Predicting Traffic Congestion in the Construction of Smart Cities in the Northwest Industrial Center: A Case Study in Lanzhou of China

Li Ma [1,2,4,a*], Xiaoshi Lv [3,b], Wude Xu[4,c], Lixia Chang [5,d], Yongshun Wang [1,e], Lei Pan[2,f]

[1] School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;

[2] School of Electronic and Electrical Engineering, Lanzhou Petrochemical Vocational and Technical University, Lanzhou 730060, China;

[3] CCCC (Xi'an) Railway Design & Research Institute Co., Ltd, Xi'an 710065, China;

[4] Institute of Sensingg Technology,Gansu Academy of Sciences, Lanzhou 730000, China;

[5] Lanzhou Municipal Administration of Big Data Social Service, Lanzhou 730000, China;

[a] mali@lzpcc.edu.cn, [b] 17793192475@163.com, [c] wude_xu@qq.com,

[d] lzsdsjjclx@163.com, [e] wangysh@mail.lzjtu.cn, [f] 2207942184@qq.com

Abstract. Intelligent transportation is a crucial component in the construction of smart cities. This paper conducts a case study on the traffic conditions at the intersection of Fuli East Road in Xigu District, Lanzhou, an industrial hub in northwest China, during the evening peak hours from 17:30 to 19:00 in June 2022, and discusses the traffic situation in the city. The paper describes the geographical location of the intersection, presents the data sources, and ensures the robustness of the data. Based on 30 days of evening rush hour traffic data, conclusions are drawn about the actual traffic conditions in the area. Using the Random Forest prediction method, the traffic speed and volume for future time periods are forecasted, and the level of traffic congestion during the evening rush hour in the area is assessed. The paper also proposes solutions to alleviate traffic congestion in the construction of smart cities. This paper provides a theoretical foundation for strategy formulation for urban planning and smart transportation development in the region.

Keywords: Smart City; Intelligent transportation; Traffic flow prediction; Random Forest.

1. Introduction

The era of big data has brought new vitality and energy to the development of world transportation, especially the acquisition and prediction of urban transportation big data provides support for the development of intelligent transportation and smart city. Scholars around the world have conducted a wealth of research on smart cities and smart transportation. Some scholars have studied the negative impacts of transportation caused by urban development [1,2,3,4]. And some scholars have also studied the impact of transportation factors on urban development and socio-economics[5,6,7,8]. Other scholars proposed new frameworks to promote smart transportation and smart cities[9,10,11].Recognizing the profound impact of urban transportation on the construction of smart cities, scholars have begun to value the problem of traffic congestion. Applying machine learning methods to intelligent traffic congestion control systems for vehicular networks[12]. A balanced approach between user equilibrium and optimal flow of the system to alleviate urban traffic congestion is introduced[13,14].There are also many innovations in the mathematical and computer methods involved in the research of smart cities and intelligent transportation, including innovations in modeling methods[15,16], The innovations also include the advances in artificial intelligence in the field of intelligent transportation[17,18,19].

In this paper, we filmed a 30-day evening peak traffic video at the intersection in Xigu District of Lanzhou, an industrial hub in northwest China, and converted the video into data through vehicle tracking technology. Based on big traffic data, we analyzed the traffic conditions of this intersection in spatial and temporal dimensions. Using the Random Forest method to predict future traffic conditions, including traffic flow and speed, A comprehensive assessment of traffic congestion was conducted, and suggestions for addressing congestion in the area were proposed, contributing to the development of smart cities through transportation construction in the region.

2. Study area and data sources

2.1 Study area

Lanzhou is one of the major central cities in Northwest China, is an important industrial base and comprehensive transportation hub in northwest China, and serves as a vital node on the Silk Road Economic Belt. The intersection studied in this paper is located in Xigu District, Lanzhou, Gansu Province, China, which is the core industrial area of both Gansu Province and Lanzhou City. It is also the largest petrochemical base in western China, and home to a large refinery Lanlian Group, and a large chemical plant China Petroleum Lanzhou Chemical Industry Company. The industrial hub is based on the industrial construction requirements. The Lanzhou-Xinjiang Railway, an important communicating artery connecting the Chinese mainland and the border region of Xinjiang, is located between the industrial production area and the family living quarter, running through Xigu District from east to west. The north-south traffic in this area mainly relies on the underpass under the railroad or the overpass erected above. The Fuli East Road intersection studied in this paper is located at the north-south connection between the overpass and the underpass, which has a high vehicle flow rate and complicated traffic conditions as a traffic choke point in the region. The geographical location of the study area in China is shown in Fig. 1.

