ADVANCES IN TRANSPORTATION STUDIES
An International Journal

Guest Editor: Y.H. Wu and Z.W. Liu

2024 Special Issue, Vol. 3

Contents

aracne

Advances in Transportation Studies an international Journal 2024 Special Issue, Vol. 3

ISBN 979-12-218-1657-0 – ISSN 1824-5463 – DOI 10.53136/97912218165701 – pag. 3-12

Dual level planning method for urban diversion road network design based on environmental impact analysis

Y.G. Zhang* L. Ke

School of Civil Engineering, Southwest Jiaotong University Hope College, Chengdu 610400, Sichuan, China
**corresponding author; email: 13983639298@163.com*

Abstract

Urban diversion road network design is crucial as it addresses the growing concerns of urban traffic congestion and environmental pollution by proposing a method to reduce carbon emissions and improve traffic flow efficiency. To optimize the design of urban road networks, this paper provides an environmental impact analysis perspective and investigates a dual-level planning approach for the design of urban diversion road networks. Initially, an upper-level model is developed with the objective of minimizing carbon emissions, environmental impact, and road network congestion impedance time. Subsequently, a lower-level Logit-type random user equilibrium allocation model is constructed to achieve random allocation. Finally, by modifying the spiral position update method and enhancing the search capability of the whale algorithm, an improved algorithm is employed to solve the model. Experimental results indicate that after applying this method, daily CO2 emissions from the road network remained below 600 kg, and the traffic flow in each section was balanced, demonstrating the feasibility and effectiveness of this new approach.

Keywords - environmental impact, carbon emissions, urban diversion road network, road planning, double layered planning, revise the spiral position update, whale algorithm

1. Introduction

With the acceleration of urbanization worldwide, cities have attracted a large influx of population. The dense gathering of population has led to the continuous expansion of urban scale, and the existing urban road network is unable to meet the growing transportation demand [9]. In addition, urban expansion not only involves an increase in population, but also an expansion of urban area. New residential, commercial, and industrial areas are constantly emerging, and the functional zoning of cities is becoming more complex. This requires a reasonable road network to connect various areas and ensure the normal operation of the city [6, 19].

From another perspective, traffic congestion can lead to motor vehicles idling or driving at low speeds for long periods of time, thereby increasing exhaust emissions and causing pollution to the urban environment [8]. Reasonable diversion road network design and planning can effectively improve traffic efficiency, alleviate traffic congestion, reduce motor vehicle exhaust emissions, optimize urban air quality, which plays a crucial role in promoting sustainable urban development.

Therefore, Liu et al. [7] developed a road planning method for physical entities based on the continuous road network environment. Firstly, based on the specific environmental conditions, the continuous environment of the physical entity is simulated through fusion expansion. Then, following the principle of prioritizing road networks, skeleton extraction technology is used to extract key information of the road network in the simulation environment. Finally, based on the A * algorithm, the extracted key information is integrated with the simulation environment to obtain the road planning results. He and Ma [4] first constructed a city road network model, which is characterized by weight changes over time. Then, the time-varying road network path planning problem with weights was modeled, and state features, interactive actions, and reward functions were designed during the modeling process. Train the intelligent agent using the deep double-Q network algorithm to learn the time-varying characteristics of road network weights. Finally, based on the modeled state characteristics, effective path planning for time-varying road networks with weighted values is achieved. Bao et al. [2] introduced traffic accessibility theory to propose a hierarchical planning model for road networks based on the characteristics of road networks. This method divides the road network into a fast layer and a branch layer. For the fast layer, use weighted average travel time as the optimization objective, For the branch layer, the goal is to maximize accessibility, and the simulated annealing algorithm is used to solve the model and obtain the final road planning results.

However, in practical applications, it has been found that the above methods cannot effectively control traffic carbon emissions on roads, and the vehicle diversion effect is generally poor. In response to this issue, this study starts from the perspective of environmental impact analysis and investigates a new dual level planning method for the design of urban diversion road networks. The research in this article includes the following three aspects:

①The upper level model aims to minimize the environmental impact of carbon emissions. In the objective function of minimizing the environmental impact of carbon emissions, CO emissions are reduced by adjusting diversion schemes and optimizing travel routes, In the objective function of minimizing the impedance time of road network congestion, a road segment impedance function is constructed based on the relationship between speed and density, and multiple influencing factors are considered, while setting road segment saturation constraints,

② The lower level model aims at random allocation and constructs a Logit type random user equilibrium allocation model to describe user path selection behavior, clarifying the calculation method of path selection probability and traffic allocation principles,

③ Based on the improved whale algorithm to solve the dual layer model of urban diversion road network design, a dual layer planning model is first constructed, and then the improved whale algorithm is used to solve the model. The improvement lies in modifying the spiral position update method to enhance convergence performance and global search capability. The solving process includes population initialization, individual sorting, iterative updating, position coordinate determination, and optimal solution output. Through these steps, a reasonable road network design and optimization can be achieved to alleviate traffic congestion and reduce carbon emissions.

2. Dual level planning method for urban diversion road network design

2.1. Upper level model with the goal of minimizing carbon emissions and environmental impact

The upper level model designed in this study is designed with the optimization objectives of minimizing carbon emissions and environmental impact, and minimizing road network congestion impedance time.

① Minimum objective function for carbon emissions and environmental impact

Automobile exhaust contains many harmful gases, among which CO, nitrogen oxides, and other substances account for a large proportion. Their emissions can have adverse effects on the surrounding environment and human health [1, 14]. Therefore, this study focuses on traffic diversion strategies by improving diversion plans and optimizing driving routes.

This study selects CO emissions as the optimization objective, and uses the fitting relationship between the CO emission factor of motor vehicle pollutants and their average speed to construct a CO emission model. The calculation formula is as follows:

$$C_{aco} = 167.2 - 5.3v_a + 0.07{v_a}^2 - 0.0003{v_a}^3 \tag{1}$$

$$C_{bco} = 64.52 - 2.3v_b + 0.03{v_b}^2 - 0.0001{v_b}^3 \tag{2}$$

Among them, C_{aco} and C_{bco} respectively represent the CO emission factors of passenger cars and large vehicles, v_a and v_b respectively represent the average speed of buses and trucks on the road.

