Application of Fuzzy Quantitative Theory for Crash Rate Prediction at Roadway Sections

Liyan Qin\textsuperscript{a,*}, Yuquan Qin\textsuperscript{b}, Chao Gao\textsuperscript{a}, Chunfu Shao\textsuperscript{c}

\textsuperscript{a} Department of Transportation Engineering, Shandong Jiaotong University, Jinan, Shandong Province, 250023, China
\textsuperscript{b} Management & Information School of Shandong Transport, Vocational College, Weifang, 261206
\textsuperscript{c} College of Traffic & Transportation, Northern Jiaotong University, Haidian, Beijing, 100044,

Abstract

A traffic crash prediction model was developed to accurately estimate roadway section safety. Crash rate of a roadway section was selected as the predicting measure. Moreover, the following seven factors were selected as major influencing factors: number of years of driving, number of lanes, radius of horizontal curves, longitudinal grade, road surface status, type of intersection, and width of road surface. The traffic crash prediction model was established based on fuzzy quantitative theory. To validate the model, crash data were collected at an actual site between 490.900 km to 550.789 km on a national highway. The result from the model was in agreement with the actual field data. The conclusion is that the model can be applied to other roadway locations.

© 2012 Published by Elsevier B.V. Selection and/or peer review under responsibility of Beijing Jiaotong University [BJU], Systems Engineering Society of China (SESC) Open access under CC BY-NC-ND license.

Keywords: road safety; fuzzy quantitative theory; road section crash rate; prediction

1. Introduction

Road traffic safety has been a very serious problem in China, and the number of traffic crashes has been reaching 25 million per year for the past few years (Ministry of Public Security Traffic Management 2007). Traffic safety has been affecting the economic development and social stability. Due to the random nature of traffic crash, many factors have inextricable links to crashes. Therefore it is necessary to establish a mathematical prediction model which can effectively integrate road traffic information and spatial structure information. Such a model can make scientific predictions for future road traffic safety.

* Corresponding author. Tel.: +86-531-8068-7416; fax.: +86-531-8068-7409.

E-mail address: qly72@sina.com.
Prediction models, from different perspectives, describe a profound understanding of the basic forms and trends of road traffic crashes, some of which are based on theory and derivation, while others are based on statistical analysis and empirical studies. Probability prediction method (Pario 1997, Navin 1963) assumed that crash occurrences follow Poisson distribution. Tanner’s square root rule (Fieldwick 1987) was used to forecast intersections without traffic lights; while non-linear regression model was improved in Britain (Glauz 1985, Tarek 1998). In China, researchers at Southwest Jiaotong University began to apply traffic conflicts technique to predict crashes (Zhang 2006, Liu 2006), but it had a long cycle to evaluate, and was difficult to observe due to crash events’ randomness, so it is still at the stage of theory. Researchers at Harbin Jianzhu University proposed the construction causes model based on the salience principle (Meng 2000), which played a major role in analyzing the main reason of road crash and proposing measures for improvements, but the calculation procedures were quite complicated and could not be completed even by a typical computer. Bayesian prediction model (Lin 2003) was proposed by researchers at Beijing Jiaotong University to analyze the relationship between traffic crash and speed.

All of the above methods did not take road special construction into consideration. In this paper, a microscopic prediction model of road section crash rate was proposed based on the improved fuzzy quantitative theory, which combined traffic information with the road structure, aiming at finding the relationship between traffic crash and road structure. The model would play an important role in providing necessary information for improving road design and effectively reducing crash.

2. Index choices of road section crash prediction

Because of different objectives, the prediction measure will be different in different traffic systems. For safety management staff, it is important to know the primary and secondary influencing factors. They must also be familiar with the roadway location. Crash rate has been defined previously by Qin (2006) as the number of crashes per every ten thousand vehicles passing a 500-m section. Therefore, intersection crash rate was selected as a microscopic prediction measure.

3. Related factors of microscopic crash prediction model

Similar to road crash occurrences, a traffic system is also a dynamic system constituted by persons, vehicles and roads (Pei 2002). When analyzing crash causes, traditional methods generally started from the various influencing factors, which included person, vehicle and road. Referring to the person factor, because drivers’ driving experience, information acquisition, processing and decision-making had the greatest impact on road traffic crash, driving years was selected to represent driver's experience and driving behavior. With respect to the vehicle and road factors, traffic flow, vehicles’ proportion, lane, lane width, horizontal radius, longitudinal grade, vertical slope, road junctions and adhesion coefficient were all important factors affecting road traffic crashes. Among them, the adhesion coefficient was difficult to obtain so road condition was selected as a substitute.

4. A forecasting model based on improved fuzzy quantitative theory

Both quantitative theory and regression analysis were used in quantitative prediction of the reference variables; the former focused on the case where the variables were qualitative variables, and the latter was usually limited to the quantitative descriptive variables, in fact they could be united in quantitative theory.