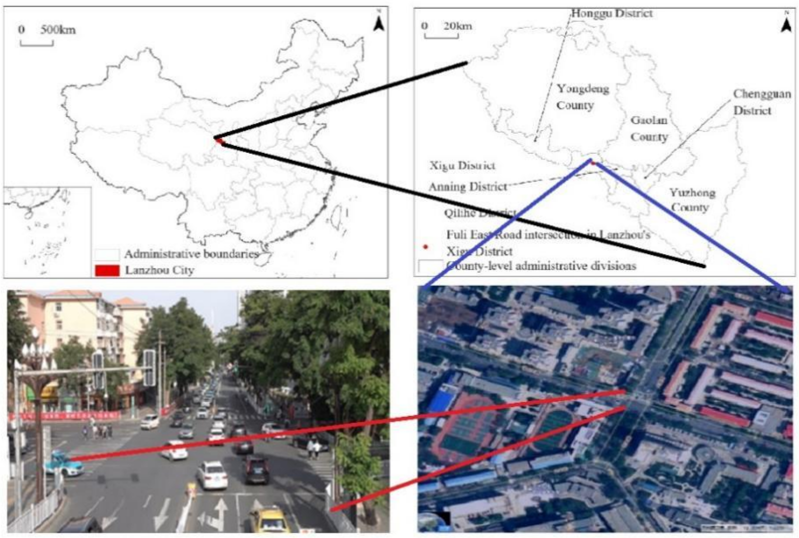


Fig. 1 Study area

2.2 Data sources

We recorded the traffic at the intersection of Fuli East Road in Xigu District of Lanzhou for a period of one month from June 4 to July 3, 2022 at the traffic evening rush hour of 17:30~19:00 every day, and obtained a total of 30 traffic videos in .MP4 format. By software Python and based on vehicle tracking technology, we obtained a total of 30 datasets by the traffic data acquisition approach, each including data features of speed, maximum speed, average speed, and traffic flow. By filtering out the outliers through numerical processing to enhance the robustness of the data, we obtained tens to hundreds of thousands of feature data records for each set.

The acquisition procedure is shown in Fig.2. The traffic data of the Fuli East Road intersection in Xigu District, Lanzhou studied in this paper were converted from the captured traffic video, showing characteristics as below: (1) The speed of the vehicle is obtained by tracking technique based on the position of the vehicle in the traffic video, i.e., the speed of the vehicle is calculated by the ratio of the change in position to the change in time, so it involves the extraction of spatial features; (2) The data are highly time dependent, but the features are relatively homogeneous in form, mainly speed information such as speed, maximum speed, average speed, and traffic flow; (3) The traffic conditions during the evening rush hour from 17:30 to 19:00 every day can be transformed into tens of thousands to more than 200,000 data records by program based on vehicle tracking technique, and the data set obtained after effective processing is still huge.

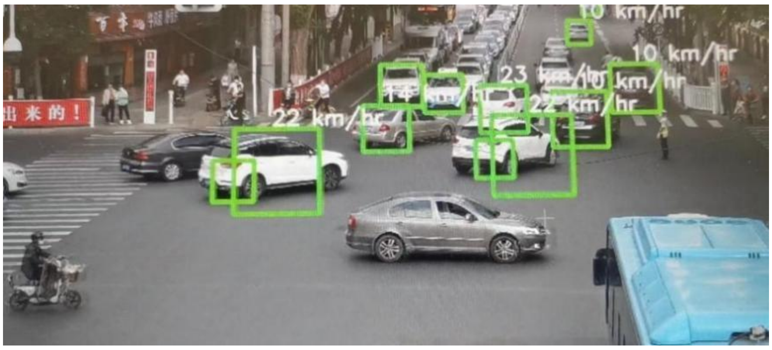
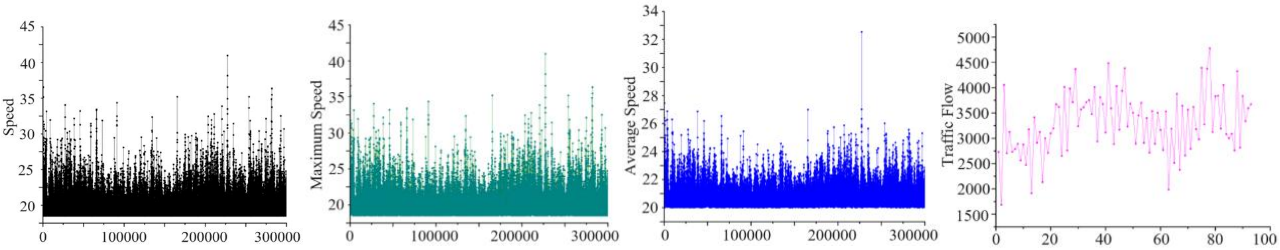


