WILEY | Hindawi

Research Article

Driver Route Planning Method Based on Accident Risk Cost Prediction

Xiaoleng Liao [ID],[1] **Tong Zhou** [ID],[1] **Xu Wang** [ID],[1] **Rongjian Dai** [ID],[1] **Xuehui Chen,**[2] **and Xiangmin Zhu**[2]

[1]*School of Qilu Transportation, Shandong University, Jinan, China*
[2]*Shandong Hi-Speed Company Limited, Jinan, China*

Correspondence should be addressed to Xu Wang; xuwang@sdu.edu.cn

The number of cars on roadways around the world continues to increase year over year. However, the imbalance between traffic supply and demand has not only brought traffic congestion but also caused serious safety problems. To reduce travel risk, this study proposes a driver route planning method based on accident risk cost prediction for connected and automated vehicles. According to the entropy weight method and an improved algorithm of K shortest paths, a route planning model with accident risk as the main optimization objective was established. Firstly, an accident risk evaluation system was built based on traffic accident data, and a quantitative prediction model of accident risk cost based on driver-, vehicle-, road-, and environment-related factors was constructed. Secondly, the entropy weight method was used to calculate the weights of each indicator to determine accident risk considering the aforementioned factors. Then, the route planning model was established, and the solution algorithm based on K shortest paths was designed to solve the optimal route by comprehensively considering accident risk cost and travel time. The accident risk index of each road section in the example road network was assigned, and the risk of the road section was quantified according to the accident risk cost model. Three candidate paths were calculated by using the path planning algorithm proposed in this study; the total risk cost is 6.19, 6.26, and 6.39, respectively; and the total travel time is 29, 29, and 31, respectively. After comparison, the optimal path and two alternative paths are obtained. The results show that the accident risk cost prediction model based on historical accident data can be used to quantify driving risk. The proposed method can help drivers in the connected and automated environment choose the optimal travel route with the lowest risk and shortest travel time and improve overall traffic safety and efficiency.

1. Introduction

With the rapid development of the global road transportation system, the imbalance between traffic supply and demand has become pronounced, which has brought severe traffic congestion and accidents and has seriously threatened people's lives and property. For example, in China, the number of traffic accidents increased year over year from 2015 to 2020, and the number of casualties was also huge. On average, about 60000 people died in traffic accidents each year [1]. Advanced driving assistance systems (ADAS) and connected and automated vehicles (CAV) can provide route planning and intelligent guidance for drivers. They are two of the most important means to improve the safety and efficiency of the traffic network, alleviate urban traffic congestion, and mitigate travel risk. However, conventional route planning takes the minimum distance or driving time as the optimization objective, and there is relatively little research based on the risk cost of road accidents. Therefore, this study quantified the accident risk and designed a route planning model based on the accident risk cost to provide drivers with a safer travel path, which can effectively reduce the risk of traffic accidents and improve overall safety.

In recent years, there have been many research achievements in traffic safety, which can be divided into two categories: (1) evaluations of traffic safety based on accident

data using the Bayesian network and the accident rate method and (2) evaluations of accident risk carried out by the analytic hierarchy process, entropy weight method, and fuzzy evaluation method in accordance with the indicator evaluation system based on the characteristics of people, vehicles, roads, and the surrounding environment. In terms of studying traffic safety using accident data, Mbakwe et al. [2] established a model combining Delphi technology and the Bayesian network to predict the accident rate and evaluate national traffic safety. Chu [3] used the ordered logit model to analyze the causes of serious accidents involving buses traveling on expressways over a long driving time. The study showed that fatigued driving, drivers or passengers not wearing seat belts, drunk driving, and other behavioral factors have a significant impact on the severity of accidents. Mohan et al. [4] used the accident rate method to evaluate urban traffic safety based on fatal traffic accident data in six cities of India. The study found that the vast majority of deaths in traffic accidents fall in vulnerable traffic subjects (pedestrians, cyclists, electric vehicle drivers, and motorcycle users). Eusofe and Evdorides [5] and Gomes et al. [6] evaluated traffic safety from the traffic management perspective based on accident data. At the same time, some scholars have established models based on the crash data to quantify the risk of road accidents and evaluate road traffic safety. Xie and Yan [7] used kernel density function to fit the spatial distribution of road network of traffic accident risk in Kentucky and Burlington, New York, and estimated its traffic accident risk status. Anderson [8] and Bil et al. [9] identified the hotspots of road accidents by K-means clustering, significance test, and other methods according to the distribution of traffic risks on the road network. However, these studies ignore the impact of driver behavior on traffic safety risks, and several studies in the US report that approximately 90% of the light-vehicle crashes involved the same type of human error such as impaired conditions, inadvertent errors, and risky driving behavior [10–13]. Therefore, the impact of driver behavior on road accident risk cannot be ignored. Jiang et al. [14] proposed a safe route mapping (SRM) model, which uses real historical collision data and driver simulation data obtained based on VISSIM simulation to score road safety and establishes safety risk heat maps for roads; drivers can use the road heat maps for situational awareness and trip planning. Arbabzadeh and Jafari [15] used the elastic net regularized multinomial logistic expression to establish a safety prediction model based on driver based data and quantified the traffic safety risk as the likelihood of adverse driving outcomes, to achieve real-time scoring of road safety risks.

In terms of establishing accident risk assessment indicators based on the characteristics of people, vehicles, roads, and environment, Li et al. [16] and Zhao et al. [17] determined the risk evaluation index system of road transportation of dangerous goods through literature research and expert consultation and used analytic hierarchy process (AHP) to evaluate the risk of transportation of dangerous goods. Fernandez et al. [18] studied 535 drivers in Manila to score and rank the factors and indicators affecting accident occurrence (such as bad driving behaviors, cognition of

traffic signs, and distracted driving) by using a questionnaire and an AHP to determine the weight of each index. The results showed that bad driving behavior is the main cause of accidents. Temrungsie et al. [19] established a road safety evaluation index system according to the United Nations' white paper on road safety. They interviewed 100 experts engaged in traffic-related industries in Thailand, scored the indexes, and analyzed the influencing factors of road traffic safety using the AHP. They found that the traffic management in Thailand was chaotic, and the implementation of traffic regulations must be strengthened. Cai et al. [20] established a traffic safety risk prediction index system based on driving behavior data. They proposed a road traffic safety entropy calculation method based on the entropy weight method and scaled the road traffic safety risk level by K-means clustering. Guo et al. [21] collected data on driver eye movement and vehicle traveling state through driving simulations. They constructed a driver behavior index system, simplified the index by principal component analysis, and finally calculated the weights of the characteristic indexes by the entropy weight method to evaluate the impact of behaviors on traffic safety.