Therefore, the objective function for minimizing the environmental impact of carbon emissions is established as follows:

$$minF_1 = \partial \sum_{l\in L}(C_{aco} \times s_l \times q_a + C_{bco} \times s_l \times q_b) \tag{3}$$

Among them, ∂ represents the CO emission cost conversion coefficient, $l \in L$ represents the diversion section, s_l represents the distance of the road segment, q_a and q_b respectively represent the traffic volume of buses and trucks on the road .

② Minimum objective function of road network congestion impedance time

In this study, the time required for vehicles to pass through a road segment in diverted traffic is taken as the impedance of the road segment, and the impedance function of the road segment is constructed based on the linear relationship between speed and density. For a fixed length road section, there is a one-to-one correspondence between vehicle speed and corresponding traffic density, The impedance of road sections and the corresponding traffic density also have a one-to-one correspondence [3].

Generally speaking, the three macro indicators that characterize the basic characteristics of diverted traffic are traffic volume, speed, and density [18, 5, 17]. By utilizing the relationship between speed and density, the correlation between traffic volume and speed/density can be derived as follows:

$$P = \rho_c \times \left(v - \frac{v^2}{v_f}\right) \tag{4}$$

$$N_l = 0.25v_f \times \rho_c \tag{5}$$

Among them, ρ_c represents the blocking density of vehicle c, v and v_f respectively represent the speeds of traffic flow and free flow, N_l represents the traffic capacity of the road section.

When the vehicle travels on a road segment of length d, its impedance is as follows:

$$t = \frac{d}{v} \tag{6}$$

According to formulas (4) - (6), the form of delay time for congested road sections can be obtained as:

$$\nabla t = 2t_0 \times \left(1 - \sqrt{1 - \frac{y_l}{N_l}}\right) \tag{7}$$

Among them, t_0 represents the time required for the vehicle to travel across the road section in the absence of other vehicles, y_l represents the allocation of traffic volume.

In the road network, the time delay caused by the passage of vehicles at intersections has a certain impact on the calculation of road impedance time [13]. Intersection traffic signal control, the time when traffic flow changes from red to green and the delay time when the green light switches to red or yellow. When the road section is in a very smooth condition, the delay time caused by vehicle queuing can be ignored, When the saturation level of the road section approaches

1, the road section is in a congested state, with slow vehicle movement and long queues. Signal cycle control has a significant impact on time delay [20, 12]. This results in deceleration delay, vehicle queue delay, parking delay when encountering red lights, and acceleration delay when starting, collectively referred to as control delay.

Based on the above analysis, establish the minimum objective function of road network congestion impedance time as follows:

$$minF_2 = \frac{(T-1)^2}{2\times\left(1-min\frac{y_l}{N_l}\right)} + 900T\left(\left(\frac{y_l}{N_l}-1\right)+\frac{8\mu y_l}{{N_l}^2\times H}\right) \tag{8}$$

Among them, T represents the time period of traffic signal lights, μ represents the correction coefficient, H represents the duration of the analysis period.

In summary, the upper level model with the goal of minimizing carbon emissions and environmental impact is as follows:

$$M_{up} = minF_1 + minF_2 \tag{9}$$

In the process of constructing the upper level model, it is necessary to comprehensively consider the carrying capacity of each road section, ensure the scientific allocation of road network resources, and prevent over saturation of certain road sections.

2.2. Lower level model aimed at random allocation

This study aims to construct a lower level model for random allocation, namely a Logit type random user equilibrium allocation model under deterministic demand. The goal of constructing the model is to describe the path selection behavior of users in the road network, so that users can reach an equilibrium state when choosing paths.

In the process of building the lower level model, whether the road is selected for diversion is the decisive factor, which is the core point of the entire modeling work [11]. After discretizing the road network, define a decision variable for each road that may be used for diversion. For example, let x_{ij} be a variable that determines whether the road from node i to node j is selected for diversion. If it is selected, then $x_{ij} = 1$, Otherwise, $x_{ij} = 0$.

The construction process of the lower level model is as follows:

Step 1: Calculate the probability of path selection.

Based on the probability definition of the Logit model, assume that U_k^{rs} is the path attribute function of the path k from the starting point r to the endpoint s. U_k^{rs} can be represented in the form of $U_k^{rs} = \vartheta_1 t_k^{rs} + \vartheta_2 c_k^{rs}$.

Here, t_k^{rs} is the travel time on path k, c_k^{rs} is the carbon emissions and other attributes on path k, ϑ_1 and ϑ_2 are the corresponding weight coefficients. The probability of selecting path k from the starting point to the endpoint is:

$$P_k^{rs} = \frac{e^{U_k^{rs}}}{\sum_{K^{rs}} e^{U_k^{rs}}} \tag{12}$$

Among them, K^{rs} represents the set of all available paths for selection.

Step 2: Traffic allocation based on selection probability.

Assuming q^{rs} is the deterministic demand for travel from the starting point to the endpoint. So, the flow allocated to path k from the starting point to the endpoint can be expressed as:

$$f_k^{rs} = q^{rs} \times P_k^{rs} \tag{13}$$

The formula for establishing the lower level model can be as follows:

$$M_{down} = \max(f_k^{rs}) \tag{14}$$

When designing the lower level model, it is necessary to follow the constraints on flow conservation and decision variables. These two constraints are as follows:

① Constraints on flow conservation:

$$\sum_r \sum_{k\in K^{ri}} f_k^{ri} - \sum_s \sum_{k\in K^{is}} f_k^{is} = 0 \tag{15}$$

② Constraints on decision variables:

$$x_{ij} \in \{0,1\} \tag{16}$$

2.3. Solving a dual layer model based on an improved whale algorithm

Based on the above analysis, a dual layer model for urban diversion road network design can be obtained as follows:

$$M = M_{up} + M_{down} \tag{17}$$

This dual level programming model belongs to the category of nonlinear programming, and its solving process is quite complex. Compared with traditional optimization algorithms, the whale algorithm exhibits superior convergence performance and can effectively avoid the trap of local optimal solutions. In view of this, this study decided to use the whale algorithm to solve the model.

