in Changchun, Northeast China, in 2017. In this paper, we analyze the risk of traffic accident in urban area, and then discover the characteristics of traffic accidents at the temporal and spatial aspect. We construct a traffic network, which takes crossings as nodes and road sections as edges and weighted by the total number of traffic accidents. In addition, we integrate road structure data and meteorological data to explore the characteristics of the traffic network.

Keywords: Traffic accident · Urban traffic network · Risk analysis

1 Introduction

With the development of urbanization and the growth of urban population and vehicles, people are facing many problems such as traffic congestion, air pollution, traffic accidents, etc. Among them, traffic accidents pose a great threat to people's lives and property. According to the World Health Organization's Road Traffic Injury Report, more than 3,400 people die in traffic accidents every day, and tens of millions of people are injured or disabled each year. Therefore, how to reduce the traffic accidents and the corresponding loss has become a hot topic [1–5].

If the early incidents and risk can be recognized and predicted, early warning or intervention could be made to reduce the possibility of traffic accident [6–8]. However, it's difficult to predict the risk of a traffic accident because the causes of road traffic accidents are complex and it can be affected by many factors [9–14], such as driving habits, road structure, bad weather, traffic condition, human mobility, etc. With the development of big data and machine learning, many researchers have focused on using new method to predict traffic accident risk and identify the key factors associated with traffic accident. For example, Chen *et al.* [15] proposed a matrix factorization method to estimate accident risk; Xiong *et al.* [6] combined SVM and HMM method to predict vehicle collision; Chen *et al.* [16] and Zhidan, L. *et al.* [17] utilized deep learning to predict traffic accident risk. However, because of the lack of data associated with traffic accidents, effective prediction of the traffic accident risk dynamically is still a challenge problem.

In this paper, we obtain a one-year dataset of traffic accidents of the city center in Changchun, Northeast China, conducted by an open-data program in Changchun Municipal Engineering Design & Research Institute. Specifically, the dataset contains 5831 traffic accident records and covers 1077 roads in Changchun city center area during one-year period since 1 January 2017. Each record in the raw file contains longitude, latitude and time of the accident. According to the location information, the accidents are divided into four types (crossing, road section, residential area and roundabout), and we match all records to the traffic network, so each record corresponds to a road ID.

In order to explore the influence of different factors and the characteristics of traffic accidents at the temporal and spatial aspect, we construct a traffic network which takes crossings as nodes and road sections as edges, and weighted by the total number of traffic accidents.

2 Original Data Sources

We obtain one-year traffic accident data in a specific space of Changchun city center. The dataset contains 5831 traffic accident records and covers 1077 roads during one-year period since 1 January 2017. Each record in the raw file contains longitude, latitude and timestamp of the accident. In order to explore the influence of different factors and the characteristics of traffic accidents at the temporal and spatial aspect, we integrate road structure data and meteorological data to construct our dataset.

We first count the total number of traffic accidents occurred on each road; And then extract the length of each road, whether it is a one-way lane and the degree of each node from the road structure data; Finally, we construct the traffic network through our dataset by taking crossings as nodes and road sections as edges, and weighted by the total number of traffic accidents.

3 Defining Traffic Network

Traffic network refers to the road system which is composed of various roads in a certain area and interweaves into a network distribution. Through OpenStreetMap platform, we get the road structure data of Changchun city, then take crossings as nodes and road

sections as edges. After that, we integrate the traffic accidents data with road structure data to extract the weight of each edge.

The traffic network is defined as a directed graph $G(V, E, W)$, where V is a set of vertices representing the crossings with accurate longitude and latitude in the road network, and E is a set of edges representing road sections, each edge connects two nodes. W is the weight of each edge, in this paper the weight is defined as the total number of traffic accidents occurred on each edge.

## 4   Spatial and Temporal Aspects Analysis

We construct the weighted traffic network to explore the influence of different factors and the characteristics of traffic accidents at the temporal and spatial aspect, it largely depends on the reliability of the source traffic accidents data. We visualize the location information of traffic accidents data, as shown in Fig. 1.
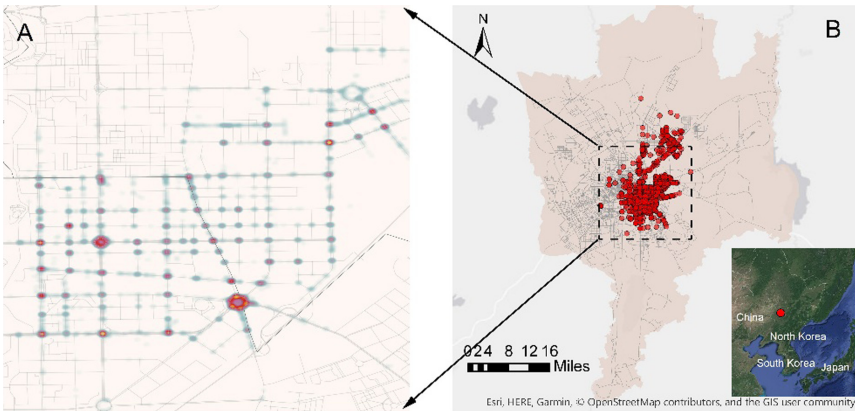
Fig. 1. Geographical distribution and heat map of traffic accidents. (A) The heat map of traffic accidents on the road network. (B) Geographical distribution of traffic accidents in Changchun city.