Most of the influencing factors of route planning in the existing research focus on minimizing the driving distance and travel time. These studies rarely consider the driver's personal characteristics, travel purpose, or the road environment. Additionally, the priority for drivers to choose a travel path is not consistent, so the meaning of "optimal" is limited and subjective. The path recommended by a given model does not necessarily meet the expectations of drivers. Therefore, many scholars have incorporated factors affecting driver choice into the route planning model: Pang et al. [22] used the fuzzy neural network method to train the driver's historical trip data to reflect the driver's travel preference and provide guidance for the travel route selection of onboard navigation equipment. Lee et al. [23] estimated the delay caused by bad driving behaviors through discrete selection analysis. Then they compared the travel time of different paths to recommend the route with the most reliable travel time. As mentioned above, there are many factors affecting driver travel choice, and as such, path guidance considering traffic safety has attracted increasing attention. Karim and Sayed [24] established the shortest path model integrating travel time and safety by analyzing the relationship between traffic conflict and collision. Payyanadan et al. [25] used the collision accident data of elderly drivers to quantify accident risk influencing factors, such as left turn, U-turn, and travel distance. They evaluated the safety of the path on this basis to help elderly drivers choose a safer route and reduce their accident risk. Zhang et al. [26] established a prediction model of route travel time and accident risk cost according to different parameters, such as traffic volume and capacity, and designed the route planning algorithm for drivers with different risk tendencies from the perspective of generalized travel cost.

It can be observed from the above literature review that many scholars have engaged in significant research on traffic safety and route planning. However, there are still some limitations:

(1) Research on traffic safety based on accident data used statistics from after the accident that has already occurred. While this can show the postaccident safety state, it lacks risk cost prediction to visualize the preaccident safety state and suggest safer routes to avoid accidents before they occur—which is obviously more critical for prevention and control.

(2) Research on travel risk according to the index system based on the physical characteristics of people, vehicles, roads, and the surrounding environment mostly combined objective analysis and subjective evaluation to evaluate risk through surveys, expert scoring, and the AHP. However, surveys are highly subjective and greatly influenced by the risk preference and experience of scoring experts, and the evaluation is neither accurate nor objective due to the lack of data support.

Additionally, conventional route planning methods often take the minimum driving distance or travel time as the optimal goal, rather than driver preference and driving safety. Although there are route planning studies that do consider preference and safety, there are still few studies on constructing a route planning model based on the accident risk cost specifically. Therefore, based on accident data, this study constructed a risk evaluation system using accident characteristics, established an accident risk quantification model based on the entropy weight method, and designed a route planning algorithm that comprehensively considers accident risk cost and travel time to provide the safest and shortest route for drivers and improve overall safety.

2. Research Method

To realize route planning based on accident risk, it was first necessary to quantify the accident risk of the road section. This study used accident data and physical characteristic indexes of drivers, vehicles, roads, and the surrounding environment, all factors that could have influenced the severity of the accident in the data set, to construct the risk evaluation system. Then, the quantitative model of accident risk cost was established using the entropy weight method to calculate the index weights. Based on this, the real-time risks of road sections were calculated, the travel time of the road sections was loaded, and the example network was constructed. Then, the algorithm of restricted loopless K shortest paths was improved and applied to the model. Taking accident risk cost as the main goal and simultaneously considering travel time, the optimal path was found in the network. The overall workflow of this study is shown in Figure 1.

2.1. Accident Risk Quantification

2.1.1. Accident Risk Quantification Model.
In the traffic system, the temporal and spatial changes of drivers, vehicles, roads, and the surrounding environment will affect the occurrence and severity of traffic accidents at any given time. There are many existing studies on traffic safety analysis through the construction of a risk index system, but there are
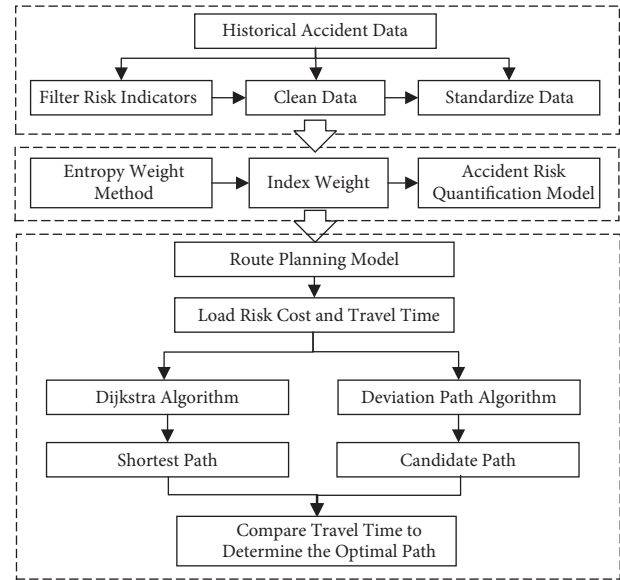
Figure 1: Research flowchart.

many subjective assumptions about the selection of indicators and the determination of weights. Therefore, this study selected risk evaluation indicators from a traffic accident data set and calculated the index weights and the comprehensive scores according to the real-time data. It defined the comprehensive score as the accident risk, which can reflect the impact of various factors on the accident risk.

$$Z_i = \sum_{j=1}^{m} d_{ij} w_j \quad i = 1, \ldots, n. \tag{1}$$

where Z_i is the accident risk cost of the ith road section; d_{ij} is the actual data of the jth index corresponding to the ith road section; w_j is the weight of the jth index.