Fig. 2 Conversion of video information to speed by program

3. Analysis traffic conditions

We plotted the daily peak hour traffic at the Fuli East Road intersection in Xigu District, Lanzhou, based on the obtained traffic data to present the day's traffic by speed, maximum speed, average speed, and traffic flow. Fig.3(a) shows the image formed by 230,945 speeds, Fig.3(b) shows the image formed by 77,000 maximum speeds, Fig.3(c) shows the image formed by 57,736 average speeds, and Fig.3(d) shows the image formed by 90 traffic flows. These types of traffic infographics help determine the traffic conditions at the intersection in both the spatial and temporal dimensions.

As for the quantitative definition of traffic congestion, according to the standards published by the Ministry of Public Security of China in 2002, the average speed of motor vehicles on the main roads is mainly used to describe the degree of congestion. Specifically, the average speed of motor vehicles on the main roads of the city not lower than 30 km/h indicates a clear road; the average speed lower than 30 km/h but higher than 20 km/h indicates mild congestion; the average speed lower than 20 km/h but higher than 10 km/h indicates congestion; the average speed lower than 10 km/h indicates serious congestion. In this paper, we classified the traffic conditions into clear road, congestion, serious congestion based on the actual congested road conditions at the intersection during the evening peak hour, and used the vehicle speed as a measurement parameter to make the following definitions: a vehicle speed higher than 23 km/h for a clear road, a speed higher than 22 km/h but lower than or equal to 23 km/h for a congested road; a speed lower than or equal to 22 km/h for a very congested road. In the daily traffic big data, the proportion of clear road data to the total of the day was defined as the "clear traffic" parameter, the proportion of congested condition data to the total of the day was defined as the “congested traffic” parameter, and the proportion of very congested condition data to the total of the day was defined as the “very congested traffic” parameter.



(a)Speed (b) Maximum Speed (c)Average Speed (d) Traffic Flow

Fig. 3 Representation of traffic information

3.1 Traffic control in the region from the perspective of nationally significant days

We studied the traffic data for the whole month of June 2022, when there was one of the most important days for Chinese people the college entrance examination. The traffic measures taken during the 2022 college entrance examination in Xigu District, Lanzhou are shown in Fig 4, the traffic in the region was very congested throughout the month of June during the evening peak (17:30 to 19:00). With the adoption of traffic control on June 6th （The annual college entrance examination in China takes place from June 7th to 9th） to regulate the parking and travel of non-motorized vehicles, the very congested conditions underwent a marked ease. Traffic congestion eased significantly on June 7th and 8th, with very congested conditions returning starting on June 9th. The above conditions indicate that the traffic management measures taken during the college entrance examination period have a significant effect on improving the traffic conditions, but the impact is in real time with a short retardation time.

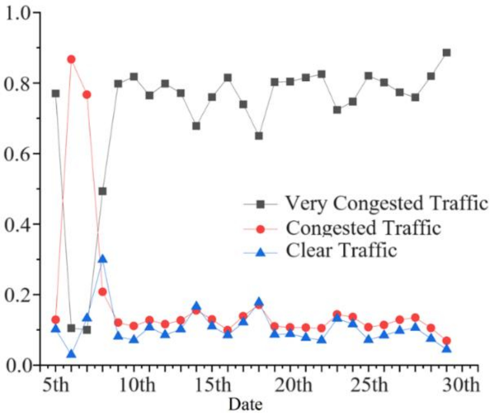
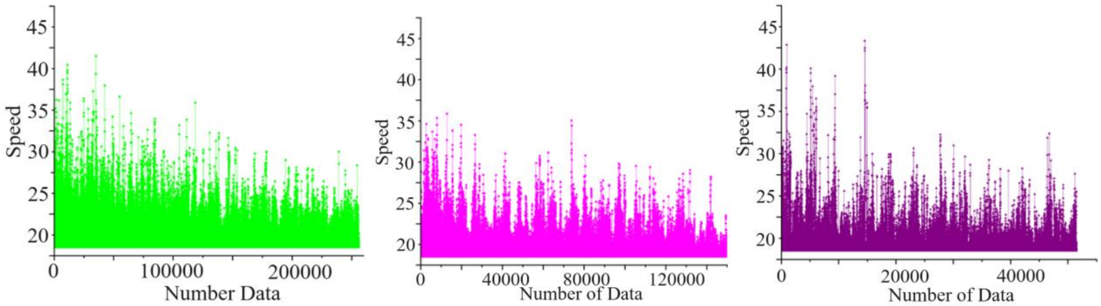


