Analysis of Freeway Incident duration for ATIS applications

- A Case study in the state of Maryland, USA

Woon Kim, Suhasini Natarajan, and Gang-Len Chang

***Abstract*—This paper presents a methodology for developing a model to identify the variables influencing incident duration to estimate and predict incident duration in the state of Maryland. The incident information from years 2003 to 2005 from the Maryland State Highway (MDSHA) database was used for model development, and year 2006 for model validation. Classification Trees (CT) were employed for a preliminary analysis to understand the influence of the variables associated with an incident. Based on the findings from CT, this study employed the Rule-Based Tree Model (RBTM) to develop the primary prediction model. The overall confidence for the estimated model was over 80% with several remarkable findings regarding the association between the identified factors and incident duration. Although the estimated results from RBTM were quite acceptable, in cases where RBTM did not provide incident duration within a desirable short range, a discrete choice model was developed as a supplemental model. It is deduced that supplemental models along with better quality database are required to improve the prediction accuracy of the duration of a detected incident.**

I. Introduction

INCIDENTS on congested highways, regardless of involving personal fatalities, injuries, or property damages cause considerable reduction in capacity [1], followed by heavy congestion and delay, and thus result in enormous socioeconomic loss. With reliably predicted incident duration, responsible agencies can convey the necessary information to travelers by updating the Variable Message Signs, estimate the resulting queue length and assess the need to implement detour operations and any other control strategies. Thus, an effective model to predict incident duration can be a valuable tool to mitigate congestion due to incidents.

Due to lack of data, incident duration was usually estimated based on field experience rather than rigorous statistical models. Improvements in reporting techniques and incident information database have facilitated the detailed analysis of critical variables influencing incident duration and hence its prediction. Past research in this field has resulted in different prediction methods and models. However, it must be noted that the prediction models are developed based on different sets of data. The data used to develop each model is derived from a different source. Information related to an incident varies with every database. It is also observed that incident duration is influenced by various location specific factors. Hence, to develop a reliable and efficient model for incident duration prediction for an area, the model must utilize information from a well-designed database which includes vital information specific to the area. Such a model can be confidently used to implement appropriate mitigation measures.

The objective of this study is to develop a methodology for estimating and forecasting incident duration and identify the variables influencing the incident duration in the state of Maryland using the Maryland State Highway (MDSHA) incident database collected from year 2003 to year 2006. Note that in this study, the prediction of incident duration is based on the range of time interval it may fall, such as between 15-30 minutes.

This paper begins with a brief review of previous research. This is followed by the description of the data used for the study and preliminary analyses to compute average incident duration and to identify high influence variables in Maryland State. The procedures adopted for model development and evaluation are described in detail along with the results of estimation and potential application for prediction. The final section draws conclusions based on the performance of the model and discusses perspectives for future research.

II. LITERATURE REVIEW

Various methodologies have been employed by researchers to study incident duration in the past decades. The most representative approaches are (1) Probabilistic Distributions [2], (2) Conditional Probabilities [3], (3) Linear Regression Models [4], (5) Time Sequential Models [5], (6) Decision Trees and Classification Trees [6], and (7) Discrete Choice Models [7]. Although there are a variety of existing techniques with acceptable results, they cannot be directly applied to incidents that occurred at any other locations.

Manuscript received June 1st, 2008.

Woon Kim is Research Assistant at Dept. of Civil and Environmental Engineering, University of Maryland at College Park, College Park, MD 20742 USA, (corresponding author) Phone: 301-405-2638, Fax: 301-405-2585, (e-mail: wkim2@ umd.edu)

Suhasini Natarajan is Research Assistant at Dept. of Civil and Environmental Engineering, University of Maryland at College Park, College Park, MD 20742 USA (e-mail: suhasini@umd.edu).

Gang-Len Chang is Professor at Dept. of Civil and Environmental Engineering, University of Maryland at College Park, College Park, MD 20742 USA (e-mail: gang@glue.umd.edu).

III. PRELIMINARY ANALYSIS

A. *Data Description*

For this study, highway incident data extracted from CHART-II Database (Coordinated Highways Action Response Team) were utilized. The CHART II database, maintained by MDSHA contains information about any incident occurring in Maryland, including nature of incident, time of detection/occurrence, response time, vehicles involved, type and number of involved vehicles, number of lanes/shoulders blocked, responded unit, pavement conditions, etc. The model was developed based on the data from year 2003 to year 2005 and validated based on the year 2006 database.

