Real-Time Estimation of Freeway Accident Likelihood

UCI-ITS-TS-WP-00-8

Cheol Oh ¹
Jun-Seok Oh ¹
Stephen G. Ritchie ¹
Myungsoon Chang ²

¹ Department of Civil and Environmental Engineering and Institute of Transportation Studies
University of California, Irvine
cheolo@uci.edu, jun@translab.its.uci.edu, sritchie@uci.edu

² Department of Transportation Engineering, Hanyang University
1271 Sal-dong Ansan-shi; Kyunggi-do 425-791, Korea
hytran@hitel.net

December 2000
Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu
Real-time Estimation of Freeway Accident Likelihood

Cheol Oh
cheolo@uci.edu

Jun-seok Oh
jun@translab.its.uci.edu

Stephen G. Ritchie
sritchie@uci.edu

Department of Civil and Environmental Engineering and
Institute of Transportation Studies
522 Social Science Tower
University of California
Irvine, CA 92697-3600
USA
Voice: (949) 824-6571
Fax: (949) 824-8385

Myungsoon Chang
hytran@hitel.net

Department of Transportation Engineering
Hanyang University
1271 Sa1-dong Ansan-shi
Kyunggi-do 425-791 Korea
Voice: +82-031-400-5151
Fax: +82-031-406-6290

TRANSPORTATION RESEARCH BOARD
80th Annual Meeting
January 7-11, 2001
Washington, D.C.
Real-time Estimation of Freeway Accident Likelihood

Cheol Oh, Jun-Seok Oh, Stephen G. Ritchie
Department of Civil and Environmental Engineering and
Institute of Transportation Studies
522 Social Science Tower
University of California
Irvine, CA 92697-3600
USA

Myungssoon Chang
Department of Transportation Engineering
Hanyang University
1271 Sa1-dong Ansan-shi
Kyunggi-do 425-791
Korea

ABSTRACT
In contrast to conventional traffic safety studies, this research focuses on the use of real-time freeway traffic data for potentially preventing traffic accidents, by integrating advanced traffic management and information systems (ATMIS) capabilities. This study deals primarily with traffic conditions leading to an accident identified from both real-time traffic data, obtained from inductive loop detectors, and past accident profiles. This study uses data from the I-880 freeway in California. Statistical analysis based on a non-parametric Bayesian modeling approach was conducted to estimate the real-time accident likelihood. The approach shows promise in identifying in real-time traffic conditions under which an accident could occur.
INTRODUCTION

Even though many studies have asserted that traffic conditions affect the occurrence of traffic accidents, we are not aware of any study that has investigated whether or not real-time traffic data can be used as a measure of accident likelihood. Earlier studies usually analyzed long-term historical data such as annual average daily traffic (AADT) and hourly volume. In addition, they have identified relationships between traffic variables or geometric elements and accidents. On the other hand, incident detection studies have focused on the change of traffic conditions after an incident occurrence.

With the advent of advanced traffic management and information system (ATMIS), much attention has been paid to incident detection and incident traffic management; however, less effort has been devoted to accident prevention under the ATMIS environment. This study is concerned with accident pre-identification by estimating accident likelihood.

The innovative feature of this study is to quantify the measures of accident likelihood using real-time traffic data from inductive loop detectors. This study is based on the concept that disruptive traffic conditions contribute to traffic accidents. Such disruptive traffic conditions, which are unstable and undesirable, can be represented by high temporal and spatial variations in traffic parameters. Environmental factors such as weather and inadequate geometric design might also be one of the reasons leading to high variations in traffic conditions. In this sense, unlike other static factors, measures of traffic parameters could be good indicators that dynamically represent the level of traffic instability. While detailed vehicle movement data in a section would be the best data source, traffic data from several consecutive detectors in a section can be a good surrogate to identify traffic dynamics that may lead to accidents.