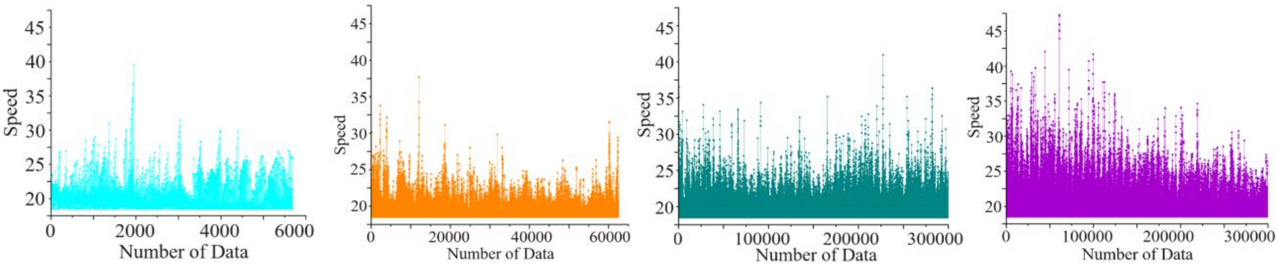
Fig. 4 Summary of traffic conditions in June

3.2 Daily traffic rules based on weekly traffic data

The traffic data for a one-week period from June 5th to June 11th, 2022, and compared them to reveal daily traffic rules, as shown in Fig5. According to the distribution of traffic speed on the pictorial diagram, from June 5th to 11th the speed showed a gradual decrease, indicating that the traffic congestion was getting more serious from 17:30 to 19:00 every day. Fig5 (a) shows that from 17:30 to 18:00 the vehicles ran fast, with speeds higher than 30 km/h concentrated, indicating a clear road and unblocked traffic, from 18:00 to 18:30, the maximum speed of vehicles was decreasing and the traffic condition was changing to congestion, from 18:30 to 19:00, the vehicle showed a tendency of further decline in traveling speed, and they ran at a low speed, indicating that the traffic condition get worsened, mainly shown as congestion and serious congestion. Fig5(b), Fig5(e) and Fig5(f) all show traffic speeds in a wave-like decline. According to the daily traffic speed decline, it can be judged that from 17:30 to 19:00, the traffic got more congested, and the wave-like speed presents that the very congested traffic conditions were characterized by intermittent distribution.



(a)June 5th （Sunday） (b) June 6th（Monday） (c) June 7th（Tuesday）



(d) June 8th（Wednesday）(e) June 9th（Thursday） (f) June 10th（Friday） (g) June 11th（Saturday）

Fig.5 Comparison of traffic conditions for each day of the week

3.3 Potential rules of traffic conditions based on traffic data over one month

According to the data for each day of the 1-month period from June 5th to July 3th, 2022 to the total traffic data, the probability of very congested traffic was greater on Sundays (except July 3th) than on Mondays immediately following, and it showed consistency for four consecutive weeks. According to Fig.6 of comparing the four consecutive weeks of very congested traffic in the month of June 2022, the traffic rules showed no significant changes from Sunday to Wednesday due to the special circumstance of college entrance examination in the first week of the month. The very congested traffic conditions for the second and fourth weeks of the month were distributed in a wavy pattern, and both increase and decrease showed continuity, with the maximum values on Sundays and Thursdays. And the regular changes in traffic conditions in the third week were consistent with the wavy distribution, with the maximum values on Sundays and Wednesdays. The very congested traffic congestion on Thursday and Friday eased significantly, with good traffic on the roads. The specific data are shown in Table 1.