The whale population contains N individuals from multiple dimensions, and each whale individual corresponds to a dual level planning result of urban diversion road network design. In the process of updating the position of individual whales, the optimal individual position and the current position of random individuals are used as the search targets for the optimal individual in the model solution. And during the process of whale hunting, there is a problem of gradually narrowing the encirclement. The solution space of the searched urban road network planning gradually shrinks, improving the local search ability of the whale algorithm.

However, during the whale hunting process, there is a situation where the group is too concentrated and lacks the navigation coordinates of the optimal individual. Whale individuals are prone to deviate from the search direction of the optimal solution, which affects the convergence speed of the whale algorithm and reduces the global convergence of the double-layer planning model for urban diversion road network design [10].

In response to the above issues, this study adopts the modified spiral position update method to improve the conventional whale algorithm and complete the update of the optimal position of individual whales.

When searching for the optimal solution of the dual layer planning model for urban diversion road network design, whale individuals are sorted in ascending order based on their fitness values, arranged as $X_1, X_2, \cdots, X_N$.

Based on the sorting results of individual whales, they are divided into the optimal individual X_i^b, the intermediate individual X_i^m, and the deviated individual X_i^t, with the proportions of each level of individuals in the set being 0.05, 0.9, and 0.05, respectively. Based on the segmentation results of whale individuals, the expression for updating the position of whale individuals searching for the optimal solution is determined as follows:

$$X_i(t+1) = X_i^b + D_h e^{hw} cos(2\pi w) \tag{18}$$

Among them, $X_i(t+1)$ represents the position of the whale individual at iteration times $t+1$, h and w represent the spiral shape parameters and random numbers within the range of [-1,1], respectively, D_h represents the distance between the current search individual and the optimal solution when the spiral shape parameter is h.

The traditional whale algorithm sets the spiral shape parameter as a constant, and when solving the model, the update method of the whale's individual hunting posture and position is single, which is prone to premature convergence [16]. Improve the fixed search posture during whale hunting by using a dynamically changing sine function to represent the spiral shape parameter h. The parameter h is continuously adjusted as the number of iterations increases, allowing individual whales to automatically adjust their spiral posture during the hunting search process. By improving the parameter h, the problem of local convergence in the whale algorithm can be alleviated, and the global search ability of the whale algorithm can be improved, thereby enhancing its convergence performance.

The calculation formula for the improved spiral shape parameter h is as follows:

$$h = -\varepsilon \sin\left(2\beta\sqrt{\frac{t_{max}-t}{t_{max}}}\right) \tag{19}$$

Among them, ε and β respectively represent the spiral update parameters and the attitude influence quantity, t and t_{max} represent the current population iteration count and maximum iteration count, respectively.

The process of using the improved whale algorithm to solve the dual level planning model for urban diversion road network design is as follows:

Step 1: Initialize the whale population position and parameters in the solution space of the model. The parameters to be initialized include the dimensionality of the urban road network structure planning problem, population size N, spiral shape parameter h, nonlinear decreasing convergence factor α, step size adjustment parameter A, initial population iteration times t, and maximum population iteration times t_{max},

Step 2: Sort the whale individuals in the population based on their fitness values and classify them into different levels. Record the optimal individual X^* and its position $\{X_1^*, X_2^*, \cdots, X_N^*\}$ within the whale population,

Step 3: Perform iterative updates on the population. When $t < t_{max}$, continue updating α, A, and h. After completing the parameter update, sort the whale individuals within the population again based on their fitness values, classify them into whale classes, and obtain a new whale population,

Step 4: Use formula (18) to determine the position coordinates of the current whale individual,

Step 5: Record the coordinates and fitness values of the optimal individual during the current iteration process. When $t \geq t_{max}$, proceed to the next step. Otherwise, increase the number of iterations by 1 and return to step 3 until the termination condition is met,

Step 6: Output the current optimal individual position and its corresponding fitness value, and terminate the algorithm iteration. The solution corresponding to the optimal individual is the final dual level planning result of the urban diversion road network design.

3. Experiments and results analysis

3.1. Experimental design

To verify the feasibility of the method of this paper, M city was selected as the research object and simulation testing was conducted. M city is a regional center for technology, politics, economy, and culture, connecting the transportation of surrounding cities. In the experiment, Matlab 7.0 simulation software was selected to simulate the urban diversion road network.

This experiment selects the classic Nguyen & Dupuis network to test the constructed road network model, which includes a total of 13 road nodes, as shown in Figure 1.

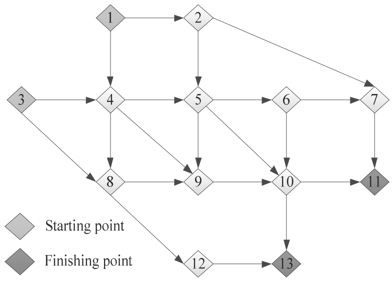

Fig. 1 - Simulation model of urban diversion road network

To avoid the singularity of experimental results, the methods of Liu et al. [7] and He and Ma [4] were compared and simulated simultaneously with the method of this paper.

3.2. Indicators and results analysis

The experimental results display section takes the 1 → 4 → 9 → 10 → 13 line as an example. The experimental indicators are the CO emissions after applying different methods and the traffic flow in each road section.

① Analysis of CO emissions using different methods

CO is one of the main pollutants in automobile exhaust, which not only has a negative impact on air quality, but also poses a threat to human health. CO emissions are an important indicator for measuring the impact of transportation on the environment. This indicator is usually calculated based on factors such as vehicle type, speed, traffic flow, and vehicle emission factors. The calculation formula is as follows:

$$n_{CO} = \frac{\sum(n_c \times \eta \times \psi_{CO} \times n_S)}{n_t} \tag{20}$$

Among them, n_c represents the number of vehicles, η represents fuel consumption rate, ψ_{CO} represents the CO emission factor, n_S represents the distance traveled, n_t represents the total time of statistics.