This study demonstrates the potential capability of identifying traffic conditions that lead to accidents from real-time traffic data.
LITERATURE REVIEW

Accident analysis and prevention is one of the most important aspects of transportation studies because it is associated with human life. Numerous existing studies on traffic accident analysis and prevention can be divided into the following several classes:

Geometric Design and Safety

Studies focused on geometric design and safety aim to improve highway design and to eliminate hazardous locations. The effects of design elements such as horizontal curvature, vertical grade, lane width, etc. on safety have been studied. For example, Zeeger(1) evaluated the effects of cross-section design for two-lane roads. Squires(2) compared median type by accident rate. Krammes(3) evaluated the effect of geometric inconsistency on safety. Knuiman(4) examined the effect of median width on the frequency and severity of the accident. Vogt and Bared(5) and Council(6) studied traffic safety on two-lane rural roads. Anderson, et al(7) examined the relationship between safety and geometric design consistency.

Traffic Conditions and Accident Rate

Existing studies in this area have tried to identify the relationship between traffic conditions, as represented by long term traffic data, and accident rates. Gwynn(8) examined the hourly accident rate on a four-lane divided highway in New Jersey, and reported that the highest accident rate occurred when traffic volume was low and a U-shaped function would display the observed relationship. Since then much research on this topic has been performed. Ceder, et al(9), Frantzeskaki, et al(10) and Hall and Pendleton(11) have, tried to clarify this relationship. Recently, Oh and Chang(12) reported U-shaped models, comparing the relationships between volume to capacity ratio (V/C) and accident rates on freeways. This study showed the effect of capacity reduction by facility type, such as tunnel and toll gate, on safety. These studies show that traffic condition would be a good variable associated with accident occurrence.
Speed Management

Speed management studies usually deal with the effects of speed limits on safety. Thorton and Lyles(13) reported that higher speeds do not lead to more numerous or serious accidents by examining speed limits, particularly 55 versus 65 mph, in Michigan. On the other hand, Raju et al(14) showed that fatal accidents increase with high speed limit. Analysis by Lave(15) revealed that the major factor leading to an accident is not speed itself but the variation of speed. That is, when most cars travel at the same speed, whether it is a high or not, the traffic condition is safer because the probability of an accident will be lower.

Incident Management

Incident detection and management have been several of the primary research topics in the field of ATMIS to date. Incident detection includes detecting an incident based on the change in the traffic conditions after the incident occurs. Traffic dynamics after the incident occurs are used as input variables in a detection algorithm. In one of the most recent studies, Abdulhai and Ritchie(16) developed a Bayesian based neural network for freeway incident detection.

A review of the literature shows that no other study has analyzed accident likelihood using real-time traffic data. This paper presents a method for estimating accident likelihoods using real-time traffic data. Under an ATMIS environment, the approach can be helpful in preventing traffic accidents and reducing the likelihood of accidents.

TRAFFIC DYNAMICS AND ACCIDENTS

The major causal factors in accidents on the highway can be divided into four categories such as the environment, traffic conditions, vehicles and drivers. Figure1 shows a conceptual traffic chain associated with an accident.
As we can see in Figure 1, four major causal factors have a mutual relationship connected by an interaction link. If one of these factors becomes unstable, this traffic chain "vibrates", and this vibration amplifies, resulting in an accident. This implies that if we somehow maintain the stability of this chain, traffic safety can be enhanced. This study is based on the concept of removing instabilities in the traffic dynamics in order to maintain the stability of this chain.

We assume that the traffic dynamics before an accident will give us some idea of the potential for an accident occurrence. This implication will appear as a variation of the indicator defined by the traffic dynamics.

We classify traffic conditions into two patterns, disruptive and normal. A disruptive traffic condition is defined as that potentially leading to an accident occurrence and a normal traffic condition means a traffic dynamics pattern which is not involved in an accident. As illustrated in Figure 2, under the normal traffic
condition, the indicator described by the traffic dynamics index is stable. However, under the disruptive traffic condition, it is unstable and increases from the T-x point onwards.

INDICATOR TO CAPTURE ACCIDENT LIKELIHOOD

Data Description

In order to obtain accident data and corresponding real time traffic data, we used a database for the I-880 freeway in Hayward, California. Data were available from February 16 through March 19, 1993. The study section used for collecting data was about 9.2 miles long. This section of freeway had 4 to 5 lanes with a high occupancy vehicle lane (HOV). There were 17 and 18 detector sections for northbound and southbound lanes, respectively, as shown in Figure 3.