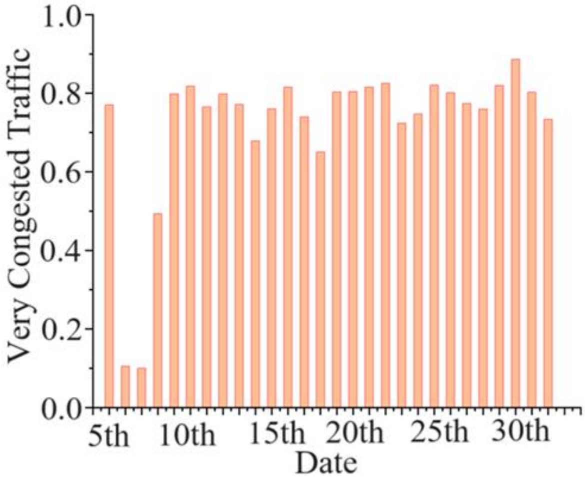


Fig. 6 Columnar comparisons of very congested traffic conditions for four weeks in June

Table 1 Statistical chart of the proportion of the day's traffic data to the total traffic data in June

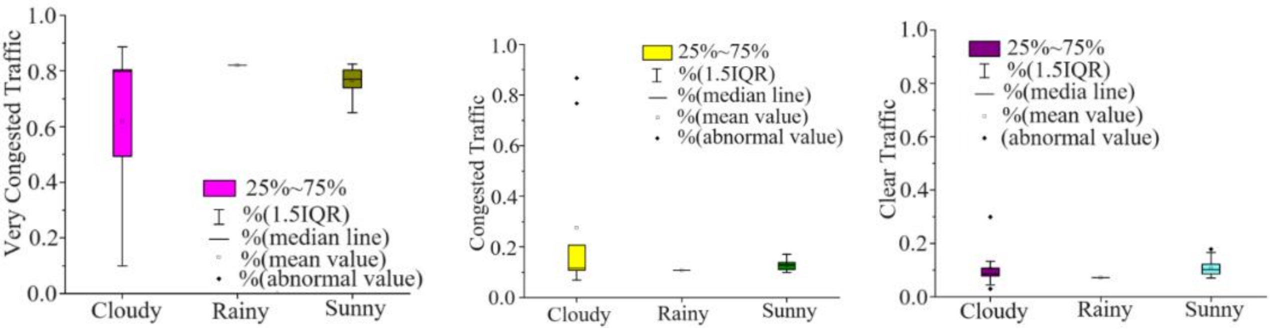
Date	Sunday			Date	Monday			Date	Tuesday			Date	Wednesday		
	A	B	C		A	B	C		A	B	C		A	B	C
5th	77.01	12.84	10.15	6th	10.49	86.70	2.97	7th	10.00	76.71	13.29	8th	49.33	20.77	29.90
12th	79.82	11.60	8.58	13th	77.14	12.71	10.15	14th	67.84	15.54	16.62	15th	76.01	12.95	11.04
19th	80.29	11.02	8.69	20th	80.40	10.73	8.87	21st	81.54	10.67	7.79	22nd	82.50	10.46	7.04
26th	80.15	11.40	8.45	27th	77.36	12.88	9.76	28th	75.92	13.46	10.62	29th	81.95	10.56	7.49
3rd	56.16	19.30	24.54												

Date	Thursday			Date	Friday			Date	Saturday		
	A	B	C		A	B	C		A	B	C
9th	79.80	12.06	8.14	10th	81.80	11.08	7.12	11th	76.49	12.75	10.76
16th	81.53	9.92	8.55	17th	73.94	13.90	12.16	18th	65.03	17.13	17.84
23rd	72.38	14.38	13.24	24th	74.72	13.63	11.65	25th	82.07	10.74	7.19
30th	88.62	6.90	4.48	1st	80.23	10.75	9.02	2nd	73.33	14.40	12.27

A: Very congested traffic B: Congested traffic C: Clear traffic

3.4 Potential rules of traffic conditions based on weather

In June 2022, the evening rush hour period from 17:30 to 19:00 experienced 1 rainy day, 9 cloudy days, and 20 sunny days in Xigu District of Lanzhou in northwest China. According to Fig.7 (a), traffic conditions were poor and very congested during cloudy weather (except the college entrance examination period), including the only time when traffic conditions were very congested during moderate rain. Fig.7 (c) has a very small percentage of clear traffic conditions during the cloudy weather, indicating that the traffic conditions were terrible on cloudy days, and the deterioration of the traffic conditions was even more pronounced during moderate rainy weather. For the reason of circumstances above, despite the cloudy weather contributed nothing to the slippery road surface, it did have an effect on the vision of all motorized and non-motorized vehicles and pedestrians, thereby reducing their traveling speed and ultimately decreasing the overall traffic volume at the intersection. In the arid climate of northwest China, moderate rainy days are infrequent, but the rain will affect the drivers' vision and slippery road surfaces will decrease the vehicle speed. These combined factors will lead to very congested traffic conditions in rainy weather.