B. *Average Incident Duration*

Prior to model development, the average incident duration under different classifications was computed to investigate its relationships with explanatory variables. The results show that the incident duration increases with the number of heavy vehicles (e.g. tractor-trailers, single unit trucks, or pickup vans) involved or the number of blocked lanes. The incident duration on weekends and at night are generally longer than the duration on weekdays and in the daytime due to the longer response and clearance times.

It is noticeable that incidents which occurred in the four major freeways, I-495, I-95, I-695, and I-270, have relatively shorter duration than others. This can be attributed to the fact that most of the operations centers are located near these four major roads, which results in short response time, therefore decreasing incident duration.

It was also found that remarkable difference in incident duration exists between different types of incidents. On the basis of the distribution of incident duration frequency by each incident nature, it was found that incidents involving disabled vehicles and property damage are likely to have a shorter duration, while incidents causing personal injuries and fatalities are likely to have longer duration. These results are consistent with the observations that the distribution of incident durations varies with its nature. Therefore, incident nature emerges as one of the most significant factors for classifying incidents of different durations.

C. *Classification and Regression Tree (CART)*

The preliminary model was developed using CART, which is a type of decision tree technique introduced and popularized by Breiman et al. in 1984 [8]. Overall, CART performed quite well for short (5-20 minutes) or middle (20-70 minutes) ranges of incident duration, especially, for those between 5 to 10 minutes. However, it did not provide satisfactory results for incidents of long duration (e.g., longer than 1 hour). Although the results from CART were not satisfactory enough to be recommended for the incident management, they confirmed that the incident nature is the most significant variable for classification of incident duration.

IV. METHODOLOGY AND ANALYSIS

A. *Rule-Based Tree Model (RBTM)*

Based on the findings from preliminary analysis, this study has redesigned a classification tree, named as Rule-Based Tree Model (RBTM), using the following procedures. Note that incident duration, which was grouped into 5-minute intervals, was used in this approach.

Step 1: Set the incident nature as the first splitter.

From our preliminary analysis and earlier researches, it is found that the incident duration shows significantly different distributions by incident nature. This is convincible because each incident type has different characteristics that define it and these characteristics play a key role to determine its duration. Thus, the incident nature is placed at the first level as the first splitter, and then using successive splitters, the range of the estimated incident duration is shortened. In this study, incident nature was classified into five categories. Three out of them are commonly related with collisions but classified by whether or not an incident caused fatality (*Collision-Fatality*), severe injuries (*Collision-Personal Injury*), or property damage (*Collision-Property Damage*). Another category was defined as *Disabled Vehicles* and the last category was defined as *Others* which account for incident natures such as vehicle fire, debris, emergency road work, etc.

Step 2: Determine the next splitter for each node.

This step is to generate a crosstabulation table [9] to determine the next splitter for each node. A crosstabulation table displays the number of cases in each category defined by two or more specified variables. For each pair of independent and dependent variable (i.e., incident duration), this step shall create a crosstabulation table along with a bar chart for each category of each independent variable. The crosstabulation table and the corresponding bar charts show the distribution of incident durations for each category of the independent variable that is potentially associated with the incident duration. Using these, one can see how different the distribution of incident duration for the each category of an independent variable. Then, the independent variable that exhibits the most different kind of distribution in different categories shall be selected as the next splitter.

Step 3: Split the node based on the splitter determined in Step 2, in each category.

The focus of this step is to convert each splitting node in *If-then; Else-then* statement, which will constitute the set of rules for determining the incident duration for the node.

Step 4: Assign the predicted incident duration range for each split node.

This is to determine the best representative range of incident duration for each node. To achieve this, one shall first search the interval less than or equal to 30 minutes which covers at least 70% of all cases within a node. If no such interval exists in the node, then one can assign the shortest interval covering at least 60% of all cases within the node as the predicted incident duration for that node.

Step 5: Repeat Step 2 to Step 4 for all nodes until the predetermined criteria for stopping the tree growth is satisfied.

When a node satisfies one of the following criteria, the tree can be stopped at that node.

