Analysis Model of Risk Factors of Urban Bus Operation Based on FTA-CLR

Jianfeng Xi,1 Yunhe Zhao,1 Tongqiang Ding,1 Jian Tian,2 and Lianjie Li1

1College of Transportation, Jilin University, Changchun 130022, China
2China Academy of Transportation Sciences, Beijing 100029, China

Correspondence should be addressed to Tongqiang Ding; dtq8@163.com

Received 8 December 2020; Revised 27 April 2021; Accepted 13 May 2021; Published 27 May 2021

Copyright © 2021 Jianfeng Xi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to comprehensively analyze the risk factors and accurately find the high risk factors related to accidents, an analysis model of risk factors of urban bus operation is proposed, in which the advantages of the structural analysis of the Fault Tree Analysis (FTA) and the correlation analysis of the Cumulative Logistic Regression (CLR) are combined. Firstly, based on the accident data in Northeast China, FTA is used to compile the urban bus operation fault tree. In the fault tree, 16 bus operation risk factors are classified, while the risk factors are sorted and compared from three aspects: structural importance, probability importance, and critical importance. And then, the 11 higher risk factors are selected according to the discriminant principle. Secondly, bus operation accidents are divided into fatal accidents, injury accidents, and major economic loss accidents. The CLR model is used to fit the much higher risk factors that lead to urban bus operation accidents from above 11 higher risk factors. Finally, the scientific rationality and applicability of the model are verified, through the goodness of fit test and the comparison test with the actual probability of occurrence.

1. Introduction

Due to the large number of passengers transported by bus, casualties are extremely likely to occur in event of an operation accident. Furthermore, the characteristics of urban bus operation accidents are different from other forms of road traffic, so the risk factors of bus operation accidents have become people’s main concerns. Some scholars conduct research on the influence degree of road infrastructure on public transportation safety, the characteristics and weights of basic indicators of urban public transportation, factors affecting the safety of bus operations, and preventive and improvement measures [1–3]. Some scholars explore the relationship between the personality characteristics of bus drivers, driving behaviors, and accident risk [4, 5]. Other scholars explore bus accidents in the United States and find out that the severity of bus accidents is related to the age, gender, and dangerous driving behavior of bus drivers [6]. For the analysis of traffic safety and the causes of traffic accidents, the scholar uses the Fuzzy Fault Tree Analysis to evaluate the safety of the public transportation system [7].
analyse the key influencing factors of the severity of the accident with the Logistic model [15]. Due to the fact that our focus is to excavate and rank all the accident risk factors for urban bus operations as well as the insufficient data about the type and severity of injuries of urban bus operations, the FTA with the advantages of structural analysis is a suitable method.

Then, other scholars believe that it is necessary to classify traffic accidents in order to enhance the homogeneity of accident data [16]. Some scholars use the Probit model, the Binomial Logistic model, and the Multinomial Logistic model to analyse the factors affecting the severity of accidents at three-way and four-way no signal intersections [17]. Because of the good applicability of the Logistic Regression model, it can better assess risk factors of bus involved accident severity including both discrete and continuous variables. Feng et al. divide drivers in US bus accidents into three categories, no traffic violation records, traffic violation records (middle-aged), and traffic violation records (young and old), and analyse the factors affecting the severity of accidents in each category of drivers [18]. Nasri and Aghabayak investigate the underlying risk factors associated with the severity of urban transit bus accidents in Mashhad, Iran, by estimating a binary logit model, in which the accident severity is divided into PDO (Property Damage Only) and injury or fatal categories, and the effect of several risk factors on the probability of bus accidents with higher severity is examined [19]. Some scholars use the real bus operation accident data in Guangdong Province of China as an example and use the Binomial Logistic model to analyse the factors affecting the severity of bus accidents from four scenarios: overall accidents, intervehicle accidents, vehicle-pedestrian accidents, and single-vehicle accidents [20].

FTA can well dig out all the risk factors of accidents based on urban bus operation characteristics, classify all factors, and present them clearly in a structured manner. In addition, CLR would get the desirable analysis results depending on the suitable set of risk factors. In this research, it is not only necessary to dig out all the accident risk factors of urban bus operation but also it is needed to sort all risk factors and find higher risk factors by the Pearson \(\chi^2\) statistic and Deviance statistic goodness of fit test. Therefore, this paper combines FTA and CLR to find out the higher risk factors of urban bus operation accidents.