The experimental statistics show the changes in CO emissions from Line 1 → 4 → 9 → 10 → 13 during the period of October 1 to October 6, 2023. The statistical results are shown in Figure 2.

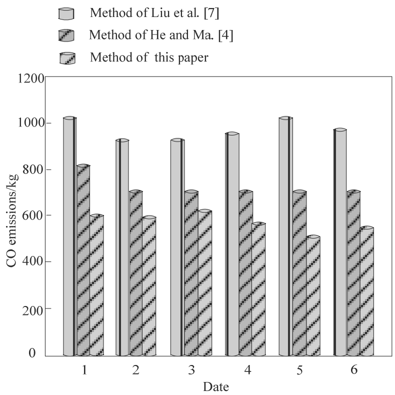

Fig. 2 - Changes in CO emissions on roads

From the analysis of the results shown in Figure 2, it can be seen that after using the method of this paper for urban diversion road network planning and design, by reasonably arranging road diversion routes, the CO emissions in the roads have been effectively reduced. Throughout the entire experiment, the daily CO emissions remained below 600kg, which was significantly lower than the two comparison methods in terms of numerical values. This indicates that the method of this paper can optimize the transportation structure, effectively reduce carbon emissions, and contribute to sustainable environmental development.

② Analysis of traffic flow in various road sections after applying different methods

Traffic flow refers to the number of vehicles passing through a certain road within a certain period of time, and is one of the key indicators to measure the rationality of road network design. High traffic volume may lead to traffic congestion, increased risk of traffic accidents, and intensified energy consumption and environmental pollution. Therefore, by optimizing road network design and allocating traffic flow reasonably, traffic pressure can be alleviated and road traffic efficiency can be improved.

The traffic flow can be obtained through traffic volume surveys, and the time periods can be 5 minutes, 15 minutes, 1 hour, or 24 hours. The daily average cross-sectional traffic flow refers to the average number of vehicles passing through each day, and is an important basis for evaluating road usage and traffic conditions. For this purpose, the experiment analyzed the changes in traffic flow in each road section during the morning rush hour (7:00~9:00) after applying different methods. The results are shown in Table 1. From the analysis of the results shown in Table 1, it can be seen after applying the two comparison methods, there is a significant difference in traffic flow between different road sections, which can easily cause traffic congestion. For method of Liu et al. [7], some road sections (such as 4 → 9, 10 → 13) have significantly higher traffic flow than others, with a large standard deviation, indicating an uneven distribution of traffic flow. For the method of He and Ma [4], there is also an issue of uneven distribution of traffic flow, especially on the 10-13 section where traffic flow is significantly higher than on other sections over multiple time periods. After using the method of this paper for urban diversion road network planning and design, the traffic flow in each section is similar, indicating the effectiveness of the method of this paper's diversion planning.

Tab. 1 - Road traffic statistics results

Time	Line	After applying the method of this paper	After applying the method of Liu et al. [7]	After applying the method of He and Ma [4]
7:00~7:30	1→4	325	380	315
	4→9	351	520	341
	9→10	362	100	171
	10→13	339	240	553
7:30~8:00	1→4	825	750	578
	4→9	846	355	96
	9→10	837	780	412
	10→13	819	940	516
8:00~8:30	1→4	850	413	789
	4→9	862	852	1031
	9→10	870	803	336
	10→13	866	945	760
8:30~9:00	1→4	495	473	537
	4→9	483	485	284
	9→10	481	206	219
	10→13	494	281	569

4. Conclusion

This article proposes a dual layer planning method for urban diversion road network design based on environmental impact analysis, aiming to reduce traffic congestion, lower motor vehicle exhaust emissions, improve urban air quality, and promote sustainable development of cities through scientific and reasonable road network design. In terms of research methods, this article constructs an upper level model with the goal of minimizing carbon emissions and minimizing road network congestion impedance time, as well as a lower level model with the goal of random allocation, and solves the model using an improved whale algorithm. Through experimental verification, the method of this paper has shown excellent performance in the planning and design of urban diversion road networks, effectively reducing CO emissions in roads and achieving balanced distribution of traffic flow in each section.

With the acceleration of urbanization and the continuous growth of transportation demand, the design of urban diversion road networks will face more challenges. The research findings of this article provide a new approach and method for urban road network planning. In the future, we will continue to conduct in-depth research on the design of urban diversion road networks, explore more efficient and environmentally friendly road network planning methods, and contribute more wisdom and strength to the sustainable development of urban transportation.

References

1. Acosta, F.C., Rengifo, S.P., García, M.L., Trondoli, M.E.A., Castillo, G.B. (2023) Road Network Planning in Tropical Forests Using GIS. Croatian Journal of Forest Engineering : Journal for Theory and Application of Forestry Engineering, 44(1), PP. 153-169.
2. BAO, D.W., CHENG, H., ZHU, T., TIAN, S.J., ZHANG, T.X. (2021) A Method for Hierarchical Planning of Ground Access Network of Airports Based on Accessibility. Journal of Transport Information and Safety, 39(03), PP. 142-151.
3. Campaña, M., Inga, E. (2023) Optimal deployment of fast-charging stations for electric vehicles considering the sizing of the electrical distribution network and traffic condition. Energy Reports, 9(1), PP. 5246-5268.
4. He, X., Ma, P. (2023) Time-varying road network path planning based on double deep Q-network. Electronic Measurement Technology, 46(17), PP. 23-29.
5. Kamga, C., Vicuna, P., Mouskos, K. (2023) Transit Network Design Problem – Impact on Network Travel Time Using a Transportation Planning Model. Procedia Computer Science, 224(1), PP. 239-249.
6. LI, R., RAN, B., QU, X. (2022) Traffic capacity enhancement strategy for urban expressway diversion area under vehicle-infrastructure cooperative environment. Journal of Traffic and Transportation Engineering, 22(03), PP. 126-138.
7. LIU, G., AN, Z.B., ZHANG, M.J., LIU, Y., LI, W. (2024) Subject objective path planning algorithm based on continuous road network environment. Systems Engineering and Electronics, 46(04), PP. 1346-1356.
8. Li, Y., Zhan, Q.L., Gao, F. (2024) A layout planning method for urban agglomeration transportation lines from the perspective of carbon reduction policies, Advances in Transportation Studies, 1(7), PP. 3-14
9. LV, B., XIE, Z.Y., KANG, Y.X., ZHAO, Y.M. (2022) Resilience Assessment on Urban Road Network by Dynamic Shunt Cell Transmission Model. Journal of Transportation Systems Engineering and Information Technology, 22(06), PP. 134-143+211.
10. Mayregger, P. (2023) Systematization of network planning procedures and network structure adjustments in urban transport network planning. Transportation Research Procedia, 72(3), PP. 2149-2156.
11. Peng, J., Chen, L., Zhang, B. (2022) Transportation planning for sustainable supply chain network using big data technology. Information Sciences, 609(10), PP. 781-798.
12. Shang, P., Yao, Y., Yang, L.Y., Meng, L.Y., Mo, P.L. (2021) Integrated Model for Timetabling and Circulation Planning on an Urban Rail Transit Line: a Coupled Network-Based Flow Formulation.