2.1.2. Index Weight Calculation.
Several different methods can calculate index weights, including the entropy weight method, AHP, and principal component analysis. Among them, the entropy weight method is an objective weighting method, which has higher reliability and accuracy than subjective weighting, and it can deeply reflect the distinguishing ability of indicators and determine better weights. According to the basic principles of information theory, information is a measure of the order degree of the system, and entropy is a measure of the disorder degree of the system. According to the definition of information entropy, entropy can be used to judge the dispersion degree of an index for a certain index. The smaller the information entropy, the greater the dispersion degree of the index, and the greater the impact of the index on the comprehensive evaluation (i.e., weight). This method is more suitable for describing the impact of abnormal values in drivers, vehicles, roads, environmental, and other indicators on the severity of accidents. For example, for several different traffic accidents, if the value of one index changes greatly, while the value of other indexes basically does not change, it indicates that the

indicator has led to the difference of accidents, and a greater weight can be taken. Therefore, this study selected the entropy weight method to calculate the index weights. The specific calculation steps are as follows:

(1) For n samples and m indexes, x_{ij} is the value of the jth index of the ith sample

(2) Normalize indexes for the homogenization of heterogeneous indexes

Positive index:

$$x_{ij} = \frac{x_{ij} - \min\{x_{1j}, \ldots, x_{nj}\}}{\max\{x_{1j}, \ldots, x_{nj}\} - \min\{x_{1j}, \ldots, x_{nj}\}}. \qquad (2)$$

Negative index:

$$x_{ij} = \frac{\max\{x_{1j}, \ldots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, \ldots, x_{nj}\} - \min\{x_{1j}, \ldots, x_{nj}\}}. \qquad (3)$$

(3) Calculate the proportion of the i$^{\text{th}}$ sample value under the j$^{\text{th}}$ index:

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}},$$

$$i = 1, \cdots, n, \qquad (4)$$

$$j = 1, \ldots, m.$$

(4) Calculate the entropy of the j$^{\text{th}}$ index:

$$e_j = -k \sum_{i=1}^{n} P_{ij} \ln(P_{ij}), \; j = 1, \cdots, m \qquad (5)$$

where $k = 1/\ln(n) > 0$, meeting $e_j \geq 0$.

(5) Calculate information entropy redundancy (difference):

$$d_j = 1 - e_j, \; j = 1, \ldots, m. \qquad (6)$$

(6) Calculate the weight of each index:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}, \; j = 1, \cdots, m. \qquad (7)$$

x_{ij} is the standardized data.

It should be noted that the entropy weight method can calculate the index weight values, but there is a problem that the entropy value of a zero index cannot be calculated in the process of practical application. Therefore, when an index value was zero, a value of 0.00001 was added to the evaluation index data of this group; adding such a small increment not only enabled the data group to be valid, but it also ensured a small impact on the difference of each index [20].

2.2. Construction of Route Planning Model.
The traditional route planning algorithm is to add the weights of each side in the network graph to find the shortest path. K shortest paths

(KSP) problem is a deformation of the shortest path problem. Different from the traditional shortest path problem, the purpose of the KSP problem is to find multiple alternative optimization paths between the start point and the end point in the network graph and form the shortest path group to meet the user's selection needs to the greatest extent [27]. Based on the improvement of the algorithm of K shortest path in existing research [28], this study designed a route planning algorithm meeting multiple objectives; that is, in the K shortest path set obtained by calculating the accident risk cost, the path with the shortest travel time T and lowest risk was chosen as the optimal path.

2.2.1. KSP Problem.
Suppose $G = (V, E)$ represents a network graph, where V is the set of n nodes and E is the set of m edges. Each edge e_k in E is represented by a node pair; that is, $e_k = (i, j)$, and $c_{i,j}$ is the length of this side. Suppose s and t are two nodes in graph G. The path p from s to t in the diagram is represented by the node sequence; that is, $p = (v_1 = s, v_2, \cdots, v_h = t)$. s and t are respectively the start node and end node of p. The length $c(p)$ of p is the total length of all sides on p; that is, $c(p) = \sum_{(i,j) \in p} c_{ij}$.

The path set from s to t is represented by P_{st}. The shortest path problem is to find the path p^* with the smallest length from s to t. The KSP problem is a generalization of the shortest path problem. In addition to determining the shortest path, it also needs to determine the second shortest path and the third shortest path until the Kth short path is found. p_k represents the Kth shortest path from s to k.

According to path constraints, KSP problems are usually divided into two types: a general KSP problem and a restricted loopless KSP problem. The general KSP problem has no restrictions on the path. The restricted loopless KSP problem requires that the obtained path is simple, and it cannot contain a loop. The network graph in this study did not contain loops, so only the restricted loopless KSP problem was analyzed.

2.2.2. Restricted Loopless KSP Algorithm.
This study used the deviation path algorithm to solve the restricted loopless KSP problem. The core of the deviation path algorithm is how to find p_{k+1} by using the shortest deviation path of the obtained $p_1, p_2, \ldots, p_k$. Firstly, the Dijkstra algorithm was used to find the shortest path from s to t and put it into the path set P_k as p_1. After calculating $\{p_1, p_2, \ldots, p_k\}$ of the previous k paths, the calculation process of p_{k+1} was as follows:

(1) Take each node v_i except the end node in p_k as the possible deviation point, and calculate the shortest path from v_i to the node t. To avoid repetition with the previously found path, the side separated from the node v_i could not be the same as the side separated from v_i on the previously found shortest path $p_1, p_2, \cdots, p_k$.

(2) Splice the shortest path from the found v_i to the node t with the path from s to v_i on the current node p_k to form a candidate path of p_{k+1} and save it in the candidate path set P_k'.

(3) Select the shortest path from the candidate path set P_k' as p_{k+1}, and put it into the path set P_k.