1. No independent variable is available as a splitter.
2. There is only one observation left in a node.

To evaluate the performance of rules for each node, this study adopts the concept of *support* and *confidence* developed for Associate Rules [9]. The *support* for the rule refers to the number of cases that satisfy the *If-Then* rule. The *confidence* of the rule is defined as the ratio of the number of cases satisfying the *If-Then* rule (i.e., the *support*) to the number of cases satisfying the *If* statement only. The indicator of *confidence* is conceptually the same as the conditional probability of the *Then* statement given the *If* statement of the rule.

For each incident nature, supplemental models have been considered for sub data sets which showed unsatisfactory results in the Rule-Based Tree Model, to enhance their performance. Since the RBTM developed for *Disabled Vehicles* and *Others* showed acceptable results for both estimation and validation process, they were excluded for the supplemental model development. For the rest, Multinomial Logit (MNL) Models were developed to estimate the relation between each category of incident duration and its associated factors. A well calibrated model will allow its users to predict the duration category of a detected incident based on the predicted probabilities of incident duration categories. The core concept of MNL is same as that used in accident severity model [10]. A detailed discussion regarding MNL models would be found in the [11], [12] and [14].

V. Results and Overall Findings

As shown in Table I, the overall estimated results from the developed model were quite satisfactory, while the overall validated results were not as satisfactory as the estimated one due to the lack of samples for validation. However, there are still remarkable findings through the model development. The estimated and validated probabilities for incident duration for each MNL model are summarized in Table II. As shown in Table II, the probabilities for the three categories of incident duration do not show large discrepancy from one another in the sub models for *Collision-Personal Injury (CPI)*. For example, for two categories (25-45 minutes and > 45 minutes) in CPI-Sub-Model I, the difference in probability is only about 2%. Similar phenomenon could also be found in CPI-Sub-Models II and III for the first two categories of incident duration. In MNL models for *Collision-Property Damage (CPD)*, the difference in probability between alternatives is large, but still no alternative dominates the entire dataset (i.e., over 70% probability). For this reason, probabilistic models, such as MNL models, are required to be applied for those subsets in which it is hard to find any short range of incident duration with high probability to satisfy given conditions. The difference between the estimated and validated probability is within 10%.

TABLE I
SUMMARY OF OVERALL MODEL RESULTS

		Incident Nature				
		Collision – Fatality	Collision – Personal Injury	Collision –Property Damage	Disabled Vehicles	Others
Number of Terminal Rules		21	32	37	5	12
Average Confidence	Estimated	98.23%	81.04%	89.71%	87.62%	92.93%
	Validated	12.19%	45.45%	39.36%	66.50%	63.33%

TABLE II
SUMMARY OF INCIDENT DURATION PROBABILITY ESTIMATED AND VALIDATED BY MNL SUB-MODELS

	Collision – Personal Injury				Collision –Property Damage			
	Incident Duration (min)	Observed Probability	Estimated Probability	Validated Probability	Incident Duration (min)	Observed Probability	Estimated Probability	Validated Probability
Sub-Model I	[5, 25]	0.276	0.265	0.328	[5, 30]	0.609	0.609	0.576
	(25, 45]	0.378	0.378	0.388	> 30	0.391	0.391	0.424
	> 45	0.346	0.357	0.284				
Sub-Model II	[5, 25]	0.481	0.483	0.494	[5, 25]	0.550	0.550	0.609
	(25, 50]	0.408	0.408	0.428	(25, 45]	0.285	0.285	0.235
	> 50	0.111	0.108	0.078	> 45	0.165	0.165	0.156
Sub-Model III	[5, 25]	0.366	0.366	0.461	N/A	N/A	N/A	N/A
	(25, 45]	0.439	0.439	0.379				
	> 45	0.195	0.195	0.160				

The overall findings can be summarized as:

1. For the categories of *Collision-Personal Injury*, *Collision-Property Damage*, *Disabled Vehicle* and *Others*, it turned out that the spatial factor, *County*, emerged as the second splitter. This implies that the duration for the same type of incidents varies significantly among different jurisdictions.
2. The sequence of splitters varies significantly among different categories of incidents, since different natures have different characteristics and are associated with different contributing factors.
3. Rule-Based Tree Models are more flexible for assigning an appropriate estimated incident duration range for given conditions (sub-dataset or node) than Classification and Regression Tree Models (CART).
4. Incidents occurring at night time or during off-peak hours generally last for a longer duration than those in daytime due to the lack of sufficient response units for incident clearance operations.
5. When incidents result i*n Collision-Fatality, or Property Damage*, the clearance operation is generally more efficient in the shoulder-lane blocked scenarios than when they are left open to traffic. This finding implies that shoulder lane blockage helps reduce the duration of severe