Networks and Spatial Economics, 21(2), PP. 1-34.
13. Van, M.M., Jafino, B.A., Lourens, L., Hüsken, L. (2023) Including equity considerations in resilient transport network planning and analysis: A flood impact perspective. Transportation Research Procedia, 72(5), PP. 3837-3844.
14. Wan, B.T., Bao, X.Y., Zhao, J.C. (2022) Evaluation Method and Application of Ecological Sensitivity of Intercity Railway Network Planning. Sustainability, 14(2), PP. 804-814.
15. Wang, Z., Lan, F.M., Lin, Z.J., Lian, L. (2021) A Heuristic Method for Bus Rapid Transit Planning Based on the Maximum Trip Service. Sustainability, 13(11), PP. 6325-6337.
16. Yedavalli, P., Kumar, K., Waddell, P. (2022) Microsimulation analysis for network traffic assignment (MANTA) at metropolitan-scale for agile transportation planning. Transportmetrica A: Transport Science, 18(3), PP. 1278-1299.
17. Zang, Z.Q., Xu, X.D., Chen, A., Yang, C. (2021) Modeling the α-max capacity of transportation networks: a single-level mathematical programming formulation. Transportation, 49(4), PP. 1-33.
18. Zhang, D.P., Li, Y.N. (2023) Network Planning for Innovative Track-Based Transportation Technologies: A Minimum Spanning Tree Algorithm to Demonstrate Network Benefits. Transportation Research Record, 2677(3), PP. 95-103.
19. ZHANG, W.H., TAO, H., CHEN, Q. (2022) The Predicting Model of Traffic Conflicts in Diverging Segments of Expressway Interchange. Highway Engineering, 47(01), PP. 149-155.
20. Ziar, E., Seifbarghy, M., Bashiri, M., Tjahjono, B. (2022) An efficient environmentally friendly transportation network design via dry ports: a bi-level programming approach. Annals of Operations Research, 322(2), PP. 1143-1166.

ISBN 979-12-218-1657-0 – ISSN 1824-5463 – DOI 10.53136/97912218165702 – pag. 13-26

Binary logistic regression analysis of factors affecting urban road traffic safety

Y.W. Chen* P. You Z.K. Chang

College of Statistics and Mathematics, Hebei University of Economics and Business, Shijiazhuang, Hebei 050061, China
**corresponding author; email address: shufa2022@163.com*

Abstract

Urban road traffic safety is of great significance as it directly affects the well-being of citizens and the efficient operation of cities. To precisely assess the influence of diverse factors on urban road traffic safety, this article employs binary logistic regression analysis. The factors related to urban road traffic safety are categorized into four aspects: human, vehicle, road, and environment. By using the occurrence of a traffic accident as the dependent variable and calculating the odds ratio to gauge the impact of independent variables, after conducting multicollinearity tests on the binary logistic regression model, the regression coefficient analysis reveals that driver behavior, weather conditions, road conditions, and lighting conditions have a substantial impact on urban road traffic safety. Experiments demonstrate that the coverage rate of influencing factors in this paper exceeds 90%, with a comprehensiveness of up to 95%. This indicates that the binary logistic model utilized in this study for factor regression analysis of urban road traffic safety exhibits high coverage, comprehensiveness, a broad regression scope, and high reliability and robustness.

Keywords - traffic safety, binary logistic regression, multicollinearity test, regression coefficient, normalization processing

1. Introduction

With the acceleration of urbanization, the population and number of vehicles in cities are growing rapidly, and traffic flow is constantly increasing [17]. However, the limited carrying capacity of urban roads has led to increasingly severe traffic congestion. Traffic congestion not only reduces road traffic efficiency, but also increases the risk of traffic accidents, making road traffic safety issues increasingly prominent. At the same time, road planning and infrastructure construction often lag behind the rapid growth of population and vehicles [4]. This leads to insufficient road network capacity and inadequate transportation facilities, such as unreasonable or missing traffic signals, signs, and markings, posing hidden dangers to traffic safety. In addition, traffic participants include drivers, pedestrians, non motorized vehicle drivers, etc. Their safety awareness, compliance with traffic regulations, and driving behavior have a significant impact on traffic safety [11]. Some drivers engage in violations such as drunk driving, fatigue driving, and speeding. Pedestrians also frequently cross the road and do not follow traffic signals. These behaviors not only endanger one's own safety, but may also lead to traffic accidents. The frequent occurrence of road traffic accidents poses a huge threat to the safety of people's lives and property [16]. The occurrence of traffic accidents not only causes casualties and property losses, but also has adverse effects on social stability and economic development. With the development of information technology, a large amount of traffic accident data has been recorded, providing valuable resources for traffic safety research. By analyzing these data, we can delve into the key factors that affect traffic safety and provide a basis for developing effective traffic safety measures.