Repeat the above steps until K paths are obtained.

2.2.3. Multiobjective Route Planning Model.

The above section describes the basic concept and solution algorithm of the KSP problem. In the next phase of the study, a multiobjective path planning model was designed based on the deviation path algorithm to obtain the optimal route. Comprehensively considering the accident risk cost and route travel time, the specific steps of the improved algorithm of KSP was as follows:

(1) Initialize network to determine the start point s and end point t of path as well as K alternative paths.

(2) Load the accident risk cost of each section and the travel time of each section calculated by (1) into the traffic network.

(3) Aiming at the accident risk cost, obtain the shortest path from s to t by using the Dijkstra algorithm, which is recorded as $(k = 1)$.

(4) If $k \geq K$, then move to the final step. If $k < K$, then all other nodes v_i except for the end point t on p_k are regarded as deviation nodes, and there are totally x pieces of $(0 \leq i \leq x)$.

(5) Traverse all deviation points, find the shortest path from each deviation point v_i to the end point t, splice the path from the start point s to v_i with the shortest path from v_i to the end point t on p_k, and save it in the set as P_k', as a candidate path.

(6) If the candidate path set P_k' is empty, go to the last step. If it is not empty, calculate the travel time of each candidate path and find the path with the shortest time; that is, p_{k+1}. Remove the path from the set P_k', and put it into the set P_k and return to step 4.

(7) From the K candidate paths, select the path with the shortest travel time as the optimal path to obtain the results.

3. Research Results

3.1. Data.

To achieve objective weighting and analyze the correlations between indicators according to the data characteristics, this study used traffic accident data from the National Highway Traffic Safety Administration (NHTSA) from the year of 2019 [29]. NHTSA uses data from many sources, including the Fatality Analysis Reporting System (FARS) which began operation in 1975. FARS provides data about fatal crashes involving all types of vehicles; therefore, it was possible to obtain indicators of impact on traffic accidents from this data set. The basic data was saved in comma separated values (CSV) files. For the 2019 data collection year, there were 23 data files. This study selected the files that describe the states of people, vehicles, roads, and the surrounding environment at the time of the accident.

3.2. Data Preprocessing.

After obtaining the data files, it was necessary to systematically screen the risk indicators in the data set according to the four elements of the transportation system: people, vehicles, roads, and environment. For the selection of indicators, firstly, this study defined the indicator factors contained in transportation risk through existing literature research, expert consultation, and analysis of accident investigation reports in recent years [3, 16, 17] and then selected the indicators in data files to delete the unnecessary indicators. For example, the relevant data files of the involved people describe the information of all persons involved in collisions, including the demographic information of the driver and passengers and the driving maneuvers before the accident. As most accidents are related to driver behaviors, this study only used data from drivers' demographic information and behaviors, so the indicators related to passengers were ignored. After screening each file, 22 indicators, including time of accident, driver gender, road alignment, and weather, were obtained to construct the indicator set of accident cost impact factors (see Table 1). The data structure in the data file was complex, and the sample size was large. It was necessary to clean the data and delete abnormal data, including null values, unreported values, and reported-as-unknown values. The original data of indicators was saved in the data table in the form of numbers. The indicator meanings corresponding to different numbers can be obtained by checking the FARS Analytical User's Manual [29] (see Table 1).

After determining the impact factors, the impact factors located in different data files needed to be integrated into one data file to facilitate the subsequent calculation by the entropy weight method. After deleting the abnormal data, different indicators included different data sample sizes. Therefore, this study used the accident number to connect the selected data files and merge them into a data table. The combined data table contained 26218 accident data points, and data on each accident contained 22 cost impact indicators. Through simple statistical analysis of the processed data (see Table 2), it was found that drivers involved in 3113 accidents were aged between 15 and 20, accounting for 11.9%. Younger drivers generally have a more aggressive driving style and usually ignore traffic regulations, which have a certain impact on accident risk. Additionally, there were 5673 accidents involving drivers aged elder than 65, accounting for 21.6%. Through comparison, it was observed that the proportion of traffic accidents among elderly drivers was higher than that among young drivers aged between 15 and 20. This is because elderly drivers generally response slowly, and it is easy to cause accidents if they do not respond to emergencies in time. Simple statistical analysis cannot accurately explain the impact of each indicator on accident risk. Therefore, in the next phase of the study, the entropy weight method was used to calculate the weight of each indicator to determine the impact of each indicator on accident risk.

3.3. Quantification of Accident Risk Cost.

This study used python to realize the entropy weight method. First, in the merged data table, the number of columns—excluding the

TABLE 1: Indicator set of accident cost impact factors.

No.	Index	Definition	Description
1	A_D15_20	Crashes involving a young driver (aged 15–20)	Young driver involved crash (aged 15–20) = 1, otherwise = 2
2	A_D65PLS	Crashes involving an older driver (aged 65+)	Older driver involved crash (aged 65+) = 1, otherwise = 2
3	A_DIST	Involving a distracted driver	Involving a distracted driver = 1, otherwise = 2
4	A_DROWSY	Involving a drowsy driver	Involving a drowsy driver = 1, otherwise = 2
5	A_RELRD	Relationship to the traffic way	On roadway = 1, off median = 2, off shoulder = 3, off roadway = 4
6	A_SPCRA	Involving speeding	Speeding involved crash = 1, otherwise = 2
7	DRYNK_DR	Number of drunk drivers	Number of drunk drivers = 1, 2, 4, no drunk drivers = 0
8	REST_USE	Restraint system use	None used = 0; otherwise = 1
9	Sex	Driver gender	Female driver involved crash = 2, otherwise = 1
10	A_LIC_S	License status	Valid = 1, invalid = 0
11	PREV_ACC	Previous recorded crashes	Number of previous recorded crashes = 1, 2, 3, 4, 5, 6, 7, 8, no record = 0
12	A_LT	Involving a large truck	Large truck involved CRASH = 1, otherwise = 2
13	HAZ_INV	Hazardous material involvement	No = 1, yes = 2
14	A_RU	Land use (rural/urban)	Rural = 1, urban = 2
15	A_INTSEC	Intersection	Intersection = 1, nonintersection = 2
16	VALIGN	Roadway alignment	Straight = 1, curve = 2
17	VSURCOND	Roadway surface condition	Dry = 1, wet = 2, mud, sand = 3, dirt or gravel = 4, snow = 5, water (standing or moving) = 6, slush = 7, ice/frost = 8
18	VTRAFCON	Traffic control device	No controls = 0, otherwise = 1
19	A_DOW	Day of week	Weekday = 1, weekend = 2
20	A_TOD	Time of day	Daytime = 1, nighttime = 2
21	WEATHER	Atmospheric conditions	Clear = 1, cloudy = 2, rain = 3, fog = 4, freezing rain or drizzle = 5, snow = 6, sleet or hail = 7, severe crosswinds = 8, blowing sand, soil, dirt = 9, blowing snow = 10
22	LGT_CON	Light condition	Daylight = 1, dark-lighted = 2, dawn = 3, dusk = 4, dark-not lighted = 5