2. Materials and Methods

Because the operating environment of urban buses is more complex and involves many influencing factors, FTA can classify the risk factors that lead to accidents of urban bus operation. The determination of the minimum cut set can present the set of each risk factor to clarify the structural importance of each risk factor. However, the analysis of structural importance only analyses the impact of each risk factor on bus operation accidents according to the structure of the fault tree. What we are concerned about is the impact of the occurrence probability of each risk factor on the probability of bus operation accidents and how to measure which is easier to reduce the risk factors with high probability or low probability. Therefore, the probability importance and critical importance of each risk factor have practical significance to the extraction of high risk factors. There are a number of requirements and assumptions in using Logistic Regression models. The most important ones of them are IIA (Independence of Irrelevant Alternatives), linearity of independent variables and log odds, little or no multicollinearity among the independent variables, and the independence of the observations. Even so, it is flexible in application and simple in calculation, which can make the results of the analysis more objective. The utilized data in this article could meet these requirements and assumptions. The existing classification of the severity of traffic accidents is mostly based on the most severe injury imposed on the passengers. In fact, heavy economic losses should also be an important consideration in the severity of the accident. Since the dependent variable is set for multiple values, and the independent variable includes both discrete and continuous variables, the Cumulative Logistic model in the multivariate Logistic model is the most appropriate.

Based on the above reasons, an analytical model of the risk factors of urban bus operation is constructed. The steps of the FTA-CLR analysis model are as follows.

The first step is to compile the fault tree. Based on the bus operation accident data in a certain area, the accident is regarded as the top event. According to accident types and causes, risk factors are classified and connected with “OR” or “AND”.

Then, the next top events are analysed until the basic events can no longer be separated.

The second step is the calculation of importance. Three kinds of importance are calculated, which are structural importance, probability importance, and critical importance.

2.1. Structural Importance. Boolean algebra [21] is used to solve the minimum cut set of the fault tree of bus operation accidents. Formula (1) [22] is used to calculate the coefficient of structural importance of risk factors. The greater the value of \(I_p(i)\), the greater the structural importance coefficient. According to the criterion of structural importance, if the minimum cut set contains only one basic event, the structural importance of the risk factor represented by the basic event is the largest. If the minimum cut set contains inconsistent basic events, the structural importance of the risk factors represented by the basic events of the minimum cut set with fewer basic events is large:

\[
I_p(i) = \sum_{X_G \subseteq G} \frac{1}{2^{n-1}},
\]

where \(I_p(i)\) represents the structural importance coefficient of the basic event \(X(i)\), \(n\) represents the minimum cut set capacity of the basic event \(X(i)\). \(G\) represents the minimum cut set containing the basic event \(X(i)\).
2.2. Probability Importance. Formula (2) [23] is used to find out the degree which means the change in the probability of the i-th basic event on the change in the probability of the top event. It is the coefficient of probability importance of risk factors. The larger the value of $I_g(i)$, the higher the probability of the basic event:

$$ I_g(i) = \frac{\partial P(T)}{\partial q_i}, \quad (2) $$

where $I_g(i)$ is the probability importance coefficient of the basic event $X(i)$. $P(T)$ represents the occurrence probability function of the top event. $q_i$ represents the probability of occurrence of the i-th basic event $X(i)$, which is calculated based on the data of bus operation accidents.

2.3. Critical Importance. Formula (3) [23] is used to measure the importance of each basic event in terms of sensitivity and its own probability of occurrence. It is the coefficient of the critical importance of risk factors. The larger the value of $I_g^c(i)$, the greater the impact on the probability of occurrence of the top event:

$$ I_g^c(i) = \frac{q_i - I_g(i)}{P(T) I_g(i)}, \quad (3) $$

where $I_g^c(i)$ represents the critical importance coefficient of the i-th basic event. $P(T)$ represents the probability of occurrence of the top event.

The third step is sorting and analysis. The risk factors are sorted by structural importance coefficient, probability importance coefficient, and critical importance coefficient. After the coefficients are normalized, higher risk factors are selected according to the discriminant principle. Discriminant principle: (1) the order of structural importance coefficient and probability importance coefficient is the main basis, and the critical importance coefficient is the auxiliary basis and (2) in the order of the main basis, the cumulative sum of the normalized index of each importance is above 90%.