There are many experts and scholars in related fields conducting research on the causes of urban road traffic accidents. For example, the method described in Chen et al. [2] preprocesses the data using rough set theory, selects and reduces the variables that affect the morphology of traffic accidents, and then uses the C5.0 decision tree algorithm to construct a decision model for predicting traffic accidents. Combined with a multi-layer perceptron neural network, it generates a set of rules for traffic accident morphology, which can effectively identify the main influencing factors such as traffic mode, driving location, illegal behavior, etc. It provides a scientific method and decision-making basis for traffic management departments to analyze and predict the morphology of urban road traffic accidents, explore the causes of accidents, and so on. However, the data relied upon by this method may be limited by factors such as source, quality, and completeness, and is mainly based on historical data for prediction, which may not fully adapt to new situations or changes that may occur in the future. The method described in Zhang et al. [20] used data from the Shandong Province Death Cause Registration and Reporting System to calculate traffic injury mortality rates. Geoda 1.18 was used for spatial autocorrelation analysis, SatScan 9.7 was used for spatiotemporal scanning analysis to explore spatiotemporal clustering characteristics, and principal component regression analysis was used to explore its influencing factors. Finally, the spatiotemporal clustering characteristics and main influencing factors of road traffic injury mortality rates in Shandong Province were obtained. Although this method uses principal component regression analysis to explore the influencing factors of road traffic injury mortality rate, the selected factors may still have certain limitations, which may vary from individual differences and are difficult to accurately quantify through large-scale data collection and analysis. Ammar et al. [1] used collision data from the Overall Estimation System (GES) from 2013 to 2015 and the Collision Report Sampling System (CRSS) from 2016 to 2018. Through logistic regression model analysis, four common significant influencing factors leading to serious injury or death of pedestrians in pedestrian vehicle collisions at intersections were identified. These factors were compared in pairs using Wald chi square statistical tests, and the results showed that pedestrians were highly likely to suffer serious injuries, which were related to pedestrians over 25 years old, dark lighting conditions, light trucks and buses, and vehicles. These factors had similar effects between 2013 and 2018. Although logistic regression models can identify significant influencing factors, they may not provide in-depth insights into how these factors interact and how they affect collision outcomes in different contexts, and may not be sufficient to reflect changes in influencing factors over longer time spans.

In the analysis of factors related to urban road traffic safety, many dependent variables are binary, such as "whether a traffic accident occurred" and "whether traffic rules were followed". Binary logistic regression analysis can directly handle such dependent variables and also consider multiple independent variables simultaneously, and evaluate their impact and direction on the dependent variable. Therefore, based on the advantages of binary logistic regression analysis, this article conducts research on factors related to urban road traffic safety.

2. Construction of binary logistic traffic safety analysis model

2.1. Determination of four dimensions of influencing factors

To ensure the effectiveness of the experiment, the International Transport Forum (ITF) official website database was used as the main source of experimental data. 5350 data points were extracted from the database to establish a test dataset, and a strict data cleaning process was carried out to remove all outliers, blank records, and data items with unclear and ambiguous definitions. Finally, 3500 data points were retained for subsequent model construction work and imported into SPSS software.

Referring to various types of traffic accident data in the database, this article systematically divides various influencing factors into four dimensions: people, vehicles, roads, and environment, in order to have a more comprehensive understanding of their potential impact on accident occurrence [5]. The specific classification and factors included are as follows:

Human factors: mainly focusing on the characteristics of drivers, including key variables such as gender, age group, education level, and driving experience (i.e. driving experience). These factors directly affect the driver's decision-making ability, reaction speed, and awareness of following traffic rules, thereby having a significant impact on road safety.

The factors of a car: focus on the driving status of the vehicle itself, including but not limited to its speed, stability, braking performance, and whether it is in good maintenance and upkeep. These factors directly affect the handling and safety of the vehicle in emergency situations [15].

Road factors: mainly examine the physical characteristics and design of roads, including whether the physical isolation measures of roads are complete, the type and effectiveness of roadside protection facilities, the types of intersections and road sections (such as highways, urban main roads, intersections, etc.), and the setting and operation of traffic signal systems. These road design elements are crucial for guiding traffic flow, reducing conflict points, and improving driving safety.

Environmental factors: covering both natural and man-made environments, including weather conditions (such as rain, snow, fog, etc.), road lighting conditions, visibility, and road surface smoothness. The changes in these external conditions will directly affect the driver's vision, judgment, and vehicle handling performance, thereby increasing the risk of accidents.

2.2. Independent variable selection and assignment

The independent variable is a variable used in a model to explain or predict changes in the dependent variable. In the binary logistic traffic safety analysis model, the independent variables usually include multiple factors that affect traffic safety, which can be divided from different dimensions such as people, vehicles, roads, environment, etc. [14]. Therefore, this model will divide the specific factors that affect road traffic safety according to different dimensions, and draw a table about the factors that affect urban road traffic safety. They will be divided into four dimensions: people, vehicles, roads, and environment, and the values assigned to each factor will be explained. The four-dimensional factors and their values that affect urban road traffic safety are shown in Table 1.

Tab. 1 - Four dimensional factors affecting urban road traffic safety and their assigned values *(continue)*