TABLE 2: Descriptive statistics.

Number	Index	Statistical value (proportion)
1	A_D15_20	1 = 3113 (11.90%), 2 = 23105 (88.10%)
2	A_D65PLS	1 = 5673 (21.60%), 2 = 20545 (78.40%)
3	A_DIST	1 = 2169 (8.30%), 2 = 24049 (91.70%)
4	A_DROWSY	1 = 538 (2.10%), 2 = 25680 (97.90%)
5	A_RELRD	1 = 16236 (61.90%), 2 = 381 (1.50%), 3 = 987 (3.80%), 4 = 8614 (32.80%)
6	A_SPCRA	1 = 6720 (25.60%), 2 = 19498 (74.40%)
7	DRYNK_DR	0 = 19424 (74.10%), 1 = 6592 (25.10%), 2 = 201 (0.80%), 4 = 1 (0.0004%)
8	REST_USE	0 = 14934 (57.00%), 1 = 11284 (43.00%)
9	Sex	1 = 17064 (65.10%), 2 = 9154 (34.90%)
10	A_LIC_S	0 = 8349 (31.80%), 1 = 17869 (68.20%)
11	PREV_ACC	0 = 19989 (76.20%), 1 = 4538 (17.30%), 2 = 1197 (4.60%), 3 = 345 (1.30%), 4 = 104 (0.40%), 5 = 33 (0.13%), 6 = 9 (0.03%), 7 = 2 (0.0008%), 8 = 1 (0.0004%)
12	A_LT	1 = 3665 (14.00%), 2 = 22553 (86.00%)
13	HAZ_INV	1 = 26116 (99.60%), 2 = 102 (0.40%)
14	A_RU	1 = 12113 (46.20%), 2 = 14105 (53.80%)
15	A_INTSEC	1 = 6260 (23.90%), 2 = 19958 (76.10%)
16	VALIGN	1 = 20020 (76.40%), 2 = 2088 (8.00%), 3 = 3628 (13.80%), 4 = 482 (1.80%)
17	VSURCOND	1 = 22060 (84.10%), 2 = 3442 (13.10%), 3 = 95 (0.36%), 4 = 239 (0.91%), 5 = 55 (0.21%), 6 = 58 (0.22%), 7 = 269 (1.00%)
18	VTRAFCON	0 = 20041 (76.40%), 1 = 6177 (23.60%)
19	A_DOW	1 = 15755 (60.10%), 2 = 10463 (39.90%)
20	A_TOD	1 = 12845 (49.00%), 2 = 13373 (51.00%)
21	WEATHER	1 = 19153 (73.10%), 2 = 4254 (16.20%), 3 = 2125 (8.10%), 4 = 288 (1.10%), 5 = 33 (0.13%), 6 = 261 (1.00%), 7 = 33 (0.13%), 8 = 44 (0.17%), 9 = 6 (0.02%), 10 = 21 (0.08%)
22	LGT_CON	1 = 12655 (48.30%), 2 = 4867 (18.60%), 3 = 553 (2.10%), 4 = 598 (2.30%), 5 = 172 (0.66%), 6 = 7373 (28.10%)

column communicating accident number—was defined as the number of indicators, and the number of rows—excluding the row communicating index names—was defined as the number of schemes, i.e., evaluation subjects. Then, the data were standardized, during which it was necessary to judge whether the indicator was a positive indicator or a negative indicator. A positive indicator is an indicator that is better when it gets larger; on the contrary, a negative indicator is an indicator that is worse when it gets larger. In this study, for the indicator A_DIST (involving a distracted driver), the value of 1 indicated that distracted driving was involved in the accident, and the value of 2 indicated that distracted driving was not involved. A higher value was better for the evaluation result. Therefore, this was a positive indicator. Similarly, each indicator was judged and the code was entered. Finally, the entropy method function was defined, and the weights of the variables were calculated. The index entropy and weights were obtained through calculation (see Table 3) to realize the objective weighting based on historical accident data. It was mentioned in Section 2.1.2 that the smaller the entropy of the index, the greater the weight. Table 3 shows that the indicator of VTRAFCON (traffic control device) was the smallest entropy and the largest weight. Therefore, whether there are traffic control measures at the location of the accident will have a great impact on accident risk. Finally, by substituting the index weights into (1), the calculation equation of the accident risk cost of vehicles driving in a certain section was determined.

3.4. Case Study of Route Planning. In the previous step, the index weights were calculated, and the calculation equation of road accident risk cost was determined. In the next step, the route planning example network was constructed to realize the route planning with accident risk cost as the optimal goal based on the improved K shortest path algorithm. The example road network constructed consisted of 10 nodes and 14 road sections. The specific road network is shown in Figure 2. By assuming the real-time driver, vehicle, road, environment, and other parameters of the road network, the corresponding attributes of each road section were given; that is, the actual values of 22 indicators included in each road section were determined. On this basis, the accident risk cost of each road section was calculated according to (1). At the same time, the travel time of each road section was assigned. The results are shown in Table 4.