Traffic flows, occupancies and speeds were collected during 2 periods, from 5am to 10am and from 2pm to 7pm, by loop detectors. Throughout the collecting period, up to 4 probe vehicles traveled on the study section. The incident database was constructed on the basis of reports submitted by probe vehicle drivers.

To accomplish the research goal, the I-880 data was reduced to generate a new data set. We defined two traffic conditions; a normal traffic condition ($\pi_1$) and disruptive traffic condition ($\pi_2$) as follows: the normal condition is a 5-minute period 30 minutes before an accident occurrence and the disruptive traffic condition is a 5-minute period right before an accident.

Flow, occupancy and speed data were collected for 10-second periods from upstream detector stations during each 5-minute period. The 10-second data for each lane were then averaged. With regard to accident data, there were data for 91 accidents in the I-880 database, but we selected 52 accidents, excluding cases that couldn’t be matched with real time traffic data.
FIGURE 3  Study site (I-880)

Indicator

We need to identify an indicator that can represent an obvious difference between normal traffic conditions and disruptive traffic conditions. Two data sets, for normal and disruptive traffic conditions as defined above, were investigated using a t-test. Double loop detectors provided not only flow and occupancy but also speed as a point measurement. Six candidates as an indicator, mean and standard deviation of occupancy, flow and speed were considered. This was done because the value of the t-statistic can give us an idea of the statistical difference between the normal traffic conditions and the conditions.
leading to an accident. The candidate that is the most statistically significant can be selected as an indicator representing the change of traffic conditions. Table 1 summarizes the t-test results.

<table>
<thead>
<tr>
<th>Cандi- Dates</th>
<th>Occupancy</th>
<th>Flow</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5min- Average</td>
<td>5min- Standard Deviation</td>
<td>5min- Average</td>
</tr>
<tr>
<td></td>
<td>Disrup-tive</td>
<td>Normal</td>
<td>Disrup-tive</td>
</tr>
<tr>
<td>Mean</td>
<td>14.96</td>
<td>11.37</td>
<td>3.46</td>
</tr>
<tr>
<td>Variance</td>
<td>40.22</td>
<td>43.53</td>
<td>2.96</td>
</tr>
<tr>
<td>t-statistic</td>
<td>3.90</td>
<td>1.87</td>
<td>3.18</td>
</tr>
</tbody>
</table>

While most of the calculated t-statistics in Table 1 are significant, the most significant was for 5-min standard deviation of speed. For simplicity, we therefore chose this indicator to show the difference in traffic dynamics between disruptive traffic conditions leading to an accident and normal traffic conditions.

**THE PROPOSED METHODOLOGY**

Our major concern is to estimate the likelihood of accident occurrence. As mentioned earlier, traffic conditions are classified into two patterns. Traffic conditions can be classified on the basis of measurements of random variables of traffic condition $X^c = [x_1, x_2, \ldots, x_p]$. Let $f_1(x)$ and $f_2(x)$ be the probability density functions (PDFs) associated with input vector $X$ for the two populations. Normal traffic and disruptive traffic can be denoted by $\pi_1$ and $\pi_2$ respectively. A reasonable classification rule that minimizes the expected cost of misclassification is to assign a new vector to either class $\pi_1$ or class $\pi_2$, as follows (17):
\[
\pi_1: \frac{f_1(x)}{f_2(x)} \geq \left( \frac{c(1|2)}{c(2|1)} \right) \frac{p_2}{p_1}
\]

\[
\pi_2: \frac{f_2(x)}{f_1(x)} < \left( \frac{c(1|2)}{c(2|1)} \right) \frac{p_2}{p_1}
\]

where,

\[C(i | j)\] is the cost of misclassifying a given traffic condition, that is, classifying X as belonging to population \(\pi_i\), while it belongs to population \(\pi_j\). \(P_i = P(\pi_i)\), the prior probability of occurrence of population \(\pi_i\). The estimated PDFs, \(f_1(x)\) and \(f_2(x)\) are used to estimate the posterior probability that \(x\) belongs to class \(\pi_i\).