The heat map of the traffic accident data is shown in Fig. 1A and the geographical distribution of traffic accidents in Changchun city is shown in Fig. 1B. We integrate one-year traffic accident records occurred on the city center in Changchun in 2017. There are totally 5831 traffic accidents occurred on 1077 roads, statistics found that, 54.91% occurred on the crossings, 15.47% occurred on roundabouts, the other occurred on road Sects. (15.29%) and residential area (14.32%) respectively. From Fig. 1B, we can intuitively find that roundabouts and crossings are high-risk areas for traffic accidents.

According to the time of each traffic accident, we count the number of accidents in each hour and find that it is highly similar to people's travel patterns. It can be found that the number of accidents happened in the daytime is significantly higher than that in the night, and there are two peaks, the morning peak (7:00–8:00) and the evening peak (4:00–5:00), as shown in Fig. 2A.
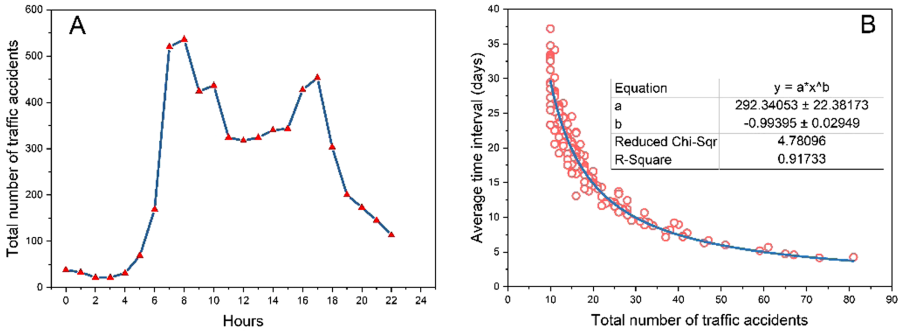
Fig. 2. Temporal aspects of the traffic accidents. (A) The number of traffic accidents happened in different hours. The X-axis represents the 24 h, and the Y-axis represents the total number of traffic accidents happened on each hour in 2017. (B) Average time interval between traffic accidents of each road. The X-axis represents the total number of traffic accidents, and the Y-axis represents the average time interval (days) between the accidents of each road.

Specially, in Fig. 2B, the X-axis represents the total number of traffic accidents, and the Y-axis represents the average time interval (days) between the accidents of each road. We fit it by using the power-law, validated by the chi-square goodness-of-fit test, and find it follows $f(x) = (292.34053 \pm 22.38173) * x^{(-0.99395 \pm 0.02949)}$ with r^2 as 0.91733.

We first analysis the weights of the traffic network. In our dataset, the traffic accidents are distributed on 1077 roads, the weights of the network are defined as the count of traffic accidents in each road. We divide the roads into 5 levels, the red edges represent the road with highest risk of accidents, which happened 41 to 81 traffic accidents in 2017, as shown in Fig. 3A. Then we calculate the percentage of roads with different weights, and fit it by using the power-law distribution, as shown in Fig. 3B. It follows $f(x) = (394.01205 \pm 3.83406) * x^{(-1.28262 \pm 0.01619)}$ with r^2 as 0.99601.