Python was used to implement the route planning model algorithm. First, the Dijkstra algorithm and deviation path algorithm function were defined. Then, the network structure and the start and end points of the network were determined. The network parameters were input, K was set to 3 (i.e., three candidate paths were calculated), and the Dijkstra algorithm function was called to calculate the shortest path $p_1 = (v_1, v_3, v_4, v_5, v_9, v_{10})$ from v_1 to v_{10} aiming at the accident risk cost, and the path risk cost was $\sum Z_i = 6.19$. Then, the deviation path algorithm was called to determine the second shortest path. The detailed calculation steps of p_2 are as Table 5.

Table 5 shows that there are five deviation points in total for finding p_2. It was necessary to calculate each deviation

TABLE 3: Weights of accident cost impact factors.

| Number	Index	e_j	w_j
1	A_D15_20	0.967	0.027
2	A_D65PLS	0.958	0.035
3	A_DIST	0.970	0.024
4	A_DROWSY	0.975	0.020
5	A_RELRD	0.938	0.050
6	A_SPCRA	0.953	0.038
7	DRYNK_DR	0.953	0.039
8	REST_USE	0.907	0.077
9	Sex	0.942	0.047
10	A_LIC_S	0.946	0.044
11	PREV_ACC	0.955	0.037
12	A_LT	0.965	0.029
13	HAZ_INV	0.977	0.019
14	A_RU	0.926	0.060
15	A_INTSEC	0.955	0.037
16	VALIGN	0.957	0.037
17	VSURCOND	0.963	0.030
18	VTRAFCON	0.852	0.122
19	A_DOW	0.936	0.053
20	A_TOD	0.918	0.067
21	WEATHER	0.952	0.039
22	LGT_CON	0.917	0.068
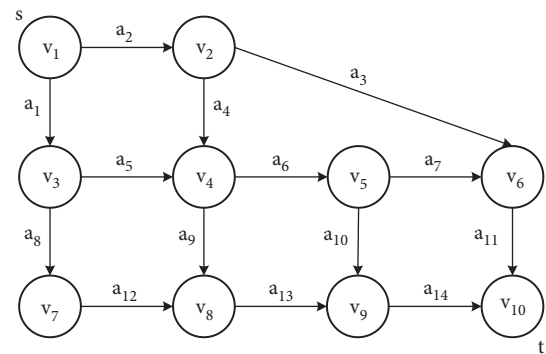
FIGURE 2: Road network.

point to ensure whether a new candidate path could be generated. After traversing each deviation point, four candidate paths were obtained. The accident risk cost of each path was 6.39, 6.58, 6.26, and 7.86. The travel time of each candidate path was calculated as 31, 31, 29, and 33, respectively. At this time, the path with the shortest travel time was selected as p_2. Therefore, the output second path was $p_2 = (v_1, v_3, v_4, v_8, v_9, v_{10})$.

Similarly, the third shortest path was calculated according to the improved deviation path algorithm. The model output of the third path was $p_3 = (v_1, v_2, v_4, v_5, v_9, v_{10})$. At this time, $k = 3$ was met, three candidate paths were obtained (see Table 6), and the program stopped calculation. Referring to Table 6, the path with the shortest travel time was selected as the optimal path. The travel time of paths p_1 and p_2 was 29, but the risk cost of p_1 was lower, so $p_1 = (v_1, v_3, v_4, v_5, v_9, v_{10})$ was selected as the optimal path of the network (see Figure 3).

Most conventional path planning methods take the shortest travel time as the optimization goal. Therefore, this

Table 4: Network parameters.

Road section	A_D15_20	A_D65PLS	A_DIST	. . .	VSURCOND	A_LIC_S	Z_i	Travel time t_0
a_1	2	2	2	. . .	1	1	1.44	6
a_2	2	2	2	. . .	2	1	1.82	7
a_3	2	2	2	. . .	7	1	2.95	10
a_4	2	1	2	. . .	1	1	1.02	6
a_5	2	1	2	. . .	1	2	1.20	5
a_6	1	2	2	. . .	1	1	1.31	7
a_7	2	2	2	. . .	2	1	1.52	8
a_8	2	2	2	. . .	1	1	1.40	5
a_9	2	2	2	. . .	1	1	1.48	4
a_{10}	2	2	2	. . .	1	1	1.19	5
a_{11}	2	2	2	. . .	2	1	2.39	7
a_{12}	2	2	2	. . .	2	0	1.60	6
a_{13}	2	1	2	. . .	1	1	1.09	8
a_{14}	2	2	2	. . .	1	0	1.05	6

Table 5: Deviation path algorithm.

Deviation point	Start to deviation	Inapplicable side	Inapplicable node(s)	Shortest path from deviation to end	Generated candidate path	Put into P_k' or not
v_1	null (0)*	$v_1 \longrightarrow v_3$	null	$v_1 \longrightarrow v_2 \longrightarrow v_4, \longrightarrow v_5 \longrightarrow v_9 \longrightarrow v_{10}$ (6.39)	$v_1 \longrightarrow v_2 \longrightarrow v_4, \longrightarrow v_5 \longrightarrow v_9 \longrightarrow v_{10}$ (6.39)	Yes
v_3	$v_1 \longrightarrow v_3$ (1.44)	$v_3 \longrightarrow v_4$	v_1	$v_1 \longrightarrow v_3 \longrightarrow v_7, \longrightarrow v_8 \longrightarrow v_9 \longrightarrow v_{10}$ (5.14)	$v_1 \longrightarrow v_3 \longrightarrow v_7, \longrightarrow v_8 \longrightarrow v_9 \longrightarrow v_{10}$ (6.58)	Yes
v_4	$v_1 \longrightarrow v_3 \longrightarrow v_4$ (2.64)	$v_4 \longrightarrow v_5$	v_1, v_3	$v_4 \longrightarrow v_8 \longrightarrow v_9, \longrightarrow v_{10}$ (3.62)	$v_1 \longrightarrow v_3 \longrightarrow v_4, \longrightarrow v_8 \longrightarrow v_9 \longrightarrow v_{10}$ (6.26)	Yes
v_5	$v_1 \longrightarrow v_3 \longrightarrow v_4, \longrightarrow v_5$ (3.95)	$v_5 \longrightarrow v_9$	v_1, v_3, v_4	$v_5 \longrightarrow v_6 \longrightarrow v_{10}$ (3.91)	$v_1 \longrightarrow v_3 \longrightarrow v_4, \longrightarrow v_5 \longrightarrow v_6 \longrightarrow v_{10}$ (7.86)	Yes
v_9	$v_1 \longrightarrow v_3 \longrightarrow v_4, \longrightarrow v_5 \longrightarrow v_9$ (5.14)	$v_9 \longrightarrow v_{10}$	v_1, v_3, v_4, v_5	None	None	No