The best classifier for a given distribution is based on Bayes decision theory and minimizes the probability of classification error. The a priori estimate of the probability of a certain class can be converted to the a posteriori probability by Bayesian classification. We can allocate a given traffic condition \(X_0\) to the population with the largest probability \(P(\pi_i | X_0)\). The posterior probabilities are:

- \(P(\pi_1 | X_0) = \frac{P(\pi_1 \text{ occurs and observe } X_0)}{P_2(\text{observe } X_0)} = \frac{P(\text{observe } X_0 | \pi_1)P(\pi_1)}{P(\text{observe } X_0 | \pi_1)P(\pi_1) + P(\text{observe } X_0 | \pi_2)P(\pi_2)} = \frac{p_1 f_1(X_0)}{p_1 f_1(X_0) + p_2 f_2(X_0)}\)
- \(P(\pi_2 | X_0) = 1 - P(\pi_1 | X_0)\)

**STATISTICAL MODEL DEVELOPMENT**

**Density Estimation**

Bayesian classification requires a PDF for each class, either parametric or non-parametric. A parametric distribution is based on a mathematical formulation and the parameters of the distribution can be estimated by fitting data to the given formulation. On the other hand, non-parametric distributions divide the data into groupings and calculate the portion of values in each group. Various kinds of non-parametric smoothing
methods can be used to obtain non-parametric distributions. In practice, it is often difficult to determine the PDF with high accuracy. When the density function is assumed, the parameters of the functions are estimated using mathematical formulations. However, when the data do not fit common density functions, nonparametric techniques are used.

To obtain a reasonable PDF of normal traffic conditions, 4787 5-min periods under normal traffic conditions were selected. In the case of disruptive traffic conditions, 52 5-min periods right before an accident were used.

First, we tried to estimate PDFs by various parametric distributions.

We tested the hypothesis that a random sample of the indicator follows a given distribution using the Chi-square goodness of fit test. The results of the Chi-square tests showed that our normal traffic data could not be explained by the parametric distributions. Most of the distributions (except normal) provided an acceptable fit to the disruptive traffic data at a 5% level of significance. However, insufficient accident data resulted in small values for the degrees of freedom.

As an alternative, we decided to obtain non-parametric density functions by kernel smoothing techniques. The kernel is a continuous, bounded and symmetric real function $K$ which integrates to one (18):

$$\int K(u)du = 1$$

The kernel density estimator used in this study was defined by

$$\hat{f}_{h_n}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_{h_n}(x - X_i)$$

where $K_{h_n}(u) = h_n^{-1} K(u/h_n)$ is the kernel with scale factor $h_n$. A variety of kernel functions can be used for the kernel density estimator. A commonly used kernel function is of parabolic shape (the so-called Epanechnikov kernel):

$$K(u) = 0.75(1-u^2)I(|u| \leq 1)$$
The non-parametric densities obtained by kernel smoothing are shown in Figure 4.

![Graph showing non-parametric densities](image)

**FIGURE 4** Non-parametric density estimation

**Bayesian Model**

Probability density functions obtained by the non-parametric method above can be used to estimate the posterior probability that the indicator defined by the 5-minute standard deviation of speed belongs to either normal or disruptive traffic conditions. If these two traffic conditions are mutually exclusive, we can get the probability that a given traffic condition $X$ might lead to an accident occurrence by the Bayesian model:

$$P(A|X) = \frac{P(A) \times f_{\text{disruptive}}(X)}{P(A) \times f_{\text{disruptive}}(X) + P(N) \times f_{\text{normal}}(X)}$$

where,

$P(A/X) = \text{Posterior probability that given traffic measurement belongs to traffic conditions leading to an accident occurrence}$

$P(A) = \text{Prior probability that given traffic measurement belong to disruptive traffic conditions}$

$= \frac{\text{number of 5-min intervals in which an accident initially occurred}}{\text{total number of 5-min intervals}}$

$P(N) = \text{Prior probability that given traffic measurement belongs to normal traffic conditions}$

$= 1 - P(A)$

$f_{\text{disruptive}} = \text{Probability density function of the traffic measurement under traffic conditions leading to an accident occurrence estimated by kernel smoothing}$
\[ f_{\text{normal}} = \text{Probability density function of the traffic measurement under normal traffic conditions estimated by kernel smoothing} \]

Figure 5 shows the plot of \( P(A/X) \) for the I-880 freeway data used in this study.

