Research on Emergency Evacuation Traffic Trip Generation Forecasting Based on Logistic Regression

Fanliang Bu Qingmei Xie
Department of Engineering of Security &Protection System
Chinese People's Public Security University
Beijing, China
e-mail:xie841103@yahoo.com.cn

Abstract—**This paper puts forward a method based on Logistic regression of emergency evacuation traffic trip generation forecasting model. Based on analysis of Xiamen residents earthquake evacuation survey data, the software of applied statistic SPSS could establish emergency evacuation traffic trip generation forecasting model by Logistic regression. This model could predict the probability of all evacuation modes of residents to obtain the better emergency processing plan of traffic evacuation in earthquake disasters. Subsequently, for the purpose of making the more reasonable and effective of city emergency management plan, the test sample would be simulated to accuracy and fitting degree of the prediction model result.**

Keywords-Traffic Evacuation; Evacuation Travel Demand; Logistic Regression Model; SPSS

I. INTRODUCTION

In recent years, especially put through the Wenchuan earthquake, China actively reform and complete the emergency management system in the organization system, the operating mechanism and the security mechanism. But in general, ability of emergency management, especially in the emergency travel evacuation management research field is very weak. However, the emergency travel evacuation management is vastly significant to city people. Under the circumstances of unproven emergency processing plan of traffic evacuation in China, earthquakes, floods, terrorist attacks, tidal waves, tornadoes - all these and many other disasters would very easily cause large acreage urban traffic jam or even paralysis, number of injuries and deaths and property loss.

When serious disaster strikes or is predicted, city emergency management administration would face a series of problems. Consequently, the emergency traffic program and management should determine the emergency traffic demand and the traffic flow of the network. The prediction of emergency traffic demand should be produced by emergency evacuation traffic demand TG (Trip Generation) including prediction of traffic capacity and trip generation under normal and emergency circumstances. Through accurate prediction of emergency traffic situation, city emergency management administration could timely prepare evacuation vehicles and rescue teams. Moreover, the limited traffic resource would be equitably distributed to avoid the traffic resource contention such as traffic jam. Therefore, dense area people could be evacuated safely in short time[1].

In the past, these emergency processes were not taken seriously enough in China, and there was little academic research of this field. In view of the above questions, this paper puts forward a method based on Logistic regression of emergency evacuation traffic trip generation forecasting model. Based on analysis of Xiamen residents earthquake evacuation survey data, the software of applied statistic SPSS could establish emergency evacuation traffic trip generation forecasting model by Logistic regression. This model could predict the probability of all evacuation modes of residents to obtain the better emergency processing plan of traffic evacuation in earthquake disasters. Subsequently, for the purpose of making the more reasonable and effective of city emergency management plan, the test sample would be simulated to accuracy and fitting degree of the prediction model result.

II. MODELING

Linear regression model is a very popular statistical analysis technique in quantitative analysis. Nevertheless, by the research of features of traffic evacuation, evacuation decision is a series of sort selections over time in emergencies. And dependent variable of it is a binary variable (such as evacuation or stay, marked as 1 and 0), but not a continuous variable. Thus, analysis of binary dependent variable of evacuation decision needs to use nonlinear functions[2]. Log-linear model is usually used to analyze classified variables. Moreover, Logistic regression model is a special form of Log-linear model. The modified Logistic regression model could be used to predict emergency evacuation traffic trip generation. It is a function which is produced of resident socio-economical conditions (e.g. residence ownership, ages and educational levels) and received emergency information (e.g. mandatory evacuation), which could predict probability of residents evacuation decision in emergencies.

A. Logistic regression model [3-5]

Evacuation of the dependent variable is set to Y ($Y=1$, evacuation. $Y=0$, no evacuation). n independent variables, which influence the values of Y, are $x_1, x_2, \cdots, x_m$. By the

* This work was supported by the National Natural Science Foundation of China (NO. 60275018)

action of n independent variables, the probability of occurrence of $Y=1$ is $p = p(Y=1|x_1, x_2, \cdots, x_m)$. Logistic regression model could be expressed as:

$$p = \frac{e^{(\beta_0+\beta_1 x_1+\cdots\beta_n x_n)}}{1+e^{(\beta_0+\beta_1 x_1+\cdots\beta_n x_n)}} \quad (1)$$

In the (1), p is probability of one respondent deciding to evacuation; $x_1, x_2, \cdots, x_m$ are independent variables; β_0 is a constant term; $\beta_1, \beta_2, \cdots \beta_n$ are coefficients of partial regression.