*The numbers in the brackets of the table indicate the accident risk.

Table 6: Shortest paths.

Shortest path	$\sum Z_i$	Travel time T
$p_1 = (v_1, v_3, v_4, v_5, v_9, v_{10})$	6.19	29
$p_2 = (v_1, v_3, v_4, v_8, v_9, v_{10})$	6.26	29
$p_3 = (v_1, v_2, v_4, v_5, v_9, v_{10})$	6.39	31

paper compares the proposed model with the model that only consider the shortest travel time. Using the Dijkstra algorithm to calculate the shortest path with travel time as the goal, the shortest path is $p_4 = (v_1, v_2, v_6, v_{10})$ and the travel time is 24. Compared with the route selection result of the model proposed in this study, it can be seen that considering the accident risk cost of the route does not have a great impact on the route selection. The travel time of path p_1 is 29; compared with path p_4, it takes more time for drivers to travel. However, because the model takes minimizing travel risk as the primary goal and minimizing travel time as the second goal, choosing p_1 greatly improves the travel safety of drivers.

—— $p1$ (Optimal path)
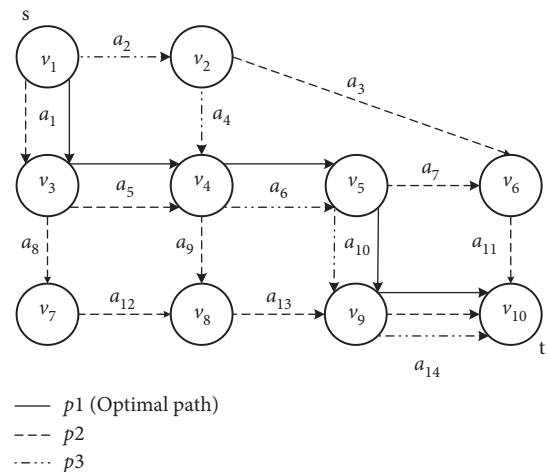--- $p2$
·-··· $p3$

Figure 3: Path selection results.

According to the above path selection results and comparison analysis, the path planning model proposed in this study takes the risk cost as the primary goal, comprehensively considers the travel time of the path, and realizes the multiconstraint path guidance. The optimal path minimizes the travel time on the basis of the lowest risk cost. It can not only ensure the travel safety but also the travel efficiency.

4. Conclusions and Future Research

This study established an accident risk evaluation system based on traffic accident data; constructed a quantitative model of accident risk cost based on drivers, vehicles, roads, and environmental factors; calculated the weight of each index by the entropy weight method; and determined the accident risk of a road section. Then, based on an improved K shortest path algorithm, a route planning model considering risk cost and travel time was designed, which can be applied to driving assistance systems to help drivers choose a better, safer, and faster travel path and improve overall traffic safety and efficiency. The conclusions of this study are as follows:

(1) Based on historical accident data, this study constructed a risk index evaluation system, used the entropy weight method to calculate the index weights, and established the accident risk cost calculation model, which realizes the objective quantification of driving risks and overcomes the disadvantage of subjectivity in traditional risk evaluation methods.

(2) Based on the improved K shortest path algorithm, a route planning model meeting multiple objectives was designed. The model considers the path risk cost and travel time simultaneously and selects the optimal travel path with the lowest risk and shortest time for drivers. The proposed method is theoretically and practically applicable for road transportation risk assessment and accident prevention.

The model proposed in this paper considers the impact of drivers, vehicles, roads, and the surrounding environment on traffic accidents. However, due to the lack of index data reflecting real-time traffic flow characteristics (such as traffic flow, traffic composition, etc.) in the data set, the impact of real-time traffic flow characteristics on accident risk is not calculated. Also, this study assumes only a small proportion of vehicles are guided by the route planning method so that the method changes neither the traffic flow states nor the driving risk. At the same time, the K shortest path algorithm used in this study outputs a large number of repeated paths, which is inefficient and can only be used in acyclic networks, so it is difficult to deal with path selection on complex networks. In future research, real-time traffic flow characteristics can be included in the index set. Combined with the accident risk quantification model proposed in this study, the real-time risk status of the road can be calculated to realize the real-time update of the travel path. At the same time, the future research can be applied in the actual road network and try to improve the path guidance algorithm to improve the efficiency and universality of the model.

Data Availability

The data that support the findings of this study are available in the FARS. These data were derived from the following resources available in the public domain: https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/FARS/2019/National/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] National Bureau of Statistics of China, *China Statistical Yearbook*, China Statistics Press, Beijing, 2021.

[2] A. C. Mbakwe, A. A. Saka, K. Choi, and Y. J. Lee, "Alternative method of Highway traffic safety analysis for developing Countries using Delphi Technique and Bayesian network," *Accident Analysis & Prevention*, vol. 93, pp. 135–146, 2016.

[3] H. C. Chu, "Assessing factors causing severe Injuries in crashes of High-deck buses in long-distance driving on Freeways," *Accident Analysis & Prevention*, vol. 62, pp. 130–136, 2014.