![FIGURE 5 Estimated probability that standard deviation of speed belongs to traffic conditions leading to an accident](image)

**REAL-TIME APPLICATION**

The successive 5-minute standard deviation of speed was used to estimate accident likelihoods to demonstrate its potential real-time application. This analysis was performed off-line, but is indicative of the potential for real-time application. Figure 6 shows two accident cases, in which accident likelihoods were estimated by the proposed methodology, with arbitrary accident thresholds of 0.001 also indicated.

As illustrated in the Figure 6, the result indicates that traffic safety could be enhanced by reducing the accident likelihood. Also, reducing the accident likelihood is equivalent to reducing the speed variation of vehicles. One means of reducing the speed variation of vehicles is through an information system to suggest to drivers to slow down or to speed up, either as part of the road infrastructure (e.g. changeable message signs), or via an in-vehicle system. In the examples in Figure 6 for accidents #1328 and #165, if 0.001 is established as the threshold value of the probability, such warning information would be provided for about 10–15 minutes per hour. Of course, the means and nature of the information displayed for reducing the speed variation needs to be studied further.
In this approach, an accident potential is identified by the probability rising above the threshold, as illustrated in Figure 6. Therefore, as the threshold varies, so may the ability of this approach to indicate the likelihood of an accident occurring. Although the approach taken in this study is preliminary, we investigated the number of accidents in our database that could be identified for different threshold levels. Table 2 presents the results. As shown in Table 2, for a threshold level of 0.0006, 65.4% of accidents were identified. However, as the threshold level is lowered, and more accidents are identified correctly, the amount of time a driver warning system is in operation would increase. For example, when the threshold in Table 2 was set to 0.001, 46.2% of the accidents were correctly identified and the warning system would have been in operation 17.4% of the time.

<table>
<thead>
<tr>
<th>Threshold</th>
<th># accidents identified</th>
<th>% accidents identified</th>
<th>% time*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0002</td>
<td>52</td>
<td>100</td>
<td>99.8</td>
</tr>
<tr>
<td>0.0004</td>
<td>50</td>
<td>96.2</td>
<td>98.2</td>
</tr>
<tr>
<td>0.0006</td>
<td>34</td>
<td>65.4</td>
<td>52.9</td>
</tr>
<tr>
<td>0.0008</td>
<td>29</td>
<td>55.8</td>
<td>27.9</td>
</tr>
<tr>
<td>0.0010</td>
<td>24</td>
<td>46.2</td>
<td>17.4</td>
</tr>
<tr>
<td>0.0012</td>
<td>19</td>
<td>36.5</td>
<td>10.6</td>
</tr>
<tr>
<td>0.0014</td>
<td>11</td>
<td>21.2</td>
<td>4.5</td>
</tr>
</tbody>
</table>

* The percentage of time when P(A/X) was above the given threshold
FIGURE 6 Real-time application by proposed methodology

CONCLUSIONS

The objective of this study was to use loop detector data in measuring the likelihood of an accident from real-time traffic conditions. One of the most important features of this study was to use real-time traffic data as the measure of accident exposure. The likelihood that the given traffic measures, described by speed variation, belong to traffic conditions leading to an accident was estimated. The Bayesian classification used non-parametric density functions estimated by kernel smoothing techniques. Accident data and corresponding real-time traffic data from the I-880 freeway were used. An implication of the
system is that reducing the speed variation is advantageous because it increases safety and reduces the accident likelihood.

This study demonstrated the potential capability of identifying traffic conditions that lead to an accident from real-time traffic data. Insufficient data led to some key assumptions and limited the scope of this study. These are being addressed in ongoing studies with more extensive data.

REFERENCES