4.1. Basic principles of quantitative theory I
Quantitative theory is a multivariate analysis method combining qualitative with quantitative, which was used to explore and process the quantity of data expressed, judged and evaluated without numerical value (Xie 2000). Therefore, microcosmic prediction crash took advantage of the qualitative and quantitative information which were hard to be collected, made the crash factors to be quantified, then regularity and relationship between the crash and factors could be thoroughly studied. Quantitative theory adopted \{1, 0\} value which have judging sense \{Yes, No\}, and could be solved and analyzed using simple method.

The model was as follows: Supposing \( n \) group random variables \( y_1, y_2, \ldots, y_n \) relied on \( p = h + m \) random variables \( x_{1(1)}, x_{1(2)}, \ldots, x_{1(h)}, \ldots, x_{(h+m)} \), which \( x_{1(1)}, x_{1(2)}, \ldots, x_{1(h)} \) were quantitative variable and \( x_{h+1}, x_{h+2}, \ldots, x_{h+m} \) were qualitative variables; and the variables \( x_{h+j} \) \( (j = 1, 2, \ldots, m) \) took \( r_j \) value (called categories): \( c_{j1}, c_{j2}, \ldots, c_{jr} \) (Table 1).

Table 1 Quantitative theory (\( i \)) data sheet

<table>
<thead>
<tr>
<th>Sample</th>
<th>Target Variable</th>
<th>Quantitative variables ( (h) )</th>
<th>Multivariate ( (P) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( X(h+1) )</td>
<td>( X(h+f) )</td>
</tr>
<tr>
<td>( i )</td>
<td>( y )</td>
<td>( c_{11}, \ldots, c_{1r} )</td>
<td>( c_{r1}, \ldots, c_{rj} )</td>
</tr>
<tr>
<td>( i )</td>
<td>( y_1 )</td>
<td>( x_{1(1)}, \ldots, x_{1(h)} )</td>
<td>( \delta_{1(1)}, \ldots, \delta_{1(r)} )</td>
</tr>
<tr>
<td>( i )</td>
<td>( y_j )</td>
<td>( x_{j(1)}, \ldots, x_{j(h)} )</td>
<td>( \delta_{j(1)}, \ldots, \delta_{j(r)} )</td>
</tr>
<tr>
<td>( i )</td>
<td>( y_n )</td>
<td>( x_{n(1)}, \ldots, x_{n(h)} )</td>
<td>( \delta_{n(1)}, \ldots, \delta_{n(r)} )</td>
</tr>
</tbody>
</table>

For generality, with the following linear model:

\[
y_i = \sum_{u=1}^{h} b_u x_{u}(i) + \sum_{j=1}^{m} \sum_{k=1}^{r_j} \delta_{j(k)}(i) b_{jk} + \varepsilon_i, i = 1, 2, \ldots, n
\]

where,

\[
\delta_{j(k)}(i) = \begin{cases} 1 & \text{qualitative variable-j of sample-i was item-k,} \\ 0 & \text{otherwise.} \end{cases}
\]

\( b_u \) \( (u = 1, 2, \ldots, h) \) and \( b_{jk} \) \( (j = 1, 2, \ldots m; k = 1, 2, \ldots r_j) \) were undetermined coefficients; \( \varepsilon_i \) \( (i = 1, 2, \ldots, n) \) was a random error.

4.2. Improved fuzzy quantification theory method

Practically during road traffic crash prediction, it was hard to use quantitative theory directly to describe road condition, driving year and other items. So membership degree of fuzzy mathematics was introduced in this paper.
Fuzzy set theory firstly proposed by Professor Chad in United States, has been greatly developed since 1965 (Xie 2003, Miao 1986). In 1980s, several Japanese scientists introduced fuzzy mathematics into quantitative theory, and formed the fuzzy quantitative theory (Liu 1991), which introduced fuzzy data and processing of fuzzy event. While target variable was defined as the fuzzy numbers, the reactive variable corresponding to the categories was represented by the subordination degree. However, the variable namely road traffic crash in this paper was not a fuzzy number, fuzzy quantitative theory could not be directly used. So, the paper made the following improvement. Here, the reactive variable \( \delta_{ijk}(j,k) \) of sample-i was defined as:

\[
\delta_{ijk}(j,k) = \begin{cases} 
\frac{W_{ijk}}{W_j} & \text{factor-k of category-j of sample-i is yes} \\
0 & \text{factor-k of category-j of sample-i is no}
\end{cases}
\]

\( W_{ijk} \) was the sample number of factor(k) of item(j) of sample(i); \( W_j \) was the total number of item (j) of sample i. Through this correction, it was more close to the actual system.

The relationship among reference variable, item and category still obeyed linear model (1).

5. Numerical example

The model above was verified with the statistical data of traffic crash happened on a national road. Three-year road traffic crash data from 2007 to 2009 was selected from Stake 490.900km to Stake 550.789km.

5.1. Determination of road length

It meant to determine the road length of black-spot. As too short or too long was likely to cover up the dangerous location of road, the principle was to make the dangerous part fully exposed. In this paper, 500m as a unit length of a road section was selected (Qin 2006).