[4] D. Mohan, G. Tiwari, and S. Mukherjee, "Urban traffic safety assessment: a Case study of six Indian cities," *IATSS Research*, vol. 39, no. No. 2, pp. 95–101, 2016.

[5] Z. Eusofe and H. Evdorides, "Assessment of road safety management at Institutional level in Malaysia: a Case study," *IATSS Research*, vol. 41, no. No. 4, pp. 172–181, 2017.

[6] S. V. Gomes, J. L. Cardoso, and C. L. Azevedo, "Portuguese Mainland road network safety Performance indicator," *Case Studies on Transport Policy*, vol. 6, no. No. 3, pp. 416–422, 2017.

[7] Z. Xie and J. Yan, "Kernel density estimation of traffic accidents in a network Space. Computers," *Environment and Urban Systems*, vol. 32, no. No. 5, pp. 396–406, 2008.

[8] T. K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots," *Accident Analysis & Prevention*, vol. 41, no. No. 3, pp. 359–364, 2008.

[9] M. Bil, R. Andrášik, and Z. Janoška, "Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation," *Accident Analysis & Prevention*, vol. 55, pp. 265–273, 2013.

[10] J. R. Treat, N. S. Tumbas, S. T. McDonald, D. Shinar, R. D. Hume, and R. E. Mayerm, *Tri-level Study of the Causes of Traffic Accidents* Rep. DOT HS-805 085, Tech, Washington, DC, USA, , 1979.

[11] D. L. Hendricks, J. C. Fell, and M. Freedman, *The Relative Frequency of Unsafe Driving Acts in Serious Traffic Crashes*, Tech, Washington, DC, USA, 1999.

[12] Y. He, S. Yang, C. Chan, L. Chen, and C. Wu, "Visualization analysis of intelligent vehicles research Field based on Mapping Knowledge domain," *IEEE Transactions on*

Intelligent Transportation Systems, vol. 22, no. No. 9, pp. 5721–5736, 2021.

[13] X. Yang, C. Wu, Y. He, X. Lu, and T. Chen, "A Dynamic Rollover prediction index of Heavy-Duty vehicles with a real-time parameter estimation algorithm using NLMS method," *IEEE Transactions on Intelligent Transportation Systems*, vol. 71, no. No. 3, pp. 2734–2748, 2022.

[14] S. Jiang, M. Jafari, M. Kharbeche, M. Jalayer, and K. N. Al-Khalifa, "Safe route Mapping of roadways using multiple sourced data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. No. 4, pp. 3169–3179, 2022.

[15] N. Arbabzadeh and M. Jafari, "A data-Driven Approach for driving safety risk prediction using driver behavior and roadway information data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. No. 2, pp. 446–460, 2018.

[16] S. Li, Y. Wu, and X. Bai, "Research on Risk Assessment of Dangerous Goods Transportation on Roads Based on AHP and Entropy Method," *Journal of Chongqing Jiaotong University*, vol. 20, no. No. 2, pp. 43–50, Social Sciences Edition, 2020.

[17] J. Zhao, S. Luo, Y. Gong, F. Li, Y. Luo, and Y. Huang, "Research on three Dimensional risk ranking model of Tanks major Hazards," *China Safety Science Journal*, vol. 25, no. No. 2, pp. 135–140, 2015.

[18] J. J. Fernandez, M. C. Paringit, J. R. Salvador, P. I. Lucero, and J. G. Galupino, "Driver's road accident factor Prioritization using AHP in relation to Mastery of traffic signs in the city of Manila," *Transportation Research Procedia*, vol. 48, pp. 1316–1324, 2020.

[19] W. Temrungsie, W. Raksuntron, S. Namee, S. Chayanan, and B. Witchayangkoon, "AHP-based Prioritization on road accidents factors: a Case study of Thailand. International Transaction Journal of Engineering, management," *& Applied Sciences & Technologies*, vol. 6, no. No. 4, pp. 135–144, 2015.

[20] X. Cai, C. Lei, B. Peng, X. Tang, and Z. Gao, "Road traffic safety risk estimation method based on vehicle Onboard Diagnostic data. Journal of Advanced transportation," *Vol*, vol. 2020, pp. 1–13, 2020.

[21] B. Guo, L. Jin, D. Sun, J. Shi, and F. Wang, "Establishment of the characteristic evaluation index system of secondary Task driving and analyzing its importance," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 308–317, 2019.

[22] G. Pang, K. Takabashi, T. Yokota, and H. Takenaga, "Adaptive route selection for Dynamic route guidance system based on fuzzy-neural Approaches," *Vehicular Technology IEEE Transactions on*, vol. 48, no. No. 6, pp. 2028–2041, 1999.

[23] H. Lee, S. Choi, and H. Jung, "A route guidance system considering travel time Unreliability," *Journal of Intelligent Transportation Systems*, vol. 23, no. No. 3, pp. 282–299, 2019.

[24] E. B. Karim and T. Sayed, "Safety Performance functions using traffic conflicts," *Safety Science*, vol. 51, no. No. 1, pp. 160–164, 2013.

[25] R. P. Payyanadan, F. A. Sanchez, and J. D. Lee, "Assessing route choice to mitigate Older driver risk," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. No. 3, pp. 527–536, 2017.

[26] H. Zhang, D. Kong, and G. Ao, "Drivers' route guidance model based on different risk Propensity," *China Safety Science Journal*, vol. 30, no. No. 5, pp. 108–114, 2020.

[27] S. Gao and F. Lu, "The Kth shortest path algorithms: accuracy and efficiency evaluation," *Journal of Image and Graphics*, vol. 14, no. No. 8, pp. 1677–1683, 2009.

[28] T. Xu, X. Ding, and J. Li, "Review on K Shortest paths algorithms," *Computer Engineering and Design*, vol. 34, no. No. 11, pp. 3900–3906, 2013.

[29] National Center for Statistics and Analysis, "Fatality analysis reporting system (FARS) analytical user's manual, 1975-2019 (report No. DOT HS 813 023)," National Highway traffic safety Administration, Washington, DC, USA, 2020.