Dimension	Influence factor	Variable assignment instructions
People	Driver's age	Assign values by age group, for example: 18-25 years old=1, 26-35 years old=2, and so on
	Driver's gender	Male=1, Female=0
	Driver experience	Assign values based on driving years, for example:<1 year=1, 1-5 years=2, 5-10 years=3,>10 years=4
	Driver behavior (such as drunk driving, fatigue driving)	None=0, Yes=1
Vehicle	Vehicle type	Small cars=1, medium cars=2, large cars=3, special vehicles (such as dangerous goods transport vehicles)=4
	Vehicle condition (such as maintenance status, age)	Good=1, General=2, Poor=3
	Vehicle safety equipment (such as ABS, ESP, etc.)	Fully equipped=1, partially equipped=2, not equipped=3
Road	Road type	Highways=1, urban roads=2, rural roads=3
	Road conditions (such as pavement damage, degree of wetness)	Good=1, General=2, Poor=3
	Road design (such as intersections, number of bends)	Complex=1, General=2, Simple=3
	Road class	Expressway=1, first-class highway=2, second-class highway=3, Class III and below highway=4
	Road width	Narrow (≤ 6m)=1, Medium (6-12m)=2, Wide (>12m)=3
	Number of lanes	Single lane=1, dual lane=2, multi lane (≥ 3 lanes)=3
	Road alignment	Straight line=1, curve=2 (can be further subdivided according to curvature), slope=3 (can be further subdivided according to slope)
	Road surface material	Asphalt=1, cement=2, crushed stone=3, others=4
	Road maintenance status	Good=1, average=2, poor=3, extremely poor=4
	Traffic signs and markings	Perfect=1, basically perfect=2, incomplete=3, missing=4
Environment	Weather conditions (such as rain, snow, fog, etc.)	Clear=1, Rain=2, Snow=3, Fog=4
	Lighting conditions (such as day, night, dusk, etc.)	Daytime=1, Dusk=2, Night=3
	Traffic flow	Low=1, Medium=2, High=3
	Temperature	Low temperature (≤ 0 ° C)=1, moderate (0-30 ° C)=2, high temperature (>30 ° C)=3
	Humidity	Low (≤ 30%)=1, medium (30% -70%)=2, high (>70%)=3
	Noise level	Low (≤ 60dB)=1, medium (60-80dB)=2, high (>80dB)=3

Air quality	Excellent=1, Good=2, Mild Pollution=3, Moderate Pollution=4, Severe Pollution=5
Surrounding buildings	Dense=1, General=2, Sparse=3 (can be further subdivided based on building density and height)

By dividing the factors that affect traffic safety into four dimensions: people, vehicles, roads, and environment, the systematic and comprehensive analysis can be ensured. Each dimension covers multiple aspects that may affect traffic safety, thus avoiding the omission of important factors. In the binary logistic model, the assigned factors can be used as independent variables to input into the model. By comparing the assigned values of different factors, the model can help predict the probability of traffic safety incidents. Reasonable assignment can significantly improve the predictive ability of the model, making it more accurate and reliable.

3. Design of binary logistic regression analysis method

3.1. Traffic data collection and preprocessing

Data collection is crucial for a comprehensive and in-depth exploration of the factors that affect urban road traffic safety. This step not only provides a foundation for accurately evaluating the actual effects of these factors, but also helps identify and analyze the underlying reasons behind them. The data collection of the system provides strong empirical support for subsequent research, enabling a more scientific understanding of the complexity of urban road traffic safety issues. In the process of data collection, various types of data related to urban road traffic safety were systematically sorted and collected from the four dimensions of people, vehicles, roads, and environment [6].

After collecting data on factors affecting urban road traffic safety, normalization can eliminate these differences as they come from different dimensions (people, vehicles, roads, environment), and different categories of data under each dimension may have different dimensions and ranges. This allows data from different dimensions to be compared and analyzed at the same scale, thereby improving the convergence speed and accuracy of the model. Meanwhile, when the data range is too large, it may lead to numerical calculation errors or overflow. Normalization processing can limit data within a reasonable range, thereby avoiding these problems [18]. Therefore, it is necessary to normalize the above data. Convert these raw data to a scale equivalent to [0,1] or [-1,1], so that all factors that affect urban road safety have the same weight in data analysis, thereby more accurately reflecting the impact of each factor on urban road safety.

The original urban road traffic safety related data is represented by ΔD_{xy}, and the calculation formula for normalizing the data affecting urban road traffic safety factors is as follows:

$$D_{xy} = \frac{\Delta D_{xy} - D_{xy}(min)}{D_{xy}(max) - D_{xy}(min)} \tag{1}$$

In the formula, D_{xy} represents the normalized data of the y-th factor affecting urban road traffic safety in the x-th category, $D_{xy}(min)$ represents the minimum value obtained from the data sample of factors affecting urban road traffic safety, $D_{xy}(max)$ represents the maximum value obtained from the data sample of factors affecting urban road traffic safety.

After completing the normalization process, in order to avoid overly relying on data from a certain category and ignoring other category factors during the analysis of factors affecting urban road traffic safety, ensure the fairness of data in each category, and preserve the main data feature information, it is necessary to remove redundant information from the normalized data that affects urban road traffic safety factors [12]. The specific calculation formula is as follows:

$$d = D_{xy} - C(\varepsilon) \tag{2}$$

In the formula, d represents the data of factors affecting urban road traffic safety after removing redundant information, C represents redundant information, ε represents the redundancy level of data on factors affecting urban road traffic safety.

3.2. Binary classification of dependent variables

Logistic regression model is a regression analysis method used to process categorical data. It is particularly suitable for situations where the dependent variable is binary or multi class, and can convert the output of linear combinations into probabilities through logical functions (such as sigmoid functions) to predict the probability of the dependent variable taking a specific category [19]. In the study of urban road traffic safety, logistic regression models can be used to evaluate the impact of different independent variables (people, vehicles, roads, environmental factors in this article) on the probability of traffic accidents. The logistic regression model is nonlinear because the sigmoid function maps the output of a linear combination to the (0,1) interval, representing the probability value [10]. This non-linear feature enables logistic regression models to capture complex relationships between independent and dependent variables. The logistic regression model can not only provide estimates of the effects of independent variables on the dependent variable (i.e. regression coefficients), but also provide statistical significance tests for these effects, helping to identify which independent variables have a significant impact on the occurrence of traffic accidents.

The dependent variable is the target variable that needs to be explained or predicted in the model. In the binary logistic traffic safety analysis model, the dependent variable is usually the binary classification result related to traffic safety, that is, whether a traffic safety event has occurred. Specifically, this dependent variable can be a binary variable [7], such as:

① A traffic safety incident occurs: assigned a value of 1 (also known as "yes", "positive class", etc.).

② No traffic safety incidents occurred: assigned a value of 0 (also known as' no ',' negative class', etc.).

The setting of the dependent variable enables the model to predict the probability of traffic safety incidents by analyzing the independent variables. If the dependent variable is Y and there are n independent variables related to the dependent variable, denoted as $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$, then the probability of an accident occurring due to the influence of the dependent variable Y can be expressed as:

$$P(Y = 1|X) = \frac{1}{1 = +\exp\left[-B_0 + \sum_{i=1}^{n} B_i x_i\right]} \tag{3}$$

In the formula, B_0 represents the regression intercept, B_i represents the regression coefficient of the i-th independent variable x_i.