Road structure is an important factor affecting traffic conditions of the traffic network. We calculated the average number of traffic accidents on each road type in 1077 roads, as shown in Fig. 3C, it is obviously higher on truck roads. Finally, we calculate the average number of traffic accidents happened in different kinds of weather conditions. As shown in Fig. 3D, the average number of accidents in moderate snow is obviously higher than others. However, it is lowest in heavy snow, one of the most possible reasons is some roads will be closed when it snows heavily in Changchun.
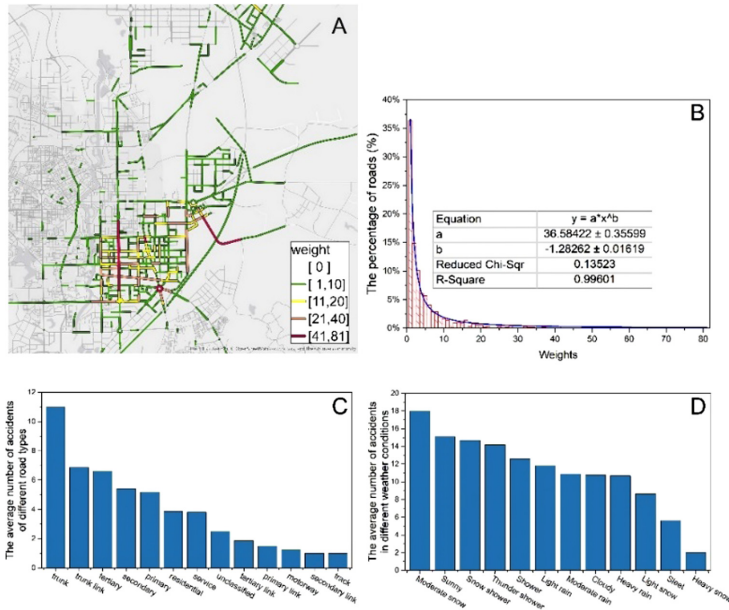
Fig. 3. Network characteristics analysis. (A) The weights of the traffic network. According to the weights, the roads are divided into 5 levels, as shown in the legend. (B) The percentage of roads with different weights in the traffic network. We calculate the percentage of roads with different weights, and fit it by using the power-law distribution. (C) The average number of accidents of different road types. (D) The average number of accidents in different weather conditions.

References

1. Ryder, B., Gahr, B., Egolf, P., Dahlinger, A., Wortmann, F.: Preventing traffic accidents with in-vehicle decision support systems - the impact of accident hotspot warnings on driver behaviour. Decis. Support Syst. **99**, 64–74 (2017)
2. Chung, Y., Recker, W.W.: Spatiotemporal analysis of traffic congestion caused by rubbernecking at freeway accidents. Ieee T. Intell. Transp. **14**, 1416–1422 (2013)
3. Deb, R., Liew, W.C.: Missing value imputation for the analysis of incomplete traffic accident data. Inform. Sci. **339**, 274–289 (2016)
4. Weiming, H., Xuejuan, X., Xie, D., Tieniu, T., Maybank, S.: Traffic accident prediction using 3-D model-based vehicle tracking. IEEE T. Veh. Technol. **53**, 677–694 (2004)

5. Ki, Y.K., Lee, D.Y.: A traffic accident recording and reporting model at intersections. IEEE T Intell. Transp. **8**, 188–194 (2007)
6. Xiong, X., Long, C., Liang, J.: A new framework of vehicle collision prediction by combining SVM and HMM. IEEE T. Intell. Transp. **19**, 699–710 (2018)
7. Cheol, O., Jun-Seok, O., Ritchie, S.G.: Real-time hazardous traffic condition warning system: framework and evaluation. IEEE T Intell. Transp. **6**, 265–272 (2005)
8. An, J., Fu, L., Hu, M., Chen, W., Zhan, J.: A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information. IEEE Access **7**, 20708–20722 (2019)
9. Moriya, K., Matsushima, S., Yamanishi, K.: Traffic risk mining from heterogeneous road statistics. IEEE T Intell. Transp. **19**, 3662–3675 (2018)
10. Eboli, L., Mazzulla, G., Pungillo, G.: Measuring the driver's perception error in the traffic accident risk evaluation. IET Intell Transp Syst. **11**, 659–666 (2017)
11. Yu, R., Abdel-Aty, M.A., Ahmed, M.M., Wang, X.: Utilizing microscopic traffic and weather data to analyze real-time crash patterns in the context of active traffic management. IEEE T Intell Transp. **15**, 205–213 (2014)
12. Zhang, W., Deng, H., Wang, X.: Safety factor analysis for traffic accident scene based on computer simulation. In: Proceedings of the International Conference on Computer Design and Applications, pp. 260–263 IEEE Computer Society (2010)
13. Huang, Q., et al.: The temporal geographically-explicit network of public transport in Changchun City. Northeast China. Sci Data. **6**, 190026 (2019)
14. Du, Z., et al.: The temporal network of mobile phone users in changchun municipality. Northeast China. Sci Data. **5**, 180228 (2018)
15. Chen, Q., et al.: A context-aware nonnegative matrix factorization framework for traffic accident risk estimation via heterogeneous data. In: Proceedings of the IEEE Conference on Multimedia Information Processing & Retrieval, pp. 346–351. IEEE Computer Society (2018)
16. Chen, Q., Song, X., Yamada, H., Shibasaki, R.: Learning deep representation from big and heterogeneous data for traffic accident inference. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pp. 338–344. AAAI (2016)
17. Liu, Z., Li, Z., Wu, K., Li, M.: Urban traffic prediction from mobility data using deep learning. IEEE Netw. **32**, 40–46 (2018)