(a)Very congested traffic (b) Congested traffic （c）Clear traffic

Fig. 7 Distribution of weather conditions and traffic conditions

4. Random Forest Model

Normalize the original samples[20]. The processing method is as follows:

$$X^{'} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Where X is the original sample value, $X^{'}$ is the normalized sample value, and X_{max} and X_{min} are the maximum and minimum values of any feature element sample, respectively.

Define the normalized video traffic feature dataset as X, which serves as the input variable for the Random Forest model, while speed and flow are taken as the output variables, denoted as Y. The basic steps of the Random Forest algorithm are as follows:

Step 1: Sample input. Input the normalized dataset X, which consists of N samples, each with M attributes.

Step 2: Sample classification. Use the bootstrap method to randomly sample and obtain the dataset X^{*}, consisting of N samples, which serves as the prediction data. The samples that are not selected are used as training data.

Step 3: Feature extraction. For any data sample containing M attributes, randomly select m feature elements to generate a sub-feature sample. The sub-feature sample reflects part of the characteristics of the data sample.

$$m = int\left(\sqrt{M}\right) + 1 \tag{2}$$

Step 4: Generate decision trees. Using the recursive partitioning method on the sub-feature samples, split each node into two child nodes until each child node contains only one sample attribute.

Step 5: Repeat Steps 2 to 4. Each decision tree grows into a forest, and the sum of the contributions of all decision trees in the forest is the prediction result of the Random Forest model.

Verify the accuracy of the prediction results based on the Gini importance index G, where, $p(j)$ represents the proportion of samples in the node that belong to category j.

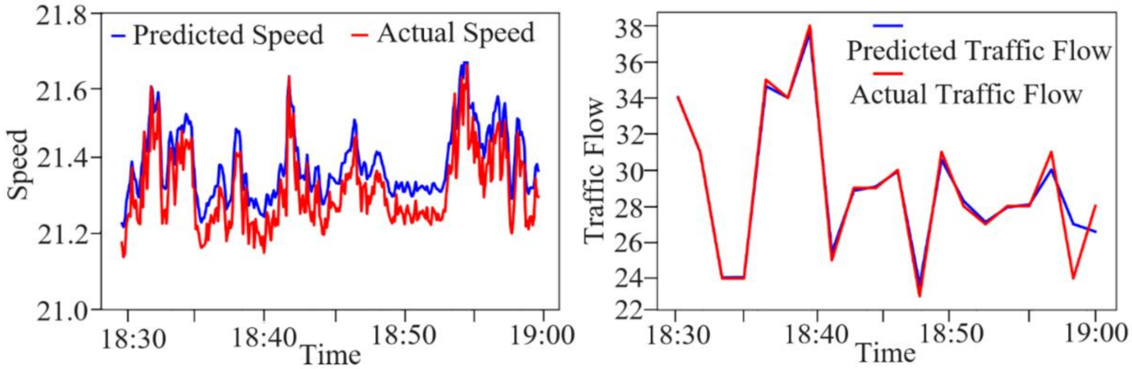
$$G = 1 - \sum_{j} p(j)^{2} \tag{3}$$

To better represent the attribute characteristics of the child nodes, a method for improving sample purity is proposed, as shown in formula (3):

$$\Delta G = G_{parent} - \left(P_{left} \cdot G_{left} + P_{right} \cdot G_{right}\right) \tag{4}$$

Where P_{left} and P_{right} represent the proportions of samples in the two child nodes, and ，G_{parent}、G_{left} and G_{right} are the Gini coefficients of the parent node and the two child nodes, respectively. The comprehensive Gini importance value of the sample is:

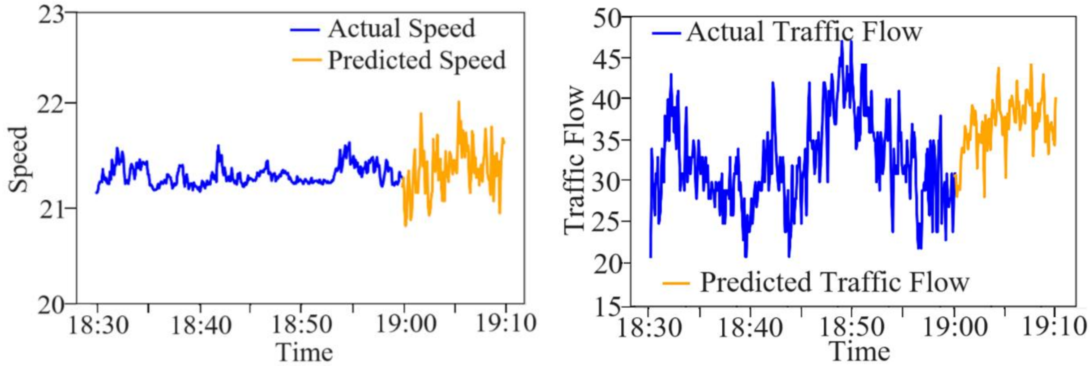
$$\Delta G_{total} = \sum_j \Delta G_j \tag{5}$$