3.3. Advantage ratio calculation

In the binary logistic regression analysis of factors related to urban road traffic safety, it is necessary to calculate the ratio of the probabilities of accidents occurring and not occurring. This ratio is called the odds of expecting an event (odds) in statistics, denoted as $P/(1-P)$, where P represents the probability of accidents occurring and $1-P$ represents the probability of accidents not occurring. The calculation formula for odd is as follows:

$$odds = \frac{P(Y=1|X)}{P(Y=0|X)} = \frac{P}{1-P} \tag{4}$$

Furthermore, in order to conduct binary logistic regression analysis, logarithmic transformation is usually performed on this ratio, i.e. $\ln[P/(1-P)]$ is calculated. Through this transformation, the probability P originally ranging from (0,1) can be transformed onto the entire real number set, making regression analysis easier. The occurrence rate of accidents can be represented by $EXP(B)$, which is the ratio of the probability $odds_1$ of accidents caused by the experimental group to the probability odds_2 of accidents caused by the reference group. In binary logistic regression analysis, odds ratio is an important indicator to measure the degree of influence of the independent variable on the dependent variable (usually binary, such as whether an accident occurred or not). The specific calculation formula is as follows:

$$EXP(B) = \ln[P/(1-P)]\frac{odds_1}{odds_2} = \frac{P(Y-1|X-1)/P(Y-0|X-1)}{P(Y-1|X-0)/P(Y-0|X-0)} \tag{5}$$

The odds ratio can be used as an indicator to estimate the size of the effect, measuring the degree of influence of a certain independent variable. There are three specific situations:

① In the case of an advantage ratio of 1:

If the odds ratio of a certain independent variable (such as "whether or not to use seat belts") is equal to 1, it means that regardless of whether the driver uses a seat belt, the probability ratio of their occurrence in a traffic accident is the same. In other words, in this particular analysis, the use of seat belts did not significantly affect the probability of traffic accidents [13]. However, this does not mean that seat belts are not important, as in practical situations, seat belts are often seen as a protective factor to reduce accident injuries, but may not be significantly reflected in this particular dataset due to other reasons such as sample size, data quality, etc.

② The situation where the advantage ratio is greater than 1:

If the odds ratio of an independent variable (such as "whether or not drunk driving") is greater than 1, it indicates that drunk drivers have a higher probability of traffic accidents compared to non drunk drivers. In this case, whether or not drunk driving can be considered a risk factor as it increases the risk of traffic accidents.

③ The situation where the advantage ratio is less than 1:

If the odds ratio of an independent variable (such as "road lighting conditions") is less than 1, it indicates that drivers driving under better lighting conditions have a lower probability of traffic accidents compared to drivers driving under poorer lighting conditions. Therefore, 'road lighting conditions' can be considered a protective factor as it reduces the risk of traffic accidents.

3.4. Model factor testing

The use of binary logistic regression analysis can help understand which factors have a significant impact on road traffic safety, and thus develop effective safety measures. However, when applying binary logistic regression models, attention should be paid to the correlation between model factors [8]. Although binary logistic regression models do not strictly require complete exclusion or no correlation between independent variables, when there is a high degree of collinearity between the independent variables, it can affect the significance test of the coefficients, making it difficult to determine which variables truly have a significant impact on the results [3]. And in the case of high collinearity, even small changes in the sample data may result in significant fluctuations in the estimated values of the coefficients. Therefore, multicollinearity test was used to analyze the relationship between different independent variables.

$$T(O_i) = 1 - R_i^2 \tag{6}$$

$$V(F_i) = \frac{1}{1-R_i^2} \tag{7}$$

In the formula, $T(O_i)$ represents the tolerance of the i-th independent variable, R_i^2 represents the square of the complex correlation coefficient between the i-th independent variable and other independent variables, $V(F_i)$ represents the variance inflation factor of the i-th independent variable.

As R_i^2 approaches 0 and $V(F_i)$ approaches 1, the multicollinearity between independent variables becomes weaker, The closer R_i^2 is to 1, the larger $V(F_i)$, and the stronger the multicollinearity between independent variables. The results of the collinearity test are shown in Table 2.

After conducting a collinearity test on the independent variables, the results showed that the $V(F_i)$ values between the 25 independent variables were all around 1, indicating that there was no obvious collinearity between the independent variables and they could be included in the model for analysis as independent variables.

3.5. Binary logistic regression analysis

After completing the collinearity test analysis of the independent variables, this article conducted a binary logistic regression analysis on the factors related to urban road traffic safety. Further reveal and quantify the impact of these factors on urban road traffic safety.

Tab. 2 - Results of collinearity test

Dimension	Independent variable	$T(O_i)$	$V(F_i)$
People	Driver's age	0.71	1.03
	Driver's gender	0.75	1.02
	Driver experience	0.79	1.15
	Driver behavior (such as drunk driving, fatigue driving)	0.99	1.55
Vehicle	Vehicle type	0.69	1.13
	Vehicle condition (such as maintenance status, age)	0.75	1.16
	Vehicle safety equipment (such as ABS, ESP, etc.)	0.74	1.25
Road	Road type	0.72	1.15
	Road conditions (such as pavement damage, degree of wetness)	0.79	1.25
	Road design (such as intersections, number of bends)	0.74	1.25
	Road class	0.72	1.15
	Road width	0.71	1.16
	Number of lanes	0.70	1.04
	Road alignment	0.71	1.08
	Road surface material	0.70	1.06
	Road maintenance status	0.71	1.12
	Traffic signs and markings	0.76	1.02
Environment	Weather conditions (such as rain, snow, fog, etc.)	0.87	1.25
	Lighting conditions (such as day, night, dusk, etc.)	0.81	1.12
	Traffic flow	0.79	1.08
	Temperature	0.65	1.02
	Humidity	0.72	1.04
	Noise level	0.64	1.00
	Air quality	0.62	1.00
	Surrounding buildings	0.57	1.00