The fourth step is to analyse the severity of bus operation accidents based on the CLR. The probability calculation model is shown in formula (4) [15]. The mixed stepwise selection method [15] is used to gradually eliminate independent variables with insignificant coefficients. That is independent variables with a significant level of Sig. < 0.05, and through correlation analysis to determine significantly related risk factors:

$$ P(Y \leq j/X) = \frac{e^{\alpha_r + \sum_{j=1}^{I} \beta_{ij}X_i}}{1 + e^{\alpha_r + \sum_{j=1}^{I} \beta_{ij}X_i}}, \quad (4) $$

where $\sum_{j=1}^{I} P(y = j/X) = 1$, $j = 1, 2, \ldots, j - 1$, $P(Y \leq j/X)$ is cumulative probability. $X$ contains $I$ independent variables, such as $X_1, X_2, \ldots, X_i$. $I$ is the number of independent variables. $\alpha_r$ is a constant term. $\beta_{ij}$ is the regression coefficient.

The fifth step is the model test. According to the Pearson $\chi^2$ statistic and Deviance statistic goodness of fit test [24], if the significance level of the Pearson $\chi^2$ statistic and Deviance statistic is more than 0.05, it indicates that the model fits well. Then, compare the predicted probability of the accident type calculated by the model with the actual probability to determine its accuracy.

The FTA-CLR model analysis process is shown in Figure 1.

3. Results and Discussion

3.1. Compilation of Fault Tree. Taking the data of bus operation accidents in Northeast China as an example, the basic events of urban bus operation accidents are summarized, and the fault tree of urban bus operation is compiled.

Because the cause of urban bus operation accidents is related to many factors, accidents of urban bus operation are divided into two types: traffic accidents and service accidents. Therefore, the accident is regarded as the top event; the fault tree of traffic accidents and the fault tree of service accidents are compiled separately. There are three main forms of bus operation accidents: vehicle-fixed object, vehicle-pedestrian, and vehicle-vehicle. Any one of them will cause accidents, so the OR door is used to connect. Continue to use this method to analyse subtop events until the basic events can no longer be separated. At this point, the construction of the fault tree of traffic accidents is completed. The construction method of fault tree of service accidents is the same. The fault tree of traffic accidents and the fault tree of service accidents are shown in Figures 2 and 3, respectively.

3.2. Calculation of Importance. Because service accidents of the urban bus generally do not cause serious accidents, this paper only analyses the risk factors that may cause serious accidents in urban bus operation. The minimum cut set of the fault tree of traffic accidents in urban bus operation is obtained as follows:

- $G_1 = \{X_1\}$
- $G_2 = \{X_2\}$
- $G_3 = \{X_3\}$
- $G_4 = \{X_4\}$
- $G_5 = \{X_5\}$
- $G_6 = \{X_6\}$
- $G_7 = \{X_{12}\}$
- $G_8 = \{X_{13}\}$
- $G_9 = \{X_{14}\}$
- $G_{10} = \{X_2X_{10}\}$
- $G_{11} = \{X_9X_{10}\}$
- $G_{12} = \{X_4X_{11}X_{10}X_{15}\}$
- $G_{13} = \{X_6X_{11}X_{10}X_{15}\}$
- $G_{14} = \{X_7X_{11}X_{10}X_{15}\}$
- $G_{15} = \{X_{16}X_{11}X_{10}X_{15}\}$

A total of 15 minimum cut sets of risk factors of traffic accidents in urban bus operation are obtained, of which there are 9 minimum cut sets of a single factor. It shows that if one of the influencing factors exists in the actual operation of the urban bus, it is very likely to cause traffic accidents. Then, it is of practical significance to find out the higher risk factors leading to bus operation accidents.