According to the given condition, evacuation probability to no evacuation probability ratio is the odds of experiencing an event, *odds* for short,

$$odds = p/(1-p) \quad (2)$$

It means the probability of evacuation would be "*odds*" times as much as the probability of no evacuation.

The below equation is obtained by the logarithm of each side of (2):

$$\log it(p) = \ln(\frac{p}{1-p}) \quad (3)$$

Logistic function is nature logarithm translated in (3), which is called the logit form or $\log it(p)$. By using p translation, the linear function could be gained:

$$\log it(p) = g(\mathrm{X}) = \beta_0 + \beta_1 x_1 + \cdots \beta_n x_n \quad (4)$$

The importance of the translation is that $\log it(p)$ has an amount of useful characters of linear regression model. To its parameter, $\log it(p)$ is linear, which depends upon value of X. The range of $\log it(p)$ is between minus infinity and plus infinity. According to Y is binary but not continuous, its error distribution is binomial distribution but not normal distribution, in addition, all of analysis should be established on binomial distribution. Hence, estimation of Logistic regression coefficient should use maximum likelihood estimate.

In the (1), the dependent variable is evacuation (the value is 1) or no evacuation (the value is 0). The below elements affect the independent variable:

- Type of building (villa or flat).
- Ownership of residence (own it or not).
- Emergency experience of the resident(experienced emergency and serious loss, experienced emergency and a little loss, or not).
- Mandatory evacuation order (received or not).
- Distance to disaster-affected area (in center of disaster-affected area, in disaster-affected surrounding area, or safety area).
- Age of evacuated resident.
- Educational level of evacuated resident (the junior secondary school level or below it, the senior middle school level, the undergraduate level, the postgraduate level or above it).
- Marriage state of evacuated resident (single and no marriage history or married and cohabitation).

The variable of Logistic regression model is limited by character of available data. All above needed information is supported by survey of families in Xiamen. The personal information would be independent variable to simulate decision of resident evacuation, for example, ages, educational level, marriage state, etc.

B. Selection of independent variables and dependent variables

The collected data is divided into two section, sample data and test data. Each of them is including all information of each respondent. 250 sample data is used by regression analysis, which is 100% of total sample data. The 250 sample data is training sample; on the other hand, it is sample for estimating Logistic regression model. After model building, the new model would reassess and rejudge the training sample. In order to check prediction result of model, 60 test data would substitute in the new model to compare to the initial data, at the same time, applicability of model would be tested too.

The probability demarcation point of this model is 0.5. When probability is greater than 0.5, prediction of the sample is evacuation. Conversely, when probability is less than 0.5, prediction of the sample is no evacuation. If there are insufficient independent variables which are inducted, the dependent variable change would not be explained very well. Nevertheless, too many independent variables are not good for modeling result, by reason that some independent variables are meaningless for explanation of the dependent variable changing. Furthermore, maybe there is a strong linear dependency between one independent variable and others, which is multicollinearity, so regression equation could not induce all independent variables. Therefore, the independent variable should be filtered. Stepwise forward method builds the model in the continuous form, thus, the independent variable is filtered by stepwise forward method.

Before regression analysis, the constant term of regression model should be tested, and the constant term test result is given by table Ⅰ, the coefficient is 0.306. Considering it is a coefficient, there is little statistic significance; the degree of freedom is 1, the standard error is 0.128. Value of Sig. is smaller, this independent variable is more important. Value of Wald is smaller, it is less significant. According to results in the table Ⅰ, the coefficient is obvious significant.

TABLE I. VARIABLES IN THE EQUATION

		B	*S.E.*	*Wald*	*df*	*Sig.*	*Exp(B)*
step 0	constant	0.306	0.128	5.731	1	0.017	1.358

In this case, some independent variables are polytomies variables, for instance, age and educational level. Therefore, infinite polytomies variables should be translated to dummy variables firstly, and each dummy variable only represents difference of two or more than two levels. Furthermore, regression result would be right.

The involved data is introduced by Score test. There is Score Test of non-induced variables in the table II; its purpose is to test this variable regression coefficient is 0 or not, when induce some variables (e.g. residence) to current model. As shown in the table II, the value p of most variables is close to 0, which indicates its coefficient is not 0.