accidents as it provides a wider space for emergency response units to do the work.
6. Similarly, during the *Collision-Fatality* incidents, if the emergency response unit can close more lanes in the same direction, it generally results in a shorter duration.
7. The impact of wet pavement, a proxy variable for rainy days, on the efficiency of incident response operations is not definitive for the existing data records. It shows a positive correlation with the incident duration for those resulting in *Collision-Property Damage*, but a reverse relation for the category of *Collision-Fatality* incidents. For all other types of incidents, the impact of wet pavement on the resulting incident duration is not statistically significant.
8. Particular locations (exits) on I-495 and I-270 cause longer incident duration. This was reflected in several MNL models with negative coefficients of the related variables in short incident duration alternatives, e.g., 5-25 or 25-45 minutes. The reason attributed to this is the complexity of geometric configuration around these exit areas or their greater distance from the traffic operation centers.
9. Response time is proportional to the incident duration and this relation exhibited a negative coefficient for the shortest incident duration alternative in supplemental models for *Collision-Property Damage*.

VI. CONCLUSIONS AND FUTURE RESEARCH

Although the estimated results from the Rule-Based Tree Model (RBTM) are quite acceptable, one shall not expect the above RBTM to capture all embedded relations and provide operationally acceptable performance for real-world applications owing to the complex nature of incidents and response operations. Grounded on the promising information generated from RBTM, supplemental models can improve the prediction accuracy for the duration of a detected incident. In addition, it is essential to integrate the CHART database with police accident records to construct a dataset of better quality in order to calibrate models for incident duration, especially for incidents involving fatalities.

REFERENCES

[1] Transportation Research Board (TRB), *Special Report 209: Highway Capacity Manual*, National Research Council, Washington, D.C., 1994.
[2] T. F. Golob, W. W. Recker, and J. D. Leonard, "An Analysis of the Severity and Incident Duration of Truck-Involved Freeway Accidents", *Accident Analysis and Prevention,* Vol. 19, No. 4, pp. 375-395, August 1987.
[3] D. Nam, and F. Mannering, "An Exploratory Hazard-Based Analysis of Highway Incident Duration", *Transportation Research – A*, Vol. 34A, No. 2, pp. 85-102, 2000.
[4] A. Garib, A. E. Radwan, and H. Al-Deek, "Estimating Magnitude and Duration of Incident Delays", *Journal of Transportation Engineering*, Vol. 123, No. 6, pp. 459-466, November/December 1997.
[5] A. J. Khattak, J. L. Schofer, and M-H. Wang, "A Simple Time Sequential Procedure for Predicting Freeway Incident Duration", *IVHS Journal*, Vol. 2, No. 2, pp. 113-138, 1995.
[6] K. Smith, and B. Smith, "Forecasting the Clearance Time of Freeway Accidents", Research Report, STL-2001-01, Center for Transportation Studies, University of Virginia, Charlottesville, VA, 2001.
[7] P-W. Lin, N. Zou, and G-L. Chang, "Integration of a Discrete Choice Model and a Rule-Based System for Estimation of Incident Duration: a Case Study in Maryland", CD-ROM of Proceedings of the 83rd TRB Annual Meeting, Washington, D.C., 2004.
[8] L. Breiman, J. H. Friedman, R. A. Olshen., and C. J. Stone, *Classification and regression trees*
[9] T. Hill, and P. Lewicki, *Statistics: Methods and Applications*
[10] Khorashadi, P. E. "Analysis of Driver Injury Severity: Logit Models of Truck Involvement/Truck Causation." Ph.D. dissertation, 2003.
[11] Koppelman, F. S. and Bhat, C. "A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models.",2006.
[12] Ben-Akiva, M. and Lerman , S.R. "Discrete choice analysis: Theory and application to travel demand." MIT Press, Cambridge, Massachusetts, 1985.
[13] Ulfarsson, G. F. "Injury severity analysis for passenger car, pickup, sport utility vehicle and minivan drivers: Male and female differences." Ph.D. dissertation, 2001.
[14] Washington, S., Karlaftis, M. and Mannering, F. "Statistical and Econometric Methods for Transportation Data Analysis." CRC Press, Boca Raton, 2003.