5.2. Variable selecting

A large number of measured data of road traffic crash and road attributes needed to be collected for applying the quantitative theory. Factors affecting crash rate were divided into seven, including driving year, lane number, horizontal curve radius, longitudinal slope, road condition, intersection and driveway width. Each factor was divided into several categories (Table 2). Because the road width selected was more than 3.4m, the seventh factor was ignored and not included in equation.

<table>
<thead>
<tr>
<th>Influence factor(item)</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driving years</td>
<td>(0,3]</td>
</tr>
<tr>
<td>Lane number(two-way)</td>
<td>4-lane</td>
</tr>
<tr>
<td>Horizontal curve radius</td>
<td>R≤500 m</td>
</tr>
<tr>
<td>Longitudinal grade</td>
<td>[0, 2%]</td>
</tr>
<tr>
<td>Road surface status</td>
<td>Moisture</td>
</tr>
<tr>
<td>Intersection and section</td>
<td>Three-way junction</td>
</tr>
<tr>
<td>Lane width</td>
<td>≤3.4m</td>
</tr>
</tbody>
</table>
5.3. Calculation of prediction model

(1) Some sections of the road were randomly sampled, taken as \( n = 20 \), road crash occurring on the defined road were calculated respectively and were divided by total traffic volume (i.e., the total number of vehicles passed through the road section within the operating time), and then the actual statistics of the crash rate \( y_i \) \((i=1, 2 \ldots N)\) was obtained.

(2) Corresponding to each sample of road section, theoretical value of crash rates \( \hat{y}_i \) could be obtained according to formula (1), the form was:

\[
\hat{y}_i = \sum_{j=1}^{r} \sum_{k=1}^{s} \delta_{ij} b_{jk}, \quad i = 1, 2, \ldots, n
\]

(3) According to the least square principle, road crash prediction model was obtained by computations:

\[
\hat{y} = 0.4802\delta_1 + 0.0152\delta_2 + 0.0009\delta_3 + 0.0076\delta_4 + 0.0098\delta_5 + 0.3256\delta_6
\]

\[
+ 0.17928\delta_7 - 0.0468\delta_8 + 0.0025\delta_9 + 0.0799\delta_{10} + 0.1301\delta_{11} + 0.0031\delta_{12}
\]

\[
+ 0.06021\delta_{13} + 0.0089\delta_{14} - 0.0020\delta_{15} + 0.6003\delta_{16} + 0.5008\delta_{17} + 0.03908\delta_{18}
\]

\[
+ 0.3117\delta_{19} + 0.07971\delta_{20}
\]

where \( \delta_{1r}, \delta_{2r}, \delta_{3r}, \delta_{4r}, \delta_{5r}, \delta_{6r} \) was eigenvalue respectively of driving years, lanes number, horizontal curve radius, longitudinal slope, road condition and type of road junction and section.

5.4. Model validation

(1) Mean square residual index was calculated as follows:

\[
r = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - m - 1} = \frac{\sum_{i=1}^{20} (y_i - \bar{y})^2}{20 - 6 - 1} = \frac{1.24809}{13} = 0.096
\]

It can be seen that the accuracy of crash rate prediction was high.

(2) Variance ratio test

\[
\frac{\sigma^2_j}{\sigma^2_y} = \frac{\sum_{i=1}^{n} (y_i^{(j)} - \bar{y}^{(j)})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

\[
\sigma^2_y = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{20 - 6 - 1}
\]

The variance of various factors calculated according to Eq. (5) are shown in Table 3.

Variance test of the model showed that six factors had different effect on crash rate, which includes; type of road section and intersection, driving years, horizontal curve radius and longitudinal slope being the remarkable factors, followed by road conditions, and then lane number.

<table>
<thead>
<tr>
<th>Table 3. Variance ratio of each type of affecting factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>( var(n \sigma^2_y) )</td>
</tr>
<tr>
<td>( var(\sigma^2_y) )</td>
</tr>
</tbody>
</table>
5.5. Model analysis

In practice, according to the annual average daily traffic on certain national road section, the annual total crash could be calculated. By analyzing the factors’ weight, the paper concludes that three-way junction had the greatest impact on crash, followed by 3 years of driving (including 3 years), and then four-way junction. The result was reasonable, and in accordance with the current status of national road No. 309.

6. Conclusions

(1) The randomness of road traffic crash and the factors affecting the crash make the road traffic crash prediction more difficult. In this paper, an improved fuzzy quantitative microscopic theory of road traffic crash prediction model was developed.

(2) Through the improvement and enhancement, the microscopic prediction model of traffic crash rate based on fuzzy quantitative theory may also be extended to other roads. During the application process, the appropriate choice of factors is the key according to the local road conditions.

(3) Limited by the data that can be collected, the microscopic prediction model is not yet complete, and there are at least three factors involved in the model that need to be further studied: traffic volume, speed and vehicle component. Apparently, since these factors have a direct impact on road traffic incidents, the statistical work should be performed and such data should be incorporated into the model.

Acknowledgment

This paper was sponsored by the Natural Science of Shandong #ZR2011GM013 program and the Shandong Jiaotong university science foundation under grant #Z201013 and #Z201109 program.

References