In the paper, $x2$ is Residence, $x3$ is the ownership of residence, $x4$ is experience of earthquake, $x5$ is received mandatory evacuation order, $x6$ is less than 1 mile to disaster-affected area, $x7$ is respondent live by oneself, $x8$ is married or cohabitation, $x9$ is age, and $x91$, $x92$, $x93$, $x94$ are its dummy variables. $x10$ is educational level of respondent, and $x101$, $x102$, $x103$ is its dummy variables.

TABLE II. VARIABLES NOT IN THE EQUATION

			Score	*df*	*Sig.*
step 0	variables	x2	6.781	1	0.009
		x3	0.170	1	0.680
		x4	0.166	1	0.683
		x5	5.345	1	0.021
		x6	15.941	1	0.000
		x7	65.988	1	0.000
		x8	70.940	1	0.000
		x9	11.880	4	0.018
		x91	0.647	1	0.421
		x92	3.079	1	0.079
		x93	11.562	1	0.001
		x94	2.342	1	0.126
		x10	12.936	3	0.005
		x101	0.246	1	0.620
		x102	9.006	1	0.003
		x103	12.852	1	0.000
	Overall Statistics		118.703	14	0.000

III. *MODEL REGRESSION*

Stepwise forward method is used to choose independent variables, which means $x8$, $x6$, $x5$, $x9$, $x2$ and $x4$ are chosen. Results in Table III are the most significant part of regression analysis results, in consideration of limited space, parametric estimation of the last step regression is given only, which includes the latest parametric variable, β_i, standard error, Wald χ^2, degree of freedom, value of p and OR. In these results, the higher value of Wald χ^2, the more importance of this variable in the model; the higher value of Sig., the less importance of this variable. There is no standardized parameter estimates as linear regression in Logistic regression, therefore, significance of the variable in model should be sized up by value of Wald χ^2 and Sig. From this, it can be seen that independent variables such as married or cohabitation and less than 1 mile to disaster-affected area are extreme significant in the model.

Hereby, $\log it(p)$ model could be estimated:

$$\begin{aligned}\log it(p) &= -0.659 + 1.365\times x2 + 1.021\times x4 + 1.859\times x5 \\ &+2.582\times x6 - 4.517\times x8 - 0.08\times x91 \\ &-2.188\times x92 - 0.221\times x93 - 0.836\times x94 \end{aligned} \tag{5}$$

Therefore, the prediction model of emergency evacuation traffic is:

$$p = \frac{e^{\log it(p)}}{1+e^{\log it(p)}} \tag{6}$$

The (5) illustrates that some independent variables would make remarkable positive influence to evacuation decision, such as experience of earthquake, received mandatory evacuation order, distant to disaster-affected area; in contrast, other independent variables would make negative influence to evacuation decision, such as age, married or cohabitation. The further analysis of questionnaires shows, the closer to disaster-affected area and higher probability of evacuation; at the same time, if the respondent has been married or lives with family, the probability of evacuation would be lower.

TABLE III. VARIABLES IN THE EQUATION

		B	*S.E.*	*Wald*	*df*	*Sig.*	*Exp(B)*
step 6[f]	x2	1.365	0.658	4.297	1	0.038	3.916
	x4	1.021	0.510	4.007	1	0.045	2.777
	x5	1.859	0.477	15.160	1	0.000	6.415
	x6	2.582	0.538	23.002	1	0.000	13.225
	x8	4.517	0.607	55.453	1	0.000	0.011
	x91	0.080	1.380	0.003	1	0.954	0.923
	x92	2.188	0.713	9.407	1	0.002	0.112
	x93	0.221	0.610	0.132	1	0.717	0.802
	x94	0.836	0.645	1.682	1	0.195	0.433
	Constant	0.659	0.846	0.606	1	0.436	0.518

IV. *MODEL TESTING*

A. *Logistic regression model*

Table IV indicates the model alteration after inducing independent variables, the model adjustment, the predictive results of infinite training sample and accuracy of model prediction. Considering to Table IV, when 6 independent variables are induced, accuracy of evacuation prediction would be 85.4%, at the same time, accuracy of no evacuation prediction would be 74.5%, the average precision is 80.8%.

Accuracy of model prediction is one of methods to judge a model. However, calculation of predictive accuracy only simply divides different situations which are provided by model prediction information into two sections. Furthermore, numerous information would be lost, and model fitting results would not be perfect by predictive accuracy judgment. Thus, Probability of model prediction is also induced to judge fitting result of Logistic regression model by ROC curve, as shown in Figure 1. Consequently, if the current probability of model prediction is calculated, area under ROC curve is 0.902, its 95% credibility interval is between 0.886 and 0.938, which means result of the model prediction is excellent.