(a)Speed (b) Traffic flow

Fig. 8 The predicted and actual value at different times

Fig. 8(a) and Fig.8(b) shows the fitted curves for the true and predicted values of speed on the test set by the Random Forest method, the curve demonstrates good fitting, indicating that the model has effective predictive performance and high robustness.



(a)Speed (b) traffic flow

Fig. 9 Actual and predicted on future

After collecting traffic data during the evening peak hour (17:30-19:00) on June 30, 2022, it is predicted by the Random Forest method that the speed in the next 10 minutes will increase from Fig. 9(a). This indicates that the speed of vehicles passing through the intersection is accelerating and the traffic conditions are good, further suggesting that traffic congestion is gradually easing. According to the predicted traffic flow in Fig.9 (b), the traffic flow will increase in the next 10 minutes. The combination of traffic speed and traffic flow indicates that traffic congestion is gradually easing and the evening peak is coming to an end.

5. Conclusions and suggestions

5.1 Conclusions

We systematically analyzed the traffic conditions at a typical crossing - the intersection of Fuli East Road in the Xigu District of Lanzhou City, an old industrial hub in northwest China, during the evening peak (17:30-19:00), based on a 30-day traffic dataset. Findings are as follows:

(1) Rectification of the parking and driving routes of non-motor vehicles without traffic control showed the same traffic effect as traffic control for road passage without restricting arbitrary parking of non-motorized vehicles.

(2) According to the speed distribution of traffic conditions, the wave-like and gradually increasing traffic congestion revealed that the traffic at this intersection was clear during the evening peak hour from 17:30 to 18:00, and the traffic started to be congested at 18:00, while the intensive passage of pedestrians had a significant impact on the traffic condition during the evening peak.

(3) During the regular work week, the intersection experienced more congested traffic during Wednesday and Thursday evening peaks compared to other evening peaks; traffic flowed much more freely on weekends compared to during the weekdays; and the Sunday evening peak was very special, with congestion equal to or even a little worse than the maximum level during the weekday evening peak.

(4) Traffic congestion at the intersection in Xigu District, Lanzhou City gradually intensified as the clock ticked into June, and then eased off a cliff by the third week, meaning traffic conditions suddenly turned around.

(5) The drivers' vision and slippery roads had a significant impact on the traffic conditions at the intersection.

5.2 Suggestions

We made a short-term prediction of the traffic conditions by the Random Forest method during the evening rush hour on June 30, 2022, and the results showed that the evening peak essentially came to an end at about 19:10. To further promote smart cities, the suggestions for urban development planning are given as follow:

(1) For Xigu District of Lanzhou City, a major industrial hub in northwest China, urban roads are impossible to be expanded due to its special construction history, the speed of urban traffic can be improved in the spatial dimension by formulating non-motorized vehicle driving guidelines to specify their traveling lanes, parking locations and parking modes. These measures will optimize the overall traffic conditions and increase the traffic speed in the city.

(2) This intersection is the traffic choke point for east-west and north-south traffic in Xigu District, Lanzhou City, and the period from 17:30 to 18:00 is the best choice for passage during the evening rush hour.

(3) The traffic at the intersection is congested on Wednesdays and Thursdays, but clear on Mondays, Tuesdays, and Fridays during a regular week, a finding that provides a basis for setting dates for major events.

(4) Congestion at this intersection gets worse in the first three weeks of June, but eases off a cliff in the fourth week. Therefore, events with long lead times, such as major chemical celebrations in the Northwest, can be timed to take place in July after June.

(5) According to the short-term prediction of traffic flow, the congestion at this intersection essentially ends at 19:10 in the evening rush hour, which provides a methodology and basis for predicting conditions of transportation in the future.

Acknowledgements

National Natural Science Foundation of China project(6136606), Innovation Fund Project for University Teachers of Gansu Provincial Education Department(2023B-294)

References

[1] Jiang Y,Ding Z,Zhou J, Wu P,Chen B.Estimation of traffic emissions in a polycentric urban city based on a macroscopic approach[J].Physica A,2022, 602:127391.