The structural importance coefficients of the risk factors affecting accidents of urban bus operation are obtained. As shown in the second column of Table 1, the order is as follows:
The probability importance coefficients of the risk factors affecting accidents of urban bus operation are obtained. As shown in the third column of Table 1, the order is as follows:

\[ I_g(13) > I_g(2) > I_g(10) > I_g(1) > I_g(4) > I_g(14) > I_g(5) > I_g(3) = I_g(8) > I_g(9) = I_g(12) > I_g(15) > I_g(11) > I_g(6) = I_g(7) = I_g(16). \]
The critical importance coefficients of the risk factors affecting accidents of urban bus operation are obtained. As shown in the fourth column of Table 1, the order is as follows:

\[
I_c^g(13) > I_c^g(2) > I_c^g(10) > I_c^g(9) > I_c^g(1) = I_c^g(4) > I_c^g(14) > I_c^g(5) > I_c^g(11) = I_c^g(15) > I_c^g(3) = I_c^g(8) > I_c^g(6) > I_c^g(12) > I_c^g(7) > I_c^g(16).
\]  

3.3. Sorting and Analysis. 16 risk factors are obtained by classifying the basic events determined by the bus traffic fault tree. The analysis of structural importance, probability importance, and critical importance is carried out. According to the sorting order of the main basis, the cumulative sum of the normalized index of each importance is shown in Table 2.

It can be seen that the cumulative sum of the normalized index of each importance of the first 11 risk factors is above 90%. At the same time, from the first 11 to 12 risk factors, the increase of the cumulative sum of the normalized index based on the critical importance is not obvious. Therefore, the first 11 risk factors are selected as higher risk factors. They are 16–25 years old or older, have less than 5 years of...
driving experience, female driver, nonsunny and low visibility, slippery road, no street lighting at night, no signal control, no protective facilities on the roadside, damaged road surface, driving straight, and driving at intersections.

3.4. Severity Analysis Based on CLR

3.4.1. Dependent Variable. When studying the significantly related risk factors of urban bus accidents, this paper lists only major economic losses into the classification of urban bus accidents. This paper comprehensively divides the severity of bus operation accidents into three levels. The code is $Y = 1$ for fatal accidents, $Y = 2$ for injury accidents, and $Y = 3$ for only major economic loss accidents.

3.4.2. Independent Variable. Based on the analysis of the importance of basic events in the bus operation fault tree compiled in this paper, from the perspective of people,

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Structural importance coefficient</th>
<th>Probability importance coefficient</th>
<th>Critical importance coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age $X_1$</td>
<td>1</td>
<td>0.0115</td>
<td>0.0028</td>
</tr>
<tr>
<td>Driving experience $X_2$</td>
<td>1.5</td>
<td>0.0386</td>
<td>0.0257</td>
</tr>
<tr>
<td>Gender $X_3$</td>
<td>1</td>
<td>0.0093</td>
<td>0.0005</td>
</tr>
<tr>
<td>Weather $X_4$</td>
<td>1.125</td>
<td>0.0113</td>
<td>0.0020</td>
</tr>
<tr>
<td>Road table conditions $X_5$</td>
<td>1</td>
<td>0.0103</td>
<td>0.0015</td>
</tr>
<tr>
<td>Visibility $X_6$</td>
<td>0.125</td>
<td>0.0007</td>
<td>0.0004</td>
</tr>
<tr>
<td>Terrain $X_7$</td>
<td>0.125</td>
<td>0.0007</td>
<td>0.0002</td>
</tr>
<tr>
<td>Lighting conditions $X_8$</td>
<td>1</td>
<td>0.0093</td>
<td>0.0005</td>
</tr>
<tr>
<td>Traffic signal mode $X_9$</td>
<td>0.5</td>
<td>0.0091</td>
<td>0.0034</td>
</tr>
<tr>
<td>Roadside facilities protection type $X_{10}$</td>
<td>1.5</td>
<td>0.0168</td>
<td>0.0127</td>
</tr>
<tr>
<td>Road physical isolation $X_{11}$</td>
<td>0.5</td>
<td>0.0010</td>
<td>0.0007</td>
</tr>
<tr>
<td>Road condition $X_{12}$</td>
<td>1</td>
<td>0.0091</td>
<td>0.0003</td>
</tr>
<tr>
<td>Driving state $X_{13}$</td>
<td>1</td>
<td>0.0441</td>
<td>0.0357</td>
</tr>
<tr>
<td>Intersection or section $X_{14}$</td>
<td>1</td>
<td>0.0106</td>
<td>0.0018</td>
</tr>
<tr>
<td>Driving period $X_{15}$</td>
<td>0.5</td>
<td>0.0051</td>
<td>0.0007</td>
</tr>
<tr>
<td>Road alignment $X_{16}$</td>
<td>0.125</td>
<td>0.0007</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