TABLE IV. CLASSIFICATION TABLE[A]

Observed			Predicted		
			Y Evacuation or not		Percentage
			0	1	Correct
step 6	*Y* Evacuation or not	0	79	27	74.5
		1	21	123	85.4
	Overall Percentage				80.8

Effectivity and goodness of fit of model fitting result has been further tested, test results in Table V shows $p = 0.223 > 0.1$. There is no marked difference between predicted value and observed value, which indicates the model fits data well at acceptable level ($p = 0.1$). A good model has high likelihood ratio value and low -2LL (-2 Log likelihood) value. As shown in Table VI, once a new independent variable is induced to the model, the value of -2LL would decrease; the value of Cox&Snell R2 and Negelkerke R2 would increase. The lower value of -2LL indicates higher fitting of the model which means better result of the model fitting.

TABLE V. HOSMER AND LEMESHOW TEST

step	*Chi-square*	*df*	*Sig.*
6	10.638	8	0.223

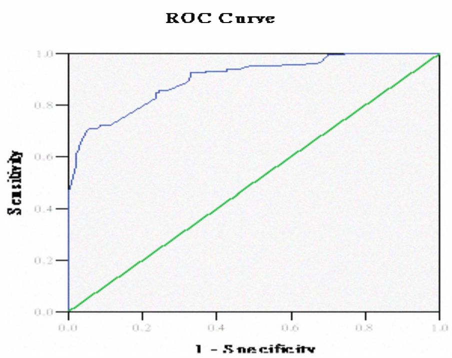

Figure 1. ROC curve

TABLE VI. MODEL SUMMARY

step	*-2 Log likelihood*	*Cox&Snell R Square*	*Nagekerke R Square*
1	265.034	0.261	0.351
2	232.360	0.352	0.473
3	214.773	0.396	0.532
4	197.113	0.437	0.587
5	193.116	0.446	0.599
6	188.918	0.455	0.612

B. *Model Testing*

For the purpose of further testing predictive accuracy and applicability of the model, testing samples are used to test the model. If probability demarcation point is 0.5, great than 0.5 is decide to evacuate, less than 0.5 is decide to stay. The test result shows there are 41 samples to evacuate, 31 samples are predicted to evacuate, the accuracy is 75.61%; there are 19 samples to stay, 14 samples are predicted to stay, the accuracy is 73.68%, moreover, average accuracy is 75%. Predictive ability of model is improved to be very good, as same as the former result of model prediction.

V. *CONCLUSION*

The analysis result demonstrates if there is an earthquake happening in Xiamen, strong influencing factors to evacuate are residence, earthquake experience, received mandatory evacuation order, distant to disaster- affected area, married or cohabitation and age. Furthermore, if residents are closer to disaster-affected area, their probability of evacuation is higher, on the other hand, living with others such as family would lead lower probability of evacuation. Emergency management should application of Population Geographic Information System (PGIS) to this model to supply more accurate information of population distribution and emergency traffic information. If evaluation of emergency evacuation traffic situation needs to be more accurate, emergency happening time should be analyzed, such as working time, the purpose of traveling would be evacuation or going back home. The next step will induce data of emergency happening time to prediction model to test and amend.

ACKNOWLEDGMENT

I would like to express my appreciation to a number of individuals who helped make the completion of this thesis possible.

REFERENCES

[1] J.X. Lin, X.L. Wei, and H.Y. Wu, "Time-variation Demand Forecast of Emergent Traffic Evacuation Based on S-curve," Journal of transportation information and safety, Vol 27, No.3, pp.92-96, 2009.

[2] L.Y. Yu and J.H. Zhan, "A Research on Probability of Default prediction Based on Logistic Regression Analysis," Journal of finance and economics, Vol 30, No.9, pp.15-23, 2004.

[3] J.C. Wang and Z.G. Guo, Logistic Regression Models: Methods and Application, Higher Education Press: Beijing, 2001.

[4] Bing Mei, "Development of Trip Generation Models of Hurricane Evacuation", Master Dissertation, Louisiana State University, Baton Rogue, Louisiana,2002.

[5] Haoqiang Fu, "Development of Dynamic Travel Demand Models for Hurricane Evacuation", PhD Dissertation, Louisiana State University, Baton Rogue, Louisiana, 2004.