[2] Khomenko S,Cirach M,Barrera-Gomez J,Pereira-Barboza E,Velise,Iungman T,Mueller N, Foraster M,Tonne C,Thondoo M,Jephcote C,Gulliver J,Woodcock J,Nieuwenhuijsen M.Impact of road traffic noise on annoyance and preventable mortality in European cities: A health impact assessment[J].Environment International,2022,162:107160.

[3] Sahu A K,Pradhan P K,Naya S K,Pradhan M,Mohanty C R.Evaluation and Modeling of Traffic Noise in Berhampur City, India[J].Fluctuation and Noise Letters, 2020,19.4:2050044.

[4] Guo M, Ni M Y, Shyu R J, Ji J S,Huang J X. Automated simulation for household road traffic noise exposure: Application and field evaluation in a high-density city[J]. Computers, Environment and Urban Systems,2023,104:102000.

[5] Das D K.Exploring the significance of road and traffic factors on traffic crashes in a South African city[J].International Journal of Transportation Science and Technology,2023,12:414-427.

[6] Fattah M A, Morshed S R,Kafy A A.Insights into the socio-economic impacts of traffic congestion in the port and industrial areas of Chittagong city, Bangladesh[J]. Transportation Engineering,2022,9:100122.

[7] Peters A,Hernandez D,Kioumourtzoglou M A,Johnson M,Chillrud S,Hilpert M.Assessing neighborhood-scale traffic from crowd-sensed traffic data:Findings from an environmental justice community in New York City[J].Environmental Science and Policy,2022,133:155-163.

[8] Liu Y l, Fang F G,Jing Y. How urban land use influences commuting flows in Wuhan, Central China: A mobile phone signaling data perspective[J].Sustainable Cities and Society,2020, 53:101914.

[9] Singh S,Singh J,Goyal S B,Sehra S S,Ali F, Alkhafaji M A,Singh R.A novel framework to avoid traffic congestion and air pollution for sustainable development of smart cities[J].Sustainable Energy Technologies and Assessments,2023,56:103125.

[10] Kavitha T,Pandeeswari N,Shobana R,Vinothini V R,Sakthisudhan K A,Jeyam,Malar J G A. Data congestion control framework in Wireless Sensor Network in IoT enabled intelligent transportation system[J].Measurement:Sensors,2022,24:100563.

[11] Cai M,Hong L,Xiong C.Data-driven traffic zone division in smart city: Framework and technology[J].Sustainable Energy Technologies and Assessments,2022,52:102251.

[12] Saleem M,Abbas S,Ghazal T M,Khan M A,Sahawneh N,Ahmad M.Smart cities: Fusion-based intelligent traffific congestion control system for vehicular networks using machine learning techniques[J].Egyptian Informatics Journal,2022,23:417－426.

[13] Noor M A,Ashrafi S, Fattah M A, Morshed S R,Rahman S. Assessment of traffic congestion scenario at the CBD areas in a developing city: In the context of Khulna City, Bangladesh[J].Transportation Research Interdisciplinary Perspectives,2021,11:100435.

[14] Chellapilla H,Sivanandan R,Chilukuri B R,Rajendran C.A Trade-off between User-Equilibrium and System Optimal Traffic Assignments for Congestion Mitigation in City Networks[J]. Transportation Research Procedia,2023,69:384–391.

[15] Wang J H,Lv W,Jiang Y J,Huang G C.A cellular automata approach for modelling pedestrian-vehicle mixed traffic flow in urban city[J].Applied Mathematical Modelling,2023:1151-33

[16] Chen Y T,Mao J N,Zhang Z,Huang H,Lu W K,Yan Q P,Liu L.A quasi-contagion process modeling and characteristic analysis for real-world urban traffic network congestion patterns[J]. Physica A,2022,603:127729.

[17] Chen G,Zhang J.Applying Artificial Intelligence and Deep Belief Network to predict traffic congestion evacuation performance in smart cities[J].Applied Soft Computing,2022,121108692.

[18] Du W D,Zhang Q Y,Chen Y P,Ye Z L. An urban short-term traffic flow prediction model based on wavelet neural network with improved whale optimization algorithm[J]. Sustainable Cities and Society,2021,69:102858.

[19] Chauhan V,Patel M,Tanwar S,Tyagi S,Kumar N.IoT Enabled real-Time urban transport management system[J].Computers and Electrical Engineering,2020,86:106745.

[20] Li Zhi-hui, Sun Ya-qian, Tao Peng-fei, et al. Prediction method of traffic operation risk level after traffic accident[J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(01): 127-135.