### Table 1: Importance coefficient of risk factors of urban bus operation accidents.

<table>
<thead>
<tr>
<th>Number of risk factors</th>
<th>Structural importance (%)</th>
<th>Probability importance (%)</th>
<th>Critical importance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>86</td>
<td>91</td>
<td>97</td>
</tr>
<tr>
<td>11</td>
<td>90</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td>12</td>
<td>93</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

### Table 2: The cumulative sum of the normalized index.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Coding and assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>Age</td>
<td>16–25 years old = 1; 26–35 years old = 2; 36–45 years old = 3; 46–55 years old = 4; 56 years old and above = 5</td>
</tr>
<tr>
<td>$X_2$</td>
<td>Driving experience</td>
<td>1–5 years = 1; 5–10 years = 2; 10–15 years = 3; 15–20 years = 4; more than 20 years = 5</td>
</tr>
<tr>
<td>$X_3$</td>
<td>Gender</td>
<td>Female = 0; male = 1</td>
</tr>
<tr>
<td>$X_4$</td>
<td>Weather</td>
<td>Sunny = 0; nonsunny (rain, snow, fog, overcast, strong wind, etc.) = 1</td>
</tr>
<tr>
<td>$X_5$</td>
<td>Road table conditions</td>
<td>Dry = 0; not dry (rain, snow, ice) = 1</td>
</tr>
<tr>
<td>$X_6$</td>
<td>Lighting conditions</td>
<td>Daytime = 1; street light at night = 2; No street light at night = 3</td>
</tr>
<tr>
<td>$X_7$</td>
<td>Traffic signal mode</td>
<td>Signal control = 0; no signal control = 1</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>Roadside facilities protection type</td>
<td>Without roadside protection = 0; With roadside protection = 1</td>
</tr>
<tr>
<td>$X_{12}$</td>
<td>Road condition</td>
<td>Road surface is intact = 0; Road surface damage = 1</td>
</tr>
<tr>
<td>$X_{13}$</td>
<td>Driving state</td>
<td>Nonstraight = 0; straight = 1</td>
</tr>
<tr>
<td>$X_{14}$</td>
<td>Intersection or section</td>
<td>Road section = 0; intersection = 1</td>
</tr>
</tbody>
</table>

### Table 3: Coding and meaning of independent variables.
vehicles, roads, and environment in the transportation system, 11 independent variable factors are determined. The independent variable coding is shown in Table 3.

3.4.3. Fitting. The independent variables with significance level Sig. < 0.05 are gradually eliminated. Through correlation analysis, 4 significantly related risk factors are determined. There are driving experience ($X_2$), lighting conditions ($X_8$), traffic signal mode ($X_9$), and driving state ($X_{13}$). The model calibration results are shown in Table 4.

3.5. Model Test. In this way, the Cumulative Logistic probability prediction model of urban bus fatal accidents and injury accidents is obtained. The results of goodness of fit test show that the Pearson $\chi^2$ statistic has a significant level of 0.998 > 0.05, and the Deviance statistic has a significant level of 0.969 > 0.05. It shows that the model fits well.

Then, 1390 cases of the data of urban bus accidents with complete information in Northeast China are selected. The Cumulative Logistic probability prediction model constructed in this paper is used to predict and calculate. The predicted probabilities of fatal accidents, injury accidents, and only major economic loss accidents are compared with the actual probability of occurrence, as shown in Table 5. From the comparative analysis, it can be seen that the errors are less than 5%, which shows that the accuracy of the constructed model for predicting the severity of urban bus operation accidents is relatively high.

4. Conclusions

(1) Combining the respective advantages of the Fault Tree Analysis and Cumulative Logistic Regression, an analysis model of risk factors of urban bus operation based on FTA-CLR is proposed. It could analyse the risk factors of urban bus operation accidents in more detail.

(2) Taking the accident data of urban bus operation in Northeast China as an example, the risk factors are analysed according to the steps of the FTA-CLR analysis model, in which the driving experience, lighting conditions, traffic signal mode, and driving state are the first 4 higher factors. Through the goodness of fit test and the actual probability of occurrence comparison test, the scientific rationality and applicability of the model are verified.

(3) The analysis results of the model could provide a reference for identifying high risk factors of urban bus operations in different regions. The analysis of risk factors can also provide some bases for taking the countermeasures of urban bus accidents.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was funded by the National Natural Science Foundation of China (Grant no. 51308249).

References


